GERİYE YAYMA TİPİNDEKİ YAPAY SİNİR AĞLARININ STOKASTİK VERİYİ İŞLEMESİNİN ÇEŞİTLİ YÖNLERİ ÜZERİNE

NİHAT YILDIZ

DOKTORA TEZİ
FİZİK ANABİLİM DALI
Ocak 1996

GERİYE YAYMA TİPİNDEKİ YAPAY SİNİR AĞLARININ STOKASTİK VERİYİ İŞLEMESİNİN ÇEŞİTLİ YÖNLERİ ÜZERİNE

NİHAT YILDIZ

DOKTORA TEZİ

FİZİK ANABİLİM DALI

Ocak 1996

Danışman : Prof.Dr. **RAUF EMİROV**

FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Bu çalışma jürimiz tarafindan, Fizik Anabilim Dalı'nda Doktora tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. Bünyamin ÖZBAY
Üye: Prof. Dr. Ali Fettah ŞAHBAZOV
Üye: Prof. Dr. İsmail SÖKMEN
Üye: Prof. Dr. Rauf EMİROV
Üye: Yrd. Doç. Dr. Mustafa YILDIRIM

ONAY

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

11/12/1996

FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRÜ

Prof.Dr. **Fuat ÖNDER**

Bu tez, Cumhuriyet Üniversitesi Senatosunun 05.01.1984 tarihli toplantısında kabul edilen ve daha sonra 30.12.1993 tarihinde C.Ü. Fen Bilimleri Enstitüsü Müdürlüğünce hazırlanan ve yayınlanan ' Yüksek Lisans ve Doktora tez yazım Kılavuzu' adlı yönergeye göre hazırlanmıştır.

Sevgili Hatice Gül ve Yosun'a . Anne ve babama.

İÇİNDEKİLER

ÖZET

Doktora Tezi

GERİYE YAYMA TİPİNDEKİ YAPAY SİNİR AĞLARININ STOKASTİK VERİYİ İŞLEMESİNİN ÇEŞİTLİ YÖNLERİ ÜZERİNE

NİHAT YILDIZ

Cumhuriyet Üniversitesi
Fen Bilimleri Enstitüsü
Fizik Ana Bilim Dalı
Danışman: Prof.Dr. Rauf Emirov

Bu çalışmada stokastik bir ortamdaki Geriye Yaymalı Yapay Sinir Ağları (GYYSA) nın performansının aşağıdaki yönleri FORTRAN 77 programlama dilinde orijinal olarak yazılan programlar aracılığıyla gerçekleştirilen bilgisayar simülasyonlarıyla araştırıldı.
(i) Çok Boyutlu Normal Dağılım (ÇBND) ve ÇBND'den farklı diğer bazı dağılımlardan üretilen verinin YSA tarafından asimptot altı (n:(gözlem sayısı)$\leq$100) ve asimptotik (n=250) bölgede dönüştürülme yeteneği.
(ii) Gizli tabaka nöronu sayısının dönüştürme yeteneğine etkisi
(iii) Verilen bir r_{XY} korelasyon yapısı ile (**X**:input ve **Y**:hedef vektörü) YSA nın hata fonksiyonunu minimize eden final ağırlık vektörleri bileşenlerinin korelasyonları arasındaki ilişki.

Asimptot altı bölgede verilen bir r_{XY} için çok sayıda veri matrisinin kullanılması nedeniyle standart GYYSA nın düşük yakınsama hızını yenmek için literatürde az tanınan Genişletilmiş Delta Bar Delta Metodu (GDBDM) kullanıldı. GDBDM stokastik bir ortamda YSA literatüründe ilk kez bu çalışmada kullanıldı .

r_{XY} sıfırdan bire doğru arttıkça YSA nın ÇBND verisini dönüştürme yeteneğinin kesin olarak iyileştiği analitik olarak türetildi ve deneysel olarak kanıtlandı. Bundan başka, asimptotik limitte korelasyon yapılarının tümü için YSA hata bileşenlerinin (dizilerinin) teorik hata limitlerine başarıyla yakınsadıkları gözlendi. Ayrıca ÇBND'den farklı dağılımlardan gelen veriler içinde YSA nın dönüştürme yeteneğinin teorik öngörülerle (White (1989a)) uyum içinde olduğu gözlendi. Bu gözlemlerin ve GY ağlarının rasgele bir çevredeki dönüştürme yeteneğine ilişkin bugüne kadar deneysel bir çalışma olmaması gerçeğinin bir sonucu olarak şu söylenebilir. Bu çalışma GY ağlarının stokastik bir ortamdaki dönüştürme becerisinin araştırılması için deneysel olarak öncü bir adım atmıştır.

Verilen bir r_{XY} korelasyon yapısı ile YSA nın hata fonksiyonunu minimize eden final ağırlık vektörlerinin bileşenlerinin korelasyonları arasında özellikle r_{XY}=0 ve r_{XY}=1 durumlarında benzerlikler olabileğine ilişkin sonuçlar elde edildi.

YSA nın stokastik bir ortamdaki genelleştirme yeteneğinin yeterince iyi olduğu çok sayıda korelasyon yapısı için gözlendi..

Anahtar Sözcükler: Yapay sinir ağları, Standart geriye yayma metodu, Genişletilmiş delta bar delta metodu, Stokastik veri, Korelasyon yapısı, Dönüştürme yeteneği, Asimptotik davranış.

SUMMARY

Ph.D. Thesis

ON THE BACK PROPAGATION TYPE NEURAL NETWORK'S PROCESSING OF VARIOUS ASPECTS OF STOCHASTIC DATA

NİHAT YILDIZ

Cumhuriyet University

Graduate School of Natural and Applied Sciences

Department of Physics

Supervisor: Professor Dr. Rauf AMIROV

In this work the following aspects of the performance of Back Propagation Type Neural Network (BPNN) were investigated through computer simulations carried out by means of an original program written in the programming language FORTRAN 77.

(i) At subasymptotic region (where n (number of observations)$\leq$100) and at asymptotic region (n=250) network's processing ability of data generated from Multivariated Normal Distribution (MVND) and that of generated from some other (different from MVND) distributions.

(ii) Possible affection of the number of hidden layer neurons to the network's processing (fitting) ability.

(iii) For a given correlation structure r_{XY} (X:input and Y: target vectors) the relationship between r_{XY} and the correlations of the components of final weight vectors which minimizes the error function of the (artificial) neural network (ANN).

At subasymptotic region as we are analyzing a vast number of data matrices of a given r_{XY} in order to overcome the problem of standard BP's low convergence rate a less known BP algorithm Extended Delta Bar Delta Method (EDBDM) was employed. In ANN literature this work is the first which has ever used EDBDM in a stochastic context.

It was analytically derived and experimentally proved that network's fitting ability of MNVD data *definitely* improves as r_{XY} approaches from zero to one. Additionally, for all correlation structures employed it was observed that the error components of the ANN have been successfully converged to their theoretical limits. Furthermore, it has been another observation that the for data from some other (non-normal) distributions network's fitting ability is well in line with theoretical predictions. As a result of these observations and of the fact that no experimental study related to BP network's fitting ability in a random environment has appeared so far, it can be said that this work took a pioneering experimental step for the investigations of BP network's fitting capability in a stochastic medium.

Some experimental results that support that there can be some similarities (in particular in r_{XY}=1 and r_{XY}=0 cases) between for a given r_{XY} (correlation structure) and the components of weights minimizing the error function have been obtained.

It was also observed that the generalization ability of the network is sufficiently good in a stochastic medium.

Key Words: Artificial neural network, Standard Back Propagation Method, Extended Delta Bar Delta Method, Stochastic data, Correlation structure, Fitting (or transformation) ability, Asymptotic behaviour.

TEŞEKKÜR

Bu tezin üretiminde emeği geçen çok sayıda insan var kuşkusuz, tıpkı yaşamın öteki üretim alanlarında olduğu gibi. Tezde kullanılan verilerin sağlanması ve bazı ön deneylerin yapılması İngiltere'nin "University of Portsmouth" üniversitesinde Dr. D.W.Salt'ın gözetiminde gerçekleşti. Onu saygıyla anarım. Tezdeki çalışmaların büyük bir bölümünü inanılmaz bir sevgi ve hoşgörüyle yöneten, çok değerli teorik ve pratik katkılar sağlayan bilim adamı, danışmanım Prof. Dr. Rauf Emirov'a bütün içtenliğimle teşekkürlerimi arz ediyorum. Onun katkısı olmaksızın bu tez hiç bir zaman gün ışığı görmeyebilirdi.

Sevgili eşim Hatice Gül'e ve biricik kızımız Yosun bu tezin oluşumu sırasında bana sağladıkları sevgi için mavi gökyüzünün enginliği ölçüsünde minnet borçluyum.

Ayrıca emeği geçen Uzman Suat Bilgin'e ve öteki insanlara teşekkür etmek benim için zevkli bir görevdir.

ŞEKİLLERİN LİSTESİ **Sayfa no**

Sayfa no

Sayfa no

Sayfa no

TABLOLARIN LİSTESİ **Sayfa no**

SİMGELER VE KISALTMALAR DİZİNİ

r: İnput uzayı boyutu
p: Output uzayı boyutu
n_h: Gizli tabakadaki nöron sayısı
N_s: Ağırlık uzayı boyutu
n: Bir veri matrisi için gözlem sayısı
n_v: nod vektörü
λ ya da λ_{th}: Teorik hata
C_1: Kısıtlayıcı hata bileşeni
C_2: Serbest hata bileşeni
λ_n: λ nın örneklem analogu
C_{1n}: C_1 bileşenenin örneklem analogu
C_{2n}: C_2 bileşeninin örneklem analogu
g: Koşullu beklenen değer
E: Ortalama değer (expectation) lineer operatörü
X: İnput rasgele değişkeni
Y: Hedef rasgele değişkeni
r_{XY}: X rasgele değişkeni ile Y rasgele değişkeni arasındaki korelasyon katsayısı
(x,y,z): (r_{X1X2}, r_{X1Y}, r_{X2Y}) korelasyon üçlüsü (yapısı)
iw: Başlangıç vektörü indisi
w: Ağırlık vektörü
w*: Optimum ağırlık vektörü
$\mathbf{Z} \approx N(\mu,\Sigma)$. **Z** rasgele değişkeni N(normal) dağılımdandır. Ortalama vektörü (μ) ve kovariyans matrisi (Σ) dir. ***R*** (korelasyon matrisi) , R ise Euclidean uzayıdır.
η: Öğrenme adımı

YSA: Yapay sinir ağları
ÇBND: Çok boyutlu normal dağılım
D.T.: Dağılım tipi
GYYSA: Geriye yaymalı standart yapay sinir ağı.
SGYM: Standart geriye yayma metodu
GDBDM: Genişletilmiş delta-bar-delta metodu
ark. (arkadaşları)
KBD: Koşullu beklenen değer
bkz: Bakınız.
Kes.: Kesim
denk: Denklem (Bu tez boyunca denklem kısaltması tanım, ifade, matematiksel denklem vb. için ortak olarak kullanılmaktadır)
min: minimum
var: variyans
İUB: İnput uzayı boyutu
AF1: Aktivasyon fonksiyonu 1
AF2: Aktivasyon fonksiyonu 2
YN: Yapı no

BÖLÜM 1: GİRİŞ: MOTİVASTON ve KAPSAM (Yapay Sinir Ağları ve Fiziksel Bilimler)

1.1 Motivasyon

*Şekillenim tanıma*nın insan beyninin olağanüstü gücünü anlamada anahtar olacağına inanılmaktadır (Stillings 1987). Konuşma ve görüntü işleme; insan beyninin daha önceden öğrenilmiş ilişkileri giriş sinyaline uyuşturmaya çalıştığı bir çeşit şekillenim tanımadır. Bugün geleneksel bilgisayarlar kullanılarak algoritmik şekillenim tanıma üzerine yürütülen çalışmalar uzman sistemler olarak anılmaktadır. Bu sistemler; özel bir alanda çalışan bir uzmanın bilmesi gereken bilgiyi ortaya çıkarması açısından iyi olup ancak sağ duyu denilen yetenekten yoksundurlar (Minsky 1987).

Eksik olan bu sağ duyuyu yakalamak için, araştırmacılar son on yıl boyunca çabalarını potansiyel olarak daha güçlü sistemleri anlamaya yönelttiler. Bu sistemlere Yapay Sinir Ağları (YSA) denir. Bu ağların ve öğrenme tekniklerinin şekillenim tanıma, kontrol, sınıflandırma, kestirme vb. alanlara uygulanmasıyla önemli başarılar elde edildi. Bu dramatik başarının sırrı , sistemlerin adından da anlaşılacağı üzere, oldukça açıktır. Bu sistemler; oldukça yüksek derecede biribirleriyle bağlantılı olarak paralel işlem birimlerinden oluşan beynin yapısını taklit ederler.

Bu çalışmanın ereği insan beyni ile YSA nı karşılaştırmak olmadığı için, beynin nasıl organize olduğuna hiç dokunulmayacaktır. Bununla birlikte, böylesi karşılaştırmaları yapan çalışmaların varlığına gönderme yapılmalıdır (örn.bkz. Harth 1982).

YSA nın fiziğin ve fizikçilerin dünyasına girmesi Hopfield'in 1982 yılındaki araştırma programı ile olmuştur (Hopfield 1982). Ünlü istatistik fizikçi Amit'e göre (Amit 1986), Hopfield programı sosyal bir fenomendir ve bulaşıcı bir hastalık gibi California, Paris, Roma, Moskova, Edinburgh, Londra vs. de bulunan her önemli fizik merkezine sıçramıştır. Oradaki fizikçilerin çalışmaları uygun bütün disiplinlerle iç içe geçti. MIT'de biyologlarla, Pasteur enstitüsünde kompüter bilimcileriyle, Paris'te mühendislerle vs. Bugün bu ortak çalışmaların sonucunda Nörofizik denilen bir bilim dalı doğmuş ve olgunlaşma sürecine girmiştir.

Hopfield'in çalışma programı bu tezin de bir anlamda esin kaynağıdır. Bu tezin yazarının yaptığı daha önceki bir çalışmada Hopfield ve Little modellerinin çeşitli termodinamik özellikleri bazı yönleriyle incelenmiştir (Yıldız 1987). Bununla birlikte bu çalışmanın odak noktası bugün evrensel düzeyde oldukça yaygın olarak kullanılan geriye yayma tipindeki YSA olduğundan Hopfield modelinin ayrıntılarına hiç değinilmeyecektir ama Kesim 1.2 de bazı noktalar yeniden vurgulanacaktır.

1.2 Kapsam (Yapay Sinir Ağları ve Fiziksel Bilimler)

Hopfield yaklaşımını başarıya götüren konjektür karmaşıktır. Başarıda en önemli etken 'zeka' nedir sorusunu matematiksel olarak inceleyebilmek için formülasyonlarda önemli ölçüde basitleştirmelere gidilmesiydi: Sinaptik ağırlıkların (formal nöron çiftlenimlerinin) rasgele ve keyfi olarak seçilmiş şekillenimleri öğrenebileceği bir bir küme seçilebilir mi? (Öğrenilmiş bir şekillenim ağın dinamik davranışının bir atraktörü (limit noktası) gibi düşünülür).

Hopfield modelinde N tane +1 ya da -1 değerlerini alan yapay nöronlar vardır. Yani yapay bir nöronun aksiyon potansiyeli $S_i=\pm1$ dir. i. ve j. nöronlar biribirlerine J_{ij} ağırlıklarıyla bağlıdırlar. Böylece konfigürasyon uzayında 2^N tane olanaklı konfigürasyon ortaya çıkar. i.nöron $V_i=\sum_j J_{ij}S_j$ ile tanımlı bir post sinaptik potansiyel elde edilir. Eğer $V_i>T_i$ ise $S_i=+1$, $V_i<T_i$ ise $S_i=-1$ olur. Burada T_i nöronun eşik potansiyelidir. $S_i=+1$ ise nöron aktif -1 ise pasiftir. Böylece S_i lerin ve J_{ij} lerin başlangıç durumu verildiğinde sistem 2^N boyutlu konfigürasyon uzayında bu kurala göre evrimini sürdürür.

Bu tezin kapsamı geriye yayılmalı YSA ların stokastik bir ortamdaki performansının çeşitli yönlerinin araştırılması ile sınırlıdır. Tezde belirli bir düzeyde genelliğe ulaşabilmek için somut ve tanıdık fiziksel terimler ve kavramlar yerine soyut matematiksel modeller kullanıldı. Bu yaklaşım bu çalışmanın sonuçlarını fizik,kimya, biyoloji vb. gibi deneysel bilimlerin değişkenlerine uygun dönüşümler yaparak uygulamaya daha yönelik yapamak için seçildi.
Bu çalışma hedeflerini özet olarak şöylece gerçekleştirdi.

Spesifik bir olasılık dağılımından çekilen çok sayıda örneklem örnekleri korelasyon yapılarıyla Standart Geriye Yayma Metodu (SGYM) (Rumelhart ve ark.1986) ve Genişletilmiş Delta-Bar-Delta Metodu (GDBDM) (Minai ve ark. 1990) {Bu ikinci metod stokastik anlamda literatürde ilk kez bu çalışma tarafından uygulandı} tiplerindeki YSA'nın stokastik veriyi dönüştürmesinin kimi yönleri arasındaki ilişki araştırılarak ortaya çıkarıldı. Bu kimi yönler şunlardır.

1) Ağın stokastik veriyi asimptot altı bölgede fit etme yeteneği.,

2) Gizli tabaka nöronu sayısının iyileştirme etkisi,

3) Başlangıç ağırlık vektörünün etkisi,

4) Son durumdaki ağırlık vektörlerinin korelayon matrislerinin seçilmiş elemanlarının değerleri

5) Ağın asimptotik davranışı,

6) Sigmoid output aktivasyon fonksiyonu yerine lineer output fonksiyonu kullanılarak 2-5 adımlarının araştırılması,

7) Ağın genelleştirme yeteneği,

8) Öğrenme adımı ve iterasyon sayısı ,

9) Normal dağılımdan farklı bazı olasılık dağılımlarından gelen verinin fit edilmesi,

1.2.1 Denel fizik ve yapay sinir ağları

Bütün deneysel bilimlerde olduğu gibi fizikte de bağımlı ve bağımsız değişkenler arasındaki fonksiyonel bağıntının bulunması deneycinin temel ereğidir. Özellikle bağımsız değişken sayısının birden çok olduğu durumlarda bu bağıntının bulunabilmesi zor bir çabayı gerektirir. Fiziksel sistemlerin çevresiyle her an etkileşmeleri nedeniyle bu değişkenlerin ölçülmesi her zaman az ya da çok olasılıksal bir süreçtir. Yani değişkenlerin deneysel olarak ölçülen değerlerinin her zaman bir belirsizliği vardır. Bir başka deyişle, eğer değişkenler X_1, X_2,...,X_n ve Y iseler bunların ortak bir olasılık dağılımı vardır ve bu dağılım deney öncesinde (a priori) çoğu kez bilinemez. Bu nedenle genellikle yapılan varsayım bu değişkenlerin mükemmel bir kesinlikle ölçülebilir olduklarıdır (en azından klasik fiziğin yaklaşımı budur). Labaratuvarda, değişkenler arasında bir ilişki bulmaya çalışan bir denel fizikçi tipik olarak şu süreci izler. X bağımsız değişken (örneğin $K=1/2mv^2$ kinetik enerji formülünde m=sbt varsayılarsa X=v(hız)) ve Y bağımlı değişken (Y=K (kinetik enerji)) olmak üzere Y=f(X) biçiminde bir f fonksiyonunu X ve Y üzerine N tane (örneğin N=15) yaparak keşfetmeye çalışır. Bunu yapmak için $A=\{(X_1,Y_1), (X_2,Y_2),...,(X_N,Y_N)\}$ kümesindeki noktaları grafik kağıdına geçirir. Ve fonksiyonu sezmeye çalışır. Bunun için genellikle fonksiyonun tipi hakkında (lineer, üstel, trigonometrik vb.gibi) varsayımlar yapmak zorundadır. Bunu bulduktan sonra ölçüm yapılmayan noktaları da bu fonksiyonla denemek zorundadır. Eğer fonksiyonun ne olduğu konusunda belirli bir güvenilirliğe ulaştıysa bu bağıntı öteki araştırmacıların bilgisine ve yararlanmasına sunulur. Süreç bu kez şöyle sürer. Öteki araştırmacılar bu fonksiyonu kendi deneysel verileriyle test ederler. Kimi durumlarda fonksiyon iyi sonuçlar verir ve evrensel kabüle doğru yol alır. Bazı durumlarda ise alternatif fonksiyonlar ortaya çıkar. Ve deneyler sürdürülür. Kimi fizik deneylerinde ise bütün ölçüm aralığı için tek bir fonksiyon bulmak fiziksel fenomenin içsel doğası yüzünden oldukça zor hatta olanaksız olabilir. Bu gibi durumlarda fonksiyon, X bağımsız değişkeninin değer bölgesi bölünerek f fonksiyonu çeşitli aralıklarda değişik biçimde tanımlanır.

Oysa bu işlem matematiksel olarak göz önüne alınırsa ilginç sonuçlar ortaya çıkar. Weierstrass teoremine göre [a,b] kapalı aralığında sürekli bir f fonksiyonuna istenilen bir $\in$ duyarlıkta n. derece bir polinomla yaklaşılabilir. Bir başka deyişle, $\in$ duyarlığı saptandıktan sonra (örn. $\in=10^{-4}$) deneycilerin aradığı f fonksiyonu tek bir fonksiyon tipinde (polinom) bulunabilir. Burada $P_n(\mathbf{a})$ gibi bir fonksiyonun bulunması amaçlanır. **a** parametre vektörü olarak bilinir ve n+1 tane bilinmeyen (polinomun katsayıları) içerir. Weierstrass teoremi normlu uzaylarda genel yaklaşım (approximation) teoremi çerçevesinde şöyle açıklanabilir. *C*[a,b], [a,b] kapalı aralığında tanımlanan sürekli fonksiyonlar kümesi olsun. *P*[a,b] ise [a,b] kapalı aralığında tanımlanan bütün polinomların kümesi olsun. Verilen bir $\in>0$ için en az bir $p\in P$ vardır ve $\forall x\in[a,b]$ için $\max|p(x)-f(x)|<\in$ dır (yani *P*[a,b] kümesi *C*[a,b] de yoğundur) Burada p polinomunun derecesi verilen $\in>0$ bağlı olarak değişir. $\in>0$ bir kez belirlendikten sonra f fonksiyonuna bu duyarlıkta yakın olabilecek sonsuz sayıda p polinomu bulunabilir. Bu polinomun nasıl bulunabileceği ise başka bir konudur. Bunlardan bir tanesi Bernstein polinomları yardımıyladır (Haaser 1971, p194). Ancak f fonksiyonu yeterince yalın değilse bu işlem de analitik olarak oldukça güçtür. Hornik ve ark.(1989) Weierstrass teoreminin genellemesi olan Stone-Weierstrass teoremini kullanarak *ileri beslemeli tek gizli tabakalı yapay sinir ağlarının* ilgi duyulan hemen hemen her fonksiyona keyfi derecede yaklaşım yapabileceğini matematiksel olarak yetkin bir biçimde kanıtladılar. Bu anlamda ileri beslemeli YSA evrensel yaklaşımcılardır. Bu yaklaşım için iki koşul vardır. (i) Yaklaşımda bulunulacak f fonksiyonu Borel ölçülebilir (bkz.Kes.5.2.3) olmalıdır. (ii) Yeterli sayıda gizli tabaka nöronu olmalıdır. İleri beslemeli YSA (Şekil 1.1) bir input tabakası (r nöronlu), bir gizli tabaka (n_h nöronlu) ve bir output tabakalı (p nöronlu) yapay bir sistemdir. Bütün tabakalardaki nöronlar biribirlerine ağırlıklar denilen bağlantılarla bağlıdırlar. İnput tabakasına öğrenilecek giriş verisi out tabakasına ise bulunması istenilen (hedef) verisi sunulur. YSA giriş versini önce ınputtan gizli tabakaya sonra gizli tabakadan output tabakasına dönüştürür. Öğrenilecek M tane örnek varsa bu M tane örnek üzerinden toplam hatayı hesaplar. Sonra bir iterasyon süreci yardımıyla ağırlıkları uygun biçimde değiştirerek toplam hatayı minimize etmeye çalışır. Ve toplam hatayı minimum yapan ağırlıklar kümesi bulunarak öğrenme (eğitme) süreci sona erer.

Not: Buradan sonra ileri beslemeli tek gizli tabakalı YSA nı kısaca YSA olarak anılacaktır. Bu tezde kullanılan YSA nın hepsi ileri beslemeli ve tek gizli tabakalıdır.

Hornik ve ark (1989) da ulaştıkları sonuçlardan birisi de şudur (Theorem 2.5). $X=\{x_1,x_2,...,x_n\}$ kümesi verilsin. Burada X in bütün elemanları birbirinden farklı olup herbiri R^r uzayının elemanlarıdır. $g:R^r \rightarrow R$ keyfi fonksiyonu verilsin. Eğer Ψ squash fonksiyonu (eksi sonsuzda 0 artı sonsuzda 1 değerini alan ve artmayan herhangi bir $R \rightarrow [0,1]$ fonksiyon) 0 ve 1 değerlerini asimptotik olarak değil ama gerçekten alırsa $\underline{\forall i \in \{1,...,n\}}$ için $\underline{f(x_i)=g(x_i)}$ olacak şekilde bir $f \in \Sigma^r(\Psi)$ (n tane gizli tabaka nöronu bulunan tek gizli tabakalı YSA) fonksiyonu vardır. Burada $\Sigma^r(\Psi)$ kümesi ileri beslemeli YSA için özel olarak tanımlanır. Bu teoreme göre n tane elemandan oluşan bir X eğitme kümesinin bir YSA ile tam (sıfır hata ile) temsili mümkündür.

Hornik ve ark.(1989). ulaştıkları sonuçların deneysel fizik açısından pratik önemi şudur: X bağımsız değişkeni (birden çok olabilir) ve Y bağımlı değişkeni üzerine yapılan N tane ölçüm verilsin. $\in$ duyarlığı saptandıktan sonra YSA kullanılarak deneycinin bulmayı umduğu eğriye kuramsal olarak eşdeğer ($\in$ ile sınırlı) pratikte ise oldukça yakın ($\in$ ile sınırlı olmanın yanı sıra bilgisayarın yuvarlama hatalarının getirdiği hata katkısı ve ayrıca gizli tabaka nöronları sayısının yeterli olmayışının indüklediği hata katkısı nedeniyle) elde edilebilir. {Yukarıdaki satırda sözü edilen eşdeğerlik şöyle anlaşılmalıdır. Deneycinin bulmayı umduğu eğri f olsun ve $\in$ duyarlığı verilsin. Uygun bir norma göre $|f-g_f(\in)|_{norm}<\in$ ise $g_f(\in)$ fonksiyonu f fonksiyonuna eşdeğerdir. Benzer olarak $a \in A=\{h:|f-h_f(\in)|_{norm}<\in\}$ biçiminde tanımlanan a fonksiyonu da f ye eşdeğerdir}. Yani eşdeğer fonksiyonların sonsuz sayıda olması nedeniyle deneyci burada eğrinin tipi hakkında bilgi sahibi olmak zorunda değildir. YSA nın deneyciye sunduğu ikinci (ve belkide en önemlisi) olanak şudur. Deneycinin uğraştığı eğriler genellikle non-lineer tipte olup eğrinin istenilen parametrelerini elde etmesi oldukça yorucu bir süreçtir. Oysa YSA optimum ağırlık bileşenlerini bulmakla bu parametreleri de bulmuş olmaktadır. Bu işlem ise bilgisayar tarafından otomatik olarak yapılmaktadır. Örneğin bu tezin yazarının da yer aldığı bir çalışmada (Salt ve ark.(1992)) YSA nın performansı araştırıldı ve çeşitli model fonksiyonlar için YSA nın fonksiyon bulma işleminde yararlı olabilecekleri gösterildi.

Son olarak deneycinin gerçek ortamına YSA nın nasıl uygulanabileceği hakkında bir kaç söz söylemekte yarar vardır. Hornik ve ark(1989) Teorem 2.5 e göre bulunması istenen f eğrisini bulmak amacıyla N ölçüm yapıldığında YSA nın g fonksiyonu f ye bütün ölçüm noktalarında çakışır. N ölçüm sayısı artırıldğında yeni bir g fonksiyonu elde edilir. Böylece N yeterince büyük olduğunda (örneğin N=40) hemen hemen bütün noktalar için geçerli olabilecek bir g fonksiyonu elde edilir.

1.2.2 Yapay sinir ağlarına istatistik mekaniksel yaklaşım

YSA na istatistik mekaniksel bir yaklaşım fizikçi Tisby ve ark.(1989) tarafından getirildi. Bu yaklaşımda YSA üzerine tutarlılık koşulu getirilerek aynı ağ kompleksitesine (input-gizli-output tabaka nöronları sayısı aynı) sahip ağların bir Gibbs kanonik dağılımına ulaşılır. Burada tutarlılık koşulu şudur: Örneklem üzerinden toplam hatayı minimize etmek ağın parametreleri üzerinden eğitme kümesinin olasılığını maksimize etmeye eşdeğerdir. Yaklaşımın ayrıntıları şöyledir.

Bir olasılık dağılımından $\{P(x_i,y_i|\ F)\equiv\overline{p}(\mathbf{x}_i)\}$ çekilen istatistiksel olarak bağımsız $\mathbf{x}_i=(x_i,y_i)$ $(i=1,2,\ldots,m)$ örneklemi verilsin. Burada $\overline{p}(\mathbf{x}_i)$ x_i ile y_i nin biribirlerine bir F fonksiyonuyla bağlı bulunduğuna ilişkin öznel (subjektif) inancımızı yansıtır. $\overline{p}(\mathbf{x}_i)$ iki farklı elemanı içerir: Deneysel verinin içsel duyarlığı ve hatalara karşı hoşgörümüz. Her nekadar $\overline{p}(\mathbf{x}_i)$ prensipte var olmalıysada YSA eğitme sürecinde ender olarak kullanılır. Bunun yerine, verilen m tane $\mathbf{x}_i$ (input-output çiftleri) üzerinden $(\mathbf{x}^{(m)}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\})$ ve $\mathbf{w}$ ağırlık vektörüne bağlı olarak

$$E^{(m)}(\mathbf{w})\equiv E(\mathbf{x}^{(m)}|\mathbf{w})=\sum_{i=1}^{m}e(\mathbf{x}_i|\mathbf{w}) \tag{1.1}$$

$E^{(m)}(\mathbf{w})$ toplam hatası eğitme sürecinde minimize edilir. Burada $e(\mathbf{x}_i|\mathbf{w})$ bireysel hata fonsiyonu çoğu kez Euclidean metriğidir. Bu tanımlardan sonra incelemeye geçilebilir. Ağın eğitme süreci $\mathbf{w}$ parametreleri üzerinden olabilirliği (likelihood) maksimize ederek gerçekleşir. Olabilirlik fonksiyonu bütün olasılıkların çarpımıdır.

$$\max_{\mathbf{w}\in W} P(\mathbf{x}^{(m)}\big|\mathbf{w}) = \max_{\mathbf{w}\in W}\prod_{i=1}^{m}p(\mathbf{x}_i|\mathbf{w}) \tag{1.2}$$

Buradaki $\underset{\rightharpoonup}{p}$ olasılığı koşullu olasılık olup $\mathbf{w}$ verildiğinde $\mathbf{x}$ in gerçekleşme olasılığıdır. Bu olasılıkların var olduğunu varsayıyoruz. Yukarıda sözü edilen tutarlılık koşulunun yerine getirilebilmesi için (yani her iki optimizasyon kriterinin biribirine eşdeğer olabilmesi için

$$\prod_{i=1}^{m}p(\mathbf{x}_i|\mathbf{w})=\varphi\Big(\sum_{i=1}^{m}e(\mathbf{x}_i|\mathbf{w})\Big)\ ;\varphi \text{ monoton ve sürekli türevlenebilir olmalıdır.} \tag{1.3}$$

Önerme: Sürekli türevlenebilir bir $e(\mathbf{x}|\mathbf{w})$ pozitif hata fonksiyonu için denk.(1.3)'in çözümü

$$p(\mathbf{x}|\mathbf{w})=(1/z(\beta))\exp[-\beta e(\mathbf{x}|\mathbf{w})];\quad z(\beta)=\int_{X\otimes Y}\exp\big[-\beta e(\mathbf{x}|\mathbf{w})\big]d\mathbf{x} \tag{1.4}$$

ile verilir. Bu çözüm tektir. Burada β, $p(\mathbf{x}|\mathbf{w})$ olasılığının hatanın değerine karşı duyarlığını saptayan pozitif bir sabittir. Ortalama hata $\overline{e}=\int e(\mathbf{x}|\mathbf{w})p_{\mathbf{w}}(\mathbf{x})d\mathbf{x}$ kabul edilebilir hata düzeyinin bir ölçüsüdür ve β' ya

$$\overline{e}=-\frac{\partial \log z}{\partial\beta};\quad \frac{\partial e}{\partial\beta}\prec 0 \tag{1.5}$$

ile bağlıdır. Burada normalizasyon katsayısı $z(\beta)$ nın ya da eşdeğer olarak $\overline{e}$ nın özel bir w sının açık bir fonksiyonu olmadığı varsayılır. Denk.(1.4) deki integralin mümkün bütün input-output çiftleri üzerinden alındığını göz önüne alırsak bu varsayım gerçekleşir.

1.2.2.1 Gibbs kanonik küme dağılımı

İnput-output çiftleri $\overline{p}(x)$ olasılık fonksiyonundan çekilmiş olsun. Ağın **w** konfigürasyonu verildiğinde bu $\mathbf{x}^{(m)}$ örneklemlerinin **w** ile biribirlerine bağlı olmasının olasılığı yani $\mathbf{x}_i \in \{\mathbf{x}: \mathbf{x}=(x,net(x,w))\}$ olması denk(1.4) kullanılarak

$$P(\mathbf{x}^{(m)}|w) = \prod_{i=1}^{m} p(\mathbf{x}_i|w) = (\frac{1}{z^m})\exp\left[-\beta\sum_{i=1}^{m} e(\mathbf{x}_i|w)\right] \tag{1.6}$$

ile verilir. Denk.(1.6) daki koşullu olasılık ağın W konfigürasyon uzayında bir olasılık dağılımı oluşturacak bir biçimde tersine dönüştürülebilir. Bunu yapmak için verilen bir $\mathbf{x}^{(m)}$ kümesi verildiğinde aşağıdaki Bayes formülü (denk.(1.7)) kullanılır. Burada $\rho^{(0)}$ konfigürasyon uzayında non-singüler prior (deney öncesi) olasılık dağılımıdır.

$$\rho^{(m)}(w) \equiv P(w|\mathbf{x}^{(m)}) = \frac{\rho^{(0)}(w)P(\mathbf{x}^{(m)}|w)}{\int_W dw\rho^{(0)}(w)P(\mathbf{x}^{(m)}|w)} \tag{1.7}$$

Denk.(1.7), $E(\mathbf{x}^{(m)}|w)$ hatasının terimleri cinsinden yazıldığında YSA üzerine dağılmış Gibbs kanonik dağılımı elde edilir. Bu dağılım denk.(1.8a) daki gibidir.

$$\rho^{(m)}(w)=(1/\Omega^{(m)})\,\rho^{(0)}(w)\exp[-\beta E^{(m)}(w)] \tag{1.8a}$$

Burada $\Omega^{(m)} = \int_W dw\rho^{(0)}(w)\exp[-\beta E^{(m)}(w)]$ (1.8b)

ile verilir ve hata bölüşüm fonksiyonu olarak bilinir ve konfigürasyon uzayında erişilebilir hacim üzerinden ağırlıklı hatadır. Öte yandan $E(\mathbf{x}^{(m)}|w)=\sum_{i=1}^{m} e(x_i|w)$ ise, **w** verildiğinde eğitme noktaları için toplam hatayı verir. Denk.(1.8) W uzayında bir dağılım olarak sezgisel bir anlama sahiptir: m tane $\mathbf{x}^{(m)}$ üzerinde eğitme yapıldıktan sonra herbir w noktasının olasılığı ağın eğitme kümesinde üzerindeki hataya göre üstel olarak azaltılır..

Bu terimlerle bağımsız bir $\mathbf{x}_{m+1}$ örnekleminin eklenmesinin eşdeğer işlemi ; $\rho^{(m)}$ dağılımını $\exp(-e(x_{m+1}|w)$ ile çarpıp dağılımı yeniden normalize etmektir. Yani

$$\Omega^{(m+1)} = \int_W \rho^{(0)}(w)\exp\{-\beta E^{(m)}(w) - \beta e(x_{m+1}|w)\}dw \leq \int_W \rho^{(0)}(w)\exp\{-\beta E^{(m)}(w)dw=\Omega^{(m)}. \tag{1.9}$$

olur. Böylece ağı eğiterek(öğreterek) konfigürasyon uzayındaki erişilebilir hacimi küçültmüş oluruz ya da eşdeğer bir deyişle kümenin "serbest enerjisi" $-\log\Omega^{(m)}$ i eğitme kümesinin büyüklüğü m ile monoton artırmış oluruz. Burada $\rho^{(m)}(w)$ eğitim sonrası dağılımdaki tek parameterenin β ya da eşdeğer olarak; eğitim sürecinde doğrudan kontrol edebileceğimiz

$$\left\langle E^{(m)} \right\rangle = \int_{w} \rho^{(m)}(w) E^{(m)}(w) dw = -\frac{\partial \log \Omega^{(m)}}{\partial \beta} \geq 0 \tag{1.10}$$

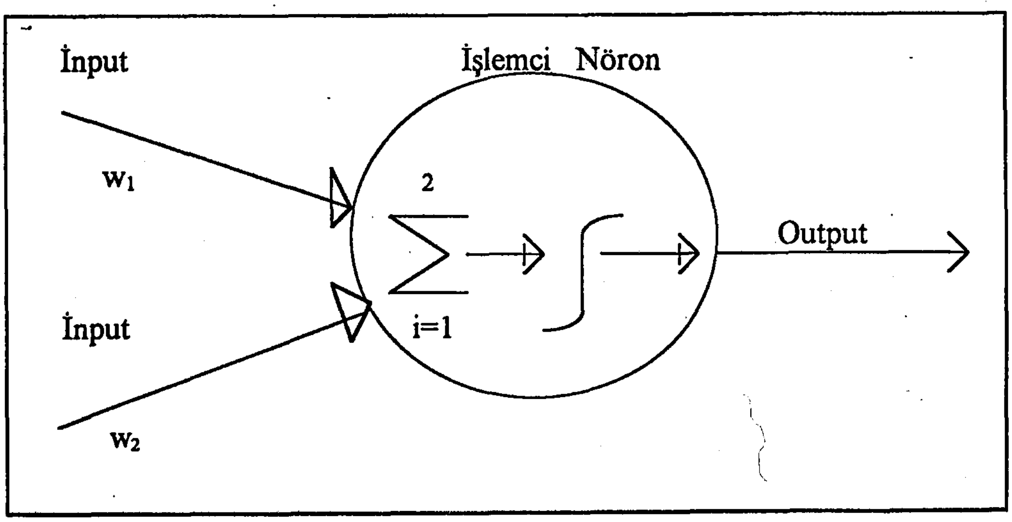
olduğuna dikkat edilmelidir.

1.3 YSA na Kısa Bir Giriş

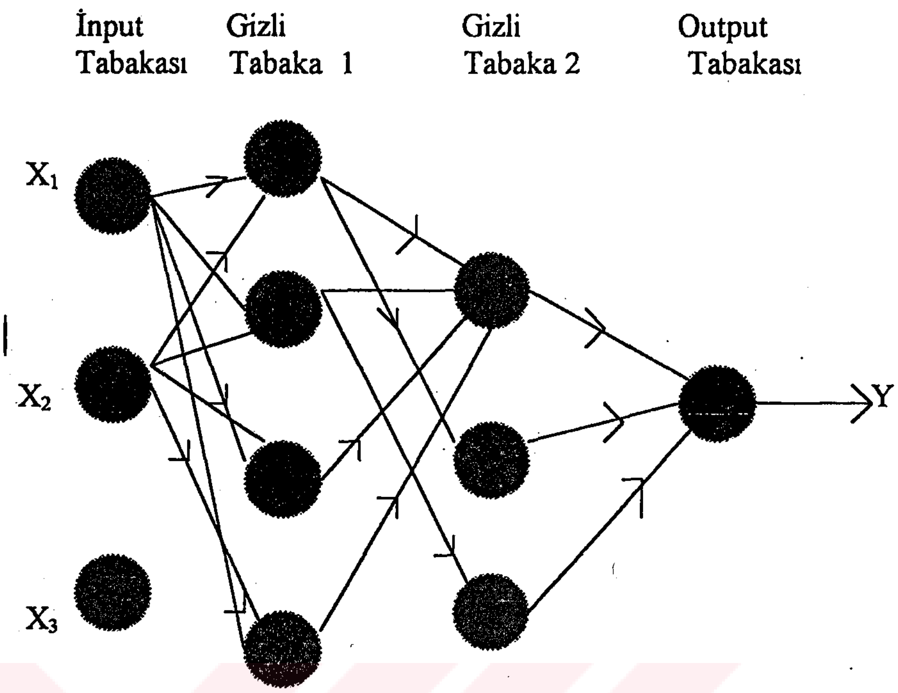
Bugünün YSA beyinle karşılaştırıldığında henüz çok basittir. Genellikle tek tip nöronlardan oluşurlar. Bu yapay nöronlar bazen yalnızca 0 ya da 1 değerini alan tek çıkışlı ve çok girişlidir. Ancak bugünlerde artık çıkışı sürekli olan nöronlar kullanılmaktadır. Bölüm 3 de yeniden sözü edileceği gibi her giriş bir ağırlıkla çarpılır ve bu çarpılan girişler toplanır daha sonra bir transfer fonksiyonundan geçerek nöronun çıkışını oluşturur (Şekil 1.1)

Yapay nöronlar genellikle basit tabakalı bir yapıdan oluşur. Örneğin bu tezin biricik ilgi odağı olan Geriye Yayma tipindeki YSA giriş ve çıkış tabakalarına ek olarak birde ara tabakadan (ayrıca gizli tabaka da denir) oluşur. Bu tabaka ağın uygun bir iç temsil yapma yeteneğinin en önemli gerekenidir (Şekil 1.2). Bu tabakalı ağlar spesifik bir inputu spesifik bir hedefe dönüştürmeyi öğrenebilirler. Ağlar dönüştürmeyi bir kez öğrendikten sonra daha önce hiç görmemiş olduğu bir input istenilen hedefe dönüştürebilir.

Bu ağların öğrenmesi , peşpeşe tabakalar arasındaki ağırlıkları uygun bir algoritmaya göre modifiye ederek gerçekleşir. Bu modifikasyon işlemine *eğitme periyodu* adı verilir.



Şekil 1.1 Tipik bir yapay nöronunun şematik gösterimi



Şekil 1.2 Çok Tabakalı Geriye Yaymalı Tipik Bir Yapay Sinir Ağı. (Gizli tabaka sayısı aslında n tane olabilir ancak şekilde yalnızca iki tanesi gösterilmiştir. Ayrıca bazı bağlantılar şeklin yalın olması için elenmiştir. Bias nöronlarının input değeri 1'e eşittir ve bias nöronları şekilde gösterilmemiştir).

1.4 Konu İle İlgili Daha Önce Yapılan Çalışmalar: Literatür

Geriye Yaymalı YSA (GYYSA) nın stokastik ortamdaki performansının sistematik deneylerle sınanması anlamında en son literatür verilerini de göz önüne alarak bu tezdeki araştırmanın literatürde ilk olduğu rahatlıkla belirtilebilir. Bununla birlikte, YSA nın dönüştürme yeteneklerinin içsel istatistiksel özellikleri teorik olarak araştırılması çeşitli araştırmacılar tarafından yapıldı (ayrıntılar için bkz. Bölüm 2). Bunlar arasında en dikkate değer olanı White (1989a) tarafindan geliştirilen analizlerdir. White (1989a) da YSA nın stokastik bir çevrede ağ öğrenmesinin objektifinin (hedefi) ağın $\mathbf{w}^*$ ile gösterilen bir optimum ağırlık vektörü bulması olduğu gösterildi. Ağ bu $\mathbf{w}^*$ yardımıyla, öğrendiği örnekleri "inşa yoluyla" genellemeyi öğrenmektedir. İnşa yoluyla neyin anlatılmak istendiği şöyle açıklanabilir. $\mathbf{w}$ rasgele (random) ağırlık vektörünün bir fonksiyonu olacak biçimde bir $\lambda(\mathbf{w})$ ortalama (beklenen) hata fonksiyonu tanımlanır. Bundan sonra $\lambda(\mathbf{w}^*)$ fonksiyonu, eğitme ve test etme olarak ikiye bölünen veriyi doğru olarak kestirir (yani genellemeyi öğrenir. Bölüm 2 de bu konuda daha fazla şey söylenecektir. YSA nın araştırılmaya değer bir başka özelliği , istatistiksel tutarlılık, yine White (1990) tarafindan çalışıldı. İstatistiksel tutarlılık ağın çevresiyle olan deneyimi arttıkça (yani eğitme kümesindeki noktaların sayısı arttıkça) ağın verilen herhangi bir $\epsilon>0$ sayısından daha fazla hata yapmasının olasılığının sıfıra gitmesini garanti eder. Bir başka yakın konu ağın ağırlık bileşenlerinin asimptotik limitteki (n eğitme kümesi elemanları sayısı sınırsız arttıkça) dağılımı da White(1989a) tarafından çalışıldı (bu tezde ise bu dağılımla ilgilenilmedi ancak ağırlık bileşenlerinin asimptotik korelasyon matrisleri üzerine bazı gözlemler yapıldı). White(1989a) GYYSA kestirimcilerinin (estimator) beklenen hata fonksiyonunu minimize eden parametreye (w*) hemen hemen kesin olarak yakınsadıklarını kanıtladı.. Bu tezde YSA nın eğitme verisinin korelasyon yapısıyla ağ dönüştürmesinin çeşitli yönleri arasındaki ilişkinin araştırılmasında GYYSA nın tercih edilmiş olmasının nedeni işte az önce sözü edilen kanıttır.

Asimptotik istatistiksel özellikler çeşitli araştırmacılar ((White 1989a,1989b,1990) ve Barron (1989) tarafından çalışılırken bağlantılı (connectionist) sistemlerin istatististiksel doğasına ilişkin birleşik bir teori Golden(1988) tarafından sunuldu.

YSA nın istatistik özellikleri değerlendirilirken göz önüne alınması gereken iki önemli konu daha vardır. Birincisi "fazlalık input hipotezidir". Bu durum bir ya da daha çok sayıda input, öteki inputların lineer kombinezonu olduğunda ortaya çıkar. İkincisi ise " bağlayıcı (gerekli) olmayan gizli tabaka nöronları"dır. Fazlalık olan inputların eliminasyonu görece kolaydır. Asimptotik durumda bu problem White (1992) ve Gallant ve White (1988) tarafından yapılan çalışmaların sonucunda ele alınabilir. Daha az karmaşık ağlarda "budama" denilen işlem uygulanabilir (Sietsma ve Dow 1991).

Gereksiz gizli tabaka nöronu hipotezini test etmek çok daha zordur. Güçlük şuradadır: Sıfır hipotezi verilen bir gizli tabaka nöronunun gereksiz olduğunu öne sürdüğünde (yani bu nörondan outputa giden ağırlık bileşeni sıfır olduğunda) bu nörona gelen ağırlık bileşenleri tanımsızdır ve tek (unique) değildir. Bu durumu değerlendirebilmek için daha yetkin bir teori gerekir (Hansen 1992). Ayrıca Davis (1977,1987) , White(1989c) ve Lee ve ark(1989) tarafından yapılan çalışmalar uygundur.

Son olarak bu çalışmaya doğrudan hiçbir etkileri olmasada ağların istatistiksel özelliklerine dokunan öteki çalışmaların varlığından da söz edilmelidir. Genel bir regresyon probleminin elektronik bir bağlamda ele alınması (Specht 1991); ve öğrenme fazına bir de unutma fazı eklenerek output ile inputun korelasyonun sağlanması (Aoyama ve Ichikawa (1991)) buna verilecek örneklerdir. Ayrıca Mielnizcuk ve Tyrcha (1993) da ise çok tabakalı ağların tutarlı kestirimlerine ilişkin analitik bazı sonuçlar verilmiştir. Ayrıca Sungur (1995) tarafından yapılan bir çalışmada rasgele YSA larda öğrenmeye sorunu pratik bir bakış açısıyla ele alınmıştır.

1.5 Tezin Bölümlerinin Organizasyonu

Bölüm 2 de YSA dönüştürmesi ile istatistiksel çıkarsama arasındaki içsel ilişkinin teorisi sunulmuştur. Bu bölümün önemi bu tezde elde edilen simülasyon sonuçlarının tümü White (1989a) tarafından ortaya konulan teorik değerlendirmeleri test etmek için yapılmış olmasından gelir. Bölüm 3 de kullanılan veri ile GYYSA ve standart olmayan GYYSA hakkında bilgi verilir. Bölüm 4 de asimptotik altı sonuçlar AF1(aktivasyon fonksiyonu 1) ile verilir. Bölüm 5 de asimptotik sonuçlar (Aktivasyon fonksiyonu 2 ile) analitik ve nümerik olarak çok sayıda sonuç için sunulur. Bölüm 5 deki son kesim genel sonuçlarla birlikte araştırmayı kapatır. Kapanışta gelecekteki olası çalışmaların yönü hakkında da kimi önerilerde bulunulur.

BÖLÜM 2: TEORİK DEĞERLENDİRMELER: YAPAY SİNİR AĞLARI ve İSTATİSTİKSEL ÇIKARSAMA

2.1 Girişsel açıklamalar

Bu bölüm yapay sinir ağları (YSA) ile istatistik bilimi arasında aslında çok derin bir içsel ilişki bulunduğunu teorik olarak ortaya koymak için ayrılmıştır. Bu ilişki ortaya çıkarıldığında YSA ile ilgili aşağıdaki zor ama önemli sorulara doyurucu yanıtlar verme olanağına kavuşacağız. "Hangi koşullar altında ağın genelleştirme yeteneği iyidir?". "Genelleştirmeyle kastedilen nedir?". "Verilen bir ağ için uygun bir düzeyde ağ karmaşası (ağırlık bileşenleri sayısı) nasıl saptanır?". Son soruya bugüne dek doyurucu teorik bir yanıt sağlanamamıştır. Bugünkü YSA pratiğinde bu sorun genellikle deneme-yanılma yoluyla üstesinden gelinmeye çalışılmaktadır.

Bölümün planı şöyledir. Kes.(2.2) de YSA nı eğitmek(öğretmek) ile istatistiksel çıkarsama arasındaki içsel ilişki tartışılır. Kes.(2.3) de YSA eğitiminin olasılıksal yönü üzerinde durulur ve koşullu beklenen değerin teorik önemi vurgulanır. Kes.(2.4) de öğrenme sürecinin hedefleri anlatılır ve öğrenmenin erekleri istatiksel doğasına bağlanır. Son olarak, Kes.(2.5) de ağın performansını örneklem üzerinden optimize ederek öğrenmenin amaçlarına nasıl ulaşılabileceğini gösteririz.

2.2 YSA nı Eğitme Metodları ile İstatistiksel Çıkarsama Arasındaki İlişki

Araştırılan bir fenomeni yöneten değişkenler X ve Y olsun. Bu iki değişken arasında nasıl bir ilişkinin bulunduğu yaşamda sıkça merak edilen bir konudur. Bu X ve Y arasındaki ilişkiyi açıklayan bir teori çoğu kez vardır ya da üretilebilir, ama öngörülerin doğruluğu eninde sonunda deneysel gözlemlerle sınanmak zorundadır. Yine öyle durumlar vardır ki, fenomenin karmaşık doğası yüzünden ortada ne bir teori vardır ne de üretilebilir. Her iki durumda da YSA deneysel bilginin depolanıp işlenebileceği kullanışlı bir araçtır.

YSA nın dönüşüm (bir uzaydan öteki uzaya) yapma sürecinde istatistiksel analizin gerekliliğini daha derin bir düzeyde anlayabilmek için X ve Y üzerine n tane gözlem(ölçüm) yapılmış bulunduğunu varsayacağız. Onları $x_1,...,x_n$ ve $y_1,...,y_n$ ile gösterelim. Eğer X ve Y nin her ikisi de vektör niceliklerse, ölçülen x_t ve y_t değerleri de vektör nicelikler olacaklardır (t=1,2,...,n). X in rx1 ve Y nin px1 boyutlu olduğunu varsayacağız. [Bu tez boyunca gerek vektör nicelikler (deterministik ya da rasgele) ve gerekse rasgele nicelikler (vektör olsun ya da olmasın) siyah kalın harflerle (**a** gibi) gösterimlenirler]. Notasyonel uygunluk için $\mathbf{Z}=(\mathbf{X}',\mathbf{Y}')'$

ve $z_t = (x_t' , y_t')'$ yazılır. Tersi belirtilmedikçe bir üs işareti " ' " bir vektörün ya da matrisin transpozunu sergileyecektir. Uygun olarak , $z^n = (z_1, ..., z_n)$ ile göstereceğimiz n tane gözlemden oluşan ya da YSA diliyle "eğitme kümesi" bulunacaktır. Ölçme sürecinin sonsuza dek yinelendiğini de düşünmek olanaklıdır. Böylece $(z_t) = (z_t, t=1,2,...)$ ile gösterilen stokastik vektörler dizisi elde edilir. Tanım gereği, örneklem uzayı (olasılık uzayı) herhangi bir ölçülebilir fonksiyona *istatistik* denir. Aynı durum (z_t) dizisinin herhangi bir ölçülebilir fonksiyonu $(f(z_t): t=1,2,...)$ dizisi için de geçerlidir. İstatistiğe en tanınmış örnek $\hat{y} = (\sum_{t=1} y_t)/n$ ile tanımlı ortalama değerdir. Her ne kadar oldukça az yaygınsada, örneklemin kendisi bile bir istatistik örneğidir. Ancak örneklemin tümüyle uğraşmak uygun bir yol olmadığından, genellikle özet bilgilere başvurmayı yeğleriz. YSA bu deneysel bilginin depolanıp işlenebileceği (bilgiyi YSA ların ağırlık vektöründe depolayarak) güçlü bir araçtır. Bu gerçek, YSA ların gizli ve içsel istatistiksel doğasının anahtarını sunar.

2.3 Olasılık ve YSA

Şeyleri ölçme yöntemimiz, herhangi bir istatistiği analiz ederken temel önem taşır. Ölçümlerimizi tam olarak kontrol edebileleceğimiz gibi, kısmi bir kontrolun bile olmadığı zamanlar vardır.

Denetleme yeteneğimiz ne olursa olsun, her zaman bir çevresel olasılık dağılımı μ_e den (kısaca çevre) söz edilebilir. Bu kavram yararlıdır çünkü ölçümlerin üretildiği koşulların tam bir betimlemesini sağlar. Bu anlamda, μ_e , x_t yi saptamada bizim ve doğanın rolünün bilgisini sağlar.Matematiksel olarak, μ_e: $A \rightarrow [0,1]$ dir. $(A \subset R^p)$. y_t nin x_t ile tamamen belirlenebildiği yani $y_t = g(x_t)$ olduğunu varsayalım. g: R^r den R^p ye bir fonksiyondur. g bize x_t ile y_t arasındaki ilişkiyi tümüyle sağladığından, ilgimiz doğal olarak g üzerine odaklanır. Öteki durumlarda y_t bazı başka rasgele değişkenlerden de etkilenir. Bu durumda **X** ile **Y** arasında yalnızca koşullu olasılıktan söz edilir. Yani x_t ile y_t , $\mathbf{X}_t$ ve $\mathbf{Y}_t$ gibi rasgele iki değişkenin gerçeklenimi olurlar. $z_t = (x_t', y_t')'$. z_t için tanımlanır ν_j ile gösterilen r+p boyutlu rasgele bir birleşik (joint) olasılık dağılımı tanımlanır. Bütün t ler için x_t ve y_t nin **X** ile **Y** nin aynı olasılık dağılımına haiz olduğunu düşünerek t indisini düşürürüz. **X** verildiğinde **Y** nin koşullu olasılık dağılımı (KOD) $\varphi(.|x)$ dir. Yani $\varphi(A|x) = P[Y \in A | \mathbf{X} = x]$ $(\forall A \subset R^{p*})$ için. ν_j, μ_e ve φ gibi iki bileşene ayrılabildiği için , ilgimizi φ üzerine veririz. Çünkü, φ bilindiğinde μ_e yi bilmek yoluyla ν_j nin bilgisine erişebiliriz.

YSA tarafından neyin öğrenilmiş olduğunu bilmek istediğimizde φ önemlidir. Ancak YSA için bilmemiz gereken φ den çok onunla yakından ilişkili olan koşullu beklenen değer (KBD) g(X)=E(Y|X) ile gösterilsin. E(Y|X) p- boyutlu bir rasgele vektördür ve $g:R^r \rightarrow R^p$ olarak tanımlanır. g bize **X** üzerine gözlemler yapıldığında **Y** nin deneysel (ampirik) bilgisini verir. **Y** nin her gerçeklenimi g(X) den her zaman farklı olacağı için $\in$=Y-E(Y|X) biçiminde beklenen bir sapma olacaktır. Bu sapmayla, **Y** rasgele vektörü **Y**=g(X)+$\in$ olarak yeniden yazılır. Verilen bir **X**=x için ortalama sapma her zaman sıfırdır. Bu sonuç $\in$ nin tanımından ve iterated beklenen değerler yasasından çıkar. Yani $E(\in|X)=E(Y|X)-E(E(Y|X)|X)=E(Y|X)-E(Y|X)=0$

2.4 Öğrenme Sürecinin Hedefleri

Her nekadar φ KOD'u soyut olarak önemliysede, YSA için asıl önemli olan spesifik bir çevrede, verilen bir amaç için optimal bir çözüme ulaşmaktır. Ancak, ağ tarafından elde edilen şey eninde sonunda φ fonksiyonuna bağlanmak zorundadır.

X ile **Y** arasındaki ilşkiyi bilmek önemlidir, çünkü **Y** yi açıklayabilmek için **X** kullanılacaktır. Bu ilişkiyi öğrenmenin yolu, ağın ortalama performans fonksiyonunu (hata fonksiyonu) minimum yapacak optimal ağırlık vektörlerini bulmaktır. Formal olarak, $\pi:R^p x R^p \rightarrow R^p$ biçiminde tanımlanan bir fonksiyonu minimize edilmelidir. **y** hedef değeri ve ağın outputu **o** verildiğinde $\pi(\mathbf{y},\mathbf{o})$ bu performansı ölçer. π üzerine sınırlamalar şöyledir. $\pi(\mathbf{y},\mathbf{o}) \geq 0$ ve **y**=**o** için π minimum yani sıfır. o output fonksiyonu, **x** e bağlı olduğu kadar ağın ağırlık vektörüne de bağlıdır. Bunu $f:R^r x W \rightarrow R^p$ biçiminde bir fonksiyonla gösterebiliriz. $W \subset R^{N_s}$ dir. (N_s bir tamsayı olup bir **w** ağırlık vektörünün bileşen sayısına eşittir. Böylece $(\mathbf{x},\mathbf{w}) \in R^r x W$ verildiğinde o=f(**x**,**w**) olup , verilen **y** için ağın performansı $\pi(\mathbf{y},f(\mathbf{x},\mathbf{w}))$ ile verilir. Her nekadar verilen bir (**y**,**x**,**w**) için π ağın performansını ölçebilirsede, gerçekte istenilen çok sayıda (**y**,**x**,**w**) seçimi için ağın performansının iyi olmasıdır. Ortalama performans, $\pi(\mathbf{Y},f(\mathbf{X},\mathbf{w}))$ rasgele niceliği için

$$\lambda(\mathbf{w}) = \int \pi(y, f(x, w)\nu_j(dy, dx) = E[\pi(Y, f(X, w)]; w \in W \quad (2.1)$$

olarak tanımlanır. λ ya beklenen performans fonksiyonu (hata fonksiyonu) denir. **x** ve **y** değerlerinin integrasyon yoluyla ortalaması alındığı için , λ yalnızca w nın bir fonksiyonudur. Bu integral R^{p+r} uzayında Lebesgue integralidir. ν_j olasılık ölçüsü, kesikli, sürekli ya da ikisinin karışımı olabilir.

Ağırlık vektörünü (**w**) ağın ortalama performansını en iyi yapacak biçimde seçme yeteneğimiz bulunduğu için ağ öğrenmesinin amacını $\boxed{\mathbf{w}^* \in W \text{ ve } \lambda(\mathbf{w}^*) = \text{minimum}}$ olacak biçimde w^* ağırlık vektörünü bulmak biçiminde tanımlayabiliriz. w^* vektörü tek (bir tane) olmak zorunda değildir, bizim için çözümü optimum yapabilecek herhangi başka bir w^* ağırlık vektörü de aynı derecede kabul görür. π ile φ arasındaki ilişkiyi aydınlatmak için bir örnek verelim. Örnek: Geriye Yayma tipindeki YSA nda beklenen performans fonksiyonu $\pi(y,o)=(y-o)^2$ olsun. Yani $\lambda(w)=E[Y-f(X,w)]^2$ dir. $g(X)=E(Y|X)$ olduğundan

$\lambda(w)=E([Y-g(X)+g(X)-f(X,w)])^2=E([Y-g(X)]^2)+2E([g(X)-f(X,w)][Y-g(X)])+E([g(X)-f(X,w)]^2)$

Burada eşitliğin sağ yanının ortasındaki terim sıfırdır. Bunu görmek için, $\epsilon=Y-g(X)$ alalım. $h(X)=g(X)-f(X,w)$ olsun. $\therefore E[h(X)\epsilon]=E[(E(h(X)\epsilon|X)]=E(h(X)E(\epsilon|X)]=0$, çünkü $E(\epsilon|X)=0$ dır. Bu kısıtlamayla

$$\lambda(w)=E[(Y-g(X))^2]+E([g(X)-f(X,w)]^2) \qquad (2.2)$$

olarak yazılır. Bu tezde White (1989a,1989b) tarafından ortaya konulan teorik değerlendirmeler derken genellikle denk(2.2) ye referans verilmiş varsayılacaktır. Denk.(2.2) deki w^* vektörü yalnızca $\lambda(w)$ yı minimize etmekle kalmayıp aynı zamanda denk.(2.2) nin sağ yanındaki ikinci terimi de minimize eder. Yani $w^*\in W$ için $\boxed{\int [g(x)-f(x,w^*)]^2\mu_e(dx)=\text{minimum}}$ olur. Sözcüklerle söylersek, w^* ağırlık vektörü, g koşullu beklenen değer fonksiyonuna yaklaşım yapacak $f(.,w^*)$ fonksiyonunu üretir. Bu gerçek aşağıdaki gerçekle daha soyut olarak yeniden ortaya konulabilir.

♠ ***Koşullu beklenen değerin çok çarpıcı bir özelliği***

X ve Y, $(\Omega,\Im,P)$ olasılık uzayında (bkz.Kes.5.2.3) tanımlı iki rasgele değişkenler olsunlar. $E(Y^2)<\infty$ olsun. X verildiğinde Y yi en iyi predikte eden fonksiyon koşullu beklenen değer $E(Y|X)=g(X)$ dir. Yani verilen herhangi bir $g^\bullet(X)$ için $E(Y-g)^2\leq E(Y-g^\bullet)^2$ dir. (bkz. örneğin White (1984 p55, teorem 3.73) ya da Grimmett and Stirzaker (1992)). Yani kısaca yukarıda vurgulandığı gibi $f(x,w^*)$ fonsiyonu $g(x)$ den daha iyi olamaz. Ya da ağın dönüştürme yeteneğinin en iyi olması için f fonksiyonunun g ye çok yakın olması gerekir. Bu gerçek Mielnizcuk ve Tyrcha (1993) tarafından da vurgulanmış ve kanıtlanmıştır.

2.4.1 Optimal ağırlıkların belirlenmesinde çevrenin rolü

Denk.(2.1) de verilen beklenen performans fonksiyonunu minimize etmede μ_e nin rolü önemlidir. w^* ağırlık vektörü bazı X değerleri için daha küçük bazı X değerleri içinse (gerçekleşme olasılığı küçük olan) daha büyük hatalar verir. Yani $\mu^e\neq\mu_e$ için w^* vektörü optimal performans üretmez. Bu durum özel bir durum olmayıp genellikle doğrudur (White 1989a).

2.4.2 Spesifik performans fonksiyonun rolü

π de optimal ağırlık vektörünü belirlemede önemli rol oynar. $\underline{\pi}\neq\pi$ gibi farklı bir fonksiyon kullanılırsa daha önceki w^* nın $\underline{\pi}$ için de optimal olması gerekmez. Örneğin, $\underline{\pi}$ maksimum sapma fonksiyonu ise yani $\underline{\pi}(o_i,y_i)=\max_{i\in I}|o_i-y_i|$, burada I verilen eğitim örneklemleri için indis kümesidir, ise verilen w^* için performans optimum değildir. [Bu durum bu tezin yazarı tarafından deneysel olarak gözlenmiştir. Ancak bu konu ile ilgili hiçbir sonuç bu tezde sunulmadı] . Özetle, π ve μ_e nin ağın performansını doğru olarak yansıtacak biçimde seçilmeleri gerekir.

2.4.3 Optimal ağırlıklar ve genelleştirme

Eğer w^* nın bulunması ağın öğrenme sürecinin ereği ise, ağın genelleştirme yeteneği kolayca tanımlanabilir. O da şudur: Eğer **X** ve **Y**, ν_j birleşik olasılık dağılımından çekilmişlerse ağın outputu $f(X,w^*)$ olduğunda $\lambda(w^*)$ en iyi ortalama performansı verir. Bunu ağın performansının yaygın terminolojisiyle söyleyelim. Veri=Eğitme + Test olarak iki kısma ayrılsın. Bu durumda $\lambda(w^*)$ nın test verisini de doğru olarak predikte etmesini bekleriz. Kuşkusuz, bunun olabilmesi için ağın predikte edeceği verinin de aynı ν_j dağılımından çekilmiş olması gerekir.

2.5 Ağın Örneklem Üzerinden Optimal Performansı

Gözlenen nicelikler (**X** ve **Y**) ve ağın herhangi bir durumu (**w** vektörü) için ağın performansını ölçen L kayıp fonksiyonu verilsin. Formal olarak $L:R^{p+r}\times W\rightarrow R$ dir. Öğrenmeyi yeniden şöyle tanımlayabiliriz: Ağ durumunu öylesine ayarlanmalıdır ki (yani öyle bir **w** vektörü bulmalı ki) risk (gözlenen nicelikler üzerinden ortalama kayıp fonksiyonu) minimum olsun. Burada gözlenen niceliklere Doğa'nın seçimi de denir. $L(\mathbf{z},\mathbf{w})=\pi(\mathbf{y},f(\mathbf{x},\mathbf{w}))$ seçelim. O zaman ağın amacı $\lambda(\mathbf{w})=\int L(\mathbf{z},\mathbf{w})\nu_j(d\mathbf{x},d\mathbf{y})$ olur. ν_j bilinirse integral minimize edilip $\mathbf{w}^*$ bulunabilir. Ancak ν_j bilinmediğinde öğrenme işlemi gerekli olur. Yani her nekadar ν_j yi bilmesekte $\mathbf{Z}=(\mathbf{X}',\mathbf{Y}')'$ vektörü üzerine gözlemler yapılabilir. Bu yolla ν_j nin ampirik bilgisi elde edilebilir. $\mathbf{z}^n=(\mathbf{z}_1,\ldots,\mathbf{z}_n)'$ örneklemi verildiğinde, ν_n (ν_j nin örneklem analoğu) rasgele niceliği,

$\nu_n=(\#\mathbf{z}_t\in A)/n$, $A\subset R^{p+r}$ yazılır. Burada # sembolü $\mathbf{z}_t$ nin oluşma frekansını verir. n büyük olduğunda $\nu_n(A)$, $\nu_j(A)$ ya yaklaşır.

Tanıt: Herbiri $p=\nu_j(A)$ ortak olasılıklı ve $k=\#(z_t \in A)$ ile tanımlı n deneme üzerinden k tane bağımsız olay düşünelim. Burada $t=1,...,n$ dir ve $s(z_t)=1$ eğer $z_t \in A$ ve $s(z_t)=0$ eğer $z_t \notin A$ olmak koşuluyla $k=\Sigma_{t=1}\, s(z_t \in A)$ olarak yazılabilir. Bu tanımın ışığında $\nu_n(A)$; $P[\nu_n(A)=(k/n)]=[n!/(k!(n-k)!]\{\nu_j(A)\}^k\{1-\nu_j(A)\}^{n-k}$ olasılık dağılımıyla tanımlanan rasgele bir niceliktir. $\nu_n(A)$ ya "ampirik dağılım fonksiyonu" denir ve $B(n,p)$ biçiminde bir binom dağılımına sahiptir. Borel kuvvetli büyük sayılar yasası gereğince $k/n=\nu_n$, ν_j ya olasılık olarak yakınsar. Bu $P[\nu_n \rightarrow \nu_j]=1$ olarak yazılır. Bununla birlikte, birçok istatistik amaçlar için bu yakınsama yetersizdir. ν_n nin yalnızca bir A kümesi için değil bütün kümeler için aynı anda ne olacağına bakılmalıdır. Bu gereksinim Glivenko-Cantelli teoremi ile giderilir. Yani $P[\sup |\nu_n(A)- \nu_j(A)| \rightarrow 0]=1$ $\forall A \subset R^{p+r}$ için. Burada supremum bütün R^{p+r} kümeleri üzerinden alınır ve $\nu_n \rightarrow \nu_j$ (düzgün yakınsar). Glivenko-Cantelli teoreminin ispatı için bkz. örneğin Lo'eve(1977) de bulunabilir ♦

Borel yasası ν_j yi yaklaşık hesaplamak için kullanılabilir. Buradan λ ya yaklaşım yapılabilir.

$\lambda_n(w)=\int L(z,w)\nu_n(dz)=n^{-1}\, \Sigma_t\, L(z_t,w);\ w \in W$

Denklemin sağındaki kesikli toplam, ağın eğitme kümesi üzerinden ortalama performansını verir. Bu durumda $\lambda_n(w)$ kolayca hesaplanabilir ve $\min_{w \in W} \lambda_n(w)$ ilgi odağımız olur. w_n bu problemin çözümünün kökü olsun. w_n ile w^* arasındaki ilişki için genel bir ifade bulunamaz, çünkü w_n sabit bir vektör olmayıp rasgele (random) bir vektördür. Bu rasgele vektörü elde etmek için $\tilde{\nu}_n$ biçiminde bir rasgele değişken (yani ν_n nin kestirimcisi) tanımlarız.

$\tilde{\nu}_n(A)=\#\,(Z_t \in A)/n,\ \forall A \subset R^{p+r}$ ve benzer biçimde $\tilde{\lambda}_n(w)=\int L(z,w)\, \tilde{\nu}_n(dz)=n^{-1}\Sigma_t\, L(Z_t,w)$

Bu tanımların ışığında, bulmaya çalıştığımız $\tilde{w}_n$ rasgele değişkeni

$$\boxed{\min_{w \in W} \tilde{\lambda}_n(w)=n^{-1}\, \Sigma_t\, L(Z_t,w)} \qquad (2.3)$$

probleminin çözümü olarak tanımlanır. Bu durumda, daha önce w_n olarak tanımlanan vektör $\tilde{w}_n$ nin bir gerçeklenimi olur. Sonuç olarak, ilgimizi w_n ile $\tilde{w}_n$ arasındaki ilişkiye yöneltiriz. Özel olarak $L(z,w)=(y-f(x,w)^2/2$ seçelim. Bu durumda denk.(2.3)

$$\min_{w \in W} \tilde{\lambda}_n(w)=n^{-1}\, \Sigma_t\, (Y_t-f(X_t,w))^2/2 \qquad (2.4)$$

olur. Denk (2.4) iyi bilinen nonlineer regresyon problemidir.

2.6 $\hat{w}_n$ nin Asimptotik Davranışı Üzerine

Herhangi bir rasgele değişkende olduğu gibi , $\hat{w}_n$ nin davranışı da onun olasılık dağılımıyla tamamen belirlenir. Ancak sınırlı bir n için bu olasılık dağılımını elde etmek çok zordur ve merkezi limit teoremleri uygulanarak $n^{1/2}$ ($\hat{w}_n$-w^*) nin çok boyutlu bir normal dağılıma uyduğu gösterilebilir (White 1989a). Ayrıca $\hat{w}_n \to w^*$ hemen hemen her yerde olasılık olarak yakınsar. Bunu $\hat{w}_n \to^{a.s.-P} w^*$ olarak yazarız. Bu yakınsamanın ardında yatan temel düşünce şudur. Denk.(2.1) için çözümün tek olduğunu varsayalım. $\hat{\lambda}_n(w)$ verilen herbir **w** için rasgele değişkenlerin ortalaması olduğu için $\{Z_t\}$ yi yöneten P olasılık dağılımına büyük sayılar yasası uygulanabilir. Bu durumda $\hat{\lambda}_n(w) \to \lambda(w)$. a.s.-P olur.

♠<u>Gözlem :</u> (i) $\hat{w}_n$, $\hat{\lambda}_n(w)$ i minimize eder. w^* ise λ yı minimize eder. (ii) $\hat{\lambda}_n(w)$ ve $\lambda(w)$ a.s.-P olarak yakınsar. Bu durumda $\hat{w}_n$, w^* de yakın olmalıdır. Üstelik bu yakınsama düzgündür. Yani $\sup_{w\in W} |\hat{\lambda}_n(w)-\lambda(w)| \to 0$ a.s.-P ♠

Bölüm 5 de $\hat{\lambda}_n(w)$ ile $\lambda(w)$ nın birbirlerine nasıl yakınsadıkları çok sayıda deneysel veri için ayrı ayrı gösterilmiştir. Son olarak $\hat{\lambda}_n(w)$ ile $\lambda(w)$ arasındaki ilişkiye ilişkin White (1989a) tarafından verilen bir teoremden söz etmekte yarar vardır. Her ne kadar kanıtının teknik ayrıntıları bizi pek ilgilendirmesede Bölüm 6 da teoremin sonuçları kullanılacaktır.

Teorem: (White 1989a, Theorem.1)

$(\Omega,\Im,P)$ tam bir olasılık uzayı ve $\{Z_t\}$ ise onun üzerinde tanımlı bağımsız ve özdeş dağılımlı bir dizi olsun. $\{Z_t\}=(Z_t:\Omega\to R^v$, t=1,2,...,) $v\in N=\{1,2,...\}$. $L:R^v x W\to R$ herbir $\mathbf{w} \in W$ için $L(.,\mathbf{w})$ B^v ölçülebilir ve $\forall z\in R^v$ için $L(z,.)$ W üzerinde sürekli olsun. Ayrıca $d:R^v\to R^+$ gibi $\forall w\in W$ için $|L(\mathbf{z},\mathbf{w})|\leq d(\mathbf{z})$ ve $E(d(Z_t))<\infty$ özelliğini sağlayan bir d fonksiyonu bulunsun. Bu durumda verilen $\forall n\in N$ için $\min_{w\in W}\hat{\lambda}_n(w)=n^{-1}\sum_t L(Z_t,w)$ probleminin $\hat{w}_n$ çözümü vardır ve $\hat{w}_n \to W^*$ a.s.-P. dir. Burada $W^*=\{w^*\in W: \lambda(w^*)\leq\lambda(w) \quad \forall w\in W\}$ ve $\lambda(w)\equiv E(L(Z_t,w))$ dır. {Burada $\hat{w}_n \to W^*$ yakınsamasının anlamı şudur: $\inf_{w^*\in W^*}\|\hat{w}_n-w^*\|\to 0$ a.s.-P. dir. $\|.\|$ Euclidean ya da eşdeğeri herhangi bir başka normdur}.

Kanıt: Bkz.(White 1989a).

BÖLÜM 3: MATERYAL VE METOD: SİMÜLASYON ÖNCESİ AYRINTILAR

3.1 Ön Açıklamalar

Bu bölüm simülasyon deneylerinde kullanılan verinin ve hata foksiyonunun betimlenmesine ayrılmıştır. Bölümün organizasyonu şöyledir. Kesim 3.2 de simülasyonda kullanılan veri matrisleri (örneklem matrisleri) tanıtılır. Kesim 3.3 de bu çalışmaya uygun olacak biçimde ağırlık vektörlerinin bir yorumu verilir. Kesim 3.4 de ise bu çalışmada kullanılan minimizasyon algoritmaları ve eğitme periyodunu durdurmak için kullanılan hata kriterleri tartışılır

3.2 Materyal: Veri Matrisleri

3.2.1 Üretilen verinin kaynağı üzerine bazı bilgiler : Korelasyon kuvantizasyonu

Bu çalışmada kullanılan YSA, bir ya da iki input nöronlu input tabakası -değişken sayıda nöronu bulunan gizli tabaka-bir nöronlu output tabakasından oluşur. Kullanılan veri genelllikle 2 ya da 3 boyutlu normal dağılımdan üretilir. Bölüm 5 de normal dağılımdan farklı diğer bazı dağılımlar da kullanıldı. Verilerin üretimi NAG (Numerical Algorithm Group,1990) hazır programları ile gerçekleştirildi.

Çok boyutlu normal bir dağılım μ (ortalama vektörü) ve Σ (kovariyans matrisi) ile tümüyle belirlenir ve N≈ (μ,Σ) ile gösterilir. Verilen bir (μ,Σ) çifti için üretilecek olacak veri matrisi nxN boyutlu **Z** matrisi olsun. Burada N:**Z** matrisini oluşturan değişken sayısı, n ise bu değişkenler üzerine yapılan gözlem sayısıdır. Bu matrisin bilgisayar üretici rutini tarafindan üretilebilmesi için (μ,Σ) parametrelerinin bu rutine önceden bildirilmesi gerekir.

Bu çalışmada $\vec{\mu}$**=0** (sıfir vektörü) ve Σ ise diyagonal elemanlarının hepsi 1 olan simetrik bir matristir. Bu durumda Σ, efektif olarak **R** korelasyon matrisine dönüşür. Yani Σ= **R** olur (bu çalışma için). Bunun nedeni: bu çalışmanın esas ilgisi verinin **R** korelasyon yapısıyla YSA nın veriyi işleme yeteneği arasındaki ilişkiyi ortaya çıkarmak olmasıdır. O nedenle veri üretilmeden önce **R** korelasyon matrisinin bilgisayar rutinine verilmesi gerekir. Ancak normal dağılımın bu rutin tarafindan üretilmesi için yalnızca μ ve Σ nin verilmesi gerekir. Yani **R** matrisi doğrudan bu rutine verilemez. Bununla birlikte Σ nin diagonal elemanları 1 olursa (σ_{ii}=1 i=1,...,N) **R** matrisine indirgenmiş olur. Böylece **R** matrisi dolaylı yoldan bilgisayar rutinine verilimiş olur.

Anlaşma: vari=σ_{ii}=1 ve var$_j$=σ_{jj}=1. O halde r_{ij}=1=σ_{ii} (i=j için) ve r_{ij}=σ_{ij} (i≠j) için. Burada r_{ij} ve σ_{ij} sırasıyla **R** ve Σ matrislerinin ij elemanıdırlar. Not: Tanım olarak r_{ij}=σ_{ij} (vari)$^{1/2}$ (varj)$^{1/2}$ dir.

Bununla birlikte R=Σ eşdeğerliği de problemimizi tümden değil ancak kısmi olarak çözer. Sorunun kaynağı, Σ nın en az pozitif semi-definit olması gerektiğidir (bkz. hipotez Kesim 5.2.3). Bunun için gerek ve yeter koşullardan biri Σ nın bütün asal alt matrislerinin non-negatif determinantları olmasıdır. İki rv.li (rasgele vektörlü) Σ bu koşulu herzaman sağlar, ancak üç ya da daha çok sayıda rv için yukarıdaki anlaşmayı her zaman geçerli kılınabilir ama ilgi duyulan çoğu korelasyon yapıları üretilemez.

Önerme: k>2 için k-boyutlu bir ÇBND (Çok boyutlu normal dağılım) bazı kovariyans yapıları için mevcut değildir.

Kanıt: k=3 olsun. Yani $\Sigma=(\sigma_{ij})$ 3x3 lük simetrik bir matris olsun. Σ nın asal matrislerinden birisi de Σ nın kendisidir. Bu koşul,

det $\Sigma=\sigma_{11}(\sigma_{22}\sigma_{33}-\sigma_{23}^2)-\sigma_{12}(\sigma_{21}\sigma_{33}-\sigma_{31}\sigma_{23})+\sigma_{13}(\sigma_{21}\sigma_{32}-\sigma_{31}\sigma_{22}) \geq 0$ biçiminde açıkça yazılabilir. Basitleştirmelerden sonra,

$\sigma_{11}\sigma_{22}\sigma_{33}+2\sigma_{12}\sigma_{13}\sigma_{23} \geq \sigma_{11}\sigma_{23}^2+\sigma_{12}\sigma_{33}+\sigma_{22}\sigma_{13}$ elde edilir. x,y,z sırasıyla $\sigma_{12},\sigma_{13},\sigma_{23}$ kovariyanslarına karşılık gelen non-negatif korelasyon katsayıları olsunlar. Yalınlaştırılan son eşitsizlik x,y,z cinsinden

$$\sigma_{11}\sigma_{22}\sigma_{33}\,(1+2xyz) \geq \sigma_{11}\sigma_{22}\sigma_{33}\,(x^2+y^2+z^2) \qquad (*)$$

olarak yazılır. (*) eşitsizliği bireysel variyanslardan bağımsızdır. O nedenle biz hepsini 1 olarak atarız. Denk (*) de örneğin x=0 ise y,z yalnızca belirli değerler alabilirler. Örneğin, (x,y,z)=(0,0.8,0.8) üçlüsü (*) koşulunu sağlamaz. Korelasyonlardan herhangi biri 1 ise ötekiler biribirine eşit olmak zorundadır. Örneğin, (x,y,z)=(1,0.6,0.8) yasaktır. Kısaca, her türlü kovariyans matrisleri üretilemez. Bu olanaksızlığa biz, kuvantum fiziğinde bazı niceliklerin kuvantalaşmasına benzeterek ***korelasyon kuvantizasyonu*** adını vereceğiz. Bu kuvantizasyon nedeniyle her türlü korelasyon üçlüsünü oluşturma şansımız olmamakla birlikte, üretilebilenlerin yeterince zengin bir yapı oluşturduğu kanısındayız. Bölüm 4 ve 5 de bu korelasyon yapılarının listeleri yeri geldikçe verilecektir.

3.2.2 Araştırmada kullanılan aktivasyon fonksiyonları: Veri matrislerinin transformasyonu

Hornik ve ark.(1989), tek gizli tabakalı YSA nın output tabakasındaki nöronun aktivasyon fonksiyonunun herhangi bir iyi davranışlı (sürekli, ölçülebilir) bir fonksiyon olabileceğini matematiksel olarak kanıtladılar. Onların önerisi gizli tabaka aktivasyon fonksiyonunun sigmoid, output tabakasındaki fonksiyonun ise afin (lineer + sabit) olması yönünde idi. Ancak bugün uygulamada en yaygın olarak kullanılan output tabakası aktivasyon fonksiyonu sigmoiddir. Bu çalışmada iki tip aktivasyon fonksiyonu kullanıldı.

Aktivasyon fonksiyonu 1 (AF1): Sigmoid (Bölüm 4 de kullanıldı)

Aktivasyon fonksiyonu21 (A21): Afin (Bölüm 5 de kullanıldı)

Tanım : AF1. x , gizli tabaka ya da output tabakasındaki herhangi bir nöronun toplam inputu ise y=1/(1+exp(-x) o nöronun çıkışıdır. Bu fonksiyona sigmoid denir. $y \in (0,1)$

Tanım: AF2: x toplam input olmak üzere y=ax+b (a ve sabitler) çıkışıdır. $y \in (-\infty,+\infty)$

Böylece AF1 li ya da AF2 li YSA nın performanslarının karşılaştırılma olanağı doğar. AF1, bilindği gibi (0,1) aralığında değerler alır (0,1) aralığı hedef vektörünün alacağı değerler olduğundan üretilen verinin hedef vektörünün değer aralığı (0,1) olacak biçimde dönüşüme uğratılmalıdır.

Üretilen veri, **Z** rasgele vektörü (rv) , $A = X_{L=1}^{k}[a_1, b_1]$ kümesinde değerler alır. ($A \subset R^k$ ve X kartezyen çarpımdır). **Z** rv si, fonksiyonuyla dönüşüme uğrar. $\mathbf{Z_t}$ olur. $Z_t \in X_{l=1}^{k}[0,1]$. Pratikte rv lerden çok onların örneklem üzerinden gerçekleşen matrisleriyle ($\underline{\mathbf{Z}}$, $\underline{\mathbf{Z_t}}$) ilgiliyiz. Not: Z rv sini **Z** ve onun örneklem matrisini $\underline{\mathbf{Z}}$ ile gösteceğiz). $\underline{\mathbf{Z_t}}$, $\underline{\mathbf{Z}}$ den şöyle elde edilir. $\underline{\mathbf{Z}}$ nin l. sütunu; X_l ; α_l ve β_l ise X_l nin max. ve min. değerleri olsunlar. Eğer X_{li}, (i=1,2,..,n) , X_l nin I. gözlem (pattern) değeri ise $X^t_{li} = (X_{li}-\beta_l)/(\beta_l-\alpha_l$, $\underline{\mathbf{Z_t}}$ nin l.sütununun i. gözlemdeki değeridir. Burada $X^t_{li} \in [0,1]$ olduğu açıktır. **Z** ile $\mathbf{Z_t}$ arasındaki dönüşüm daha açık yazılabilir.

f: $\mathbf{Z} \rightarrow \mathbf{Z_t}$; $f(Z)=\Lambda Z+\gamma$. Burada Λ(kxk) boyutlu deterministik diyagonal bir matris, γ ise (kx1) boyutlu deterministik bir sütun vektörüdür. Λ nın elemanları (l=1,...,k ve m=1,...,k) $d_{lm}=(1/\beta_l-\alpha_l)\,\delta_{lm}$ (δ: Kronecker deltası) gibi yazılır. γ nın bileşenleri $\gamma_l = -\beta_l/(\beta_l-\alpha_l)$ ile verilir. Böylece l. bileşen için $(\mathbf{Z_t})_l = d_{ll}\,\mathbf{Z}_l + \gamma_l$ elde edilir.

♠**Tranformasyonun (dönüşümün) nedenleri:**

(i) AF1, (0,1) aralığında değerler aldığı için X_k=Y (hedef vektörü) nın herbir bileşeninin bu aralıkta değerler alacak biçimde dönüştürülmesi gerekir. Yukarıdaki dönüşüm bunu sağlar. Bununla birlikte biz input vektörlerini de aynı dönüşüme uğratmayı uygun bulduk.

(ii) **Z** ve $\mathbf{Z_t}$ nin korelasyon matrisleri özdeştir. Bizim ilgimiz korelasyon matrisi üzerine olduğu için **Z** yerine $\mathbf{Z_t}$ alınmakla birşey yitirilmiş olmaz.

(iii) Afin dönüşüm altında normal dağılım korunur. Yani $\mathbf{Z_t}$ rv'si de çok boyutlu bir normal dağılıma sahiptir (bu istatistikte çok iyi bilinen bir teorem olduğu için ispatı verilmeyecektir)

3.3 Ağırlık Vektörü

Bu kesim amacımıza uygun w ağırlık vektörünü tanıtır. N_s bağlantı sayısı ise, w N_s- boyutlu bir vektördür. İnput tabakası r nöronlu ise, w^1_{ij} gizli tabakanın j.nöronundan input tabakasının i. nöronlarına giden rx1 boyutlu bir vektördür. Burada '1' üst indisi input tabakasını simgeler. Eğer gizli tabakada n_h tane nöron varsa w^1 rxn_h boyutlu bir vektördür. Benzer biçimde w^2_{ij} vektörü tanımlanır. Bu vektör gizli tabakasını output tabakasına bağlayan nöronlar arasındaki ağırlık bileşenlerinden oluşur ve $n_h xp$ boyutludur (p:output tabakasındaki nöron sayısıdır). Benzer biçimde b^1_j ve b^2_j bias vektörleri tanımlanır. Bu çalışmada p=1 alındı. Böylece N_s= $(rxn_h+ n_h x1)+ (n_h +1)=n_h (r+2)+1$ adet bileşeni vardır. Son olarak nod vektörü tanımı yapalım.

Tanım : <u>Nod vektörü</u>: $n_v=(r,n_h,p)$. r-input nöronu sayısı, n_h- gizli tabaka nöronu sayısı, p-output tabakası nöronu sayısı.

3.4 Beklenen Performans (Hata) Fonksiyonunu Minimize Etme Metotları

Stokastik bir çevrede YSA nın bir örneklem üzerinden hedefinin $\min_{w\in W} \lambda_n$ (w) olduğu belirtilmişti. Bu problemi çözebilecek herhangi bir metot bizim için uygundur. Problemin non-lineer karakteri yüzünden yaklaşımlar çok çeşitli ve son derece boldur. Öteki metotların yanı sıra, araştırmacılar tarafindan kullanılan metotlar şunlardır.

(i) Simulated annealing (Kirkpatrick ve ark. 1983, Cerny 1985)

(ii) Genetik algoritma (Holland 1975, Davis 1987)

(iii) Çoklu başlama metotları (Rinney Kan ve ark. 1985)

Ancak biz bu olası metotları bir yana bırakıp geriye yayılma (GY) tipi metotlarla ilgileneceğiz. Bunun nedeni (i), (ii) ve (iii) metotlarının hepsinin de son derece karmaşık ve zahmetli oluşlarıdır. Aslında bu metotlar bile optimizasyon gibi devesa bir alanın yalnızca sınırlı bir bölgesini kapsarlar. Bu çalışmada kullanılan GY metotları şunlardır.

<u>Metot 1</u>: Standart Geriye Yayma Metodu (SGYM)

<u>Metot 2</u>: Genişletilmiş Delta Bar Delta Metodu (GDBDM)

3.4.1 Hata fonksiyonunu minimize etmek için bu çalışmada kullanılan metotlar

(i) SGYM

Bu metodun adımları aşağıdaki gibidir.

(i) Bir input değeri (n tane gözlemden biri) input tabakası nöronlarına sunulur.

(ii) Bu tabakanın çıkışı bir sonraki tabakaya (gizli tabaka) giriş olarak girer. Gizli tabakasından çıkış ise output tabakasına giriş olarak girer.

(iii) İnput tabakasındaki dışındaki nöronlardan j. nöron kendine ulaşan bütün o_i inputların ağırlıklı bir toplamını (i_j) hesaplar.

$i_j=\sum_i w_{ji}\, o_i + b_j$ (b_j, j. nörona gelen bias ağırlık bileşenidir) (3.1)

w_{ji} ise j. ve i. nöronlar arasındaki ağırlık bileşenidir. Burada o_j terimi her nekadar output olarak kullanılsada bu terimin anlamı jeneriktir. Yani o_j j.nörondan k. tabakaya outputu simgeleyebilir ama aynı zamanda k+1. tabakaya ise bir inputu simgeler. Örneğin denk.(3.1) de o_i ler bir önceki tabakadan gelen output sinyalleri olabilir ama aynı zamanda o_i ler bir sonraki tabakanın j. nöronuna input sinyalleri olurlar. o_j nin bu dual karakterini akılda tutarak aktivasyon fonksiyonuna dönelim. Aktivasyon fonksiyonu (a.f.) herhangi bir iyi davranışlı fonksiyon olabilsede bu çalışmada kullanılan a.f. (dolayısıyla j.nöronun outputu)

$o_j=f(i_j)=1/(1+\exp(-i_j)$ (3.2)

ile verilir. Outputlar bir tabakadan bir sonraki tabakaya olacak biçimde input tabakasından output tabakasına gittikleri ve bir daha geriye hiç dönmedikleri için bu YSA'ya ileri beslemeli YSA denir.

Eğitme (yani YSA ya spesifik bir input verildiğinde spesifik hedef vektörünü bulmayı öğtertme) başlamadan önce bütün w_{ji} ağırlık vektörü bileşenleri rasgele olarak atanırlar (genellikle (-1,1) arasında olacak biçimde). Durum böyle olunca output tabakasının output değerleri başlangıçta, istenilen hedef değerlerinden farklı olacaktır. İşte output tabakasındaki bu output değerleriyle YSA nın bulması istenilen hedef değerleri arasındaki bu fark (*genel olarak hata*) geriye yayılma algoritmasında (öğrenmesinde) ağırlıkları modifiye etmenin temelini oluşturur Bu *modifikasyonun aşamaları* şunlardır.

(i) Bütün gözlem değerleri (eğitme kümesindeki elemanlar) YSA'ya sunulduktan sonra (inputlar input tabakasına hedef değerleri ise output tabakasına olacak biçimde) λ ile simgelenen kümülatif hata (denk.(3.3)) hesaplanır.

$\lambda=(\sum_m\sum_j(y_{mj}-o_{mj})^2)/2$ (m=1,2,...,n; j=1,2,...,p) (3.3)

n eğitme kümesindeki eleman sayısı (yani input için n gözlem ve hedef için n gözlem) ve p ise output tabakasındaki nöron sayısıdır. y_{mj} ve o_{mj} ise sırasıyla m.gözlem için istenilen output (hedef) ve output tabakasındaki j. nöronun hesaplanan outputudur. Eğer λ_m m.gözlem için λ değerine yapılan katkıyı simgelerse k. iterasyonda (ağırlık vektörlerini değiştirme sayısı) w_{ji} bileşenindeki değişme

$\Delta w_{ji}= -\eta[\partial\lambda_m(k)/ \partial w_{ji}(k)]$ (3.4a) ile verilir. Burada η öğrenme adımı olarak bilinen bir sabittir. Denk.(3.4a) daha açık olarak denk. (3.4b olarak yazılır).

$$w_{ji}(k+1)-w_{ji}(k)= -\eta[\partial\lambda_m(k)/ \partial w_{ji}(k)] \qquad (3.4b)$$

Bu kısmi türev aşağıdaki gibi hesaplanabilir. Zincir kuralından

$$(\partial\lambda_m/ \partial w_{ji})=(\partial\lambda_m/\partial i_{mj}) (\partial i_{mj}/\partial w_{ji}) \qquad (3.5)$$

burada k indisi notasyonu sadeleştirmek için ihmal edilmiştir. Şimdi, $-(\partial\lambda_m/\partial i_{mj})=\delta_{mj}$ tanımı ile

$$-(\partial\lambda_m/ \partial w_{ji})= \delta_{mj}\, o_{mi} \qquad (3.6)$$

olur. Böylece k.iterasyonda w_{ji} deki değişme

$$\Delta w_{ji}(k)=\eta\, \delta_{mj}\, o_{mi} \qquad (3.7)$$

olur. Denk.(3.7) ye momentum denilen bir başka terim daha eklenebilir, bu terim ıraksak osilasyonları engellemeye yarar. Bu terimin eklenmesiyle denk.(3.7) denk.(3.8) e dönüşür.

$$\Delta w_{ji}(k)=\eta\, \delta_{mj}\, o_{mi} + \mu w_{ji}\, (k-1) \qquad (3.8)$$

burada μ momentum sabiti (ya da kısaca momentum) olarak bilinir. $-(\partial\lambda_m/\partial i_{mj})=\delta_{mj}$ ile tanımlanan δ_{mj} kısmi türev terimi bilgisayarda uygulanmak için bu haliyle uygun değildir. O nedenle ilk önce onu bilgisayra uygulanabilecek forma getirmeliyiz. δ_{mj} için analitik bir ifade

$$\delta_{mj}=-(\partial\lambda_m/\partial i_{mj})= (\partial\lambda_m/\partial o_{mj})\, (\partial o_{mj}/\partial i_{ji}) \qquad (3.9)$$

olduğuna dikkat edilerek elde edilebilir. Denk.(3.9)'un sağ yanı denk.(3.2) ve (3.3) kullanılarak elde edilir. Sonuç denk.(3.10) olarak bulunur.

$$\delta_{mj}=(y_{mj}-o_{mj})\, f'(i_{mj}) \quad (' \text{ türev işaretidir}). \qquad (3.10)$$

Denk.(3.10) yalnızca output tabakası nöronları için geçerlidir. Eğer j. nöron gizli tabakaka nöronu ise (o zaman belirli bir hedef yoktur) δ_{mj} denk.(3.11) ile verilir.

$$\delta_{mj}=f'(i_{mj})\, \Sigma_z\, \delta_{mz}\, w_{zj} \qquad (3.11)$$

burada z, j. nöronun outputunun geldiği nöron sayısını gösterir. Denk.(3.11) den görülür ki gizli tabakadaki hata δ_{mj} bir sonraki tabakadaki nöronların δ_{mz} hatasıyla orantılıdır. Böylece önce denk.(3.10) daki hata hesaplanmalı daha sonra bir önceki tabakanın ağırlıklarını değiştirmek üzere (3.11) denkleminde kullanılmalıdır. Yani hata geriye doğru yayılmalıdır. İşte bu nedenden dolayı bu YSA algoritmasına *Geriye Yayma Algoritması* denir (Rumelhart ve McClelland 1986). Eğitme periyodu, daha önceden belirlenen bir iterasayon sayısına kadar ya da denk.(3.3) ile verilen (toplam) hata kabul edilebilir bir düzeye düşünceye dek sürer. Bu konuyla ilgili biraz daha bilgi bu bölümün sonundaki *hata kriteri* kesiminde verilecektir.

(ii) GDBDM

SGYM'nin çok yaygın olarak kulllanılmasına karşın birçok istenilmeyen yönleri verdır. Bunlardan belki de en önemlisi $\lambda_n(w)$ fonksiyonunu minimize ederken itersayon sürecinin çok yavaş işlemesi ve dolayısıyla istenilen doğrulukta bir hata düzeyine erişmek için kullanıcının uzun bir zaman beklemek zorunda kalamasıdır. Bunun nedeni, Minai ve Williams (1990) tarafından da vurgulandığı gibi SGYM'nin yalnızca lokal gradiyent kullanması nedeniyle, W-ağırlık uzayının istenmeyen bölgelerine sıçramayı engellemek için η öğrenme adımının çok küçük seçimesi gerektiğidir. Oysa, bu durum süreci önemli ölçüde yavaşlatır. Ağırlık uzayının uzun ve düzgün (plato) bölgelerinde bu durum özellikle daha belirgin hale gelir. Böylece, sorun biribiriyle çelişkili gibi gözüken iki kavaram (karalılık ve verimlilik) arasında nasıl uzlaşma sağlanacağıdır. Yani, kullanıcı tarafından saptanan parametrelerle daha hızlı yakınsama nasıl sağlanır. İşte Minai ve Williams (1990) tarafından geliştirilen GDBDM, SGYM'nin istenilmeyen yönlerini önemli sayılabilcecek ölçüde iyileştirmek için bir alternatiftir. Ancak GDBDM' den söz etmeden önce onun kaynağı olan ve Jakobs (1988) tarafından geliştirilen Delta bar delta metodu (DBDM)ndan söz edilmelidir. DBDM özünde standart olmayan bir GY metodudur. DBDM aşağıdaki modifikasyonları yapar.

(m_1): Herbir ağırlık bileşeninin kendi öğrenme adımı η vardır.

(m_2): Bu ağırlıklar hata yüzeyi bilgisine dayanarak modifiye edilir.

(m_3): λ nın w_{ji} kısmi türevi birçok iterasyon için aynı işarete sahipse bu ağırlığın η sı artırılır, çünkü bu ileride bulunan bir minimuma doğru gidildiğine işaret eder

(m_4): Eğer kısmı türevler birçok peşpeşe adımda işaret değiştirirse, η azaltılır, çünkü bu durum bir minimumun üzerinden atlanmış olduğuna işaret eder. Bu gözlem ışığında, η aşağıdaki gibi modifiye edilir.

$$\eta_{ij}(k+1)=\eta_{ij}(k)+\Delta\eta_{ij}(k) \tag{3.12}$$

$$\Delta\eta_{ij}(k)=\kappa \quad \text{eğer} \quad \bar{\delta}_{ij}(k-1)\delta_{ij}(k) > 0 \tag{3.13a}$$

$$= -\phi\eta_{ij}(k) \quad \text{eğer} \quad \bar{\delta}_{ij}(k-1)\delta_{ij}(k) < 0 \tag{3.13b}$$

$$= 0 \quad \text{öteki durumlarda} \tag{3.13c}$$

Burada $\eta_{ij}(k)$ k.iterasyonda , peşpeşe iki tabaka arasındaki i. ve j. nöronlar arasındaki öğrenme adımı, $\delta_{ij}(k)$ ise k anında λ nın w_{ji} ye göre kısmı türevidir. κ ve ϕ kullanıcı tarafından atanan parametrelerdir. $\bar{\delta}_{ij}(k)$ niceliği ise gradiyent değerlerinin üstel olarak azalan bir miktarı olup

$$\bar{\delta}_{ij}(k) = (1-\theta)\,\delta_{ij} + \theta\,\bar{\delta}_{ij}(k-1) \tag{3.14}$$

ile verilir. Burada θ kullanıcı parametresidir. Bu modifikasyonlarla temel öğrenme denklemi

$$w_{ji}(k+1)=w_{ji}-\eta_{ji}(k)\,(\partial\lambda(k)/\partial w_{ji}(k)) \tag{3.15}$$

ile verilir. DBDM birçok test problemlerinde denendi ve yararlı bulundu. Ancak onun da limitleri vardır. Bu limitler Minai ve Williams (1990) tarafından şöyle özetlenir. κ küçük olsa bile bazen vahşi sıçramaları önlemek için yeterince küçük değildir. İkinci olarak eğer ϕ artırılırsa, problem oldukça karmaşıklaşır, çünkü hata yüzeyi üzerindeki kötü noktalardan geriye dönmeyi zorlaştırır. Son olarak, DBDM momentum kullanmadığı için verimlilik azalır. Bütün bunları göz önüne alan Minai ve Williams (1990) aşağıdaki genişletmeleri yaptılar.

(g_1): η sabit (κ) olacak yerde $\delta(k)$ nın üstel azalan bir fonksiyonu biçiminde değiştirilir. Bunun anlamı η nın cok düzgün alanlarda hızlı artması ve büyük eğimli yerlerde yavaşlamasıdır.

(g_2) Momentum faktörü algoritmaya eklendi.

(g_3): η ve μ için üst limitler saptandı. Bu durum her ikisini de önemli ölçüde artırma olanağı sağladı. Çünkü şimdi her ikiside sınırlıdır.

(g_4): Algoritmaya hafiza ve geri çağırma eklendi. Böylece, şimdiki iterasyona kadar görülen en iyi sonuç saklandı. Bunun için Tl tolerans parametresi kullanıldı. Eğer $\lambda>$ Tl ise iterasyon süreci zayıflatılmış η ve μ parametreleriyle yeniden başlatıldı. Bu durum yeni bir noktadan başlamayı engellemek için stokastik olarak yapıldı. (Bu çalışmada (g_1-g_3) adımları aynen g_4 adımı ise kısmen izlendi.

GDBDM' in denklemleri aşağıdaki gibidir.

$$w_{ji}(k+1)=w_{ji}-\eta_{ji}(k)\,(\partial\lambda(k)/\partial w_{ji}(k))+\mu_{ji}(k)\Delta w_{ji}(k-1) \tag{3.16}$$

Burada $\mu_{ji}(k)$ k.iterasyonda peşpeşe tabakalardaki j. ve i. nöronların momentum faktörüdür.

$$\eta_{ji}(k+1)=\min\,[\eta_{max},\ \eta_{ji}(k)+\Delta\,\eta_{ji}(k)] \tag{3.17}$$

$$\mu_{ji}(k+1)=\min\,[\mu_{max},\ \mu_{ji}(k)+\Delta\,\mu_{ji}(k)] \tag{3.18}$$

$$\begin{aligned}\Delta\eta_{ji}(k) &= \kappa_l \exp(-\gamma_l\,|\bar{\delta}_{ij}(k)|) && A>0\\ &= -\phi_l\,\mu_{ji}(k) && A<0\\ &= 0 && A=0\end{aligned} \tag{3.19}$$

Burada $A=\bar{\delta}_{ij}(k-1)\delta_{ij}(k)$ olarak tanımlanır.

$$\begin{aligned}\Delta\mu_{ji}(k) &= \kappa_m \exp(-\gamma_m\,|\bar{\delta}_{ij}(k)|) && A>0\\ &= -\phi_m\,\mu_{ji}(k) && A<0\\ &= 0 && A=0\end{aligned} \tag{3.20}$$

3.4.2 Bu çalışmada kullanılan eğitmeyi durdurma kriteri

GDBDM parametrelerinin uygun bir kümesi seçildikten sonra hatanın evrimi düzgün iterasyon aralıklarıyla bilgisayarda izlendi. (100,200, 300 vb gibi). Birçok GDBDM parametre kümesi denendi (her minimizasyon süreci için en az 10 değişik parametre kümesi test edildi) ve hatada en etkin azalmayı sağlayan parametreler kullanılarak gerçek minimizasyon süreci başlatıldı. İterasyonlar için her zaman bir üst limit belirlendi. Bu üst limitin saptanması genel olarak şöyle idi. Eğer iterasyonu artırmak hatada önemli ölçüde bir azalmaya neden olmuyorsa belirli sayıdaki iterasyon üst limit olarak alındı ve YSA nın bu iterasyona kadar minimizasyonu sürdürmesine izin verildi.

BÖLÜM 4 : SONUÇLAR VE TARTIŞMA I : GERİYE YAYMALI STANDART OLMAYAN YAPAY SİNİR AĞLARININ ASİMPTOTİK OLMAYAN DAVRANIŞI

4.1 Ön Açıklamalar

Bu bölüm öz olarak Bölüm 5 deki sonuçlara bir giriş niteliği taşır. Bu bölümde YSA nın asimptotik olmayan ($n \leq 100$) performansını verilen bir korelasyon yapısı için çok sayıda örneklem matrisi (100 tane) üzerinden araştırırız. Bir örnek vermek uygun olacaktır. Belirli bir korelasyon katsayılı **X** (input) ve **Y**(output) değerleri için iki boyutlu normal dağılımdan 100x2 lik (gözlem sayısı x değişken sayısı) 100 adet matris üretilir. Bu üretilen veri matrisler üzerinden ağın ortalama performansı araştırılır. Aynı işlem üç değişkene de genişletilir. Bu denli çok sayıda matris üzerinden işlem yapıldığı için SGYM yerine GDBDM kullanıldı. Bu yeni metot stokastik bir ortamda ilk kez bu çalışma tarafından kullanıldı (Yıldız 1995a) GDBDM'nin SGYM'ye oranla stokastik bir ortamda daha uygun olabileceği sonucuna varıld (Kesim 4.2).

Bu bölümün organizasyonu aşağıdaki gibidir. Kesim 4.2 de SGYM ile GDBDM stokastik bir çevrede karşılaştırılır. Kesim 4.3 de literatürde ilk kez olmak üzere seçme kümesi tanımı yapılır ve bu küme üzerinde YSA'nın stokastik bir veriyi işlemesinde yararlı olabilecek baz fonksiyonlar tanımlanır. Bu tanımların White (1989a) daki teorik değerlerdirmelere bir katkı olduğu inancındayız. Kesim 4.4 de input ve output uzaylarının her ikisinin de bir boyutlu olduğu bir YSA için asimptot altı bölgede AF1 fonksiyonu için stokastik sonuçlar sunulur. Kesim 4.5 de ise input uzayı boyutu=2, output uzayı boyutu=1 olan YSA için asimptot altı bölgede (AF1 için) sonuçları verilir.

4.2 SGYM ile GDBDM nun Stokastik Çevredeki Performanslarının Karşılaştırılması

Bu metotlar hakkında ayrıntılı bilgi Bölüm 3 de sunulduğu için burada yalnızca sonuçlar sunulacaktır. Metotlar YSA nın veriyi işleme (dönüştürme) yeteneği ve YSA nın yakınsama hızları açısından karşılaştırıldı. Eğitme verisi olarak ÇBND kullanıldı. Başlangıç ağırlık vektörleri $[0,1]^{N_s}$ ve $[-1,1]^{N_s}$ uzayından düzgün olasılık dağılımıyla çekildi (N_s W- ağırlık vektörleri uzayı boyutudur). Tablo 4.1 de $n_v=(1,1,1)$ li YSA da $n_g=10$ için SGYM ve GDBDM nun E hata metriğinin ($\sqrt{\frac{\sum_{i=1}^{n}(y_i - o_i)^2}{n}}$) özet istatistiğini gösterir (y hedef ve o YSA outputudur). Tablodan GDBDM'nin E ortalamasının SGYM ortalamasından daha düşük olduğu açıktır. Ayrıca GDBDM nun standart sapması daha küçüktür. Bu durumun nedeni Minai ve Williams (1990) tarafından belirtildiği gibi SGYM daha yüksek bir minimumda çakılıp kalabilirken GDBDM öğrenme adımının dinamikliği nedeniyle bu yüksek minimumdan atlayabilir. Tablo 4.1 den bir başka ilginç sonuç ise her iki metot içinde r_{XY} arttıkça E nin ortalama değerinin azaldığıdır. Bu gözlemin açıklaması kısaca şöyledir (ayrıntılı açıklama Bölüm 5 deki Tablo 5.1 sonuçlarını yorumlarken verilmiştir). r_{XY} arttıkça verinin rasgeleliği azalmaktadır ve hatanın sıfıra yakın olacak şekilde elde edilebilme olasılığı artmaktadır. $r_{XY}=1$ olduğunda ise **X** ile **Y** arasında deterministik bir ilişki olduğundan YSA tarafından çok daha az (teorik olarak sıfır) hata ile dönüştürülmektedirler. Tablo 4.2 ise GDBDM nun SGYM ye oranla E'nin önceden belirlenen bir değere yakınsaması için çok daha az iterasyona gereksinim duyduğunu gösterir.

Tablo 4.1 E (hata metriğinin) çeşitli r_{XY} (**X** inputu ile **Y** hedefi arasındaki korelasyon katsayısı) ortalama değeri ve standart sapması. n_i =100 (aynı r_{XY} için kullanılan veri matrisleri sayısı), n=10. SGYM için η(öğrenme adımı)=μ=0.5 veη(momentum faktörü)=0.9. GDBDM için $\eta_{max}=8.00$, $\mu_{max}=0.2$, $\kappa_l=0.09$, $\varphi_l=0.25$, $\gamma_l=2.00$, $\kappa_m=0.1$, $\varphi_m=0.1$, $\gamma_m=1.0$ ve Tl=1.00.

r_{XY}		0.0	0.2	0.4	0.6	0.8	1.0
Ortalama	SGYM	0.285	0.283	0.269	0.243	0.208	0.130
değer	GDBDM	0.263	0.257	0.248	0.255	0.182	0.118
Standart	SGYM	0.041	0.043	0.051	0.054	0.055	0.062
sapma	GDBDM	0.038	0.043	0.048	0.048	0.047	0.002

Tablo 4.2 n_i=100 üzerinden E için önceden belirlenen değere erişmek için YSA nın gerek duyduğu ortalama iterasyon sayısı (n_v=(1,1,1)).

r_{XY}		0.0	0.2	0.4	0.6	0.8	1.0
<niter>	SGYM	40000	40000	10000	10000	10000	10000
	GDBDM	8000	4000	3630	3040	2970	580

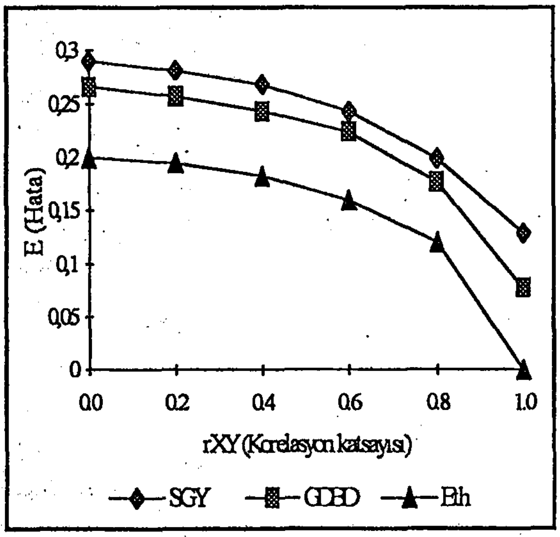
Tablo 4.3 n_v=(1,3,1) YSA için artan n değerleri için (n=20,50,100) SGYM ve GDBDM ile E hata metriği için elde edilen sonuçları gösterir. Tablodan bu kez de GDBDM nun ortalama performansının SGYM performansına oranla daha iyi olduğu açıkça görülür (yani GDBDM nun E değerleri E_{th} teorik hata değerlerine daha yakındır daha düşüktür). {Benzer sonuçların n_v=(1,1,1) ve (1,2,1) YSA için de elde edildiğini not edelim}. Özellikle n=100 değeri için elde edilen sonuçlarda GDBDM değerlerinin teorik hata değeri E_{th}'e oldukça yakın olduğu

belirtilmelidir. $\sum (E_i - E_{th})^2 = 1.408x10^{-6}$ (SGYM için) ve $\sum (E_i - E_{th})^2 = 0.9332x10^{-7}$ (GDBDM için). Burada toplam, değişik bütün r_{XY} üzerinden alınır. Bu sonuç GDBDM nin performansının daha iyi olması anlamında öemlidir ancak dahada önemli sonuç şudur. Tablo 4.3 de elde edilen sonuçlarda GDBDM SGYM'ye oranla 32 kez daha hızlı olarak teorik hata değerine yakınsar. Bunun nedeni ise şudur. Tablo 4.3 den de anlaşılacağı üzere SGYM'nin minimizasyonu gerçekleştirebilmesi için η öğrenme adımı küçük olmalıdır, bu ise yavaş yakınsamaya neden olur. Öte yandan GDBDM için bu durum söz konusu değildir, çünkü η değerleri dinamik olarak değiştirilmektedir. Tablo 4.3 den bir başka önemli sonuç ise her iki metot için de n arttıkça E hata değerlerinin E_{th} teorik hata değerine daha yakın olmalarıdır. {E_{th} hata değeri (denk 5.14b Bölüm 5) den hesaplanan değerlerin kare köküdür}. Bu durum istatistikte büyük sayılar yasasına uygundur. Yani gözlem sayısı (n) arttıkça yaklaşımlar daha iyi olmaktadır.

Tablo 4.3 E (hata metriğinin) çeşitli r_{XY} ve artan n değerleri için ortalama değeri. (n=20,50,100). SGYM için η=0.04 (çok küçük !), η=0.2. GDBDM için değerler iki farkla (η_{max}=6.00, γ_i=0.50). <niter>=12000 (SGYM için) ve =380 (GDBDM için). (n_h=3)

n / r_{XY}	20		50		100		E_{th}
0.0	0.243	0.242	0.195	0.223	0.195	0.204	0.200
0.2	0.243	0.196	0.195	0.220	0.207	0.196	0.195
0.4	0.229	0.224	0.230	0.211	0.211	0.188	0.183
0.6	0.185	0.157	0.180	0.184	0.167	0.157	0.160
0.8	0.119	0.155	0.105	0.145	0.114	0.127	0.120
1.0	0.040	0.059	0.040	0.059	0.012	0.010	0.000

n_h=1 ve 2 değerleri için de iki metot Tablo 4.3 dekine yakın performans gösterdi. Onun için bu değerlere ilişkin görsel veri sunulmadı. Son olarak n_g=10 için SGYM ile GDBDM'nin performansının E_{th} değerine yakınlıkları Şekil 4.1 de görülmektedir. GDBDM'nin SGYM'ye oranla daha iyi performans gösterdiği görülür. Gerek performansının iyi olması ve gerekse yakınsama hızının yüksek olması nedeniyle Bölüm 4 deki sonuçlar için GDBM tercih edildi. Bununla birlikte GDBDM'nin optimum parametrelerinin elde edilmesi yorucu bir çabayı gerektirir, çünkü çok sayıda parametre (toplam 11 adet) vardır.



Şekil 4.1 SGYM ve GDBDM'nin elde edilen deneysel E (hata) değerlerinin E_{th} teorik hata değerleriyle karşılaştırılması. (n=10 ve n_h=2). <niter>=12000 (SGYM için) ve GDBDM için ise =380 dir. n=10 değeri aslında çok küçük olamsına karşın her iki metodun da E_{th} değerlerine belirli bir oranda yakınlaştıkları görülmektedir. Ayrıca r_{XY} değeri arttıkça E değeri parabolik olarak azalmaktadır. Bu durum denk.5.14b dekine uygundur.
(SGYM ve GDBM'nin parametreleri Tablo 4.1 de kullanılana benzerdir)

4.3 Seçme Kümesi ve Bu Küme Üzerinde Tanımlanan Bazı Yararlı Fonksiyonlar

YSA nın stokastik bir ortamdaki performansı araştırılırken YSA parametrelerinin sayısı (gizli tabaka nöronları sayısı), veri matrisinin bileşenleri (input ve hedef) üzerine yapılan gözlem sayısı vb. gibi çok sayıda konu göz önüne alınmak zorundadır. Kısaca problem çok yönlüdür. Böylesi bir probleme sistematik bir yaklaşım sunabilmek için bir seçme kümesi tanımlamak uygun olacaktır. Seçme kümesi kavramı YSA literatürüne ilk kez bu çalışma tarafindan getirildi.

4.3.1 Ana seçme kümesi

Bir C seçme kümesi aşağıdaki seçme kümelerinin kartezyen çarpımıdır.

C= NHxNxRxIW

Burada NH, gizli tabaka nöronları sayısı kümesi, N, bir veri matrisindeki gözlem sayıları (yani input ve hedef için kullanılan eğitme ya da test kümesindeki eleman sayısı) kümesi, R, input ile hedef (output) vektörleri arasındaki korelasyon katsayılarının kümesi ve IW ise verilen bir veri matrisi için eğitme sürecinin başlangıç koşullarının kümesidir (yani W-ağırlık uzayında başlanan noktalarının indis kümesi). Bu bölümde kullanılan C ana seçim kümesinin elemanları Tablo 4.4 deki gibidir.

Tablo 4.4 C ana seçim kümesinin bileşenlerinin aldığı değerler

NH	{1,2,3,4}
N	{10,20,50,100}
R	{0.0,0.2,0.4,0.6,1.0}
IW	{1,2,3,4}

Tabl 4.4 de NH,N ve IW kümelerinin birden fazla inputu bulunan YSA için de geçerli olduğuna dikkat edilmelidir. Bununla birlikte, iki input ve bir hedef vektörlü YSA için R'nin elemanları üçlüler olacaktır.

4.3.2 Seçme kümesi tanımlı bazı rel değerli fonksiyonlar

F(C,R) C' den R'ye (reel sayılar kümesi) tanımlı bütün fonksiyonların kümesi olsun. Ayrıca I ise verilen bir r_{XY} için ağı eğitmede kullanılacak veri (örneklem) matrislerinin sayısı olsun. Bizim örneğimizde I={1,2,...,100} olacaktır. $\{\lambda_i\}_{i\in I}$ indisli ailesini aşağıdaki gibi tanımlayalım. λ fonksiyonu şöyle tanımlansın. λ:I$\rightarrow$F(C,R). Öyleyse λ_i fonksiyonu λ_i:$C\rightarrow$R dir. Yani λ_i $\in$

$R(\lambda)$ dır. Burada $R(\lambda)$, λ nın değer kümesini simgeler. $R(\lambda) \subset F(C,R)$ olduğu için $\lambda_i \in F(C,R)$ dır. Bu durunmda λ_i C üzerinde tanımlı bir operator gibidir. Şimdi, verilen bir $c \in C$ için,λ_i açıkça şöyle tanımlanabilir. $\lambda_i(c)$=Verilen bir i ve c ve w ağırlık vektörü için veri kümesi üzerinden hata ortalamasının kare kökü $E_n=(\sum (E_k)^2)^{1/2}$, $k=1,\ldots,n$)). Burada E_n , c ve w ya bağlıdır ve $E_n(c,w)$ ile gösterilir. Bu E_n (c,w) fonksiyonu Bölüm 2 de White (1989a) tarafından tanımlanan $\lambda(\mathbf{w})$ fonksiyonuyla esas olarak aynıdır. Bölüm 2 de c ve i ye açık olarak referans verilmemektedir. Bununla birlikte, bu tezde kullanılan bu yeni notasyon sabit bir ağ kompleksitesi (n_h=sabit) varsaymadan YSA veri işlemesinin farklı alanlarından özgür seçmeler yapma olanağı sunar.

Yukarıda özel bir i için tanımlanan λ_i yerine ($i=1,2,,,,n_i$) için tanımlanan $\lambda_{n_i}(c) = \frac{1}{n_i}\sum_{i=1}^{n_i}\lambda_i(c)$, $\lambda^s{}_{n_i}(c) = \{\frac{1}{n_i - 1})\sum_{;i=1}^{n_i}(\lambda_i(c) - \overline{\lambda}_{n_i}(c))^2\}^{1/2}$ fonksiyonları tanımlamak daha uygundur. Burada $\overline{\lambda}_{n_i}$ ortalama hata , λ^s_{ni} ise n_i tane veri matrisi üzerinden hatanın standart sapması olarak bilinir. $F(C,R)$ R üzerinde bir vektör uzayı oluşturduğu için $\overline{\lambda}_{n_i}$ ve λ^s_{ni} fonksiyonlarının her ikisi de bu uzayın elemanıdırlar. Böylece $F(C,R)$ nin bu tezin amaçları için uygun olduğu görülür. Bununla birlikte *C* kümesi amaçlarımız için fazla büyüktür, o nedenle *C*'nin alt kümeleriyle çalışmak uygun olacaktır. Bunları C_{ij} ile gösterceğiz.

Tanım: $C_{kl}=\{k\} \times N \times R \times \{l\}$; $k \in \{1,2,3,4\}$ ve $l \in \{1,2,3,4\}$

4.4 Bir İnput Nöronlu -Bir Output Nöronlu YSA İçin <u>Asimptot Altı</u> Sonuçlar.(C_{kl} $k \in \{1,2,3,4\}$ ve $l \in \{1,2,3,4\}$)

Bu kesimde input ve output tabakası birer nörondan oluşan YSA için asimtot altı bölgede bazı sonuçlar sunulacaktır . Gizli tabaka nöronları k=1,2,3,4 olarak değişti ve verilen bir $r_{XY} \in R$ için ağın W- uzayında eğitimin başlatıldığı başlangıç vektörlerinin seçimi iw=1,2,3,4 ile indislenecek biçimde YSA eğitimi yapıldı. Gerek bu kesim ve gerekse tezin bundan sonraki tüm kesimleri için yazar tarafından FORTRAN 77 programlama dilinde yazılan bilgisayar programı stokastik verinin YSA tarafından işlenmesinde kullanıldı (bkz. Ek: PROGRAM BACKPROPAGATION). Bölüm 4 de GDBDM kullanıldı. Bunun nedeni, verilen bir $c \in C_{ij}$ için çok sayıda (100 adet) veri matrisinin işlenecek olması ve SGYM'nin düşük yakınsama hızı nedeniyle bu amaç için uygun olmamasıdır.

Kullanılan veri matrislerinin hepsi 2xn lik matrisler olup (n=10,20,50,100), veri matrisinin sütunları (X:input ve Y: hedef) 2 boyutlu standart normal dağılıma sahiptirler. Daha sonra bu X ve Y Kesim 5.3.1 de anlatıldığı gibi $\mathbf{X}_t$ ve $\mathbf{Y}_t$ bileşenlerine dönüştürüldüler. Bu dönüşüm sonunda $\mathbf{X}_t \in [0,1]$ ve $\mathbf{Y}_t \in [0,1]$ oldu. Dönüşüm afin olduğu için $r_{XY} = r_{X_tY_t}$ dir ve bu nedenle X ile Y' nin arasındaki korelasyon katsayısı için her zaman r_{XY} notasyonu kullanıldı.

Eğitmeye başlamadan önce YSA nın minimizasyonu gerçekleştirmesi için belli bir iterasyon sayısı belirlendi. Bu iterasyon sayısından sonra eğitme sona erdirildi. {Aslında bu iterasyon sayısının belilenmesi için önce çok sayıda simülasyon yapıldı ve daha sonra iterasyon sayısına karar verildi}. YSA nın ortalama hata (performans) fonksiyonu $\overline{\lambda}_{n_i}$ denk.(4.1) deki gibi tanımlandı.

$$\overline{\lambda}_{n_i} = \frac{1}{n_i}\sum_{i=1}^{n_i}\lambda_i; \quad \lambda_i = \{\frac{1}{n}\sum_{t=1}^{n}(o_{it} - y_{it})^2\}^{1/2} \quad (i=1,2,...,n_i) \tag{4.1}$$

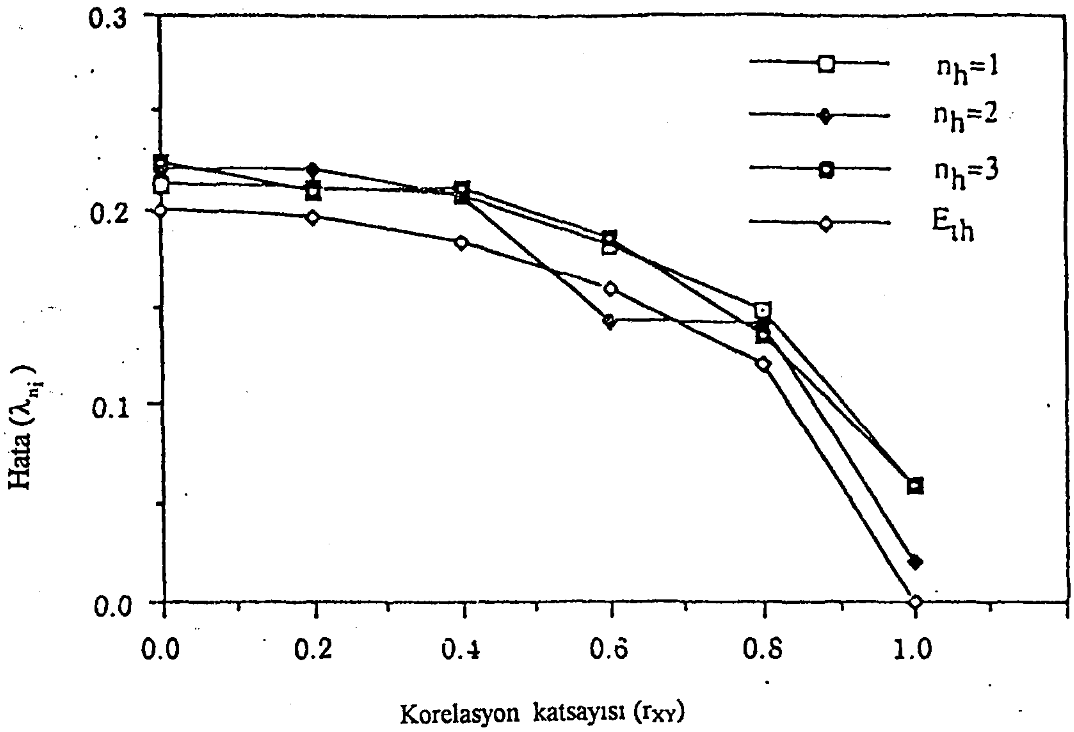
n_i=100 dür. *Kullanılan notasyon* için kısa bir açıklama yapmak yararlı olacaktır. Bu tez boyunca gizli tabaka nöronu sayısı için kullanılan pratik ve açıklayıcı notasyon n_h tır, ancak Kesim 4.3 de tanımlanan C_{kl} kümesinde n_h yerine k alındı, bu yalnızca matematiksel tanım için geçerli olup veri sunumunda ve her yerde n_h kullanılacaktır.

4.4.1 Küçük sayılı iterasyonlarla YSA nın dönüştürme yeteneği

YSA nın dönüştürme yeteneği $\overline{\lambda}_{n_i}$ ortalama hatası ile sıfır hata arasındaki uzaklık olarak tanımlanır (Yıldız 1995b). Yani $\overline{\lambda}_{n_i}$ sıfıra ne kadar yakınsa YSA nın dönüştürme yeteneği o kadar iyidir. Simülasyon deneyleri gösterdi ki , bütün i değerleri için λ_i değerleri biribirlerine oldukça yakın olduğunu gösterdi bu nedenle aslında $\overline{\lambda}_{n_i}$ yerine λ_i değeri alınabilirdi. Ancak Bölüm 4 de biz ortalama hata ya da bazen kısaca hatadan söz ederken $\overline{\lambda}_{n_i}$ notasyonu kullanılacağız.

Çeşitli r_{XY} ve n_h değerleri için $\overline{\lambda}_{n_i}$ hata fonksiyonun değerleri Şekil 4.2 de gösterilmiştir. Şekilden verilen bir n_h için r_{XY} arttıkça YSA nın dönüştürme yeteneğinin kesinlikle arttığı görülmektedir. Ayrıca verilen r_{XY} için YSA nın dönüştürme yeteneği genellikle artan n_h ile artar (n_h=3 için bazı sapmalar vardır). Ancak n_h ile bu artış hiç bir zaman dramatik olmayıp genellikle çok küçüktür.

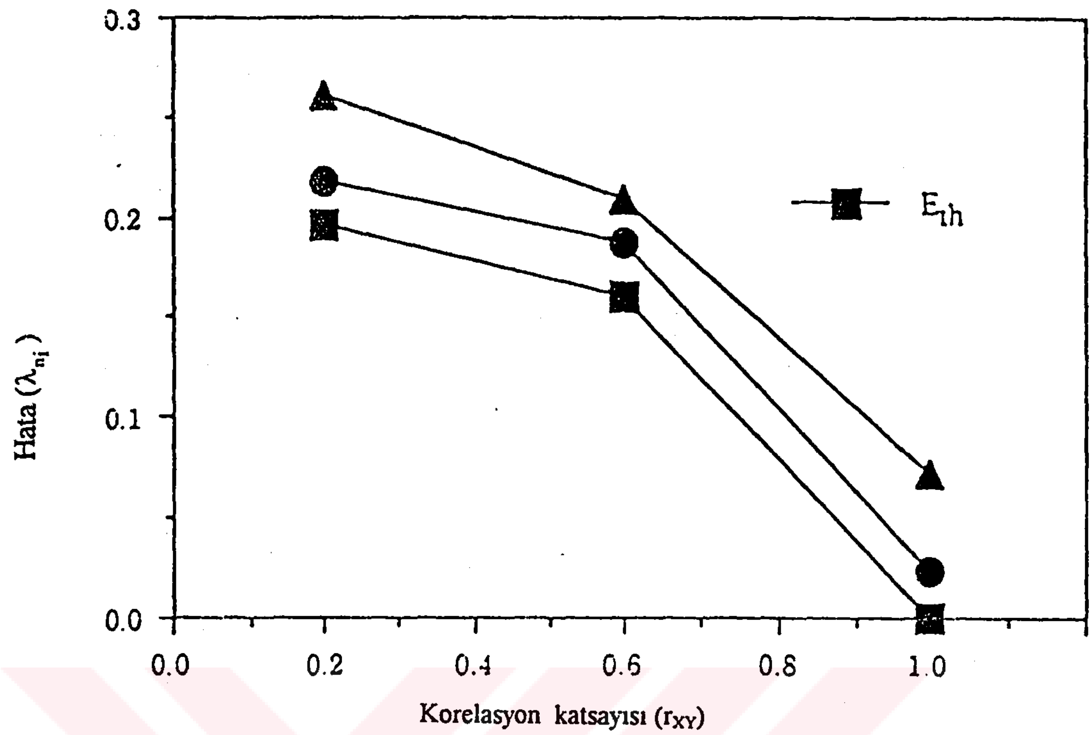
Şekil 4.2 dikkatle analiz edildiğinde söylesi bir sav ortaya atılabilir: Eğer YSA nın eğitim sürecinde yeterince büyük sayıda iterasyonuna izin verilseydi bütün r_{XY} değerleri için de $\overline{\lambda}_{n_i}$ hata fonksiyonun değerleri sıfıra çok yakın olacaktı. Ama Kesim 4.4.2 deki tartışmanın ve sonuçların açıkça gösterdiği gibi bu sav öz olarak yanlıştır.



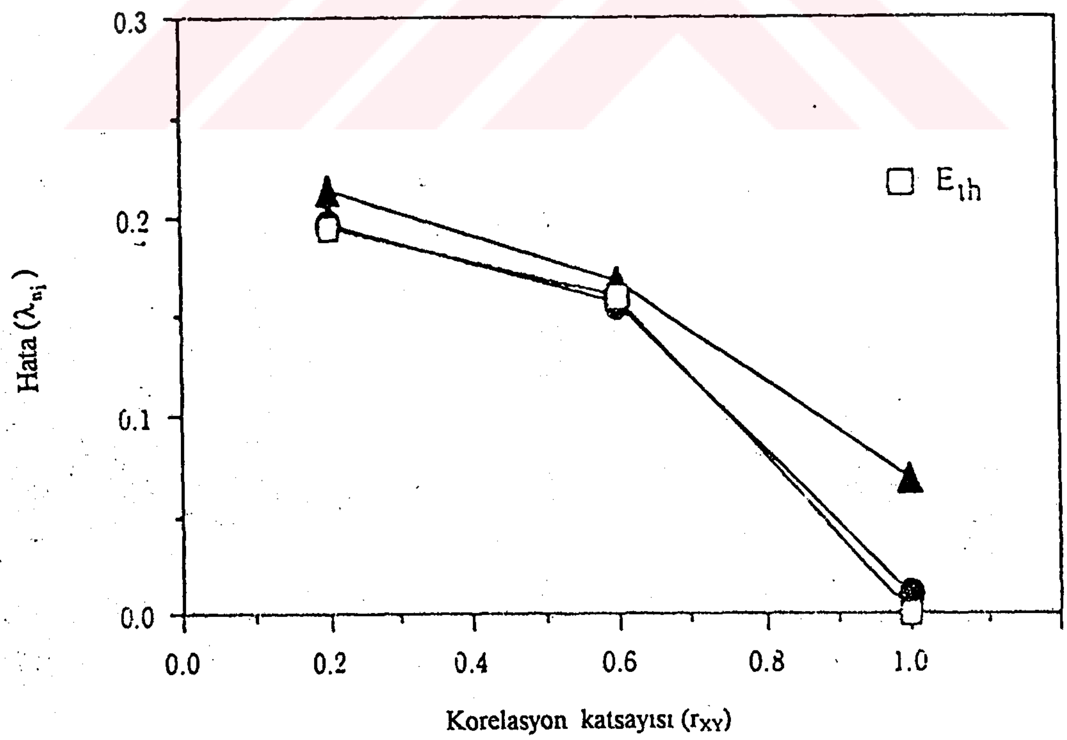
Şekil 4.2 Gözlem sayısı n=50. $\overline{\lambda}_{n_i}$ hata fonksiyonunun farklı n_h değerleri için r_{XY} ye karşı asimptot altı davranışı. (iw=2) Ortalama iterasyon sayısı=380 dir. $E_{th}=\lambda_{th}$ teorik hata değeri denk.(5.14b) den hesaplanmıştır.

4.4.2 Büyük sayılı iterasyonlarla YSA nın dönüştürme yeteneği

Yukarıda öne sürülen savın doğruluğunu test etmek amacıyla YSA nın 10 kez daha fazla iterasyonuna izin verildi. Şekil 4.3a-b n_h=3 için elde edilen $\overline{\lambda}_n$ değerlerini sergiler. Bütün r_{XY} değerleri için YSA nın dönüştürme yeteneğinde belirli bir iyileşme vardır (burada öteki n_h değerleri =1,2 ve 4 içinde benzer iyileşmelerin görüldüğü not edilmelidir). Bununla birlikte r_{XY}=1 durumu dışında bu iyileşmeler yine de dramatik olmaktan uzaktır. Buna şöyle bir açıklama getirilebilir. Denk (2.1) ile tanımlı λ fonksiyonu ya da daha doğrusu $\lambda^{1/2}$ denk(4.1) ile tanımlı λ_i fonksiyonunun popülasyon analogudur. Büyük sayılar yasası gereğince eğer n_g yerince büyükse $\lambda^{1/2}(w^*)$ olasılık olarak $\lambda_i(w_n)$ ye yakındır. Daha da ötesi $\lambda^{1/2}(w^*)$ minimum iken denk.(2.1) deki C_2 değeri sıfıra yakındır. Böylece optimum bir durumda $\lambda^{1/2}(w^*)$ değeri efektif olarak w_n değerine yakındır. Şimdi n=50 ya da 100 değerinin bu yakınlığı sağlayabilmek için yereterince büyük olduğunu varsayarsak denk.(5.14b) den her r_{XY} değeri için $\overline{\lambda}_{n_i}$ hata değerinin farklı olduğunu ve r_{XY} sıfıra yaklaştıkça YSA nın dönüştürme yeteneğinin azalacağını görürüz. Yani iterasyon ne kadar artırılırsa artırılsın YSA nın dönüştürme yeteği r_{XY} ile sınırlıdır.



Şekil 4.3a r_{XY} nin seçilmiş değerleri için iyileştirilmiş dönüştürme yeteneği. (n_h=3, n=10, iw=3, $E_{th}=\lambda_{th}$. Üçgen : Küçük sayılı iterasyon =380, Daire: Büyük sayılı iterasyon ı=4000)

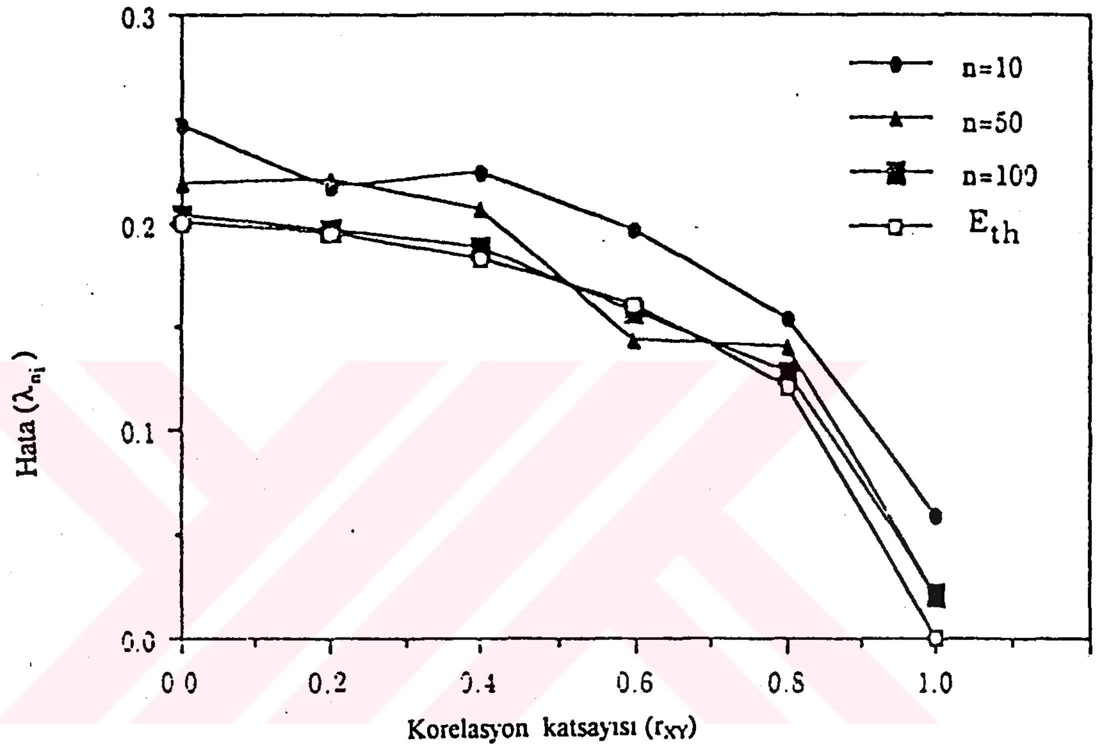


Şekil 4.3b r_{XY} nin seçilmiş değerleri için iyileştirilmiş dönüştürme yeteneği. (n_h=3, n=100, iw=2, Üçgen : Küçük sayılı iterasyon =380, Daire: Büyük sayılı iterasyon =4000)

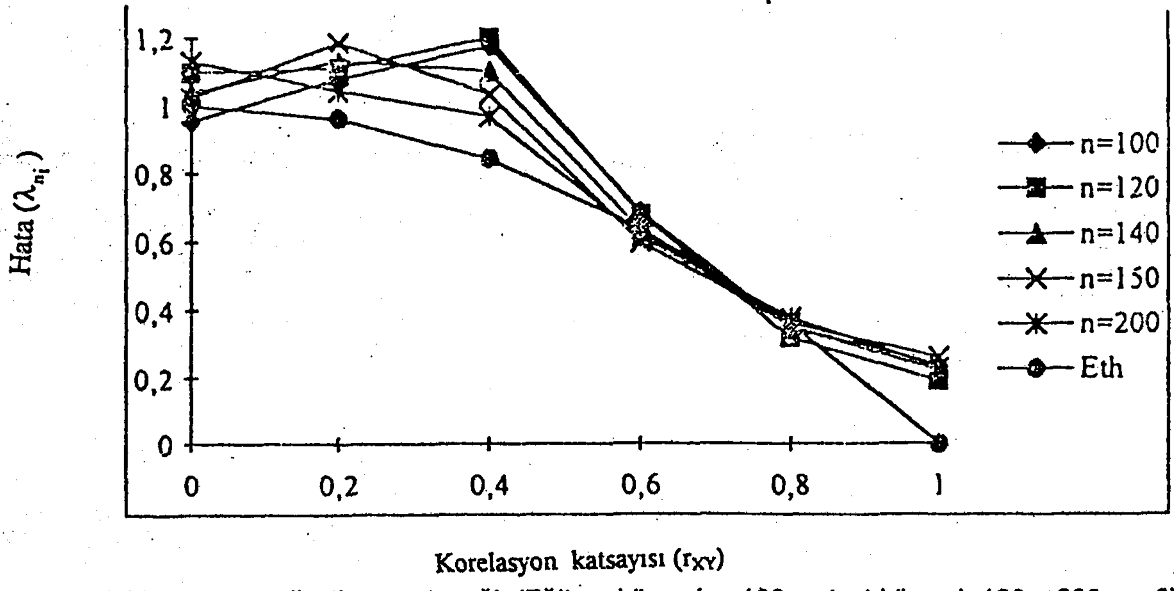
Son olarak n arttıkça ağın dönüştürme yeteneğinin arttığını belirtelim (Şekil 4.4). Büyük sayılar yasası gereğince bu durum hiçde şaşırtıcı sayılmaz. Yani gözlem sayısı arttıkça ağın hatasının teorik hata değerine daha yakın olmasının olasılığı 1 e yakınsar.

4.4.3 YSA nın dönüştürme yeteneğinin test kümesi üzerindeki performansına bir örnek

Şekil 4.5 n=100 üzerine elde edilen final ağırlık vektörü kullanılarak YSA nın n=200'e kadar genişleyen test kümesi üzerindeki performansını sergiler. Şekilden görüleceği gibi bütün r_{XY} değerleri için YSA nın genelleştirme yeteneği kabul edilebilecek ölçüde iyidir. Burada gösterilmeyen çok sayıda örnekle birlikte bu örnek YSA nın stokastik veriyi işlemede iyi bir araç olacağını göstermektedir. Bölüm 5 deki sonuçlar bu konuda daha da fazla kanıt sağlayacaktır.



Şekil 4.4 Artan n değerleri için değişik r_{XY} katsayılarına karşı $\overline{\lambda}_{n_i}$ hata değerlerinin evrimi. (n_h=4).



Korelasyon katsayısı (r_{XY})

Şekil 4.5 YSA nın ~~genelleştirme~~ yeteneği. (Eğitme kümesi n=100 ve test kümesi=120,...200. n_h=3).

4.5 İki İnput Nöronlu-Bir Output Nöronlu YSA için Asimptot Altı Sonuçlar

Tablo 4.5 araştırma için kullanılan (x,y,z) üçlülerinin listesini verir. Burada $x=r(X_1,X_2)$, $y=r(X_1,X_3)$ ve $z=r(X_2,X_3=Y)$ dir (yani bileşenler arasındaki korelasyon katsayılarını gösterirler).

Tablo 4.5 $x=r(X_1,X_2)$, $y=r(X_1,X_3)$ ve $z=r(X_2,X_3=Y)$ (yani bileşenler arasındaki korelasyon katsayıları) üçlülerinin değerleri.

No	(x,y,z)	No	(x,y,z)	No	(x,y,z)
1	(0.0,0.0,0.0)	12	(1.0,0.0,0.0)	23	(1.0,0.6,0.6)
2	(0.0,0.0,0.6)	13	(0.98,0.2,0.35)	24	(0.76,0.76,0.2)
3	(0.0,0.0,0.8)	14	(0.4,0.4,0.0)	25	(0.8,0.8,0.4)
4	(0.0,0.0,1.0)	15	(0.4,0.4,0.4)	26	(0.8,0.8,0.8)
5	(0.4,0.0,0.0)	16	(0.4,0.4,0.8)	27	(0.8,0.8,1.0)
6	(0.4,0.0,0.4)	17	(0.4,0.4,1.0)	28	(1.0,0.55,0.55)
7	(0.4,0.0,0.8)	18	(0.8,0.4,0.0)	29	(1.0,0.8,0.8)
8	(0.35,0.1,0.35	19	(0.8,0.4,0.4)	30	(1.0,0.95,0.95)
9	(0.8,0.0,0.0)	20	(0.8,0.4,0.8)	31	(1.0,0.9,0.9)
10	(0.8,0.0,0.4)	21	(1.0,0.2,0.2)	32	(1.0,1.0,1.0)
11	(0.74,0.1,0.74)	22	(1.0,0.4,0.4)		

Bu YSA nın değişik n_h ve n değerleri altında $\overline{\lambda}_{n_i}$ hatasının asimptot altı davranışı için Tablo 4.6 daki değerler elde edildi. λ_{th} teorik hata değerleri hesaplanması için denk.(5.30 Bölüm 5) kullanıldı. Tablodan görüleceği gibi en düşük hata değerini veren üçlülerin no.su 4,8,11,17,27,30 ve 32 dir. n_h atrtıkça bu hatalar biraz daha azalmaktadır. n=20 için n_h=1 ve n_h=3 değerlerini karşılaştırınız. Yukarıdaki üçlülerin dışında kalanlar için de $\overline{\lambda}_{n_i}$ değerleri λ_{th} teorik hata değerine yakındır. Tablonun son iki sütunu karşılaştırıldığında bu durum kolayca gözükür. Örneğin No=12 için (n_h=3 ve n=50) $|\overline{\lambda}_{n_i}-\lambda_{th}|=0.010$ dur. No=23 için (n_h=3 ve n=50) $|\overline{\lambda}_{n_i}-\lambda_{th}|=0.016$ dır. Bundan başka bir gözlem ise, n arttıkça $|\overline{\lambda}_{n_i}-\lambda_{th}|$ azalma eğilimi göstermesidir. Örneğin No:6 için (n_h=3) n=20 ise $|\overline{\lambda}_{n_i}-\lambda_{th}|=0.030$ n=50 ise $|\overline{\lambda}_{n_i}-\lambda_{th}|=0.020$ dir. No:15 için (n_h=3) n=20 ise $|\overline{\lambda}_{n_i}-\lambda_{th}|=0.024$ n=50 ise $|\overline{\lambda}_{n_i}-\lambda_{th}|=0.014$ dır. Bu konuyla ilgili daha fazla veri Kesim 5.3.5 de verilecektir.

Tablo 4.6 Çeşitli (x,y,z) üçlüleri için artan n_h değerleri (1,2,3) ve artan n değerleri (20,50) için iki input-bir output YSA nın asimptot altı $\overline{\lambda}_{n_i}$ değerleri.

No	n_h=1		n_h=2		n_h=3		λ_{th}
	n=20	n=50	n=20	n=50	n=20	n=50	
1	0.182	0.182	0.234	0.22	0.229	0.225	0.200
3	0.181	0.180	0.162	0.16	0.160	0.16	0.120
4	0.078	0.078	0.072	0.072	0.07	0.07	0.000
6	0.216	0.199	0.215	0.203	0.210	0.20	0.180
8	0.098	0.09	0.095	0.08	0.097	0.078	0.024
11	0.058	0.098	0.098	0.098	0.08	0.090	0.020
12	0.250	0.224	0.246	0.246	0.235	0.210	0.200
15	0.219	0.196	0.207	0.195	0.200	0.190	0.176
17	0.080	0.080	0.070	0.070	0.07	0.070	0.091
20	0.118	0.110	0.112	0.110	0.118	0.107	0.089
23	0.182	0.182	0.180	0.180	0.176	0.176	0.160
26	0.145	0.137	0.133	0.133	0.129	0.119	0.107
27	0.046	0.046	0.040	0.040	0.048	0.048	0.028
28	0.193	0.185	0.209	0.185	0.204	0.170	0.167
30	0.080	0.080	0.080	0.080	0.078	0.078	0.061
31	0.123	0.111	0.116	0.100	0.114	0.106	0.086
32	0.049	0.049	0.040	0.040	0.047	0.047	0.008

BÖLÜM 5 : STOKASTİK SONUÇLAR VE TARTIŞMA II : GERİYE YAYMALI STANDART YAPAY SİNİR AĞLARININ ASİMPTOTİK DAVRANIŞI.

5.1 Ön Açıklamalar

Bölüm 2 de yapay bir ağın genel bir olasılık dağılımından üretilen input-hedef bilgisini nasıl dönüştürdüğünün (map ettiğinin) teorisi istatistiksel bir perspektiften sunuldu. Bu bölümde ise, spesifik bir olasılık dağılımı olan *Çok Boyutlu Normal Dağılım* (ÇBND) ve *normal dağılımdan farklı diğer bazı dağılımlar* (Kesim.6.6) altında $n_v =(r, n_h , p)$ nod vektörlü [r- input, n_h - gizli, p- output tabakasının boyutlarıdır] bir *Geriye Yaymalı Standart Yapay Sinir Ağı* (GYSYSA) nın dönüştürme yeteneğinin matematiksel ve nümerik analizi yapılacaktır. Analiz, Bölüm 4 de elde edilen sonuçların kuramsal değerlendirmelerle ne derece uygunluk içinde olduğunu test etmek için de kullanılacaktır. Kes. 5.5 de görüleceği gibi Bölüm 4 de elde edilen $\overline{\lambda}_{n_i}$ değerleri çoğu korelasyon yapıları için teorik beklentilere uyar. Öte yandan bazı yapılar için bu öngörülerden sapmalar sergiler. Bu durumun birçok nedeni olmakla birlikte bu çalışma açısından en önemli bir kaçı şöyledir.

(i) Verilen bir $i \in I$ için λ_i fonksiyonu bir $c \in C$ için tanımlıdır. $c=(n_h , n, r_p , iw)$ dörtlüsünde n_h = sbt , iw=sbt ,r_p =sbt ve n=değişken olsun. Bu koşullar altında $\lambda_i(c)$ nin $\lambda(w^*)$ ye yakın olması için n sınırsız artmalıdır. Oysa Bölüm 4 de n yeterince büyük olmadığı için $(n \leq 100)$ λ_i nin kuramsal $\lambda(w^*)$ değerine yakın olması gerekmez. Bölüm 4 de sınırlı bir n için araştırma yapılmış olmasının biricik amacı verilen bir nod vektörü ve verilen bir korelasyon yapısı için λ_i nin örneklem matrisleri üzerinden ortalama davranışını öngörmektir. Bununla birlikte Bölüm 5 da i=1 (yani bir tek $\underline{\mathbf{Z}}$ matrisi) alınıp verilen bir n_h , r_p ve iw için n büyük değerlere götürülerek (n=250) ağın asimptotik davranışı incelendi. Böylece White(1989a,1989b) tarafından öngörülen teori test edilme olanağına kavuştu. En son literatür de göz önüne alındığında, GYSYSA nın istatistiksel davranışının ampirik araştırılması anlamında bu çalışma YSA literatüründe ilktir. .

(ii) Bölüm 4 deki sonuçlar GDBDM ile elde edilmişlerdir. Ancak daha önce vurgulandığı gibi (Bkz. Kes 4.2) bu metodun istenilir sonuçlar üretmesi için GDBDM parametrelerinin optimum seçilmesi gerekir. Bu seçimin optimum-altı (suboptimum) olması $\overline{\lambda}_{n_i}$ nin değerlerini etkileyebilir. O nedenle bu bölümdeki (yani Böl.5) sonuçlar yeniden SGYM ile elde edilmişlerdir. Çünkü, burada i=1 alındığı için iterasyon sayısının çokluğu artık temel bir eksiklik değildir. Bölüm 5 deki bir başka yenilik ise output tabakasındaki aktivasyon fonksiyonunun sigmoid yerine afin olmasıdır. Bu fonksiyon kısaca **AF2** dir (bkz. Kesim 3.2.2)

5.1.1 AF2 nin avantajları

AF2 nin avantajı elde edilen $\underline{\mathbf{Z}}$ örneklem matrislerinin sütun elemanlarının [0,1] aralığına dönüşümüne gerek göstermeyip $(-\infty,\infty)$ aralığında değerler almasıdır. Bu ise η nin çok küçük seçilmesi (Bölüm 6 da η=0.05) anlamına gelir. Çünkü, afin fonksiyon ancak küçük adımlı iterasyonlarla denetlenebilir. η nın daha büyük değerleriyle afin fonksiyon kullanarak ağın performansının ne olabileceğine ilişkin bazı deneysel sonuçlar ileride sunulacaktır (Kes.5.4.1). Buna ek olarak verilen herhangi bir korelasyon yapısı için λ_n hata fonksiyonun pratik olarak ne zaman (kaç iterasyon sürecinden sonra) sıfıra yakın sayılabileceği sorusuna deneysel yanıtlar da sağlanacaktır (bkz.Kes.5.4.2).

5.2 Ağın Çok Boyutlu Normal Dağılım (ÇBND) Altındaki Davranışı: Analitik İfadelerin Türetilmesi

Bu kesimde bir GYSYSA için çevre olasılık dağılımı bir ÇBND olduğunda ağın performans fonksiyonu $\lambda(\mathbf{w})$ için analitik bir ifade türetilecektir. Kuşkusuz ÇBND dışındaki dağılımlar da önemlidir. Bu nedenle bizim ilgi odağımız ÇBND üzerine olsada *ÇBND den farklı diğer bazı olasılık dağılımları* için de ağın performansı bu tezde araştırıldı.

5.2.1 Tezde kullanılan dağılımlar (ÇBND ve ÇBND den farklı diğer bazı dağılımlar) üzerine birkaç söz.

Bölüm 2 deki teorik değerlendirmelerden anlaşılacağı gibi GYSYSA'nın en temel işlevi stokastik bir çevrede çevre olasılık dağılımının koşullu beklenen değerinin minimize edilmesidir. Bu nedenle minimize edilecek koşullu beklenen değer (KBD) analitik olarak elde edilebilmelidir. Ayrıca bu tezin yönelimi korelasyon yapılarıyla ağın dönüştürme yeteneğinin ilişkisinin ortaya çıkarılması olduğundan ÇBND bu amaç için özellikle uygundur. Kesim 3.2.1 de belirtildiği gibi ÇBND için Σ kovariyans matrisi (dolayısıyla *R*) korelasyon matrisi önemli bir parametredir. Veri matrisinin üretilebilmesi için *R* gereklidir. O nedenle ÇBND bu gerekliliği yerine getirir. ÇBND nin tercih edilmesi için bir başka neden bilinen bütün olasılık dağılımlarının uygun koşullar altında merkezi limit teoremi gereğince asimptotik olarak ÇBND ye yakınsamasıdır

Ayrıca, ÇBND dışındaki *çok boyutlu diğer dağılımları* elde etmek için geliştirilmiş çok az algoritma vardır. Ronning (1977) bir çeşit çok boyutlu gamma dağılımı elde etmek için bir algoritma sundu. Ancak algoritma oldukça kısıtlı olup yanlızca sabit ve pozitif kovariyans matrisi için geçerlidir. Gerek Ronning(1977) algoritması gerekse öteki bazı algoritmaların hiçbirisi evrensel yaygınlık kazanamamıştır. Bu nedenle bu tezde bu algoritmalar göz önüne alınmadı. Bununla birlikte ÇBND dağılımı kaynak dağılım olarak kullanılıp diğer bazı dağılımlar elde edildi. Ve ağın bu dağılımlar için performansı belirli bir düzeyde incelendi. Bu yolla ağın stokastik ortamdaki performansının incelenmesi konusunda belirli bir düzeyde bir genelliğe ulaşıldığı ileri sürülebilir. Kısaca:

♣ ÇBND limit dağılım olduğundan yeterince geneldir.

♣ ÇBND dışında *çok boyutlu* bir dağılım verisi üretilebilecek bilgisayar rutini yoktur.

♣ Buna karşın, belirli düzeyde bir genelliğe ulaşabilmek için , *bu tezde*, ÇBND den farklı diğer bazı dağılımlardan elde edilen verilerle ağın davranışı araştırıldı (Kesim 5.6).

Merkezi limit teoremi:

Verilen herbir n doğal sayısı için , hepsi aynı dağılıma sahip ve hepsi biribirinden istatistiksel olarak bağımsız olan $X_1, X_2, \ldots, X_n$ rasgele değişkenler verilsin. Ayrıca bu değişkenlerin hersinin ortalama değeri μ_X ve variyansları σ^2_X olsun. Bu durumda $Z_n = \frac{(\overline{X}_n - E(\overline{X}_n))}{\sqrt{\text{var}\,(\overline{X}_n)}} = \frac{\overline{X}_n - \mu_X}{\sigma_X / n}$ ile tanımlanan Z_n rasgele değişkeni dağılım olarak N(0,1)'e yakınsar. Burada N(0,1) ortalaması sıfır ve variyansı 1 olan standart normal dağılımı gösterir.

5.2.2 $\lambda(w)$ beklenen performans (hata) fonksiyonunun bileşenlerine ayrılması

Y rasgele (rastlantısal) hedef vektörü , g(X) rasgele KBD vektörü ve f(X,w) rasgele outputu göstermek üzere [Not: Rasgele fonksiyon (değişken) ve öteki bazı tanımlar için bkz.Kes.5.2.3] ağın beklenen performans fonksiyonu (BPF)

$\lambda(w)=E[Y-g(X)]^2+E\ [f(X,\omega)-g(X)]^2$ ile verilir(Denk.2.2). $\lambda(w)$ görüldüğü gibi iki parçadan oluşur. Bu tezde bu parçalara $\lambda(w)$ nın bileşenleri olarak bakıldı. Yani

$\lambda(w)$ =Kısıtlayıcı bileşen (Bileşen 1) + Serbest bileşen (Bileşen 2)

Bileşen 1(C_1): Ağa sunulan çevre verisi, [$Z'=(X,Y)$ vektörü], ile kısıtlı olup **Z** nin olasılık dağılımı bir kez belirlendikten sonra kullanıcının serbestçe değiştirebileceği parametreler içermez.

Bileşen 2(C_2) : Ağı eğitme sürecinde serbestçe değiştirilebilecek parametreler (yani **w** ağırlık vektörü) içerir. Bir başka deyişle, C_2 ; seçilen bir f fonksiyonuna bağlı olarak ağın dönüştürme yeteneğini (optimum $w=w^*$ bulması anlamında) ölçer. Oysa, C_1 bu yeteneğin dışında kalır. Bu bileşen $\lambda(w^*)$ nın minimumdan ($\lambda_{min}=0$) ne kadar üst tarafta yer alacağını belirler. Bu nedenle C_1 daha yakından incelenmelidir. Bunun için X^1 p-boyutlu hedef ve X^2 r-boyutlu input vektörleri olsunlar. Bu iki vektörü r+p boyutlu bir **Z** vektörü $Z'=[\ X^{1'}, X^{2'}]$ gibi düşünmek uygundur. Bu noktada **Z** nin stokastik (olasılıksal) özelliklerini değerlendirebilmek için, şimdiye dek tanımlanmadan gevşek bir biçimde kullanılan bazı olasılıksal kavramların matematiksel olarak evrensel tanımlarını vermek tezin tamlığı açısından uygun olacaktır.

5.2.3 Bazı stokastik tanımlar

Tanım: Örneklem (sample) uzayı: Sonucu önceden kestirilemeyen rasgele bir deneyin olanaklı bütün sonuçlarının kümesi, Ω.

Tanım: σ(sigma) cebiri: 2^Ω, Ω nın bütün alt kümelerinin oluşturduğu küme (kuvvet kümesi) yi göstermek üzere eğer $\Im \subset 2^\Omega$ aşağıdaki üç koşulu yerine getiriyorsa $\Im$, Ω üzerinde bir σ- cebiridir.

(i) $\Omega\in\Im$ (ii) $A\in \Im \Rightarrow \Omega-A \in\Im$ (iii) $A_1, A_2,... \in\Im \Rightarrow \bigcup_{n=1}^{\infty} A_n \in\Im$

Tanım: Olasılıksal Ölçü : P: $\Im \rightarrow[0,1]$ (reel sayılar kümesinin kapalı bir aralığı) biçiminde tanımlanan fonksiyon aşağıdaki koşulları yerine getiriyorsa P ye $\Im$ üzerinde olasılıksal bir ölçü denir.

(i) $\forall A\in\Im$ $P(A) \geq 0$ (ii)$P(\Omega) =1$ (iii) $(i \neq j)\ A_i \cap A_j = \varnothing$ (I=1,2,...; j=1,2,...)

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \Sigma_{i=1}^{\infty} P\left(A_i\right)$$

Tanım: Olasılık uzayı : $(\Omega,\Im,P)$ üçlüsüne olasılık uzayı denir. $(\Omega,\Im,P)$ yukarıda tanımlandıkları gibidirler.

Tanım: Borel σ- cebiri B : R=E^1 (bir boyutlu Öklid uzayı) de bütün aralıkları (açık, yarı açık, kapalı) içine alan Borel σ- cebiridir. $B=B(E^1)$ ile gösterilir.Bu sonuçlar uygun biçimde E^n ye genişletilebilir.

Tanım: Rasgele fonksiyon (değişken) (Borel ölçülebilir fonksiyon): $(\Omega,\Im,P)$ de rasgele bir fonksiyon şöyle tanımlanır. $f:\Omega\rightarrow R= E^1$ ve $\forall B\in B$(Borel σ-cebiri) için $\{\omega:\omega\in\Omega$ ve $f(\omega)\in B\}\in\Im$ yani $f^{-1}(B)\in\Im$ ise f ye Ω üzerinde rasgele bir fonksiyon (ya da yaygın adıyla değişken) denir. Eğer f bu özellikleri sağlıyorsa buna ölçü teorisinde f Borel ölçülebilir bir fonksiyondur denir.

Tanım: Rasgele fonksiyonun dağılım (ve yoğunluk fonksiyonu): f, E^1 üzerinde negatif olmayan (≥ 0) ve $\int_{-\infty}^{\infty} f(x)dx = 1$ özelliğini sağlayan bir fonksiyon olsun. $B=B(R)$ nin alt kümelerinde tanımlı olup, $P(B)=\int_B f(x)dx$ [bütün B=[a,b] $\in$R için] özelliğini sağlayan yalnızca bir tane olasılıksal ölçü vardır. Bu Radon-Nikodym teoreminin bir sonucudur. Eğer Ω üzerinde tanımlı reel değerli bir X fonksi yonu varsa ($X(\omega)=\omega$ $\forall\omega\in\Omega$ için) $P(\omega:X(\omega)\in B)=P(B)= \int_B f(x)dx$ dır. X rasgele fonksiyonunun dağılım fonksiyonu $F_X(x)=P\{\omega:X(\omega)\leq x\}=P(-\infty, x] =\int_{-\infty}^{x} f(t)dt$ ile verilir. Burada f fonksiyonuna olasılık yoğunluğu fonksiyonu denir. Bu fonksiyon bilindiğinde X rasgele değişkenine ilişkin her türlü bilgi türetilebilir.

Tanım: n- boyutlu dağılm / yoğunluk fonksiyonları: Eğer $X_1, X_2, \ldots X_n$ aynı olasılık uzayında tanımlı rasgele fonksiyonlarsa

$F_{12\ldots n}(x_1, x_2, \ldots x_n)=P(X_1\leq x_1, X_2\leq x_2, \ldots X_n\leq x_n)$ fonksiyonuna birleşik(joint) olasılık fonksiyonu denir. Burada P, $B(E^n)$ de tanımlıdır. $(X_1, X_2, \ldots X_n)$ rasgele vektörü (r.v.) aşağıdaki koşulu sağlıyorsa mutlak süreklidir.

Koşul: $F_{12\ldots n}(x_1, x_2, \ldots x_n)= \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{12\ldots n}(u_1, \ldots u_n)du_1\ldots du_n$ olacak biçimde E^n de tanımlı negatif olmayan bir $f_{12\ldots n}$ fonksiyonu vardır. Bu fonksiyona birleşik olasılık yoğunluğu denir. $X=(X_1, X_2, \ldots X_n)'$ biçiminde yazıp n-boyutlu X rv sinden söz etmek daha uygundur. $\forall\omega\in\Omega$ için $X(\omega)=(X_1(\omega), X_2(\omega), \ldots X_n(\omega))\in E^n$ dir.

Tanım: Beklenen (ortalama) değer vektörü: X vektörünün beklenen değeri EX (ya da kısaca μ) ile tanımlanan integraldir. μ n bileşenli olup i. bileşen

$\mu_i = EX_i = \int_{-\infty}^{\infty} x_i\, f(x_i)\, dx_i$ (i=1,2,...,n) ile tanımlanır.

Burada $f_i(x_i)$, x_i bileşeninin marjinal olasılık yoğunluğu olup

$f_i(x_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_i, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$ biçiminde tanımlıdır. $f_{\mathbf{X}}$ ile yukarıda tanımlanan $f_{12\ldots n}$ aynı niceliği simgelerler.

Tanım: *Koşullu beklenen değer (KBD)*: (X,Y) iki boyutlu rasgele vektör ve h(.,.) bu iki değişkenin h: $R^2 \to R$ Borel ölçülebilir bir fonksiyonu olsun. X=x verildiğinde h(X,Y) nin kbd si { E[h(X,Y) | X=x] } ile gösterilir.

$E[h(X,Y) \mid X=x] = \int_{-\infty}^{\infty} h(x,y)\, f_{Y|X}(y|x)\, dx$ ile tanımlıdır. Burada $f_{Y|X} = f_{X,Y}(x,y) / f_X(x)$ ile tanımlanan X=x olduğunda Y nin koşullu olasılık fonksiyonudur. Bu daha önce Kesim 5.2.2 de söz edilen KBD dir. Eğer $X \in R^r$ ve $Y \in R^p$ ise KBD vektörü g: $R^r \to R^p$ dir. Burada KBD nin de bir rasgele fonksiyon olduğunu anımsatalım.

E[Y| X=x]=g(X)=g(X(ω)) yani $g: \Omega \to R^1$ (ya da R^p ye bir fonsiyon) . Son olarak g fonksiyonun esasen tek (essentially unique) olduğu belirtilmelidir. Yani $g^* = E[Y|X=x]$ biçiminde bir başka fonksiyon ise $g = g^*$ (ölçüsü sıfır olan kümenin dışında). O nedenle bir tek g den söz edilebilir (bkz. Ash(1972)).

Tanım: *Kovariyans Matrisi*: $\mathbf{X}=(X_1, X_2, \ldots X_n)'$ rv (rasgele vektör) si verilsin. $(\Sigma)_{nxn}$ kovariyans matrisinin elemanları

$\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = \int (x_i - \mu_i)(x_j - \mu_j) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$. Burada $\mathbf{X}=(X_1, X_2, \ldots X_n)'$ ve $d\mathbf{x} = \Pi^n_{i=1} dx_i$.

♠Hipotez: Herhangi bir Σ kovariyans matrisi en az pozitif semi-definittir.

Kanıt: **X** bir rv ve **a** nx1 boutlu bir vektör olsun. Var(**a'X'**)≥0.

$Var(\mathbf{a'X'}) = E[\mathbf{a'}(\mathbf{X}-\mu)']^2 = TraceE[\mathbf{a'}(\mathbf{X}-\mu)(\mathbf{X}-\mu)'\mathbf{a'}] = Trace(\mathbf{a'}[E(\mathbf{X}-\mu)(\mathbf{X}-\mu)'\mathbf{a'}]) = Trace\ a\Sigma a \geq 0$. ♠

5.2.4 ÇBND altında Bileşen1(C_1) için Analitik İfadelerin Türetilmesi

Bu kesimde C_1 in analizi için gerekli olan ifadeler adım adım türetilir ve YSA nın performansının objektif olarak değerlendirilebilmesi için sonuçlar elde edilir. Bunun için $\mathbf{Z}' = (\mathbf{X}^{1'}, \mathbf{X}^{2'})$ olsun. **Z** bir ÇBND olmak koşuluyla **Z** nin birleşik olasılık fonksiyonu

$$f(\mathbf{x}^1, \mathbf{x}^2) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{|\Sigma_{11.2}|}} \exp\left\{ -\frac{1}{2} \left[(\mathbf{x}^1 - \mu^1) - \Sigma_{12} \Sigma_{22}^{-1} (\mathbf{x}^2 - \mu^2) \right]' \Sigma_{11.2}^{-1} \left[(\mathbf{x}^1 - \mu^1) - \Sigma_{12} \Sigma_{22}^{-1} (\mathbf{x}^2 - \mu^2) \right] \right\}$$

$$x \frac{1}{(2\pi)^{\frac{r}{2}}\sqrt{|\Sigma_{22}|}} \exp\left[-\frac{1}{2}(\mathbf{x}^2-\mu^2)'\,\Sigma_{22}^{-1}(\mathbf{x}^2-\mu^2)\right] \tag{5.1}$$

(bkz. Anderson 1958, s28). Burada

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{22}^{-1}\Sigma_{22} \tag{5.2}$$

dır.

♠ *Bazı ek tanımlar*

$$(\Sigma_{11})_{pxp} = E[(\mathbf{X}^1-\mu^1)(\mathbf{X}^1-\mu^1)'] \tag{5.3a}$$

$$(\Sigma_{22})_{rxr} = E[(\mathbf{X}^2-\mu^2)(\mathbf{X}^2-\mu^2)'] \tag{5.3b}$$

$$(\Sigma_{12})_{pxr} = E[(\mathbf{X}^1-\mu^1)(\mathbf{X}^2-\mu^2)'] \tag{5.3c}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{22} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} pxp & pxr \\ rxp & rxr \end{bmatrix}_{(p+r)x(p+r)} = kxk \quad ♠$$

Koşullu olasılık fonksiyonu (kof) bilindiğinde C_1 için analitik bir yazım üretilebilir. Bu tezin ilgi odağı bu kesimde ÇBND olduğu için bu yazım ÇBND ye özel olacaktır. (ÇBND den farklı diğer bazı dağılımlar için yazımlar daha sonra sırası geldikçe türetilecekir). $\mathbf{X}^2=\mathbf{x}^2$ verilsin. $\mathbf{X}^1$ in kof'u

$$f(\mathbf{x}^1 \mid \mathbf{x}^2) = \frac{1}{(2\pi)^{\frac{p}{2}}\sqrt{|\Sigma_{11.2}|}} \exp\left\{-\frac{1}{2}\left[(\mathbf{x}^1-\mu^1)-\Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^2-\mu^2)\right]'\Sigma_{11.2}^{-1}\left[(\mathbf{x}^1-\mu^1)-\Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^2-\mu^2)\right]\right\}..$$

..(5.4).

olarak elde edilir. Denk.(5.4) den $\mathbf{X}^2=\mathbf{x}^2$ için $\mathbf{X}^1$ in kof'u $f(\mathbf{x}^1 \mid \mathbf{x}^2)$ p- boyutlu normal bir dağılımdır ve beklenen değeri

$$E(\mathbf{X}^1 \mid \mathbf{X}^2) = \mu^1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^2-\mu^2) = \mathbf{g}(\mathbf{x}^2) \tag{5.5}$$

ve koşullu variyans matrisi

$$E\left\{[\mathbf{X}^1-g(\mathbf{X}^2)][\mathbf{X}^1-g(\mathbf{X}^2)]' \mid \mathbf{X}^2\right\} = \Sigma_{11.2} \tag{5.6}$$

ile verilir. Denk.(5.6) nın C_1 terimine nasıl bağlı olduğu aşağıdaki birebir (1-1) eşlemelerle anlaşılabilir.

(i)$\mathbf{X}^1 \rightarrow \mathbf{Y}$; $\mathbf{X}^2 \rightarrow \mathbf{X}$

(ii) $g(\mathbf{X}^2) \rightarrow g(\mathbf{X})$

(iii)$E\left\{[\mathbf{X}^1-g(\mathbf{X}^2)][\mathbf{X}^1-g(\mathbf{X}^2)]'\right\} \rightarrow E[\mathbf{Y}-g(\mathbf{X})]^2$

(iii) eşlemesi ile

denk(5.6) arasındaki geçiş için aşağıdaki adımlar atılır.

$A_1 : E[E\{[...][...]'\}|\mathbf{X}^2]=E\{[...][...]'\}=E[\Sigma_{11.2}]=\Sigma_{11.2}$

A_1 adımında 'iterated beklenen değerler yasası 'ile $\Sigma_{11.2}$ nin sabit bir matris olduğu gerçeği kullanılmıştır.

$A_2=E\{[...]'[...]\}=\text{Trace}E\{[...][...]'\}$

A_2 ile (iii) eşlemesi karşılaştırılarak

$$C_1=E[\mathbf{Y}-g(\mathbf{X})]^2=\text{Trace}\Sigma_{11.2}=\Sigma^p_{k=1}(\Sigma_{11.2})_{kxk} \qquad (5.7)$$

elde edilir. C_1, denk.(5.7) yardımıyla analiz edileblir, çünkü $\Sigma_{11.2}$ bizim tarafımızdan (deneyci tarafından) atanabilen değerlerden oluşur. Bu değerler Σ kovariyans matrisidir.

5.3 Bu Çalışmada Araştırılan Dönüşüm Yeteneklerinin (ÇBND için) Analizi

Bölüm 5 de simülasyon sonuçlarına ilişkin yaptığımız yorumlar gözleme dayalı olup kuramsal bir çerçeveden sunulmamışlardır. O nedenle yeterince sağlam değildir. Daha açıkçası oradaki yorum ağın performansı üzerine analitik değerlendirmelerden yoksundur. Bu eksiklikler nedeniyle oradaki sonuçların sorgulanması bu bölümde matematiksel bir analizle yeniden yapılacaktır. Kesim(5.2) deki değerlendirmeler analiz için gereken teorik temeli sağlarlar. Bu bölümde gittikçe artan ağ kompleksitesi altında iki farklı input uzayı için ağın stokastik performansı araştırılacaktır.

Tanım: *Ağ kompleksitesi (karmaşası)* : Verilen bir input ve hedef uzayı (sırasıyla R^r ve R^p uzayları) için uzaylar arasında tanımlı f fonksiyonunun YSA yardımıyla temsil edilebilmesi için gizli tabaka nöronları sayısına ağ kompleksitesi denir. Bu temsilin en iyi olabilmesi için gerekli gizli tabaka nöronu sayısına optimum ağ kompleksitesi denir

Ancak verilen bir problem için optimum ağ kompleksitesinin ne olması gerektiği önsel (a priori) olarak bilinemez. O nedenle bu problemin çözümü gizli tabaka nöronu sayısını (n_h) uygun uygun şekilde artırarak bulunur.

Tanım: *Input Uzayı Boyutu (İUB)*: İnput vektörlerinin çekildiği uzayın (R^r) boyutu. Bu tezde bu boyut 1 ve 2 alındı. ♠Yani r=1,2 ♠

İUB 1 : $n_v=(1,n_h,1)$ Bir input - bir hedef YSA

İUB 2 : $n_v=(2,n_h,1)$ İki input - bir hedef YSA

Ya da Kes.(5.2) nin notasyonuyla

İUB1→ $Z=(X^{1'},X^{2'})'_{2x1}$; X^1=Y(hedef) ve X^2=X(input)

İUB2→ $Z=(X^{1'},X^{2'})'_{3x1}$; X^1=Y(hedef) ve $X^2=(X_1,X_2)'$ (input)

5.3.1 İUB 1 : ÇBND altında analiz

$\lambda(w)$ nın C_1'i koşullu kovariyans matrisinin trace beklenen değeridir (denk.(5.7)).

$C_1=\Sigma_{11.2}=\mathrm{trace}\Sigma_{11.2}$

$\therefore \Sigma_{11.2}=\sigma_{11}-[\sigma_{12}]^{-1}[\sigma_{21}]$; (1x1)(1x1)(1x1)=(1x1)

<u>AF2 için C_1</u>

$\Sigma_{11.2}$ (1x1) lik bir matris yani bir skalerdir. AF2 için bu Q_1 olsun. $\sigma_{12}=\sigma_{21}$ den

$$Q_1=\sigma_{11}-\sigma_{12}^2(1/\sigma_{22}) \qquad (5.8)$$

elde edilir. $\sigma_{11}=\sigma_Y$, $\sigma_{22}=\sigma_X$ ve $\sigma^2_{12}=r_{XY}\sigma_X\sigma_Y$ alınarak

$$Q_1=\sigma_Y-r_{XY}^2\sigma_X\sigma_Y(1/\sigma_X)=\sigma_Y(1-r_{XY}^2) \qquad (5.9a)$$

bulunur. (r_{XY} **X** ile **Y** arasındaki populasyon korelasyon katsayısıdır).. AF2 için **Z** İUB1 için 2-boyutlu standart bir normal dağılımdır. Buna göre $\sigma_Y=1$ olduğundan denk(5.9b) elde edilir.

$$Q_1=(1-r_{XY}^2) \qquad (5.9b)$$

<u>AF1 için C_1</u>

Denk.(5.9a) dönüşmemiş **X** ve **Y** için geçerlidir. Yani $(X,Y)\in(-\infty,\infty)x(-\infty,\infty)$ olabilir. Ancak daha önce Bölüm 3 de verinin transformasyonu kesiminde anlatıldığı gibi input ve hedef vektörleri [0,1] aralığında değerler alacak biçimde afin bir dönüşüme uğrar. Yani **Z** yerine $Z_t=(Y_t\,,\,X_t)'_{2x1}$ biçiminde bir normal rv nin neden olduğu $(Q_1)_t=(C_1)_t$ teriminin incelenmesi gerekir.

$$Y\to Y_t=\alpha Y+\beta \;;\; X\to X_t=\alpha X+\beta \qquad (5.10)$$

değişken dönüşümü olmalıdır. Burada (α,β) katsayıları Kesim(3.2.2) deki gibi

$$\alpha=(1/y_{max}-y_{min})=(1/x_{max}-x_{min}); \quad \beta=-y_{min}\ x\ (1/y_{max}-y_{min})=x_{min}\ x\ (1/x_{max}-x_{min}) \qquad (5.11a)$$

olarak atanırlar. Kuşkusuz (α,β) katsayıları teorik olarak keyfi derecede büyük/küçük olabilirler (max ve min değerlerinden dolayı). Bununla birlikte bu çalışmanın amaçları açısından bu sayıların alt ve üst limitlerini çok büyük bir olasılıkla saptayabiliriz. Öyle ki;

♠ *Varsayım*: $x_{max}=-x_{min}$ ve $y_{max}=-y_{min}$ olsun ♠

$Pr(a<\mathbf{X}<b)$ **X** rasgele değişkeninin (a,b) aralığında değerler almasının olasılığını göstermek üzere

$$Pr(x_{min} < X < x_{max}) = N(x_{max}) - N(x_{min}) \tag{5.12}$$

elde edilir. Burada $N(x)=\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = Pr(-\infty < \mathbf{X} < x)$ dir. $N(-x)=1-N(x)$ özelliği kullanılarak denk.(5.12) den $2N(x_{max})-1=0.990 \Rightarrow N(x_{max})=0.995 \Rightarrow x_{max} \approx 2.5$ elde edilir. Toparlanırsa

$$(x_{max}, x_{min})=(y_{max}, y_{min})=(-2.5, 2.5) \tag{5.13}$$

değerleri elde edilir. Denk.(5.9) ın X_t ve Y_t için yazımı

$$(Q_1)_t = \sigma_t (1-r_{X_tY_t}^2) \tag{5.14a}$$

dır. Afin dönüşüm altında X_t ve Y_t normal dağılıma uydukları için denk(5.14a) doğrudur. Ayrıca afin bir dönüşüm için

$r_{X_tY_t} = r_{XY}$ ve $\sigma_{Y_t} = \alpha^2 \sigma_Y$. Bu bilgilerle denk.(5.14a) için

$$(Q_1)_t = \alpha^2 \sigma_Y (1 - r_{X_tY_t}^2) = \frac{1}{[2.5-(-2.5)]^2} .1.(1 - r_{XY}^2) = \frac{1}{25}(1 - r_{XY}^2) \tag{5.14b}$$

elde edilir. $(Q_1)_t$ dönüştürülmüş X_t ve Y_t için $\lambda(w)$ nın C_1 terimidir. Yani

$$\lambda(w)=(1/25)(1-r_{XY}^2) + C_2 \tag{5.15}$$

Şimdi çeşitli r_{XY} değerleri için C_1 in sayısal değerlendirmesini yapmadan önce bir noktayı aydınlatmakta yarar vardır. λ_n , $\lambda(w)$ nın n tane gözlem için sample (örneklem) değeri olsun. Yani

$\lambda_n(w) = n^{-1} \sum_{i=1}^{n} (y_i - f_i)^2$ dir. Ancak Bölüm 4 deki hata metriği λ_n yerine $\lambda_n^{1/2}$ ile verilir. Yani Bölüm 4 deki bulguları değerlendirebilmek için

$$\lambda_n^{1/2} = \{(C_1)_n + (C_2)_n\}^{1/2} \tag{5.16}$$

yazılmalıdır. $x \geq 0$ ve $y \geq 0$ için $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ olduğu anımsanarak denk.(5.16) dan

$$\lambda_n^{1/2} \leq (C_1)_n^{1/2} + (C_2)^{1/2} \tag{5.17}$$

(Not. Burada $(C_1)_n$ ve $(C_2)_n$ kareli toplamlar oldukları için ≥ 0 dırlar). Demek ki $\lambda_n^{1/2}$ nın üst limiti

$$(\lambda_n)_{max}^{1/2} = (C_1)_n^{1/2} + (C_2)_n^{1/2} \tag{5.18a}$$

ile verilir. Alt limit ise (g=f hemen her yerde) olması koşuluyla

$$(\lambda_n)_{min}^{1/2} = (C_1)_n^{1/2} \tag{5.18b}$$

Kısaca f nin seçimine bağlı olarak

$(C_1)_n^{1/2} \leq \lambda_n^{1/2} \leq (C_1)_n^{1/2}+(C_2)_n^{1/2}$ (5.18c)

dir. Bu değerlendirmelerden sonra $(Q_1)_t$ çeşitli r_{XY} için saysal değerleri Tablo 5.1 de sunulmuştur. Burada yukarıdaki açıklamalarla tutarlılık sağlamak için $(Q_1)_t \rightarrow \sqrt{(Q_1)_t}$ olarak alınmıştır.

Tablo 5.1 $(C_1)_n^{1/2}=(Q_1)_t$ nin çeşitli r_{XY} katsayıları için değerleri

r_{XY}	0.0	0.4	0.2	0.6	0.8	1.0
$(Q_1)_t$	0.200	0.183	0.195	0.160	0.120	0.000

Tablo 5.1 deki $(C_1)_n^{1/2}=(Q_1)_t$ tanımı tam olarak doğru değildir. Çünkü, $(C_1)_n^{1/2}$ n tane gözlem üzerinden bir örneklem ortalamasını gösterir. $(Q_1)_t$ ise $\mathbf{Z}_t$ rv sinin bütün uzay üzerinden alınan olasılık ortalamasını gösterir. Ancak yeterince büyük n için $(C_1)_n$, $(Q_1)_t$ ye olasılık olarak yakınsar. Yani $(C_1)_n \xrightarrow{Pr} (Q_1)_t$ dir. Her ne kadar verilen bir $(C_1)_n$ terimini sayısal olarak hesaplamak olanaklıysada Tablo 5.1 de limit yaklaşımı kullanıldı. Kuşkusuz n=20 için bu yakalşımın hata payı önemsenebilecekken n=100 için limit yaklaşım daha iyidir. Bu konu üzerinden ileride asimptotik sonuçlar değerlendirilirken daha çok şey söylenecektir.

♠Bir uyarı ! Notasyonun çift anlamlı kullanımı üzerine

Tablo 5.1 deki değerler Kesim 4.4 deki $\overline{\lambda}_{n_i}$ değerleriyle karşılaştırabilmek için Bölüm 4 de tanımlanan $\overline{\lambda}_{n_i}$ fonksiyonunun n_i indisi denk.(5.16) daki $\lambda_n^{1/2}$ terimindeki n indisiyle aynı niceliğe referans vermelidir. Oysa Bölüm 4 deki $\overline{\lambda}_{n_i}$ teriminde n_i indisi örneklem matrisleri sayısı (n_i=100), öte yandan denk.(5.16)daki n indisi verilen bir n_i için gözlem sayısıdır. Bu durumda $\overline{\lambda}_{n_i}$ ile λ_n şöylesi bir gözlemle uyumlu kılınabilir. Verilen bir $n \in N$ için $\overline{\lambda}_{n_i}$ değerleri n_i=100 tane örneklem matrisi üzerinden alınan ortalamadır. Bölüm 4 deki verilere göre $\overline{\lambda}_{n_i}$ nin *standart sapması* herbir r_{XY} için yaklaşık 0.030 ya da daha küçüktür. Yani λ_i değeri pratik olarak herbir i değeri için $\overline{\lambda}_{n_i}$ ye yakındır ($\lambda_i \approx \overline{\lambda}_{n_i}$ $\forall i \in I$). O zaman $\overline{\lambda}_{n_i}$ deki n_i indisini düşürürüz ve sanki yalnızca bir tane örneklem matrisi varmış gibi ilerleriz. Bu durumda denk.(5.16) daki λ_n ile Bölüm 4 deki $\overline{\lambda}_{n_i}$ uyuşur. Yani bir tek λ_n den söz edilebilir♠

5.3.2 AF1 ve İUB1 için asimptot altı bölgede bazı sonuçlar

Yukarıdaki uyarıdan sonra Bölüm 4 deki (verilen bir $n \in N$ için) değerleri (ki bunu artık λ_n ile gösterebiliz) Tablo 5.1 deki değerlerle önemli derecedebir yakalşıklık sergiler. Bu yaklaşıklık daha önce teorik olarak öngörüldüğü üzere $(C_2)_n$ değerinin minimum (ya da ideal olarak sıfır) olması gerçeğini doğrulamada bir kanıt sağlar. Tablo 5.2 çeşitli r_{XY} katsayıları için artan n_h altında $(C_1)_n$ ile Bölüm 4 de elde edilen $\lambda_n^{1/2}$ değerlerinin bir karşılaştırmasını sunar. Burada $\lambda_n^{1/2}$ ile verilen hata (beklenen performans) fonksiyonu $w=w_n$ gibi optimum ağırlık vektörü üzerinden hesaplanmıştır. Tablodan görüldüğü gibi $\lambda_n^{1/2}$ ve $(C_1)_n$ çeşitli n ve n_h değerleri için birbirlerine yakındırlar. Daha önce vurgulandığı gibi Tablo5.2a daki değerler için n=20 sayısı aslında küçüktür. Tablo 5.2c deki n=100 ise kabul edilebilecek ölçüde yüksek bir sayıdır. Ve bu kez $\lambda_n^{1/2}$ ile $(C_2)_n$ in uyuşumu daha iyidir. Yani n arttıkça d(=Fark) nin genel olarak azalması; gözlem sayısı (n) arttıkça daha iyi fit (uyuşum) lerin oluşacağına ilişkin kuramasal öngörüyü dağrular niteliktedir. [Bu durum istatistikte "büyük sayılar yasası" olarak bilinir. İleride bu konuya yeniden değinilecektir]. Verilen bir n_h için n arttıkça fitlerin daha iyi olacağına ilişkin gözlemler Tablo 5.3 de özetlenmiştir. Tablo 5.2a,b,c den öteki bazı gözlemler şunlardır: Verilen bir (n, r_{XY}) çifti için n_h arttıkça $\lambda_n^{1/2}$ ve $(C_1)_n$ birbirlerine daha yakın olmaktadırlar. Bu eğilim Hornik ve ark(1989) tarafından öngörülen kuramsal değerlendirmeleri doğrular niteliktedir.

♠**Bir teorem (Hornik ve ark(1989)), bir açıklama**

Sigmoid aktivasyon fonksiyonu kullanan tek gizli tabakalı ileri beslemeli YSA, yeterli sayıda gizli tabaka nöronu bulunmak koşuluyla, bir sonlu uzaydan öteki sonlu uzaya herhangi bir Borel ölçülebilir fonksiyonu istenilen bir ε duyarlığında yaklaşılabilr. Bu anlamda YSA evrensel yaklaşımcılardır. Uygulamadaki başarısızlık yetersiz öğrenmeye, yetersiz sayıda gizli tabaka nöronu bulunmasına, input ile hedef vektörleri arasında deterministik ilişki yerine stokastik ilişki bulunmasına bağlanmalıdır♠

5.3.2.1 Tartışma

Yukarıdaki teoremin ışığında Tablo 5.2a,b,c sonuçlarını bir kez daha gözden geçirelim. r_{XY}=1.00 dışında **X**(input) ve **Y**(hedef) arasında deterministik bir ilişki bulunmadığı için ağın dönüştürmesi istenilir(keyfi) bir düzeyde olamaz. Bununla birlikte **X** ile **Y** arasındaki deterministik ilişki g(x) ile gösterilen koşullu beklenen değerdir. Bu durumda ağın output fonksiyonun (f) yaklaşımda bulunabileceği tek bir fonsiyon bu g(x) fonksiyonudur. Yani $\lambda(\omega)$ nın minumum olması $E[f(X,\omega)-g(X)]^2$ nin minumum olmasına denktir. Tablo 5.2 aslında f fonksiyonunun g fonksiyonuna istenilen duyarlıkta (daha doğrusu n_h ile sınırlı bir duyarlıkta) yaklaşım yapılabildiği gerçeğini ortaya çıkarmada kanıtlar sağlamaktadır. (Daha kesin kanıtlar ileride asimptotik incelemelerle verilecekir). O nedenle $(C_1)_n$ terimi ile $\lambda_n^{1/2}$ terimi herbir r_{XY} için birbirlerine oldukça yakın gözükmektedirler. Ayrıca n_h arttıkça teoremin öngördüğü gibi bu yaklaşım daha da iyileşmektedir. Son olarak Tablo 5.3 de n arttıkça $\lambda_n^{1/2}=(C_1)_n=(Q_1)_n$ yaklaşımının daha geçerli olduğuna dikkat edilmelidir.

Tablo 5.2a n=20 için $\lambda_n^{1/2}$ ve $(C1)_n$ nin artan ağ kompleksitesi (n_h) için değerleri .

$d=| \lambda_n^{1/2} -(C_1)_n |$

r_{XY}	$\lambda_n^{1/2}$			$(C_1)_n$	d=Fark		
0.0	0.247	0.247	0.242	0.200	0.047	0.047	0.042
0.2	0.227	0.218	0.196	0.195	0.032	0.023	0.001
0.4	0.224	0.224	0.224	0.183	0.041	0.041	0.041
0.6	0.198	0.196	0.157	0.160	0.038	0.036	0.003
0.8	0.148	0.153	0.155	0.120	0.028	0.033	0.035
1.0	0.059	0.059	0.059	0.000	0.059	0.059	0.059
n_h	1	2	3		1	2	3

Tablo 5.2b n=50 için Tablo 5.2a daki değerlendirmeler

r_{XY}	$\lambda_n^{1/2}$			$(C_1)_n$	d=Fark		
0.0	0.213	0.219	0.223	0.200	0.013	0.019	0.023
0.2	0.211	0.220	0.220	0.195	0.016	0.025	0.025
0.4	0.207	0.207	0.224	0.204	0.003	0.003	0.020
0.6	0.182	0.143	0.154	0.160	0.022	0.017	0.006
0.8	0.148	0.140	0.145	0.120	0.028	0.020	0.025
1.0	0.059	0.030	0.059	0.000	0.059	0.030	0.059
n_h	1	2	3		1	2	3

Tablo 5.2c n=100 için Tablo 5.2a daki değerlendirmeler

r_{XY}	$\lambda_n^{1/2}$			$(C_1)_n$	d=Fark		
0.0	0.247	0.2o4	0.204	0.200	0.047	0.004	0.004
0.2	0.196	0.196	0.196	0.195	0.001	0.00	0.001
0.4	0.192	0.188	0.188	0.183	0.009	0.005	0.005
0.6	0.168	0.157	0.157	0.160	0.008	0.003	0.003
0.8	0.139	0.127	0.120	0.120	0.019	0.007	0.000
1.0	0.020	0.020	0.020	0.000	0.020	0.020	0.020
n_h	1	2	3		1	2	3

Tablo 5.3 n_h=3 için artan n altında $\lambda_n^{1/2}$ ve $(C_1)_n$

r_{XY}	$\lambda_n^{1/2}$			$(C_1)_n$	d=Fark		
0.0	0.242	0.223	0.204	0.200	0.042	0.023	0.004
0.2	0.196	0.220	0.196	0.195	0.001	0.025	0.001
0.4	0.224	0.211	0.188	0.183	0.021	0.028	0.005
0.6	0.157	0.184	0.157	0.160	0.003	0.024	0.003
0.8	0.155	0.147	0.127	0.120	0.035	0.027	0.007
1.0	0.059	0.059	0.010	0.000	0.059	0.059	0.010
n	20	50	100		20	50	100

5.3.2.2 (Bileşen 2)$_n$=$(C_2)_n$ in büyüklüğü üzerine kabaca bir sayısal değerlendirme

n yeterince büyük olmak koşuluyla verilen bir r_{XY} için $(C_2)_n$ nin sıfira yakın olması gerektiğini vurgulanmıştı. Bununla birlikte bu öngörü için sayısal kanıtlar sunma zorunluluğu vardır. Denk.(5.5) den

$$g(x) = E(Y|X=x) = \mu_Y + r_{XY}\sqrt{\sigma_Y}\sqrt{\sigma_X}\frac{1}{\sigma_X}(x-\mu_X) = \mu_Y + r_{XY}\sqrt{\frac{\sigma_X}{\sigma_Y}}(x-\mu_X) \qquad (5.19a)$$

Burada σ_Y ve σ_X, **X** ve **Y** rv lerine karşılık gelen variyanslardır. Ya da $\mathbf{X}_t$ ve $\mathbf{Y}_t$ için

$$g(X_t) = E(Y_t|X_t) = \mu_{Y_t} + (r_{XY})_t\sqrt{\frac{\sigma_{X_t}}{\sigma_{Y_t}}}(x_t-\mu_{X_t}) \qquad (5.19b)$$

dir. Denk(5.19b) denk.(5.19a) nın terimleri cinsinden yazılırsa

$$g(x_t) = (\alpha\mu_Y+\beta) + r_{XY}\sqrt{\frac{\alpha^2\sigma_X}{\alpha^2\sigma_Y}}(x-\alpha\mu_X-\beta) \quad \text{olduğundan}$$

$g(x_t)=\beta+r_{XY}(x_t-\beta)=r_{XY}\,x_t+(2.5/5)(1-r_{XY})=r_{XY}\,x_t+0.5(1-r_{XY})$

bulunur. Denk.(5.19c) r_{XY} verildiğinde herhangi bir $\mathbf{X}_t$=x_t için $g(x_t)$ yi hesaplama olanağı sağlar. Tablo 5.4 de İUB1 ve n=100 için bulunan $\mathbf{w}_{ort}^*$ dayalı olarak $f(x_t, \mathbf{w}_{ort}^*)$ ile $g(x_t)$ nin bir karşılaştırmasını sunar. Burada $\mathbf{w}_{ort}^*=(\overline{w}_1,\overline{w}_2,\overline{b}_1,\overline{b}_2)$ olup "-" ise i=100 tane örneklem matrisi üzerinden alınan ortalamayı gösterir. Bir başka deyişle, ağ öğretim sürecinde bu örnek için 100 tane (x_t,y_t) çifti kullanılmıştır. n=100 sayısı kabaca yeterince büyük kabul edilirse $(C_2)_n\approx E(f-y)^2$ limit yaklaşımı kullanılabilir. Bu durumda $\mathbf{X}_t$=x_t için $f(x_t, \omega_{ort}^*)$ ile $g(x_t)$ arasındaki fark küçük olmalıdır. Tablo 5.4 de seçilen üç korelasyon yapısı için (r_{XY}=0.2,0.6,1.0) verilen bazı x_t değerlerine karşılık gelen (iw=1,2) için f ve g nin bir karşılaştırması sunulur. Bazı gözlemler şunlardır.

r_{XY}=0.2 ve 0.6 için $(C_1)_n$ / d oranı ortalama 2.8-3.0 civarındadır. Yani $(C_1)_n$ terimi önemli ölçüde baskındır. Aynı durum iw=1 için r_{XY}=1.0 durumunda da söz konusudur, ancak iw=2 için bir sapma vardır (iterasyon sayısı yetersiz olabilir). Ya da EDBDM nin parametreleri yetersiz seçilmiş olabilir). Kabaca $(C_2)_n$ terimi n=5 için bile sıfıra belirli ölçülerde yakındır.

Tablo 5.4 n_v=(1,2,1)' için f(output fonksiyonu) ve g(KBD) için r_{XY}=0.2,0.6,1.0 ve bazı $x=x_t$ için $d=(1/5)\sum_{ik=1}^{5}(g_k - f_k)^2$. Herhangi bir r_{XY} için birinci satır değerleri iw=1, ikinci satır değerleri iw=2 durumuna karşılıktır.

r_{XY}	f					g					d	$(C_1)_n$/d
0.2	0.370	0.456	0.610	0.710	0.410	0.400	0.460	0.500	0.540	0.600	0.067	2.985
	0.460	0.463	0.468	0.481	0.540	"	"	"	"	"	0.057	3.508
0.6	0.194	0.452	0.510	0.530	0.874	0.200	0.380	0.500	0.620	0.800	0.062	2.580
	0.288	0.393	0.456	0.506	0.720	"	"	"	"	"	0.076	2.105
1.0	0.110	0.257	0.521	0.778	0.900	0.000	0.300	0.500	0.700	1.00	0.078	
	0.280	0.346	0.425	0.621	0.905	"	"	"	"	"	0.300	
x_t	0.0	0.3	0.5	0.7	1.0	0.0	0.3	0.5	0.7	1.0		

5.3.3 İUB2 : Analiz (ÇBND ve C_1 ve C_2 için analitik ifadelerin türetilmesi)

Bu kesimde denk.(5.5) yardımıyla C_2 için , denk.(5.7) yardımıyla C_1 için üç boyutlu normal dağılım altında analitik ifadeleri açıkça türeteceğiz. Bu türetme iki amacın gerçekleştirilmesine olanak tanıyacaktır.

Amaç 1 : C_1 ve C_2 nin bilgisayara uygulanabilmesi

Amaç 2: Teorik olarak öngörülen C_1 ve C_2 değerlerinin *bilgisayar simülasyonu* sonuçlarıyla elde edilen değerlerle karşılaştırma olanağının elde edilmesi

♠ $C_1 = \sum_{k=1}^{p}(\Sigma_{11.2})_{kxk}$ **nın p+r=3 için hesabı**

Denk.(5.6) ve denk.(5.7) bir anda düşünülerek

$\Sigma_{11.2}= \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2x1}$ $\rightarrow E[Y-g(X)]^2=Tr(\Sigma_{11.2})=C_1$ yazılır. Σ_{11} matrisi (1x1), Σ_{12} matrisi (1x2), Σ_{22} (2x2) , Σ_{21} (2x1) lik matrislerdir. Bu durumda $\Sigma_{11.2}$ matrisi yine (1x1) lik matris (yani bir skaler) ve $\Sigma_{11.2}=C_1$ olur. $\mathbf{X}^1=X_1=\mathbf{Y}$ ve $\mathbf{X}^2=\mathbf{X}=(X_2,X_3)'$ ve $\mathbf{Z}=(\mathbf{Y},\mathbf{X})'=(X_1,X_2;X_3)'$ biçiminde bir rv elde edilir. Kovaryans matrisi $\Sigma_{3x3}=(\sigma_{ij})_{3x3}$

(i=1,2,3 ve j=1,2,3). Σ_{11} matrisini Σ matrisinin bileşenleri cinsinden yazalım.

$$\Sigma_{11}=\sigma_{11};\ \Sigma_{12}=[\sigma_{12}\ \ \sigma_{13}]\ ;\ \Sigma_{22}=\begin{bmatrix}\sigma_{22} & \sigma_{23}\\ \sigma_{32} & \sigma_{33}\end{bmatrix}^{-1}\ ;\ \Sigma_{21}=\Sigma_{12}.\ \text{Yani}$$

$$\Sigma_{11.2}=C_1=\sigma_{11}-[\sigma_{12}\ \ \sigma_{13}]\begin{bmatrix}\sigma_{22} & \sigma_{23}\\ \sigma_{32} & \sigma_{33}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{21}\\ \sigma_{31}\end{bmatrix} \tag{5.20a}$$

Σ_{22} matrisini hesaplamak için $k=1/\det\Sigma_{22}$ yazalım. Buna göre,

$$C_1=\sigma_{11}-[\sigma_{12}\ \ \sigma_{13}]\begin{bmatrix}k\sigma_{33} & -k\sigma_{23}\\ -k\sigma_{32} & k\sigma_{22}\end{bmatrix}\begin{bmatrix}\sigma_{21}\\ \sigma_{31}\end{bmatrix}=\sigma_{11}-[\sigma_{12}\ \ \sigma_{13}]\begin{bmatrix}k\sigma_{33}\sigma_{21} & -k\sigma_{23}\sigma_{31}\\ -k\sigma_{32}\sigma_{21}\ + & k\sigma_{22}\sigma_{31}\end{bmatrix}\begin{bmatrix}\sigma_{21}\\ \sigma_{31}\end{bmatrix}$$

$$=\sigma_{11}-k\,[\sigma_{12}\sigma_{33}\sigma_{21}-\sigma_{21}\sigma_{23}\sigma_{31}-\sigma_{13}\sigma_{32}\sigma_{21}+\sigma_{13}\sigma_{22}\sigma_{31}] \tag{5.20b}$$

$\sigma_{11}=\sigma_Y=1$; $\sigma_{22}=\sigma_{X1}=1$,$\sigma_{33}=\sigma_{X2}=1$ (daha önceki anlaşmaya göre)

Yani Σ kovariyans matrisini **R** korelasyon matrisine dönüştürmek için variyanslar 1 olarak alınırlar. Buna göre $C_1=1-k\,[\sigma_{12}-2\sigma_{12}\sigma_{13}\sigma_{23}+\sigma_{13}]^2$ olur.

$\sigma_{12}=x$, $\sigma_{13}=y$, $\sigma_{23}=z$ yazarak

$$\boxed{C_1(x,y,z)=C_1=1-\frac{1}{(1-z^2)}\left[x^2-2xyz+y^2\right]} \qquad *** \tag{5.21}$$

elde edilir. Buradaki x,y,z değerlerinin aynı zamanda korelasyon değerleri oldukları açıktır. Yani $x=r_{YX1}$; $y=r_{YX2}$; $z=r_{X1X2}$ dir. Böylece (x,y,z) korelasyon üçlüsüne dayalı 3-boyutlu standart normal dağılımdan gelen örneklem matrisleri $\underline{\mathbf{Z}}$ elde edilir. daha sonra bu bilgi YSA da işlenir. $\lambda(\mathbf{w})$ ortalama performans (hata) için $\mathbf{w}=\mathbf{w}^*$ optimum vektörü elde edilir ($n\to\infty$ limitinde) $\lambda(\mathbf{w}^*)=C_1+C_2$ denklemindeki C_1 bileşeni denk.(5.21) yardımıyla analiz edilebilir. Son olarak denk.(5.21) in ağın asimptotik davranışındaki (ÇBND altında) C_1 bileşeni olduğunu belirtelim.

♠C_2= E(g-f)^2 nin hesabı için g nin türetilmesi (p+r=3)

Denk.(5.15) den $E(X^1|X^2)=\mu^1+\Sigma_{12}\Sigma_{22}^{-1}(x^2-\mu^2)=g(x^2)$ dir. $\mathbf{X}^2$=**X**(input) olarak alınır. Bu tezde standart ÇBND kullanıldığı için $\mu^1=\mu^2=\theta$ (sıfır vektörü) dir. Buna göre $g(\mathbf{x}^2)$

$$g(\mathbf{x}^2)=\Sigma_{12}\Sigma_{22}^{-1}\,\mathbf{x}^2 \tag{5.22}$$

$$=\begin{bmatrix}\sigma_{YX_1} & \sigma_{YX_2}\end{bmatrix}\begin{bmatrix}\sigma_{X_1X_2} & \sigma_{X_1X_2}\\ \sigma_{X_2X_1} & \sigma_{X_2X_2}\end{bmatrix}^{-1}[x_1x_2]'.$$

$$\begin{bmatrix}\sigma_{X_1X_1} & \sigma_{X_1X_2}\\ \sigma_{X_2X_1} & \sigma_{X_2X_2}\end{bmatrix}^{-1}=\frac{1}{1-\sigma_{X_2X_1}{}^2}\begin{bmatrix}1 & -\sigma_{X_1X_2}\\ -\sigma_{X_1X_2} & 1\end{bmatrix}$$ olduğundan $a=r_{YX1}$, $b=r_{YX2}$,

$c=r_{X1X2}$ alınarak (burada bütün variyanslar 1 olarak alındığından a,b,c sayıları hem korelasyon hem de kovariyans değerlerini gösterirler) denk.(5.22) için ($k=1/1-r_{X2X1}{}^2$)

$$[a\ \ b]\,k\begin{bmatrix}1 & -c\\ -c & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=[a\ \ b]\begin{bmatrix}k & -kc\\ -kc & k\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\frac{1}{1-c^2}\left[(a-bc)\ x_1+(b-ac)\ x_2\right]$$

$$\boxed{g(x_1,x_2)=g(\mathbf{x}^2)=(1/1-r_{X1X2}{}^2)[x_1(r_{YX1}-r_{YX2}r_{X1X2})+x_2(r_{YX2}-r_{YX1}r_{X1X2}]^{***} \qquad (5.23)}$$

Denk.(5.23) verilen bir $\mathbf{X}=(X_1, X_2)'$ değeri için KBD yi hesaplama olanağı sağlar. Denk.(5.23) deki g fonksiyonu $R^2 \rightarrow R$ ye bir fonksiyondur. Yani p=2, r=1 koşuluna uygundur. Burada $C_2=E(f\text{-}g)^2$ nin hesaplanabilmesi için bütün $\mathbf{X}=(X_1, X_2)'$ değerleri üzerinden integral alınması gereklidir. Teorik olarak doğru olan bu olanağın bilgisayara uygulanabilmesi için E (beklenen değer) nin istatistiksel kestirimci analogu kullanılacaktır. Kuşkusuz aynı kestirim C_1 bileşeni için de yapılmak zorundadır.

5.3.4 C_1 ve C_2 bileşenlerinin örneklem analoglarının (kestirimciler) türetilmesi

Daha önce de belirtildiği gibi C_1 ve C_2 bileşenleri integral formda verildikleri için bilgisayarda doğrudan hesaplanma olanakları yoktur. Bizim yapabileceğimiz; n tane sınırlı sayıda gözlem için C_1 ve C_2 bileşenlerinin kestirimcilerini kullanmaktır. Bunlar sırasıyla C_{1n} ve C_{2n} ile gösterilsinler.

Tanım: $C_{1n}=\frac{1}{n}\sum_{i=1}^{n}(y_i-g_i)^2 \quad C_{2n}=\frac{1}{n}\sum_{i=1}^{n}(f_i-g_i)^2$ (5.24a)

Burada y_i , g_i ,f_i sırasıyla i.örneklem için hedef vektörünün, koşullu beklenen değerin ve ağ outputunun aldığı değerlerdir. y_i değeri önceden bilinmektedir. g_i değeri ise üretilen $\mathbf{X}_i=(X_{1i},X_{2i})'$ vektöründen denk.(5.23) yardımıyla elde edilebilir. f_i değeri ise i tane örneklem üzerinden λ_i hata fonksiyonunu minimize eden $\mathbf{w}_i$ ağırlık vektöründen hesaplanır. Bu düşünüşün ışığında λ_n fonksiyonu büyük n ler için

$\lambda_n = C_{1n} + C_{2n}$ (büyük n ler için) bağıntısıyla verilir. Herhangi bir n için ise

$$\lambda_n = C_{1n} + C_{mn} + C_{2n} \tag{5.24b}$$

ile verilir. Burada C_{mn} C_m bileşeninin n için değeridir. C_m bileşeni ise Kesim (2.4) de $\lambda(\mathbf{w})$ için yazılan ifadede $2E([g(\mathbf{X})-f(\mathbf{X}, \mathbf{w})][Y-g(\mathbf{X})]$ ile verilen ortadaki terimi anlatır. Ancak bu terim sıfırdır ve simülasyon deneylerinin gösterdiği gibi C_2 bileşenine oranla daha çabuk sıfıra gitmektedir (n $\approx$20 değerinde limit değeri olan sıfıra oldukça yakındır). Bu nedenle bizim ilgi odağımız C_{1n}, C_{2n} ve $C_{12n}=C_{1n}+C_{2n}$ ile tanımlı reel sayı dizileri üzerine olacaktır. $\lim\lambda_n = \lim C_{12n}$ olacağı açıktır. Buna göre

$$C_{12n} = \frac{1}{n}\sum_{i=1}^{n}(y_i - g_i)^2 + \frac{1}{n}\sum_{i=1}^{n}(f_i - g_i)^2 ;**** \lambda_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2 \tag{5.24c}$$

yazılabilir. Denk.(5.24c) bu tezin genel amaçları için son derece yararlıdır. Çünkü, λ_n, C_{1n}, C_{2n} ve C_{12n} dizilerinin herhangi bir n için davranışı bilgisayar programında kolaylıkla izlenebilir ve böylece yeterince büyük n için asimptotik davaranış açıklanabilir.

<u>λ_n, C_{1n}, C_{2n} ve C_{12n} dizilerinin bilgisayar programındaki yazımları</u>

Input ve hedef vektörlerini içeren $\mathbf{Z}$(r+p=3 boyutlu) bir ÇBND olsun (r=2 ve p=1 yani İUB2). $\mathbf{Z}=(X_1,X_2,X_3)'$, $(X_1,X_2)'$ (input) ve $Y=X_3$(hedef vektörü) olarak atanırlar. $\mathbf{Z}\sim$ÇBND olduğundan X_1,X_2 ve X_3 bileşenleri de tek boyutlu normal dağılıma sahiptirler. Daha önce belirtildiği gibi $\mu_1 = \mu_2 = \mu_3 = 0$ ve $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$ dir. $\mathbf{Z}_i=(X_{1i},X_{2i},X_{3i})'$ i.örnekleme karşılık gelmek üzere n tane örneklem matrisi elde edilir. Denk.(6.23) ve $Y=X_3$ olması gerçeğini kullanarak

$$C_{1n} = \frac{1}{n}\sum_{i=1}^{n}\left[f(X_{3i}, g(X_{1i}, X_{2i})\right]^2$$

ayrıca i.örneklem için optimum ağırlık vektörü $\mathbf{w}_i$ ise

$$C_{2n} = \frac{1}{n}\sum_{i=1}^{n}\left[f(X_{1i}, X_{2i}, \omega_i) - g(X_{1i}, X_{2i})\right]^2 ; \quad \lambda_n = \frac{1}{n}\sum_{i=1}^{n}(f(X_{1i}, X_{2i}, w_i) - X_{3i})^2 \text{ ve}$$

$C_{12n}=C_{1n}+C_{2n}$ toplamlar dizisi bilgisayar yardımıyla hesaplanabilir (Bu toplamların yazımı için Ek te verilen " PROGRAM BACKPROPAGATION " adlı ana programın hata bileşenleri kesimine bakınız. Programda $\lambda_n \to E$ olrak alınmıştır).

5.3.5 AF1 ve İUB2 (yani $n_v=(2,n_h,1)$) için asimptot altı sonuçların yeniden gözden geçirilmesi: Analitik türetme ve tartışma

Bu kesimde Bölüm 4 de elde edilen sonuçlar analitik bir çerçevede incelenir. $Z \sim N(\theta,\Sigma)$ rv si $Z_t \sim N(\theta_t,\Sigma_t)$ dönüşümüne uğrar. Böylce X_{1t}, X_{2t} ve $X_{3t}=Y_t$ bileşenlerinin hepsi [0,1] aralığında değerler alırlar. $(C_1)_t$ terimini şöyle elde ederiz. Denk(5.20b) C_1 için yazılmıştır. Buradaki σ_{ij} kovariyans değerlerini uygun şekilde $(\sigma_{ij})_t$ için yeniden düzenlensek $(C_1)_t$ terimi elde edilir. Görüldüğü gibi $(C_1)_t$ teriminin θ_t vektörüne bağlılığı yoktur. Denk.(5.20b) den

$$(C_1)_t=\{(\sigma_{11})_t-k_t\,[(\sigma_{12})_t^2(\sigma_3)_t-2(\sigma_{12})_t(\sigma_{13})_t(\sigma_{23})_t+(\sigma_{13})^2(\sigma_{22})_t]\} \qquad (5.25)$$

$(\sigma_{ij})_t=(r_{ij})_t\,(var_i)^{1/2}{}_t(var_j)^{1/2}{}_t$ yazılabilir. $(r_{ij})_t=(r_{ij})_t$ ve $(var_i)_t=\alpha^2\,var_i$ $(i=1,2,\ldots,k)$ olduğu göz önünde tutulursa

$$(\sigma_{ij})_t=(r_{ij})_t\,(\alpha^2 var_i)^{1/2}{}_t(\alpha^2 var_j)^{1/2}{}_t=r_{ij}\alpha \qquad (var_i=var_j=1) \qquad (5.26)$$

elde edilir. Burada α katsayısı denk.(5.10) daki benzer olarak $\alpha=1/25$ olarak elde edilir. Denk.(5.26) daki sonuç denk.(5.25) de yerine konulursa

$$(C_1)_t=\left[r_{11}\alpha^2-\frac{1}{\alpha^4(r_{22}r_{33}-r_{23})^2}(r_{12}{}^2\alpha^4 r_{33}\alpha^2-2\alpha^6 r_{12}r_{23}r_{13}+\alpha^6 r_{13}{}^2 r_{22}\right]=\alpha^2\,(C_1) \quad (5.27)$$

Denk.(5.27) elde edilirken denk.(5.20b) de $r_{ij}=\sigma_{ij}$ olduğu gerçeği kullanılmıştır. Denk.(6.27) yi Bölüm 5 deki sonuçlara uygulayabilmek için $C_1{}^t \to \sqrt{C_1{}^t}$ dönüşümünü yapmak gerekir. çünkü Bölüm 4 deki hata fonksiyonu $\{(1/n)\Sigma\,(y_i-f_i)^2\}^{1/2}$ ile verilir. Böylece,

$$\sqrt{C_1{}^t}=(\alpha^2\,(C_1))^{1/2}=(1/5)\,(C_1)^{1/2} \qquad (5.28)$$

olarak elde edilir. Denel verilerin ışığında $\sqrt{C_1{}^t}\approx 0$ kabul edilerek

$$\sqrt{\lambda^t}\to\sqrt{C_1{}^t}\ \text{ya da}\ \sqrt{\lambda_n{}^t}\to\sqrt{(C_1)_n{}^t}=(1/5)(C_1)_n \qquad (5.29)$$

bulunur. Denk.(5.29) eğer n yeterince büyükse $\sqrt{\lambda_n{}^t}\approx(1/5)\,C_1$ gibi değerlendirilebilir. Ancak Bölüm 4 deki n değerleri $n\leq 50$ dir. n yeterince büyük olmasada elde edilen sonuçları teorik (asimptotik limitte) sonuçlarıyla karşılaştırmak ilginç olacaktır. Bunun için (5.29) denklemini

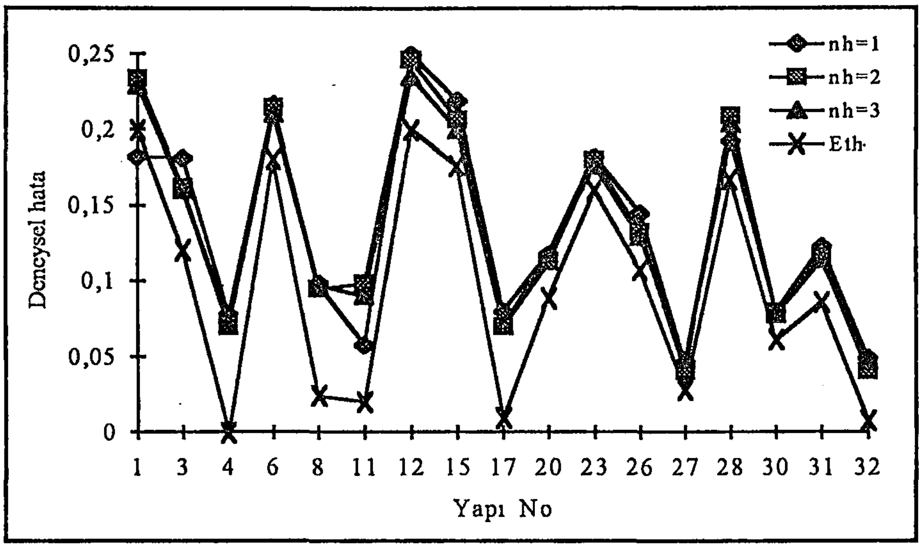
$$\sqrt{\lambda_n{}^t} = \frac{1}{5}\sqrt{C_1} = \frac{1}{5}\left[1 - \frac{1}{(1-z^2)}(x^2 - 2xyz + y^2)\right]^{\frac{1}{2}} \tag{5.30}$$

Burada x,y,z daha önceki gibi $\sigma_{12}=r_{12}=x$, $\sigma_{13}=r_{13}=y$, $\sigma_{23}=r_{23}=z$ atamalarından oluşur. Denk.(5.30) $z=1$ için sonsuz olacağı için biz $z=1$ yerine $z=1-0.01=0.99$ alacağız. Tablo 5.5 de Bölüm 4 de kullanılan korelasyon üçlülerinin (x,y,z) denk.(5.30) da elde edilen $\sqrt{\lambda_n{}^t}$ değerleriyle Bölüm 5 de asimptotik incelemeler için kullanılan korelasyon üçlülerinin denk.(5.21) ile elde edilen C_1 teorik değerleri verilmiştir.

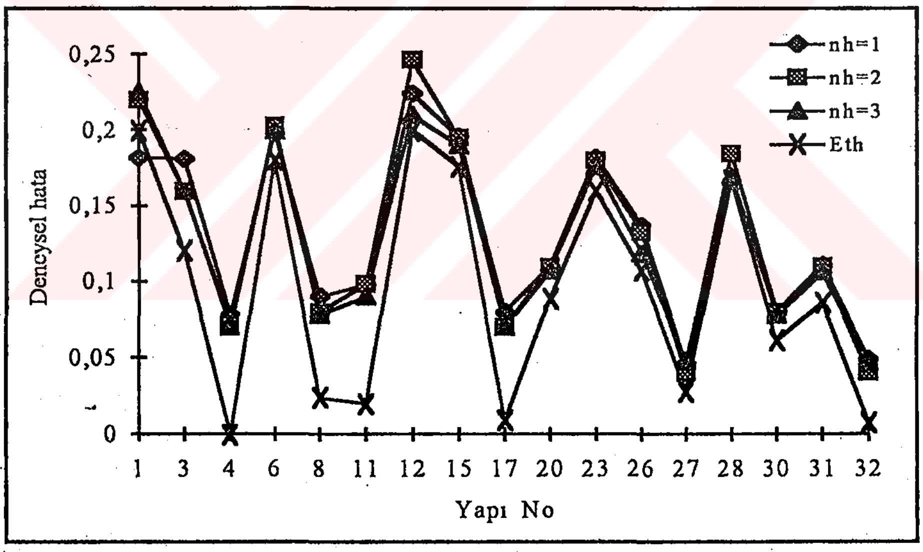
Tablo 5.5 Asimptot altı ve asimptotik bölge için bu çalışmada kullanılan korelasyon yapıları ve C_1 bileşeninin değerleri . Tabloda $1/5\ (C_1)^{1/2}$ için denk.(5.30) ve C_1 için denk.(5.21) kullanılmıştır.

Asimptot altı bölge için (Bölüm 4) AF1 ile						Asimptotik Bölge için (Bölüm 5) AF2 ile		
Yapı No	$1/5\ (C_1)^{1/2}$	Yapı No	$1/5\ (C_1)^{1/2}$	Yapı No	$1/5\ (C_1)^{1/2}$	Yapı No	(x,y,z)	C_1
1	*0.200*	12	*0.200*	23	*0.160*	1	(0.0,0.0,0.2)	*0.960*
2	*0.183*	13	*0.120*	24	*0.058*	2	(0.0,0.0,0.6)	*0.640*
3	*0.120*	14	*0.180*	25	*0.089*	3	(0.0,0.0,1.0)	*0.000*
4	*0.000*	15	*0.176*	26	*0.107*	4	(1.0,0.0,0.0)	*1.000*
5	*0.200*	16	*0.119*	27	*0.028*	5	(0.0,1.0,0.0)	*0.000*
6	*0.180*	17	*0.091*	28	*0.167*	6	(1.0,0.6,0.6)	*0.638*
7	*0.098*	18	*0.149*	29	*0.120*	7	(0.8,0.8,0.8)	*0.289*
8	*0.024*	19	*0.181*	30	*0.061*	8	(0.4,0.4,1.0)	*0.000*
9	*0.200*	20	*0.089*	31	*0.086*	9	(1.0,1.0,1.0)	*0.005*
10	*0.149*	21	*0.196*	32	*0.008*			
11	*0.020*	22	*0.183*					

Şekil 5.1 ve 5.2 asimptot altı bölgede Bölüm 4 de $n_v=(2,n_h,1)$ ($n_h=1,2,3$) için elde edilen ampirik sonuçların bu bölümdeki teorik değerlendirmelerin ışığında yeniden gözden geçirilmeleriyle elde edilmişlerdir. $\overline{\lambda}_{100}$ (Hata) sı ve $E_{th}=\lambda_{th}$ için şekil altındaki açıklamaya bakınız. Şekil 5.1 de n=20 için $\overline{\lambda}_{100}$ değerleri verilen herhangi bir n_h için $_{th}$ değerleriyle aynı eğilimi gösterirler. Ayrıca n_h değeri arttıkça bu eğilim (genel olarak) daha da belirgindir. Şekil 5.2 de ise n, gözlem sayısı, =50 dir. Burada da aynı eğilimler gözlenir. Ancak bu kez $\overline{\lambda}_{100}$ (Hata) değerleri λ_{th} hata değerlerine daha yakındır (bkz. Yapı No:6,11,12,26). Bu ise gözlem sayısı arttıkça daha iyi fit'lerin (yaklaşımların) olacağına ilişkin teorik beklentilerine uygun bir davranıştır. Bir not olarak, n_h değerinin artırılmasının yaklaşımları birazcık iyileştirdiğini, ancak teori limit (E_{th}) nedeniyle dramatik iyileştirmelerin olmadığı bir kez daha vurgulanmalıdır.

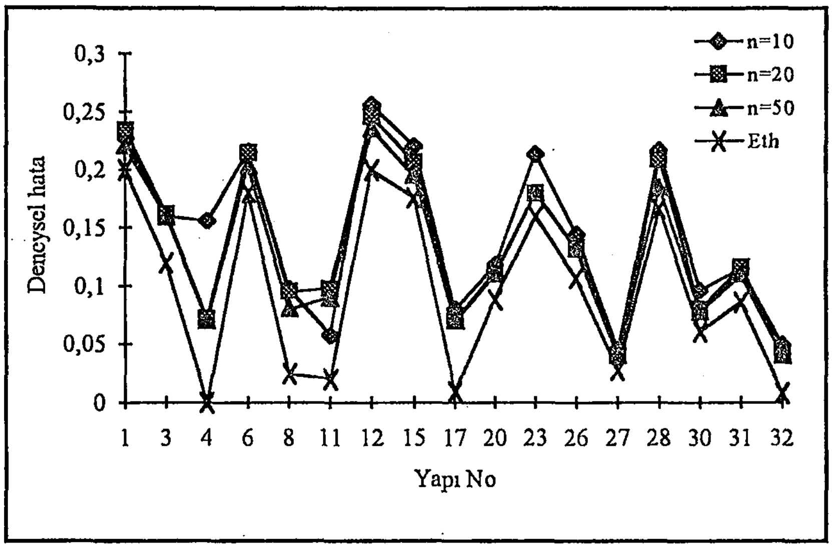


Şekil 5.1 $\overline{\lambda}_{100}$,(Deneysel hata)'nın Yapı No. ya göre artan ağ kompleksitesi ile değişimi. Burada 100 alt indisi aynı Yapı No.dan çekilen 100 tane örneklem matrisi üzerinden alınan hata ortalamasını anlatır (yani n_i=100). n(gözlem sayısı)=20 dir. $E_{th}=\lambda_{th}$ teorik hata değerleri olup Tablo 5.5 deki 1/5 $(C_1)^{1/2}$ değerlerine karşılık gelir. Not: Şekil 5.1 de Tablo 5.5 deki Yapı No.ların bazılarının atlandığına dikkat edlimelidir.



Şekil 5.2 $\overline{\lambda}_{100}$, (Deneysel hata)'nın Yapı No. ya göre artan ağ kompleksitesi ile değişimi. n(gözlem sayısı)=50 dir. $E_{th}=\lambda_{th}$

Asimptot altı bölge için son olarak Şekil 5.3 de artan n ile ağın davranışının ne olacağı (sabit n_h için, n_h=2) verilmiştir. Yapı No.ların hepsi için, n arttıkça $\overline{\lambda}_{100}$ değerleri teorik değerlerine giderek daha çok yaklaşmaktadır. Bu durum artan gözlem sayısının daha iyi tahminlere olanak vereceği yolundaki beklentiyle uyum içindedir. Ayrıca burada sunulmasada n_h=1 ve 3 değerleri için de Şekil 5.3 deki davranışın gözlendiğini belirtelim.



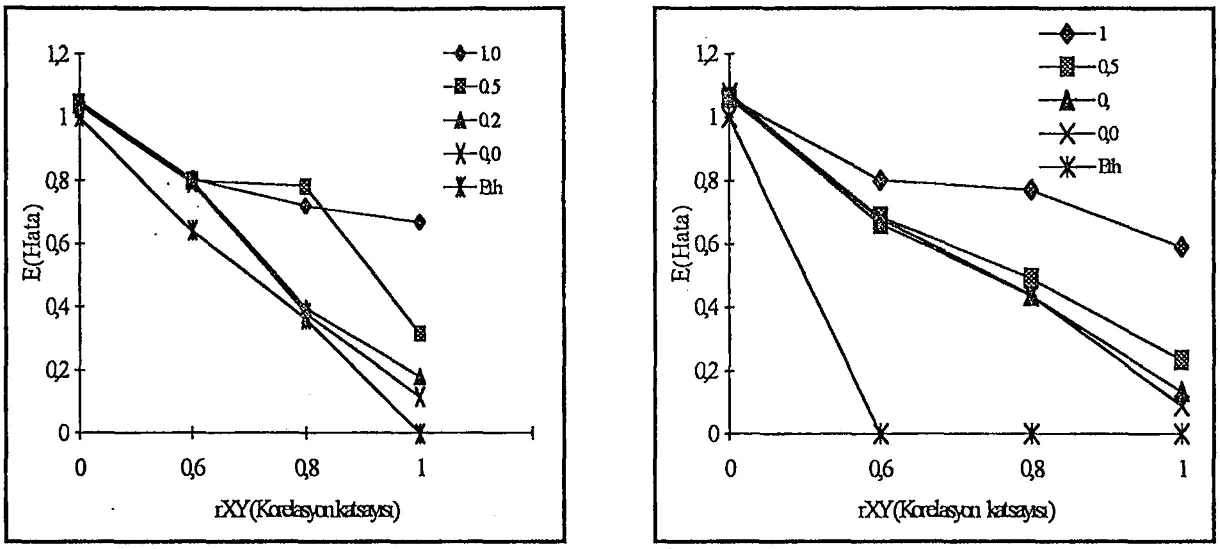
Şekil 5.3 λ_{100} (Deneysel hata)'nın sabit ağ kompleksitesi (n_h=2) altında artan gözlem sayısı,n, için davranışı.

Özetle, asimptot altı bölgede AF1 ile sınırlı sayıda gözlem için elde edilen hata fonksiyonu değerleri teorik olarak hesaplanan sonuçlarla belirli derecede uyuşum içindedir. Ancak yaklaşımların daha iyi olabilmesi için asimptotik bölgeye geçilmelidir.

5.4 Öğrenme adımı (η) ve iterasyon sayısı (niter) nın saptanması üzerine bazı deneysel sonuçlar

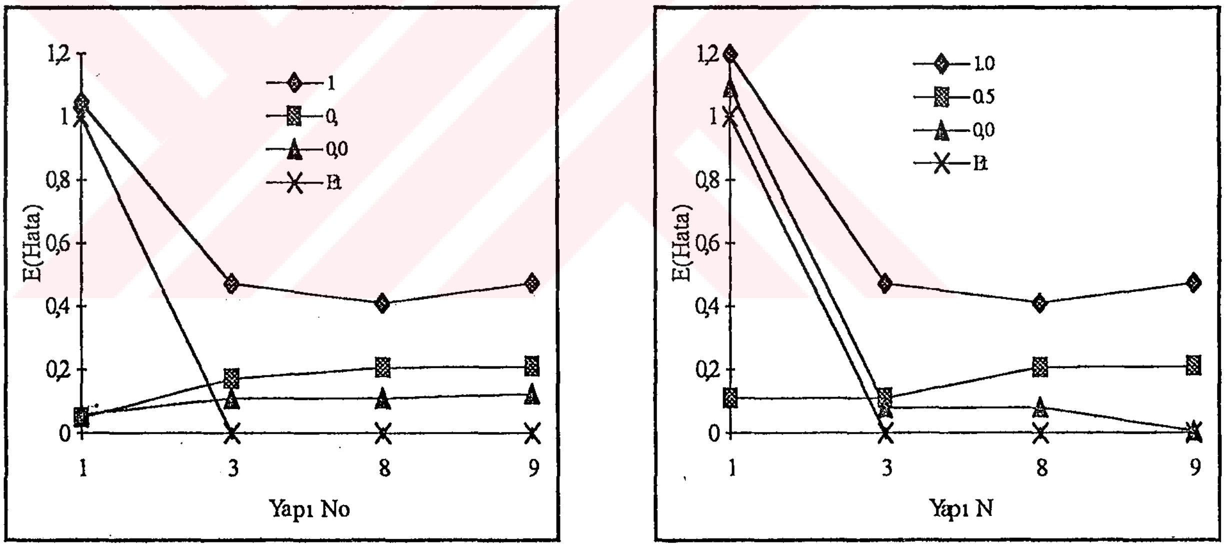
Bu kesimde η ve niter'in deneysel olarak optimum saptanması üzerine bazı sonuçlar verilir. AF2 fonksiyonu kullanıldığında λ_n hata metriğinin denetlenebilmesi için η nın görece küçük seçilmesi gerekir. Aksi durumda λ_n minimuma yaklaşmadan, minimumdan atlama durumu ortaya çıkabilir. niter (verilen bir n için) λ_n in minumum yapılması gereken iterasyon (ağırlık modifikasyonu) sayısıdır. n arttıkça niterin de artması gerektiği düşünülebilir (YSA üzerine yapılmış çok sayıdaki simülasyon deneyleri bunu doğrulamaktadır). Bunun için niter=(m)x(n) {m=20,30,40} olarak alınır. Burada m bir katsayıdır. Örneğin m=20 ve n=120 ise niter=20x120=2400 olur. Yani 2400 kez ağırlık modifikasyonu yapılmalıdır. Burada şunu da belirtmek zorundayız. Bu kesimde elde edilen sonuçlar gerek η nın gerekse niterin optimum olarak saptandığı anlamına gelemez, çünkü bizim deneylerimiz sınırlı sayıda η ve niter için yapılmıştır. Ancak görsel verinin (şekiller, grafikler) çokluğu nedeniyle burada sunulmayan verileri de göz önüne alarak bu kesimde elde etttiğimiz sonuçların yine de belirli düzeyde bir bilgi verdiği söylenebilir. Ayrıca bu tezin temel amacı optimum η ve niteri saptamak olmayıp bu kesimde ki deneyler *yalnızca belirli bir öngörü* sağlamak için yapılmışlardır. Şekil 5.4 ve Şekil 5.5, n=120 ve n=200 gözlem sayıları kullanılarak çeşitli η değerleri için λ_n deneysel hatanın korelasyon katsayısına (İUB1 için r_{XY} ve İUB2 için Yapı No.) karşı evrimini gösterir. Görüldüğü gibi η=0.05 için λ_n değeri λ_{th} (=λ) değerine öteki η değerlerine oranla daha yakındır {bu sonuç deney öncesi (a priori) beklentilerle uyumludur} ve bu nedenle Bölüm 5 deki deneylerin tümü için η=0.05 alındı.

5.4.1 η(öğrenme adımı) için bazı sonuçlar



a b

Şekil 5.4 Çeşitli η (=1.0,0.5,0.2,0.05) için E=λ_n (Hata) nın r_{XY} (Korelasyon katsayısı) na karşı evrimi. λ_{th} denk.(5.9b) den elde edilen teorik değerleri gösterir. (Aktivasyon fonksiyonu:AF2 ve n_h=2). **a**: n=120 , **b**:n=200

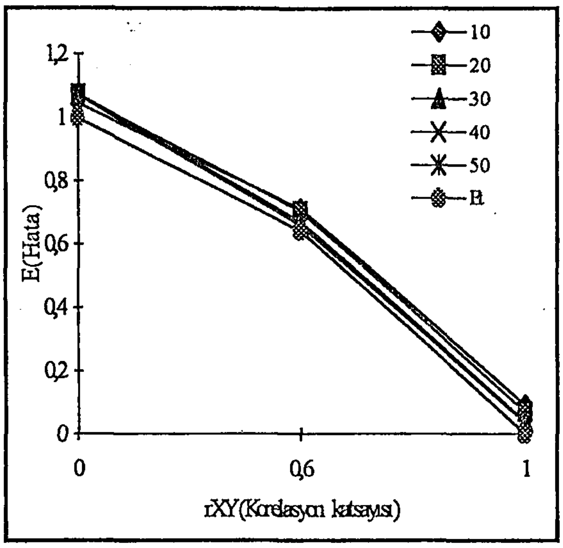


a b

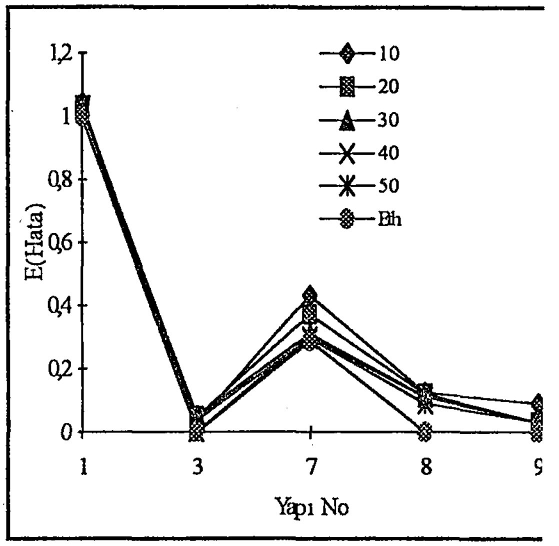
Şekil 5.5 Çeşitli η (=1.0,0.5,0.05) için E=λ_n (Hata) nın Yapı No' ya karşı evrimi. λ_{th}=C_1 değerleri Tablo 5.5 den elde edilmiştir. (Aktivasyon fonksiyonu:AF2 ve n_h=2). **a**: n=120 , **b**:n=200

5.4.2 niter için bazı sonuçlar

Şekil 5.6 ve Şekil 5.7 iterasyon sayısı çarpanı niter(=m=10,20,30,40,50) için İUB1 ve İUB2 kompleksiteleri için λ_n hata metriğinin evrimini gösterir. m=20 ile m=50 arasında dramatik bir fark gözlenmediği için m=20 olarak alındı.



Şekil 5.6 Çeşitli niter çarpanları için (10,20,30,40,50) $E=\lambda_n$ in r_{XY}' ye karşı evrimi. (n_h=2, n=200). η=0.05. λ_{th} denk 5.9b den elde edilen değerleri gösterir.



Şekil 5.7 Çeşitli niter çarpanları iç (10,20,30,40,50) $E=\lambda_n$ in Yapı No'ya ka evrimi. (n_h=2, n=200). η=0.05. λ_{th} teorik ha değeridir.

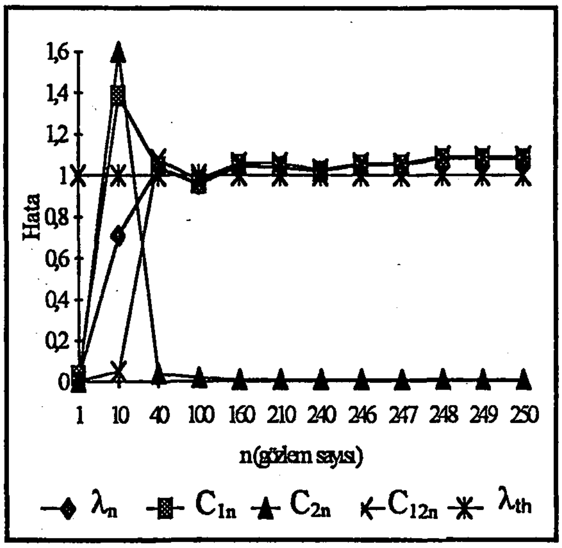
5.5 AF2 li YSA da İUB1 ve İUB2 için Asimptotik Sonuçlar

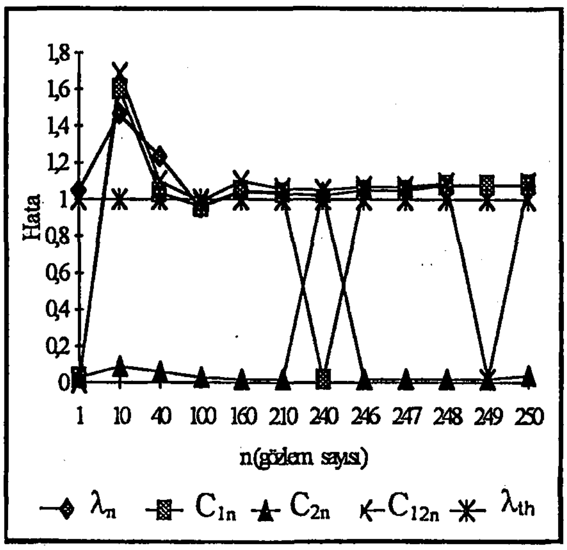
Bu kesimden başlayarak λ_n hata metriğinin evrimi büyük n ler için (n=250) araştırıl Böylece denk.(5.9b) ve denk.(5.21) ile verilen teorik limit değerlerinin sınanması olanal oldu. İUB1 için r_{XY}={0.0,0.2,0.4,0.6,0.8,1.0} ve İUB2 için Tablo 5.5 de AF2 için veril yapılar kulllanıldı. Ancak sonuçların sunulmasında bazı korelasyon yapıları elendi elenen yapılar için hatanın evrimine ilişkin sonuçlar sözlerle (şekillerle desteklenmeksizi verildi.

♠**Veri sunumunda eliminasyonun temel nedeni** : seçme kümesindeki parametrelerin çokluğudı Bunu sayısal bir örnekle değerlendirmek uygun olacaktır. n_h=1,2,3,4 olsun. (x,y, korelasyon üçlüleri sayısı =9. Dört değişik başlangıç koşulu iw={1,2,3,4} olsun. Böyle 4x9x4=144 adet , aynı işlemlerle (x,y) korelasyon ikilileri için de yapılırsa 4x6x4=9 Toplam olarak 144+96= **240** adet şekil çizilmesi gerekir. Bu ise böylesi biz te sınırlarını gereksiz yere zorlamak olacağından eliminasyon zorunludur.

5.5.1 İUB1 için asimptotik sonuçlar ve tartışma

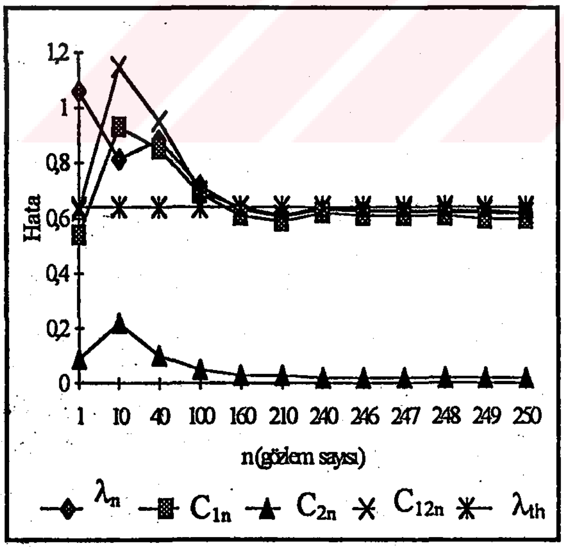
Şekil 5.8-5.11 de çeşitli r_{XY} katsayıları için çeşitli hata bileşenlerinin asimptotik evrimi değişimi sergilenmektedir (n_h=1 ve iw=1,2,4).

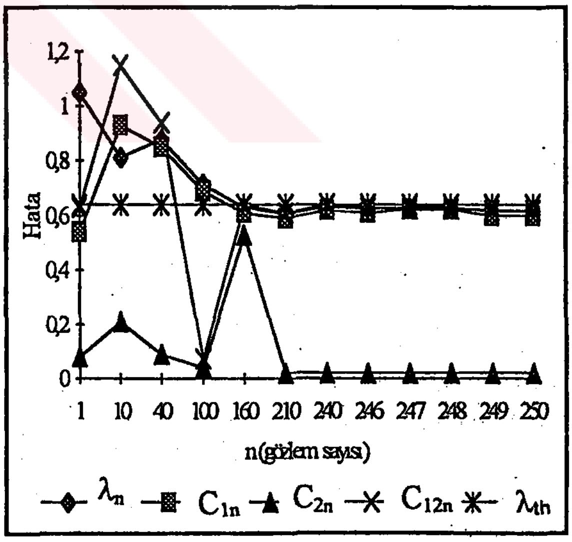




Şekil 5.8a (iw=1 ve n_h=1) r_{XY}=0.0 için çeşitli hata bileşenlerinin (ya da kısaca hata) n (gözlem sayısı) için asimptotik evrimi. λ_n ; n. gözlem için hata. C_{1n} , C_{2n}, C_{12n} (tanım için bkz. Kesim 5.3.4) ise daha önce Kesim 5.2.2 de tanımlanan C_1, C_2, C_{12} hata bileşenlerinin n. örneklem üzerinden gerçekleşen hata değerleridir. λ_{th} her zamanki gibi teorik hata değeridir

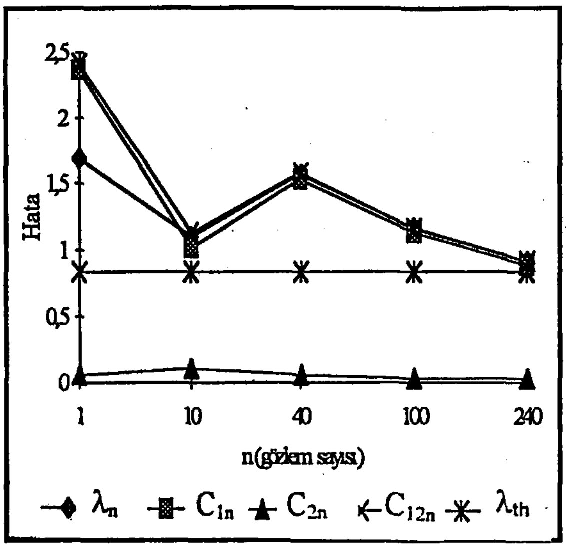
Şekil 5.8b (iw=2 ve n_h=1) r_{XY}=0.0 için çeşitli hata bileşenlerinin n (gözlem sayısı) için asimptotik evrimi



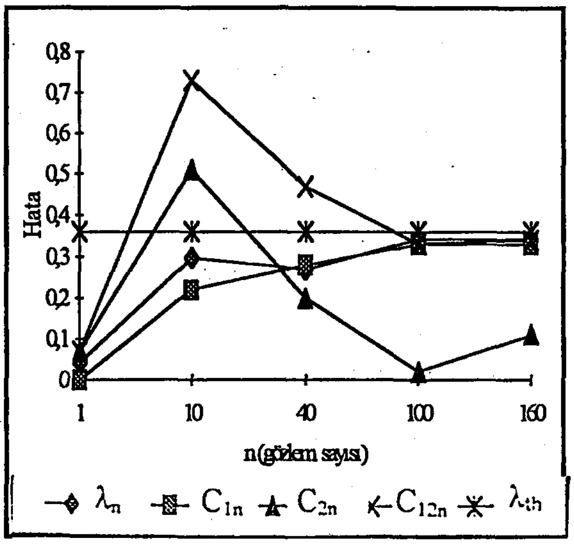


Şekil 5.9a (iw=1 ve n_h=1) r_{XY}=0.6 için çeşitli hata bileşenlerinin n (gözlem sayısı) için asimptotik evrimi

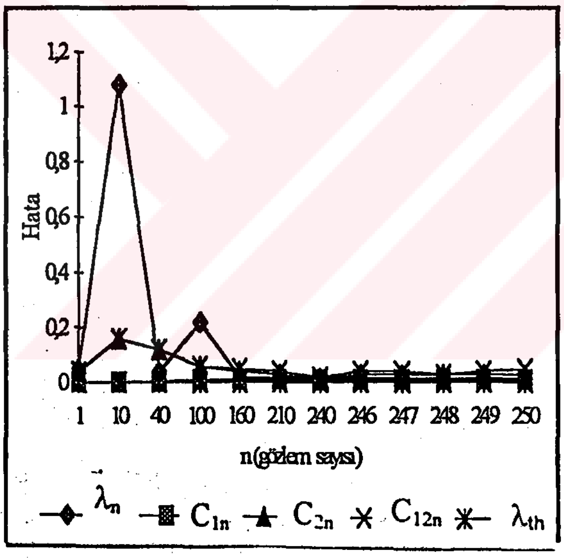
Şekil 5.9b (iw=2 ve n_h=1) r_{XY}=0.6 için çeşitli hata bileşenlerinin n (gözlem sayısı) için asimptotik evrimi.



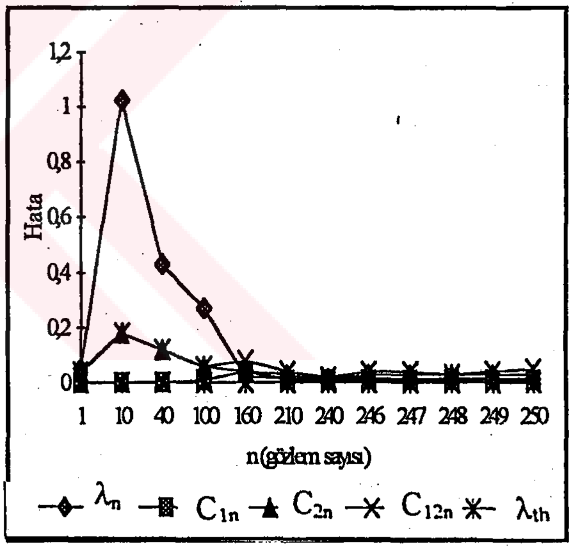
Şekil 5.10a (iw=1 ve n_h=2) r_{XY}=0.4 için çeşitli hata bileşenlerinin n (gözlem sayısı) için asimptotik evrimi



Şekil 5.10b (iw=1 ve n_h=2) r_{XY}=0.8 için çeşitli hata bileşenlerinin n (gözlem sayısı) için asimptotik evrimi.



Şekil 5.11a (iw=1 ve n_h=1) r_{XY}=1.00 için çeşitli hata bileşenlerinin n (gözlem sayısı) için asimptotik evrimi



Şekil 5.11b (iw=4 ve n_h=1) r_{XY}=1.00 için çeşitli hata bileşenlerinin n (gözlem sayısı) için asimptotik evrimi.

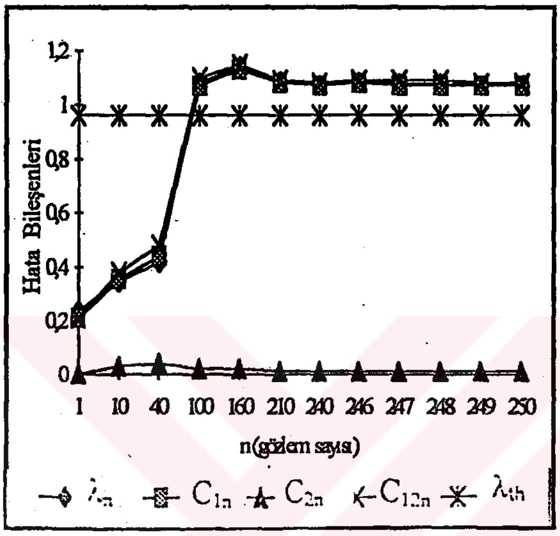
Şekil 5.8, 5.9, 5.10, 5.11 birlikte ele alınarak şu sonuçlara ulaşılabilir. C_{2n} terimi bütün r_{XY} korelasyon katsayıları için sıfıra yakınsar ve bu yakınsama genellikle çok hızlıdır (bütün örneklerde n=40 dan hemen sonra). C_{2n} teriminin sıfıra yakınsaması daha önce de vurgulandığı gibi ağın performansının iyi olduğu anlamına gelir. λ_n terimi bütün örneklerde n büyük sayılı değerler alırken (≥100) C_{1n} terimine ve λ_{th} değerine yakınsar. Bu durum teorik beklentilere uygundur. Ayrıca değişik başlangıç ağırlık vektörleri (iw indisi ile etiketlenen) için de ağın yönelimi teorik beklentilere uyar (örneğin Şekil 6.11a ve Şekil 6.11b yi karşılaştırırsak aynı eğilim açıkça görülür). İUB1 YSA'nın asimptotik davranışı n_h=3,4 değerleriyle de araştırıldı ve n_h=1 kompleksitesinde ki davranışa benzer durumlar gözlendi.

5.5.2 İUB2 için asimptotik sonuçlar ve tartışma

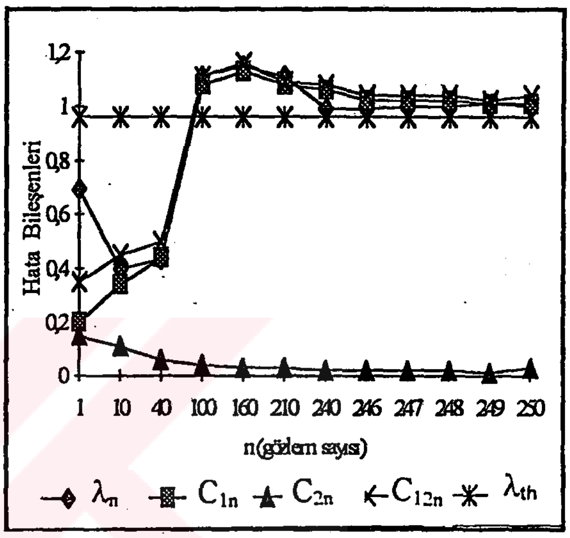
5.5.2.1 Verilen bir korelasyon yapısı ve hata bileşenlerinin evrimi

Yapı No:1 Korelasyon yapısı (x,y,z)=(0.0, 0.0, 0.2)

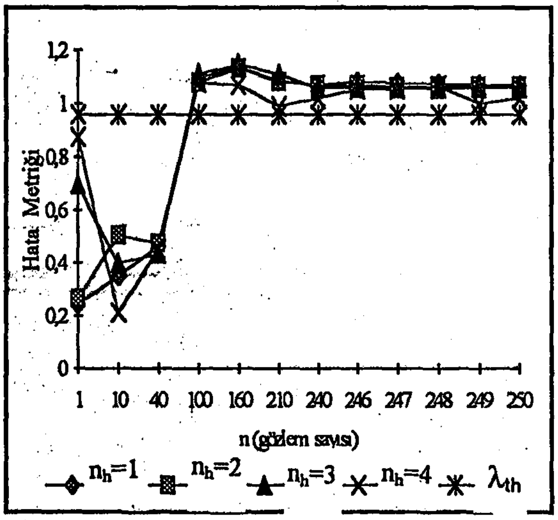
Şekil 5.12a ve 5.12b incelendiğinde λ_n ve C_{1n} hata bileşenlerinin büyük n için λ_{th} teorik hata değerine yakınsadığı görülür. Ayrıca C_{2n} değeri görece hızlı (n=40 ve sonrası) olarak sıfıra yakınsar. Bütün bu durumlar teorinin öngörüsüyle tümüyle uyumludur. Şekil 5.12c de artan n için λ_n nin evrimi unulmuştur. Ağ kompleksitesi n_h arttıkça $|\lambda_n - \lambda_{th}|$ biraz daha azalmaktadır. Ancak bu iyileşme daha önce de vurgulandığı gibi dramatik olmayıp son derece kısıtlıdır. Şekil 5.12d çeşitli iw için sabit n_h altında λ_n nin davranışını sergiler. Bütün başlangıç vektörleri için λ_n nin λ_{th} değerini yakınsaması öz olarak aynıdır.



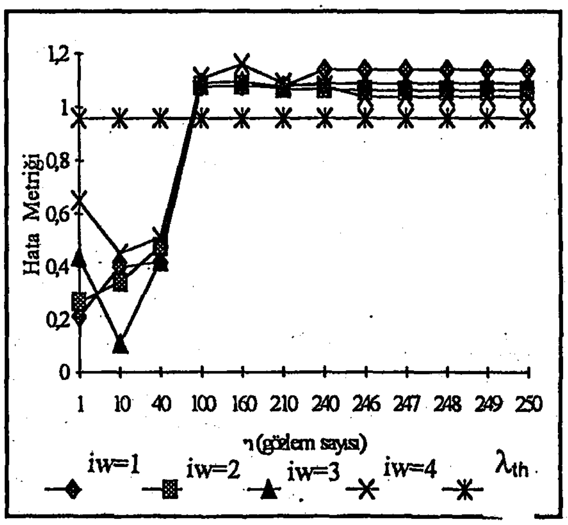
Şekil 5.12a Yapı No.(YN)=1 için çeşitli hata bileşenlerinin büyük n için evrimi (ıw=1, n_v=(2,1,1)).



Şekil 5.12b. YN=1 için çeşitli hata bileşenlerinin büyük n için evrimi. (iw=3, n_v=(2,3,1)).



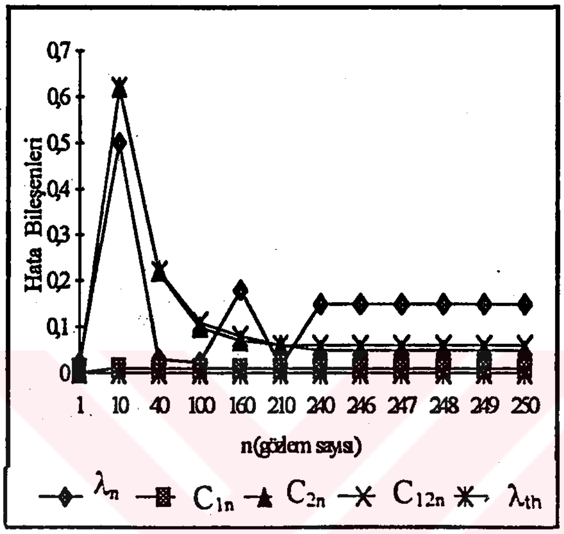
Şekil 5.12c YN=1 için hata metriğinin (λ_n) artan ağ kompleksitesi (n_h) altında büyük n için evrimi (iw=2).



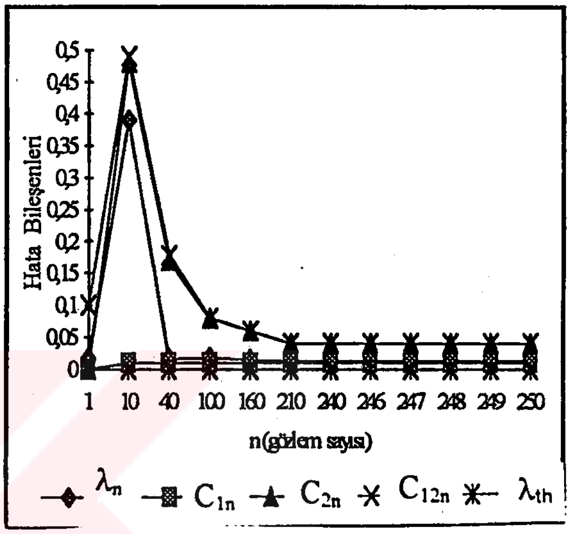
Şekil 5.12d YN=1 için hata metriğinin (λ_n) sabit ağ kompleksitesi (n_h=2) için çeşitli başlangıç vektörleri altında büyük n için davranışı

Yapı No: 3 (x,y,z)=(0.0,0.0,1.0)

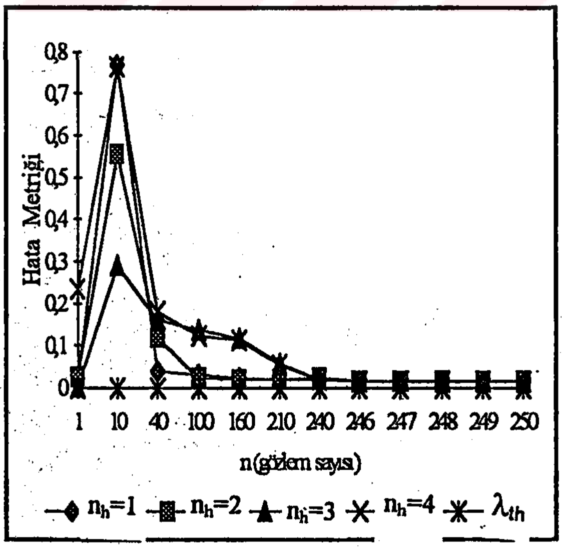
Şekil 5.13a ve 5.13b bazı gözlemlere izin verir. λ_n ve C_{1n} hata bileşenlerinin büyük n için λ_{th} teorik hata değerine yakınsar. Ayrıca C_{2n} değeri görece hızlı (n=40 ve sonrası) olarak sıfıra yakınsar. Bütün bu durumlar YN=1 yapısında olduğu gibi teorinin öngörüsüyle tümüyle uyuşur. Şekil 5.13c de artan n için λ_n nin evrimi izlenmektedir. Ağ kompleksitesi n_h arttıkça $|\lambda_n - \lambda_{th}|$ biraz daha azalmaktadır. Ancak bu iyileşme son derece kısıtlıdır. Şekil 5.13d çeşitli iw değerleri için sabit n_h li ağda λ_n nin davranışını sergiler. Bütün başlangıç vektörleri için λ_n eninde sonunda λ_{th} değerine yakınsar..



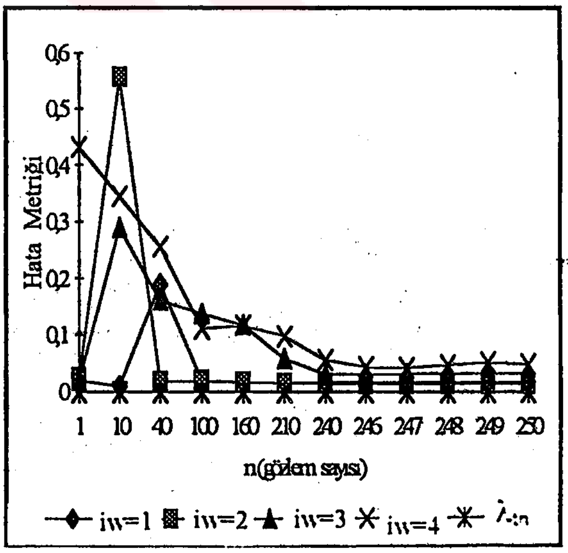
Şekil 5.13a YN=3 için çeşitli hata bileşenlerinin büyük n için evrimi (ıw=1, n_v=(2,1,1)).



Şekil 5.13b. YN=3 için çeşitli hata bileşenlerinin büyük n için evrimi. (iw=3, n_v=(2,3,1)).



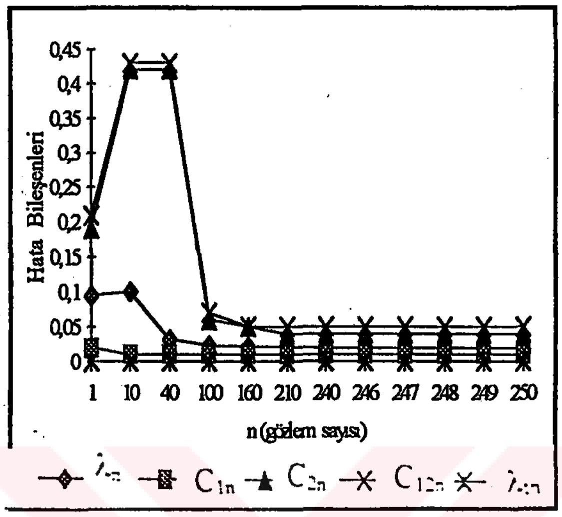
Şekil 5.13c YN=3 için hata metriğinin (λ_n) artan ağ kompleksitesi (n_h) altında büyük n için evrimi (iw=2).



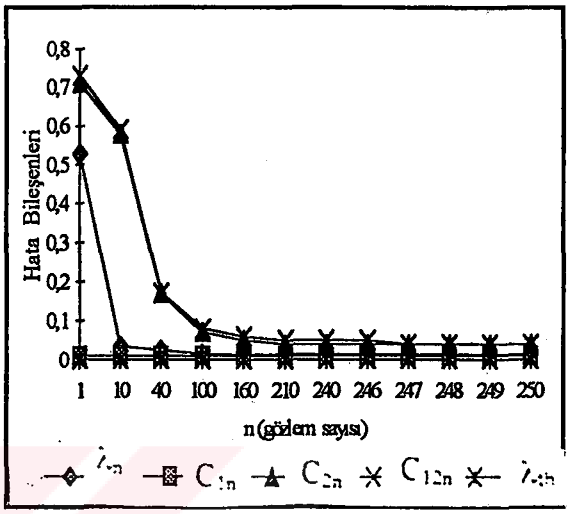
Şekil 5.13d YN=3 için hata metriğinin (λ_n) sabit ağ kompleksitesi (n_h=2) için çeşitli başlangıç vektörleri altında büyük n için davranışı

Yapı No:5 (x,y,z)=(0.0,1.0,0.0)

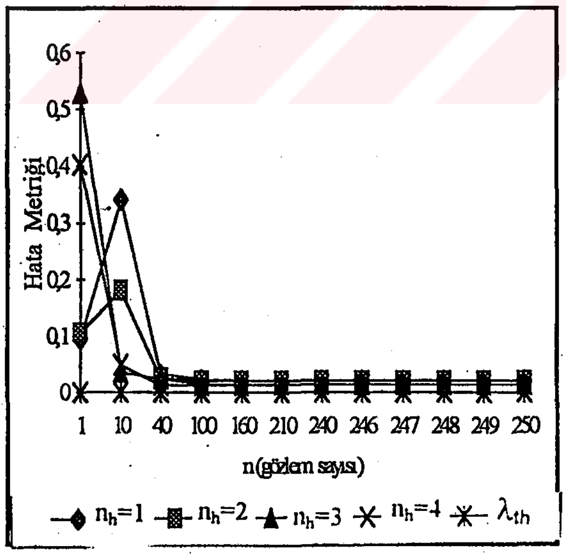
Şekil 5.14a ve 5.14b λ_n ve C_{1n} bileşenlerinin YN=5 için de büyük n değerleri için λ_{th} teorik hata değerine (0.00, bkz. Tablo 5.5) yakınsadığının bir kanıtıdır. Şekil 5.14c de artan n_h ile birlikte λ_{th} değerine biraz daha iyi yaklaşım yapılabildiğini gösterir. Şekil 5.14d bütün iw ler için λ_n hata metriğinin kendi teorik değerine er geç yakınsadığını vurgular.



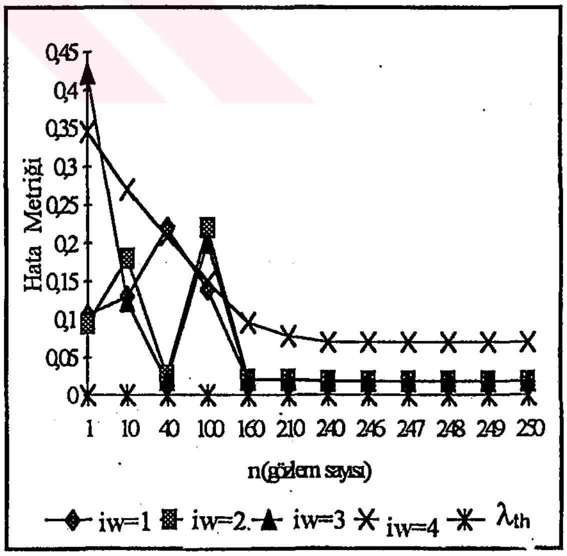
Şekil 5.14a YN=5 için çeşitli hata bileşenlerinin büyük n için evrimi (ıw=1, n_v=(2,1,1)).



Şekil 5.14b. YN=5 için çeşitli hata bileşenlerinin büyük n için evrimi. (iw=3, n_v=(2,3,1)).



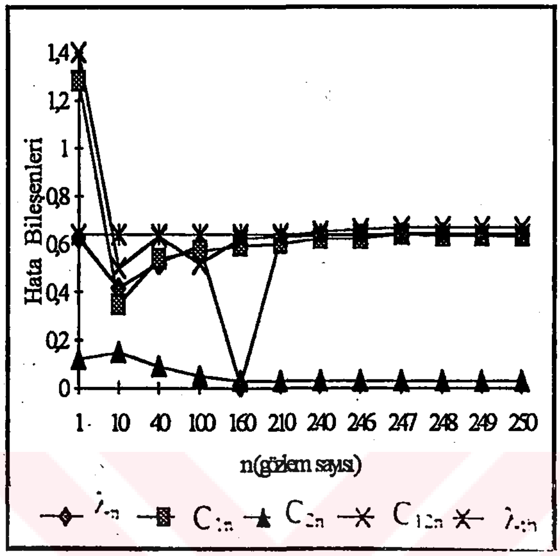
Şekil 5.14c YN=5 için hata metriğinin (λ_n) artan ağ kompleksitesi (n_h) altında büyük n için evrimi (iw=2).



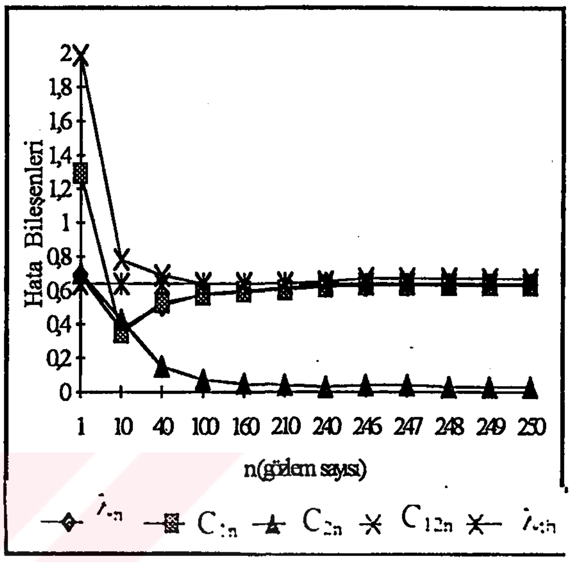
Şekil 5.14d YN=5 için hata metriğinin (λ_n) sabit ağ kompleksitesi (n_h=2) için çeşitli başlangıç vektörleri altında büyük n için davranışı

Yapı No:6 (x,y,z)=(1.0,0.6,0.6)

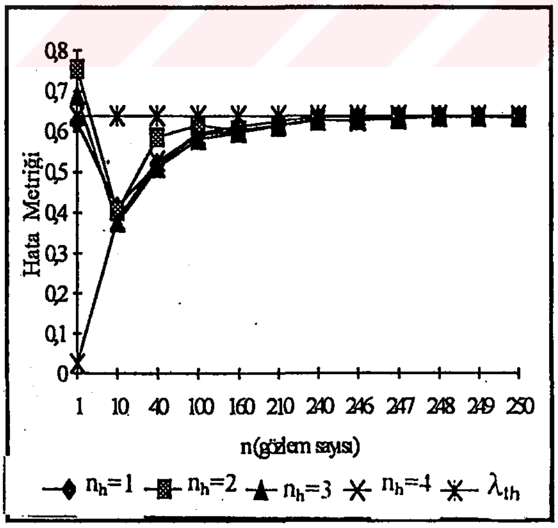
Şekil 5.15a ve Şekil 5.15b YN=6 için çeşitli hata bileşenlerinin büyük n için (asimptotik) evriminin teorik beklentilerle elele uyumlu olduğu sonucunu verir. Şekil 5.15c artan kompleksite altında λ_n hata metriğinin çok fazla değişmediğini anlatır. Şekil 5.15d ise değişik iw durumları için ağın davranışının özünü koruduğunu belirtir.



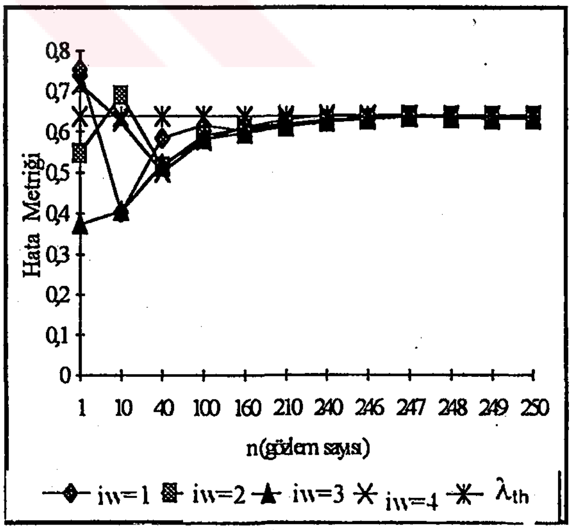
Şekil 5.15a YN=6 için çeşitli hata bileşenlerinin büyük n için evrimi (ıw=1, n_v=(2,1,1)).



Şekil 5.15b. YN=6 için çeşitli hata bileşenlerinin büyük n için evrimi. (iw=3, n_v=(2,3,1)).



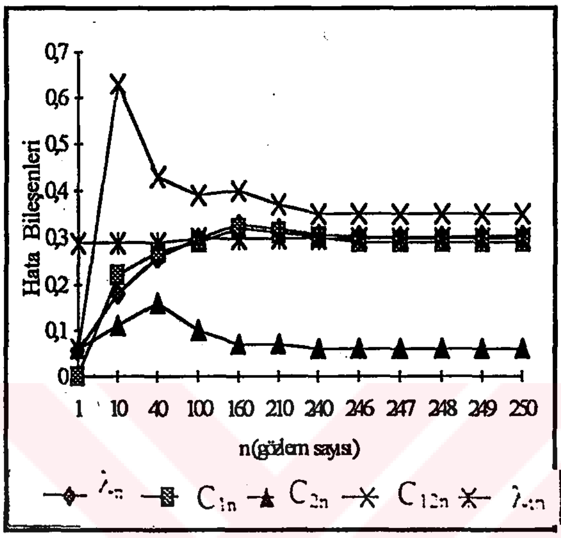
Şekil 5.15c YN=6 için hata metriğinin (λ_n) artan ağ kompleksitesi (n_h) altında büyük n için evrimi (iw=2).



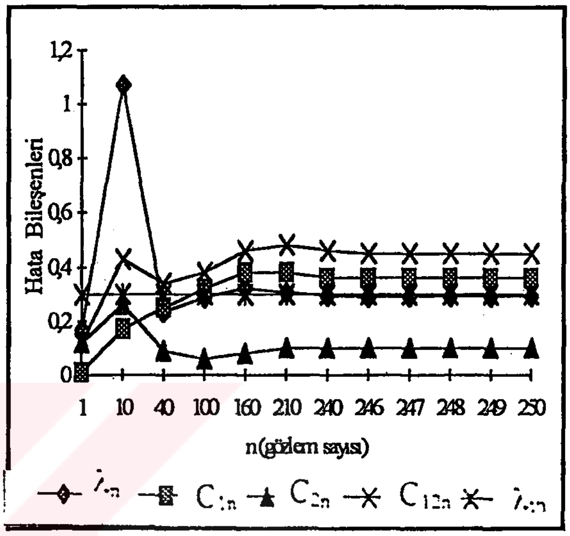
Şekil 5.15d YN=6 için hata metriğinin (λ_n) sabit ağ kompleksitesi (n_h=2) için çeşitli başlangıç vektörleri altında büyük n için davranışı

Yapı No:7 (x,y,z)=(0.8,0.8,0.8)

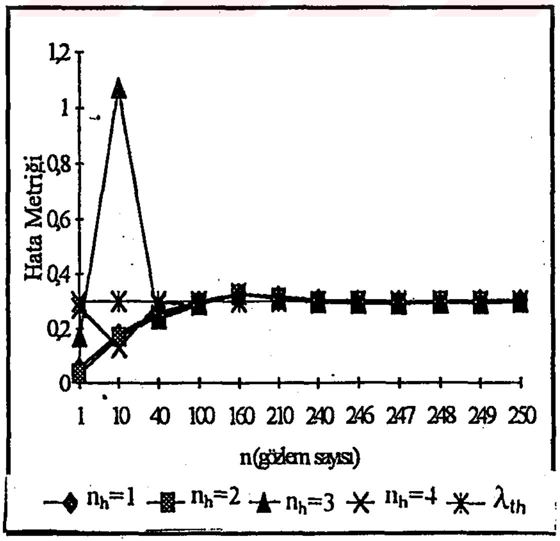
Şekil 5.16a ve 5.16b bu korelasyon yapısı için de ağın kendisinden beklenileni hemen hemen yerine getirdiğini gösterir. Şekil 5.16b ye bakılırsa C_{1n} teriminin teorik λ_{th} değerine yakınsamakta biraz geç kaldığı görülür. Bunun nedeni korelasyon yapısının görece karmaşık olmasıdır. Ayrıca daha çok sayıda iterasyon gerekir. Buna karşın genel eğilim teorik değere yakınsama yönündedir. Şekil 5.15c artan kompleksite ile birlikte sınırlı bir iyileşme olduğunu anlatır. Şekil 5.16d ise çeşitli başlangıç vektörleri için asimptotik eğilimin özdeş olduğunu vurgular.



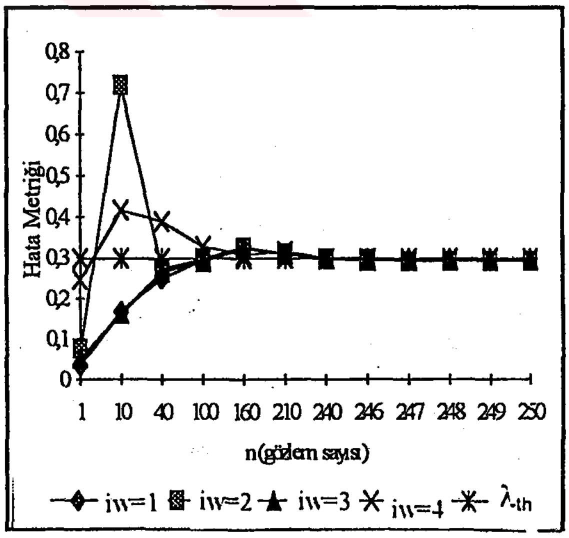
Şekil 5.16a YN=7 için çeşitli hata bileşenlerinin büyük n için evrimi (ıw=1, n_v=(2,1,1)).



Şekil 5.16b. YN=7 için çeşitli hata bileşenlerinin büyük n için evrimi. (iw=3, n_v=(2,3,1)).



Şekil 5.16c YN=7 için hata metriğinin (λ_n) artan ağ kompleksitesi (n_h) altında büyük n için evrimi (iw=2).

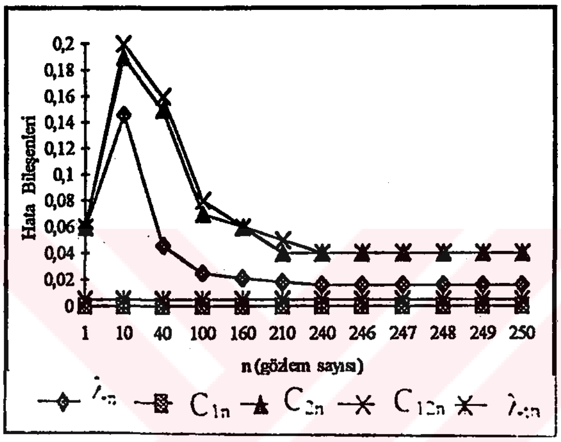


Şekil 5.16d YN=7 için hata metriğinin (λ_n) sabit ağ kompleksitesi (n_h=2) için çeşitli başlangıç vektörleri altında büyük n için davranışı

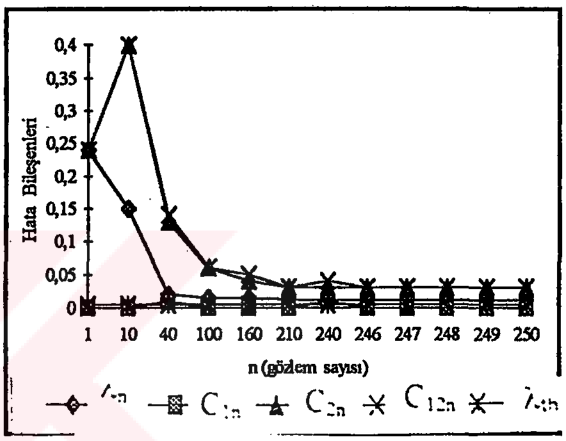
Yapı No:9 (x,y,z)=(1.0,1.0,1.0)

Çok boyutlu normal dağılım (ÇBND) için YN=9, görsel verinin sunulduğu sonuncu yapıdır. Şekil 5.17a ve 5.17b çeşitli hata bileşenlerinin YN=9 için büyük n tepkisini sergiler. Genel eğilim teorik beklentilerle oldukça uyumludur. Ayrıca Şekil 5.17b de λ_n ve C_{1n} nin teorik λ_{th} değerine yakınsaması Şekil 5.17a dakine oranla daha iyidir. Bu durum $n_h=3$ ağının $n_h=1$ ağına oranla daha iyi yaklaşım yapabileceğinin bir başka örneğidir. Şekil 5.17c ve 5.17d ağın genel eğilimiyle uyumludur. Bunlara ek olarak burada sunulmayan yapılar (YN=2,4,8) için de ağın büyük n için davranışının teorik beklentilerle oldukça uyumlu (% 1-3 hata sınırları içinde) olduğunu belirtelim.

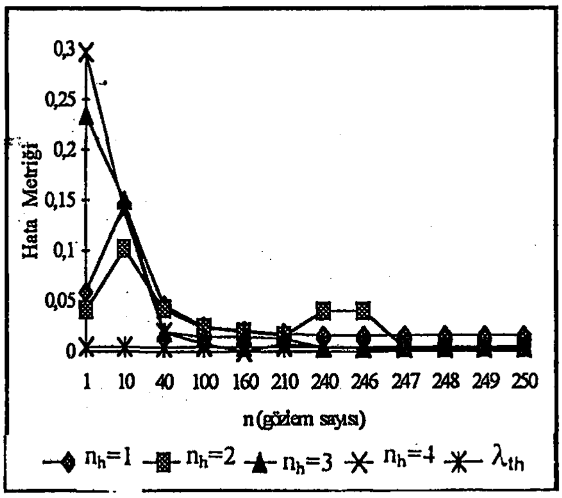
ÇBND için genel sonuç: YSA nın ÇBND verisini işlemedeki başarısı oldukça iyidir. White (1989a) da yapılan teorik analizlerin ÇBND için doğruluğu sistematik deneyler sonucu kanıtlandı. ÇBND verisi için YSA nın gözlenen yakınsama hızının ortalama değeri şudur: λ_n ve C_{1n} bileşenleri n=100 yada 160 dan sonra asimptotik limit değerlerine oldukça yakındırlar. C_{2n} bileşeni ise n=40 ya da 100 den sonra asimptotik limit değerine (=0.00) yakınsamaktadır.



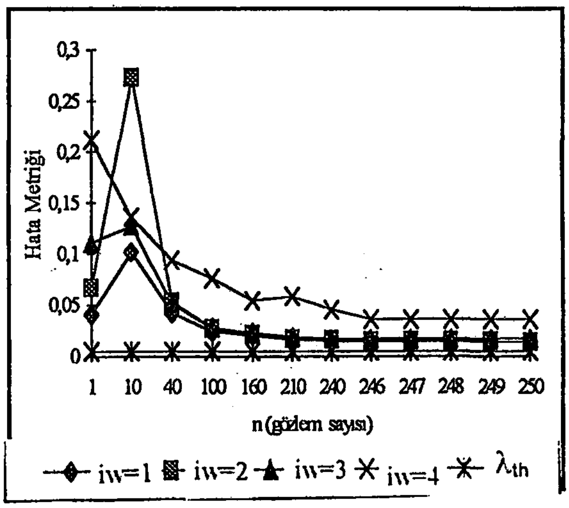
Şekil 5.17a YN=9 için çeşitli hata bileşenlerinin büyük n için evrimi (ıw=1, n_v=(2,1,1)).



Şekil 5.17b. YN=9 için çeşitli hata bileşenlerinin büyük n için evrimi. (iw=3, n_v=(2,3,1)).



Şekil 5.17c YN=9 için hata metriğinin (λ_n) artan ağ kompleksitesi (n_h) altında büyük n için evrimi (iw=2).



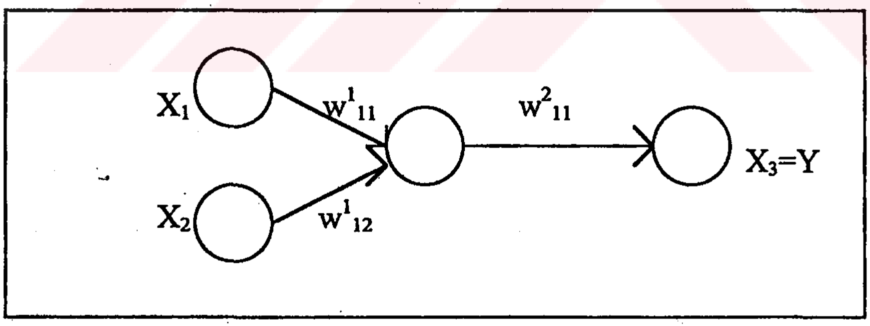
Şekil 5.17d YN=9 için λ_n hata metriğinin(λ_n) sabit ağ kompleksitesi (n_h=2) için çeşitli başlangıç vektörleri altında büyük n için davranışı

5.5.2.2 Verilen bir korelasyon yapısı için final ağırlık vektörlerinin korelasyon matrisleri

Bu çalışmayı ele alırken beklentilerimizden birisi de şuydu. Bilgi veren input bileşenlerinin bulunduğu bir YSA ile fazlalık inputların bulunduğu YSA nda ağırlık bileşenleri farklı olabilir. {Gerçekte input bileşenlerinin mükemmel korelasyonlı olması durumu dışında bu beklentiyi haklı çıkarabilecek herhangi bir teorik değerlendirme literatürde yoktur. O nedenle bu beklenti bir dereceye kadar bir varsayım olarak değerlendirilmelidir}.Bağımsız değişken (inputlar) ve bağımlı değişken (hedef) arasındaki korelasyon yapısı yalnızca yakınsama hızını (niter) etkilemekle kalmayıp aynı zamanda verinin uygun (yeterli) temsili için gereken gizli tabaka nöronu sayısını da etkileyebilir. Bölüm 4 de verilen bazı üçlü korelasyon yapılarının etkin olarak ikili yapılara indirgenebileceğine ilişkin gözlemler yapılmıştı. Bu kesimde ise n_h=1,2 için final ağırlık vektörleri bileşenlerinin n=250 adet gözlem üzerinden korelasyon matrislerinin çeşitli elemanları üzerine gözlemler ve yorumlar sunulacaktır. Burada şunu belirtmemiz uygun olacaktır. Aynı gözlemler n_h=3 ve 4 için de yapıldı. Ancak bu ağ karmaşasına sahip ağların ağırlık bileşenleri sayısı çoktur ve verilerin sunumunda oldukça fazla sayıda parametre gerekir. O nedenle onlara ilişkin veri sunumu yapılmayacak sırası geldikçe sözlerle bilgi verilecektir.

n_v=(2,1,1) YSA için sonuçlar (Ağırlık bileşeni sayısı: 3)

<u>Başlangıç vektörleri</u> : $(w^1_{11},w^1_{12},w^2_{11})$=(0,3,-0.9,0.38); (-0.8,1.5,-0.18);(1.8,-2.5,2.18)

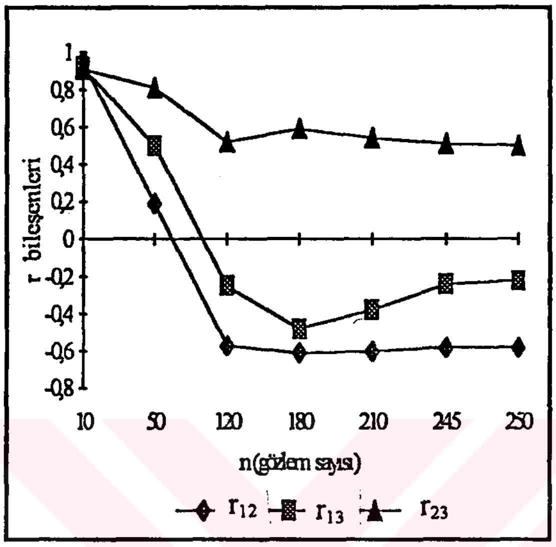


Şekil 5.18 İki input-bir gizli tabaka-bir output tabaka nöronlu YSA. Burada bileşenler sırasıyla w^1_{11}=1, w^1_{12}=2 ve w^2_{11}=3 alınacaktır.

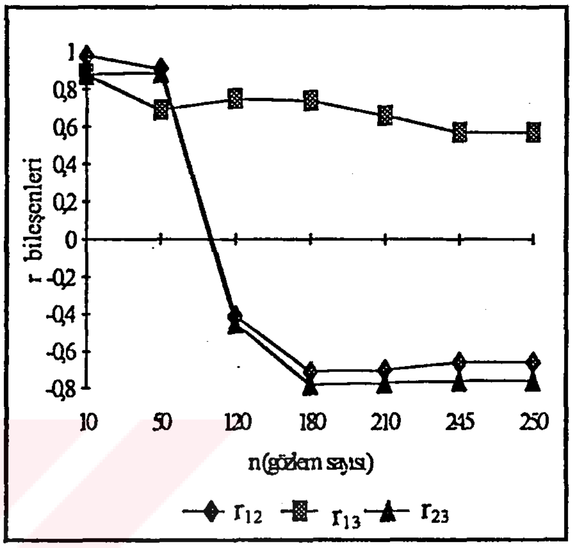
<u>Yapı No</u>: <u>1</u> (x,y,z)=(0.0,0.0,0.2)

Şekil 5.19a-d YN=1 için final ağırlık vektörlerinin artan n (gözlem sayısı) için korelasyon matrislerinin bileşenlerinin evrimini sergiler. 1,2 ve 3 no.lu ağırlık bileşenleri Şekil 5.18 deki gibidir. (Not: Korelasyon matrisleri simetriktir ve r_{ii}=1 bütün i ler için). Şekil 5.19 dan bazı gözlemler şunlardır. Hemen hemen her iw ve her korelasyon bileşeni için n=120 den sonra asimptotik değere erişilmektedir. Yani n=120 den sonra ağırlıkların değişimi kararlı bir şekilde olmaktadır. 5.19a da bütün bileşenlerin korelasyonlasrı düşüktür (≤0.5). Bu durum (x,y,z) üçlüsündeki değerlere benzerlik gösterir. Ancak aynı durum iw=3 için (Şekil 5.19b) için söylenemez. Şekil 5.19d ise r_{13} için iw=1 durumunda sıfıra yakın bir korelasyon gösterir. Bu eğilim de y=0.0 korelasyonuna benzer.Ancak r_{13} iw=2,3 için sıfırdan oldukça uzaktır. Bu durumun nedeni iw=2 ve 3 deki başlangıç vektörünün bileşenlerinin iw=1 dekine oranla oldukça büyük olmasıdır. Yani iw=1 deki değer y=0 durumuna uygun olarak daha karalı değişir.

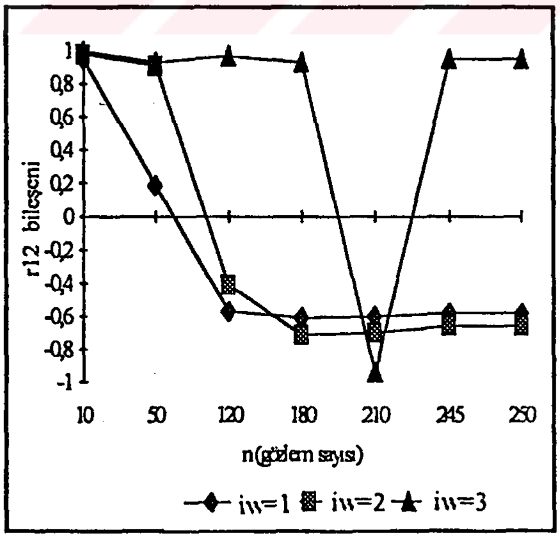
<u>Sonuç</u>: Ağırlık bileşenlerinin asimptotik korelasyonlarının (x,y,z) yapısına uygun olarak değiştiğine ilişkin bazı yönelimler var olsada genel bir eğilim bulunamadı.



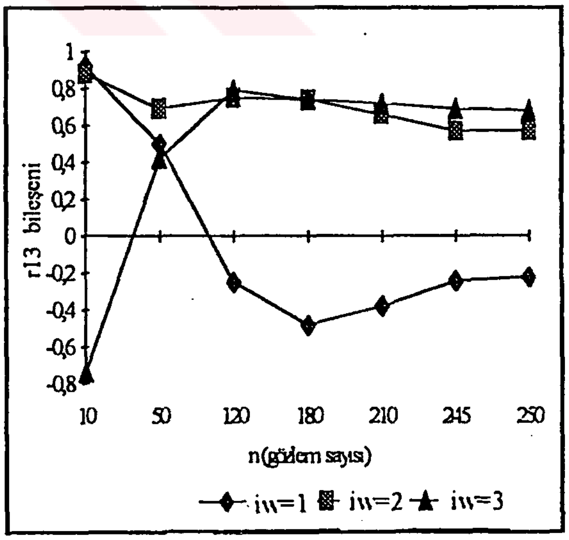
Şekil 5.19a YN=1 için final ağırlık vektörlerinin (n_v=(2,1,1) YSA için) çeşitli bileşenlerinin korelasyonlarının asimptotik evrimi (iw=1)



Şekil 5.19b YN=1 için final ağırlık vektörlerinin (n_v=(2,1,1) YSA için) çeşitli bileşenlerinin korelasyonlarının asimptotik evrimi (iw=3)



Şekil 5.19c YN=1 için ağırlık vektörünün r_{12} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

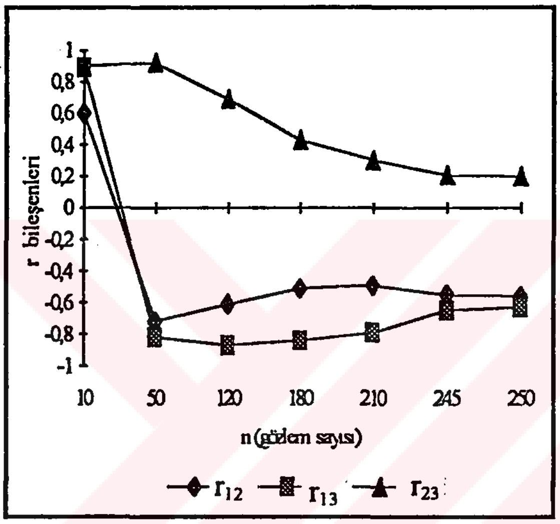


Şekil 5.19d YN=1 için ağırlık vektörünün r_{13} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

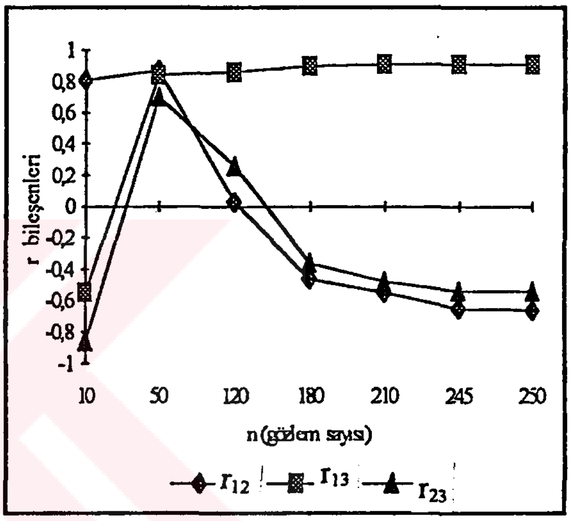
Yapı No: **2** (x,y,z)= (0.0,0.0,0,6)

Şekil 5.20a-d için de n=120 den sonra kararlı bir gidiş olduğu söylenebilir. Ayrıca r_{12} bileşeni her iw için -0.5 civarındadır (Şekil 5.20c). Bir başka gözlem: r_{23} değerinin iw=3 için (Şekil 5.20b) iw=1 dekine oranla (Şekil 5.20a) daha iyi olarak z=0.6 değerine yakındır . Bu duruma bir açıklama şudur . iw=3 deki başlangıç vektöründeki değerlerinin iw=1 dekine oranla önemli ölçüde büyük olması iw=3 deki ağırlık bileşenlerinin asimptotik evriminin z=0.6 değerine benzer biçimde olmasına zorlamıştır. Bu yorum Y.N=1 deki r_{13} değerinin eğilimi için yapılan yorumla çelişik gibidir. Daha açık söylenise Y.N.=1 de iw=1 deki başlangıç değerlerinin küçük olması r_{13} in iyi davranmasına neden olurken Y.N=2 iw=1 in küçük olması r_{23} in kötü davranmasına neden olmuştur. Bu çelişki,bir dereceye kadar, şöylece giderilebilir. Y.N=1 de r_{13} in yaklaşması istenen değer sıfır olduğundan iw=1 başlangıç değerlerinin sıfıra yakın olması denetimli bir değişim sağlayabilmek için (r_{13}) için arzu edilir bir durumdur. Ancak Y.N.=2 de r_{23} in yaklaşması istenilen değer (0.6) 1'e belli ölçülerde daha yakın olduğu için iw=3 deki değerlerin büyük olması sorun oluşturmamış olabilir.

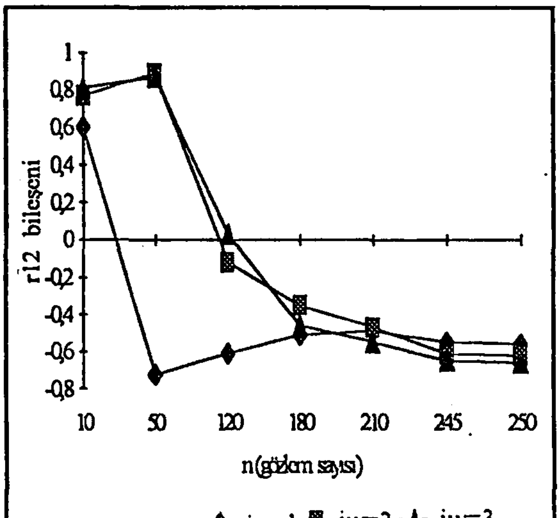
<u>**Sonuç**</u>: Y.N.=2 için gözlenenler (Şekil 5.20a-d) (x,y,z) yapısını ancak çok sınırlı ölçüde yansıtmaktadır.



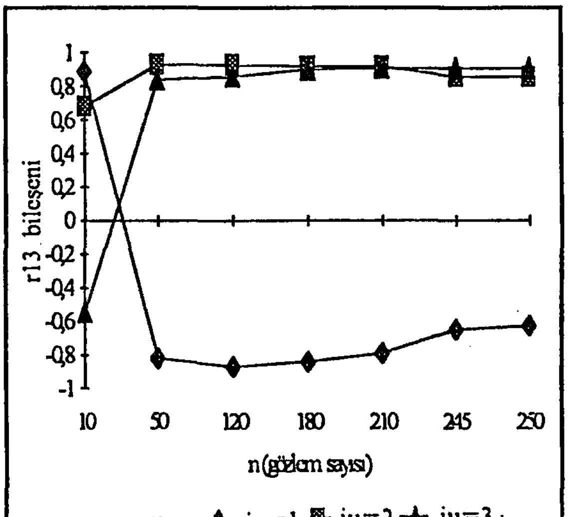
Şekil 5.20a YN=2 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=1)



Şekil 5.20b YN=2 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=3)



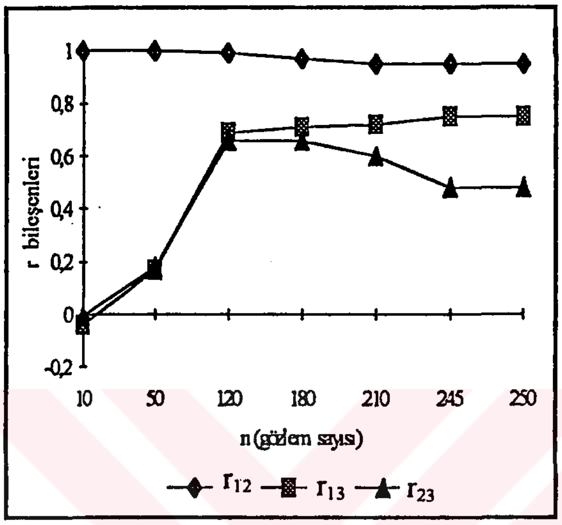
Şekil 5.20c YN=2 için final ağırlık vektörünün r_{12} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi



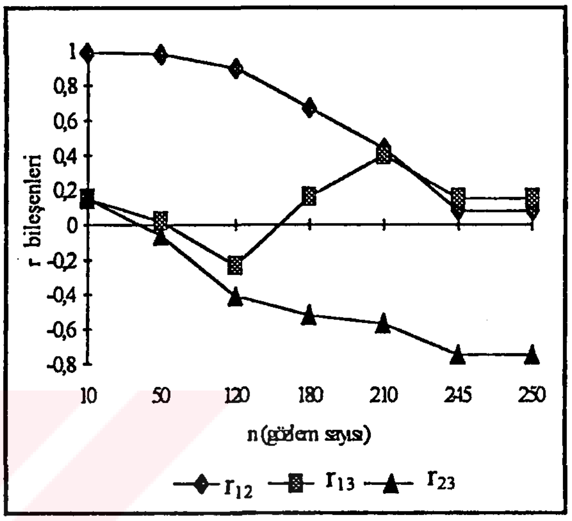
Şekil 5.20d YN=2 için final ağırlık vektörünün r_{13} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

Yapı No:4 . (x,y,z)=(1.0,0.0.0.0)

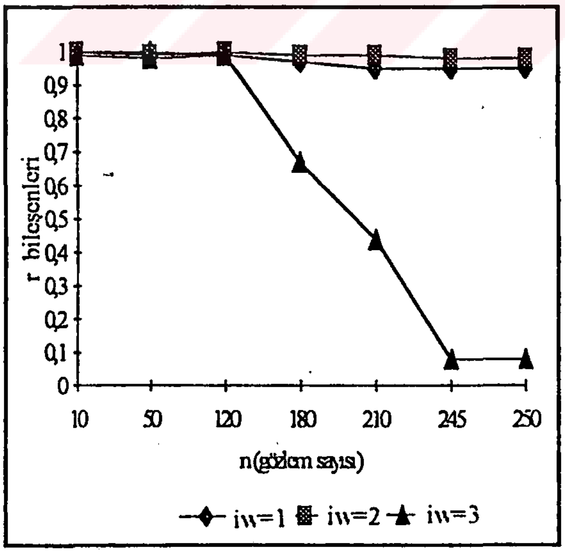
r_{12} bileşeni iw=1 ve 3 için 1 değerine oldukça yakındır. Bu durum x=1.0 değerine iyi bir eğilimin göstergesidir. Ancak iw=3 için beklenenin aksi bir durum söz konusudur (Şekil 5.21c). r_{13} bileşeni de sıfıra yakın değerler alır (Şekli 5.20d), bu eğilim y=0.0 durumunu iyi bir temsilin göstergesi olabilir. Ayrıca Şekil 5.21a da r_{23} değeri z=0.0 durumuna uygun olarak sıfır değerine yakındır.
__Sonuç__: Y.N=4 için gözlenenler özellikle x=1.0 değerini iyi temsil ederler. Ancak iw=3 için sapma olması bu sonucu biraz zayıflatır.



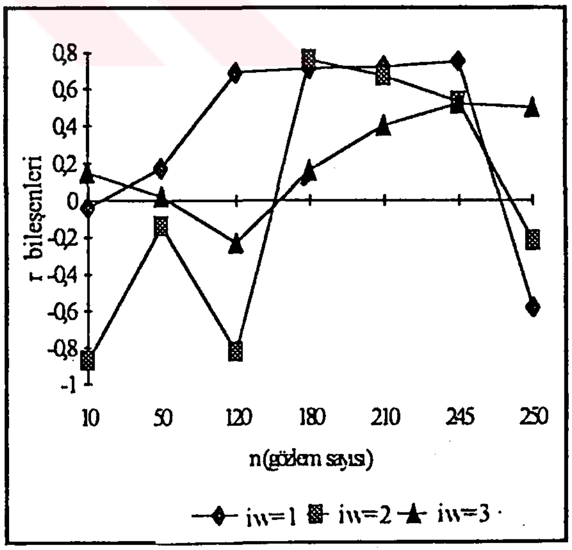
Şekil 5.21a YN=4 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=1)



Şekil 5.21b YN=4 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=3)



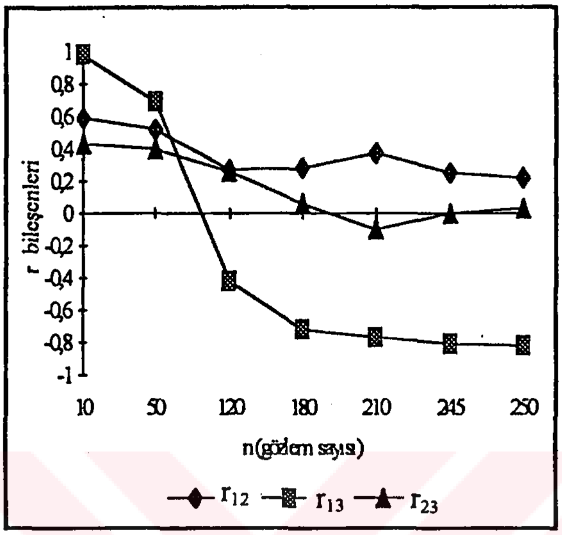
Şekil 5.21c YN=4 için final ağırlık vektörünün r_{12} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi



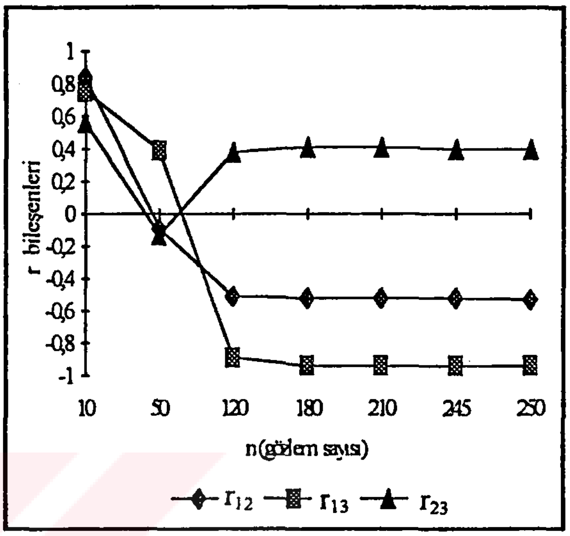
Şekil 5.20d YN=4 için final ağırlık vektörünün r_{13} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

Yapı No: 5 (x,y,z)=(0.0,1.0,0.0)

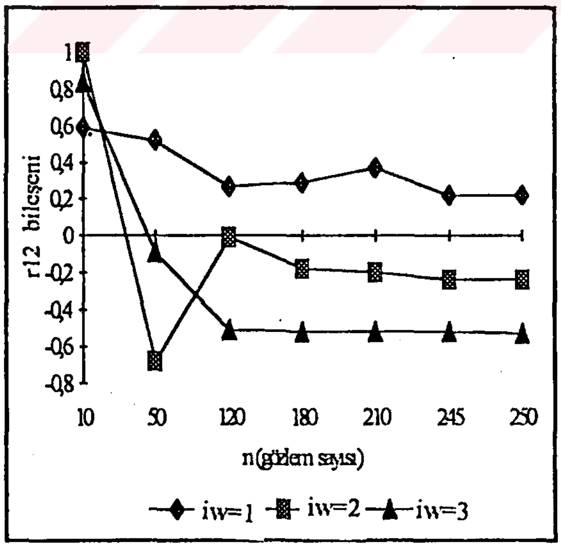
Şekil 5.21a-b (x,y,z) çiftinin değerlerine uygun bir eğilim gösterir, ancak burada r_{13}=-1 olduğuna dikkat edilmelidir. Oysa y=1.0 dır. Yani r_{13} mutlak değer olarak y=1.0 değeriyle uyumludur. Şekil 5.21 c de r_{12} mutlak değer olarak bütün iw ler için x=0.0 durumuna eğilimlidir. Benzer durum r_{13} için de sözkonusudur.
Sonuç : Y.N=5 için elde edilen değerler (x,y,z) çiftinin temsilini özellkle y=1.0 durumunu iyi yaparlar. Y.N.=6 (1.0,0.6,0.6) için elde edilen değerlerde de x=1.0 durumu çok iyi derecede temsil edildi.



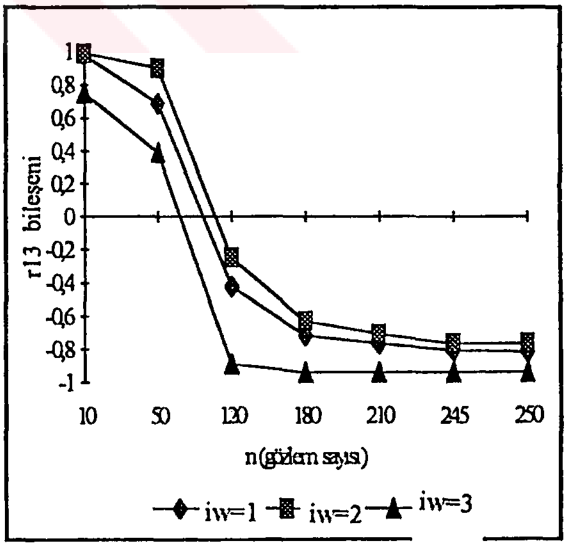
Şekil 5.22a YN=5 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=1)



Şekil 5.22b YN=5 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=3)



Şekil 5.22c YN=5 için final ağırlık vektörünün r_{12} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

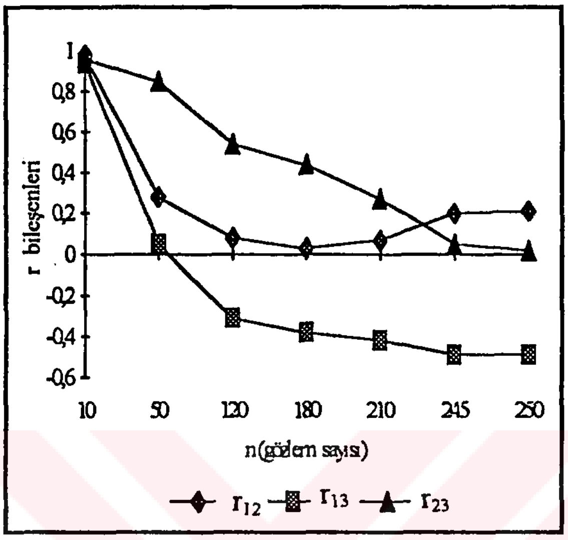


Şekil 5.22d YN=5 için final ağırlık vektörünün r_{13} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

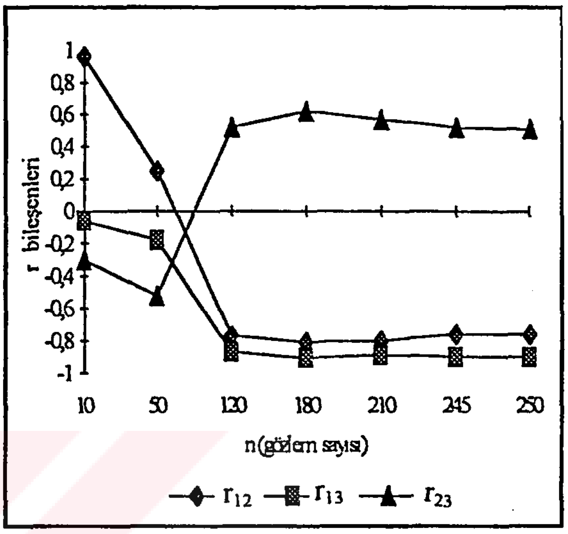
Yapı No: 7 (x,y,z)=(0.8,0.8,0.8)

Şekil 5.23c iw=1 dışında r_{12}=0.8 asimptotik değerinin x=0.8 değerine uygunluk gösterdiğini anlatır. Aynı eğilim (Şekil 5.23d) r_{13}=0.8 asimtotik değerinin y=0.8 değerine uygunluğunu gösterir. Şekil 5.23 de ki değerler de (x,y,z) değerlerine önemli ölçüde uygunluk gösterir.

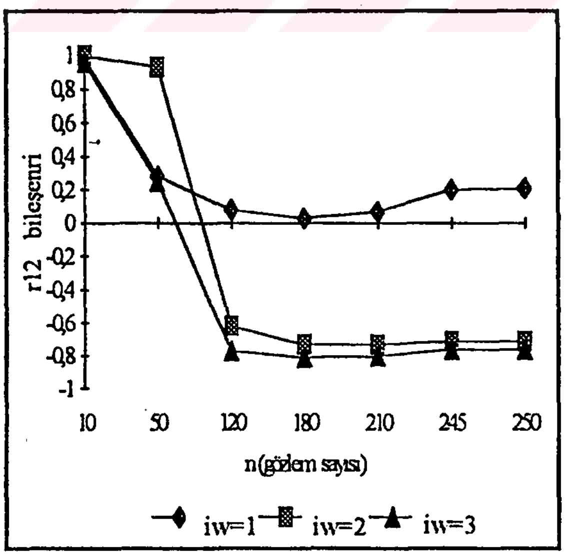
Sonuç: Y.N.=7 için final ağırlık vektörlerinin gözlenen korelayonları (x,y,z) çiftine iyi denebilecek derecede uyumludur.



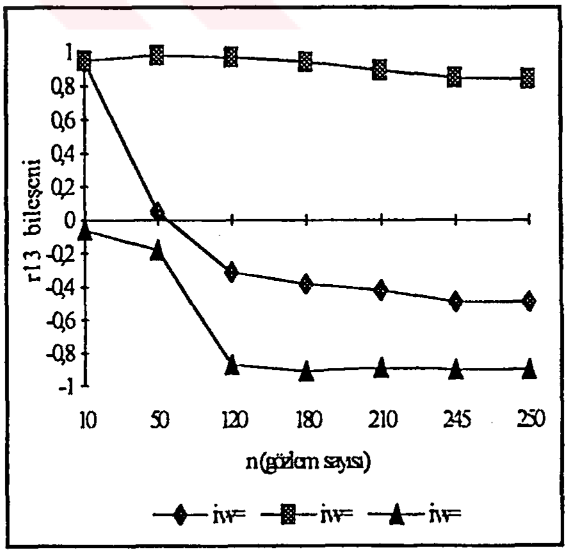
Şekil 5.23a YN=7 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=1)



Şekil 5.23b YN=7 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=3)



Şekil 5.23c YN=7 için final ağırlık vektörünün r_{12} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

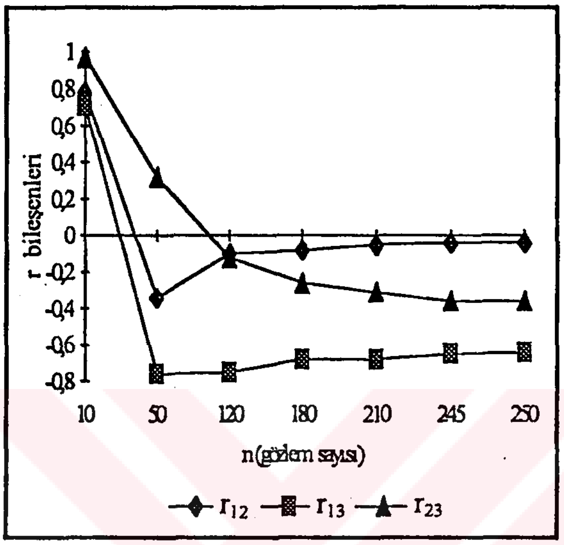


Şekil 5.23d YN=7 için final ağırlık vektörünün r_{13} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

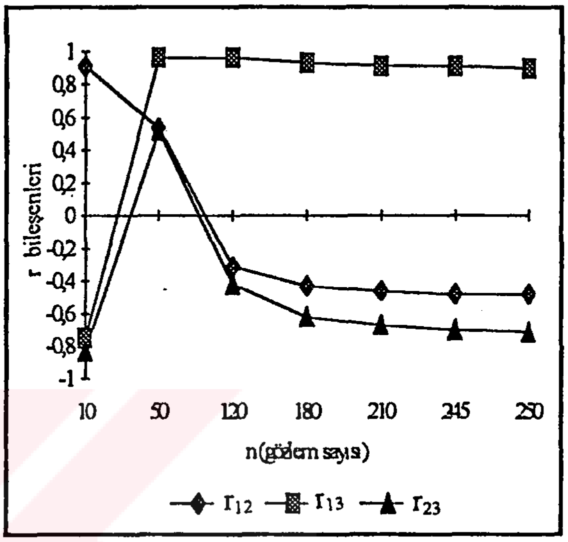
Yapı No: **8** (x,y,z)=(0.4,0.4,1.0)

Şekil 5.24a-d incelendiğinde (x,y,z) yapısına ancak önemsiz bir derecede uyum vardır. Bununla birlikte Şekil 5.24c de bütün iw ler için r_{12} değeri x=0.4 değerini uyum gösterir. Ancak z=1.0 korelasyonuna uyum gösteren bir eğilim Şekil 5.24 den çıkarılamaz.

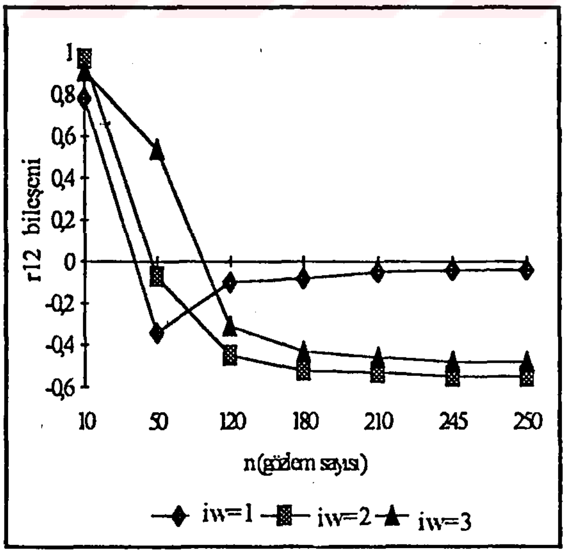
Sonuç: Y.N.=8 final ağırlık vektörleri korelasyonları (x,y,z) değerlerini ancak kısıtlı bir oranda temsil ederler.



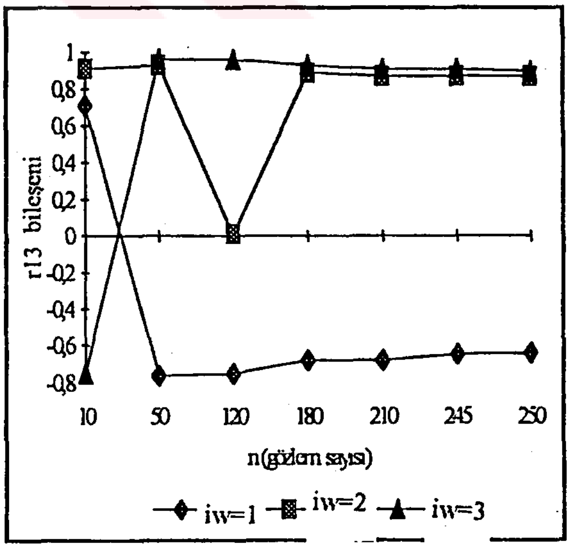
Şekil 5.24a YN=8 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=1)



Şekil 5.24b YN=8 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=3)



Şekil 5.24c YN=8 için final ağırlık vektörünün r_{12} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

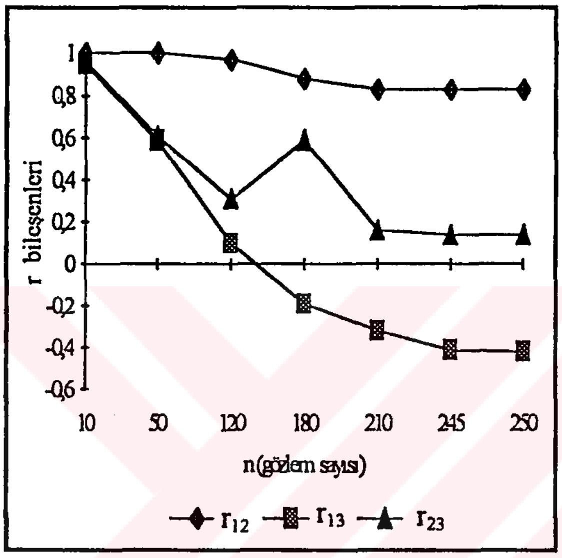


Şekil 5.24d YN=8 için final ağırlık vektörünün r_{13} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

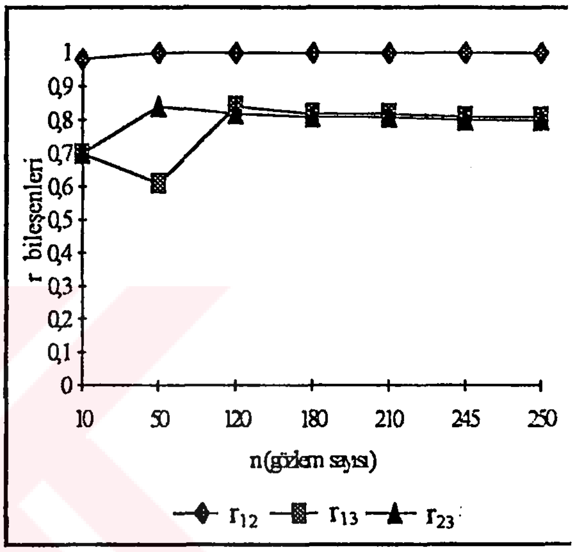
Yapı No:9 (x,y,z)=(1.0,1.0,1.0)

Şekil (5.25c) her iw için r_{12} nin değerinin 1'e yakınsadığını gösterir. Bu durum r (X_1,X_2)=x=1.0C durumuna oldukça iyi bir uyumun olduğunu gösterir. Yani input bileşenleri X_1 ve X_2 arasındaki korelasyon katsayısı 1.00 iken onların gizli tabaka nöronuna olan bağlantılarının (ağırlık bileşenleri) korelasyonlarının da 1 olduğunu gösterir. Anımsanacağı gibi r (X_1,X_2)=x=1.00 ise X_1 ile X_2 arasında hemen hemen her yerde bir afin bağıntı vardır. Yani X_1 X_2 cinsinden elde edilir ve böylelikle ağ iki inputtan bir inputa indirgenir. Yani giriş uzayının boutu efektif olarak azalır. Bu durum fazlalık input hipetizine uygundur. Şekil 5.25b iise bütün ağırlık bileşenleri korelasyon katsayılarının 1'e yakın olduğunu gösterir. Ancak aynı durum iw=1 (Şekil 5.25a) için gözlenmediğinden bu eğilimin genel olduğu söylenemez. Şekil 5.25d ise r_{13} değerinin mutlak değerce oldukça büyük olduğunu ($|r_{13}| \geq 0$) anlatır. Bu durum y=1.00 değerine yakınlık gösterir.

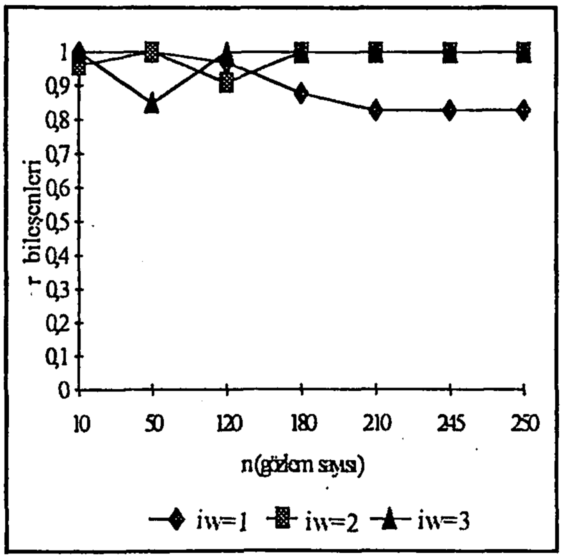
__Sonuç__: Şekil 5.25 den input bileşenleri arasında korelasyon 1 olduğunda onalrı gizli tabakanın belirli bir nöronuna bağlayan ağırlık vektörleri bileşenlerinin korelasyonlarının da 1 olduğu sonucu çıkarılabilir.



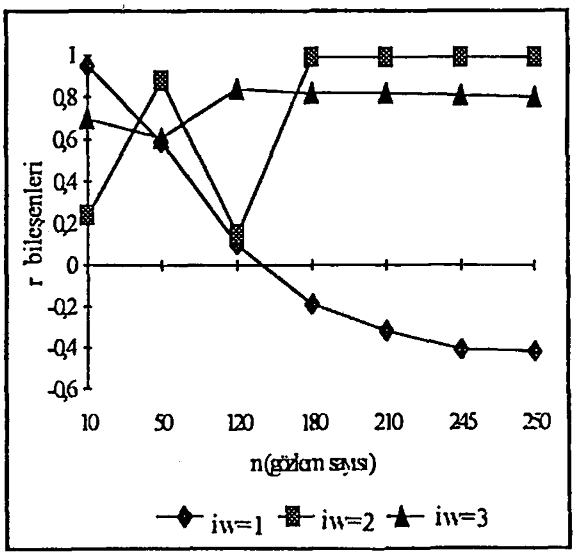
Şekil 5.25a YN=9 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=1)



Şekil 5.25b YN=9 için final ağırlık vektörlerinin çeşitli bileşen korelasyonlarının asimptotik evrimi (iw=3)



Şekil 5.25c YN=9 için final ağırlık vektörünün r_{12} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi



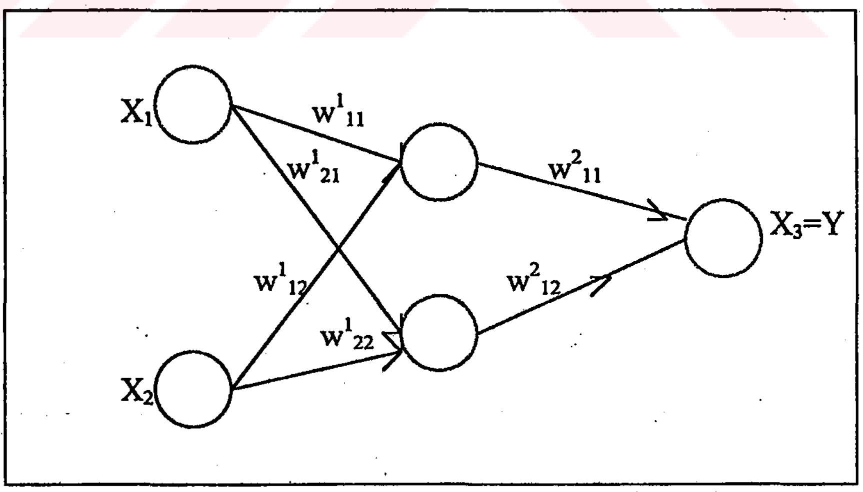
Şekil 5.25d YN=9 için final ağırlık vektörünün r_{13} bileşeninin çeşitli başlangıç koşullarına göre asimptotik evrimi

Şekil 5.19-5.25 den elde edilen gözlemlerin özeti: (x,y,z) üçlüsünde x=1.00 olduğu zaman r_{12} korelasyonu x ile uyumlu olarak 1'e hemen hemen eşitir. Ancak y ya da z değerlerinin 1 olması durumunda r_{13} ya da r_{23} değerlerinin de 1'e yakın olduğu genel olarak söylenemez. Bununla birlikte böylesi bir eğilimin olabileceğine ilişkin bazı gözlemler yapıldı. Bu uç durumların dışında yani (x,y,z) nin değerlerinin sıfırdan farklı olduğu zamanlarda r_{12}, r_{13}, r_{23} final ağırlık vektörleri bileşenlerinin korelasyonlarının (x,y,z) korelasyonlarına genel olarak benzediğini söylemek güçtür. Yine de böylesi eğilimlerin olabileceğine ilişkin bazı gözlemler vardır. Bu konu daha fazla araştırma istemektedir.

Başlangıç vektörlerinin bileşenlerinin eşit olduğu durumda final ağırlık vektörü korelasyonları:

(r_{12}, r_{13}, r_{23}) ile (x,y,z) arasında var olabilecek olası bir benzerlik başlangıç ağırlık vektörlerindeki bileşenlerin biribirinden farklı olması nedeniyle ortaya yeterince çıkmamış olabilir (Şekil 5.19-5.25 de olduğu gibi) düşüncesiyle bu farklılık ortadan kaldırılarak sonuçlar incelendi. Başlangıç ağırlık vektörleri sırasıyla (0.3,0.3,0.3), (0.9,0.9,0.9), (-0.4,-0.4,-0.4) olarak alındı. Sonuçlar bu kez de Şekil 5.19-5.25 de elde edilenlere önemli çok önemli ölçüde benzerlik gösterdi.

(n_v=2,2,1) bazı sonuçlar



Şekil 5.26 İki input-bir gizli tabaka-bir output tabaka nöronlu YSA. Burada bileşenler sırasıyla **w^1_{11}=1, w^1_{12}=2, w^1_{21}=3, w^1_{22}=4, w^2_{11}=5, w^2_{12}=6 alınacaktır.**

<u>Yapı No: 1</u> (x,y,z)=(0.0,0.0,0.2)

Şekil 5.26 daki ağırlık bileşenlerinin final değerlerinin (n=250 için) korelasyon değerleri Tablo 5.6 da verilmiştir. Dikkate değer tek özellik r_{45}, r_{46}, r_{56} değerlerinin yüksek öteki bileşenlerin

ise görece düşük oluşudur. 5 ve 6 no.lu bileşenlerin değişme yönü aynıdır. Bu sonuç gizli tabaka- output tabakası ağırlık bileşenlerinin (x,y,z) yapısından çok fazla etkilenmediğini gösterebilir. Ayrıca r_{12} ve r_{34} değerlerinin düşük oluşu x=0.0 değerine bir benzerlik sergiler. r_{15} ve r_{16} değerlerinin biribirine eşit olması ilginç bir başka gözlemdir. İlginçlik bu iki bileşenin biririne bağlantılı olmadığı halde bu denli yakın korelasyonlu olmalarıdır.
<u>Sonuç</u>: Y.N.=1 YSA nın (x,y,z) yapısını ancak kısıtlı bir ölçüde temsil edebildiği söylenebilir.

Tablo 5.6 n_v=(2,2,1), Y.N.=1 için final ağırlık vektörü bileşenlerinin hesaplanmış korelasyonları.

r	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{23}	r_{24}	r_{25}	r_{26}	r_{34}	r_{35}	r_{36}	r_{45}	r_{46}	r_{56}
iw=1	-0.44	0.88	-0.62	-0.20	-0.20	-0.57	0.55	0.65	0.65	-0.57	-0.66	-0.27	-0.27	0.85	1.0
iw=2	0.41	0.41	0.69	0.81	0.81	-0.09	0.10	0.36	0.36	-0.09	0.70	0.34	0.34	0.83	1.0

<u>*Yapı No:* **3**</u> (x,y,z)=(0.0,0.0,1.0)

Tablo 5.7 Y.N.=3 için n_v=(2,2,1) li YSA daki sonuçları verir. 5 ve 6 no.lu bileşenlerin değişme yönü aynıdır. Bu sonuç gizli tabaka- output tabakası ağırlık bileşenlerinin (x,y,z) yapısından çok fazla etkilenmediğini gösterebilir. Bir başka gözlem iw=1ve 2 için r_{46} değerlerinin birbirine eşit olmasıdır. Ayrıca r_{12} ve r_{34} değerlerinin düşük oluşu x=0.0 değerine bir benzerlik sergiler. r_{15} ve r_{16} değerlerinin biribirine eşit olması ilginç bir başka gözlemdir.
<u>Sonuç</u>: Y.N.=3 YSA nın (x,y,z) yapısını ancak kısıtlı bir ölçüde temsil edebildiği söylenebilir.

Tablo 5.7 n_v=(2,2,1), Y.N.=3 için final ağırlık vektörü bileşenlerinin hesaplanmış korelasyonları.

r	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{23}	r_{24}	r_{25}	r_{26}	r_{34}	r_{35}	r_{36}	r_{45}	r_{46}	r_{56}
iw=1	0.52	0.68	0.82	0.60	0.12	0.66	-0.14	-0.14	0.12	0.93	0.93	0.64	0.64	0.85	1.0
iw=2	-0.07	0.50	-0.59	-0.69	0.08	0.69	-0.30	-0.30	0.08	-0.66	-0.66	0.40	0.40	0.83	1.0

<u>*Yapı No:*5</u> (x,yz)=(0.0,1.0,0.0)

Tablo 5.8 Y.N.=5 için n_v=(2,2,1) li YSA daki sonuçları verir. 5 ve 6 no.lu bileşenlerin değişme yönü aynıdır. Bu sonuç gizli tabaka- output tabakası ağırlık bileşenlerinin (x,y,z) yapısından çok fazla etkilenmediğini gösterebilir. Bir başka gözlem iw=1,2,3 için r_{45} ve r_{46} değerlerinin birbirine eşit olmasıdır. Bu gözlem sistematik olarak bütün yapılar için geçerlidir ve uygun bir açıklama sunulamamıştır. Buna karşın r_{12} ve r_{34} değerlerinin düşük oluşu x=0.0 değerine bir benzerlik sergilediğini gösterir. r_{15} ve r_{16} değerlerinin biribirine eşit olması ilginç bir başka gözlemdir.
<u>Sonuç</u>: Y.N.=5 YSA nın (x,y,z) yapısını belirli bir ölçüde temsil edebildiği söylenebilir.

Tablo 5.8 n_v=(2,2,1), Y.N.=5 için final ağırlık vektörü bileşenlerinin hesaplanmış korelasyonları.

r	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{23}	r_{24}	r_{25}	r_{26}	r_{34}	r_{35}	r_{36}	r_{45}	r_{46}	r_{56}
iw=1	-0.17	0.89	0.43	-0.64	-0.64	-0.48	0.78	-0.41	-0.41	0.08	-0.29	-0.29	-0.84	-0.84	1.0
iw=2	-0.77	0.71	0.87	-0.76	-0.76	-0.32	-0.90	0.93	0.93	0.43	-0.24	-0.24	-0.96	-0.90	1.0
iw=3	0.01	0.74	0.33	-0.91	-0.91	-0.31	0.87	-0.21	-0.21	-0.15	-0.40	-0.40	-0.57	-0.57	1.0

<u>*Yapı No:* **6**</u> (x,y,z)=(1.0,0.6,0.6)

Tablo 5.9 Y.N.=6 için n_v=(2,2,1) li YSA daki sonuçları verir. r_{56} değerinin yüksek oluşu 5 ve 6 no.lu bileşenlerin değişme yönü aynı olduğunu gösterir. Bu sonuç gizli tabaka- output tabakası ağırlık bileşenlerinin (x,y,z) yapısından çok fazla etkilenmediğini gösterebilir. Bir başka gözlem iw=1,2,3 için r_{45} ve r_{46} değerlerinin birbirine eşit olmasıdır. Ayrıca r_{12} ve r_{34} değerlerinin yüksek oluşu nedeniyle x=1.0 değeri iyi bir şekilde temsil edilir. r_{15} ve r_{16} değerlerinin biribirine eşit olması ilginç bir başka gözlemdir. .
<u>Sonuç</u>: Y.N.=6 YSA nın (x,y,z) yapısını r_{12} ve r_{34} değerlerinin yüksek oluşu x=1.00 değerini iyi temsil etme yeteneğinde bulunduğunu gösterir.

Tablo 5.9 n_v=(2,2,1), Y.N.=6 için final ağırlık vektörü bileşenlerinin hesaplanmış korelasyonları.

r	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{23}	r_{24}	r_{25}	r_{26}	r_{34}	r_{35}	r_{36}	r_{45}	r_{46}	r_{56}
iw=1	0.99	0.57	0.42	-0.64	0.49	0.36	-0.63	-0.63	0.95	0.08	-0.09	-0.09	0.12	0.12	1.0
iw=2	0.93	0.92	0.42	-0.03	0.95	0.66	0.27	0.27	0.93	0.69	-0.33	-0.33	0.88	0.88	1.0
iw=3	0.86	0.51	0.33	0.47	0.05	0.01	0.81	0.81	-0.21	0.87	0.30	-0.30	-0.12	-0.12	1.0

<u>Yapı No</u><u>: **8**</u> (x,y,z),

Tablo 5.10 Y.N.=8 için n_v=(2,2,1) li YSA daki sonuçları verir. r_{56} değerinin yüksek oluşu 5 ve 6 no.lu bileşenlerin değişme yönü aynı olduğunu gösterir. Bu sonuç gizli tabaka- output tabakası ağırlık bileşenlerinin (x,y,z) yapısından çok fazla etkilenmediğini gösterebilir. Bir başka gözlem iw=1,2,3 için r_{45} ve r_{46} değerlerinin birbirine eşit olmasıdır. r_{15} ve r_{16} değerlerinin biribirine eşit olması ilginç bir başka gözlemdir. .
<u>Sonuç</u>: Y.N.=8 YSA nın (x,y,z) yapısını kısıtlı ölçüde temsil eder.

Tablo 5.10 n_v=(2,2,1), Y.N.=8 için final ağırlık vektörü bileşenlerinin hesaplanmış korelasyonları.

r	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{23}	r_{24}	r_{25}	r_{26}	r_{34}	r_{35}	r_{36}	r_{45}	r_{46}	r_{56}
iw=1	0.44	0.81	0.14	-0.83	-0.83	-0.01	0.81	-0.67	-0.67	-0.47	-0.48	-0.48	-0.14	-0.14	1.0
iw=2	0.28	0.60	0.73	0.40	0.40	-0.11	0.71	-0.54	-0.54	0.59	0.88	0.88	0.19	0.19	1.0
iw=3	0.87	-0.78	-0.42	0.02	0.02	-0.85	-0.74	-0.35	-0.35	0.75	0.40	0.40	0.86	0.86	1.0

<u>Yapı No</u><u>:9</u> (x,y,z)=(1.0,1.0,1.0)

Tablo 5.11 Y.N.=9 için n_v=(2,2,1) li YSA daki sonuçları verir. r_{56} değerinin yüksek oluşu 5 ve 6 no.lu bileşenlerin değişme yönü aynı olduğunu gösterir. Bu sonuç gizli tabaka- output tabakası ağırlık bileşenlerinin (x,y,z) yapısından çok fazla etkilenmediğini gösterebilir. Bir başka gözlem iw=1,2,3 için r_{45} ve r_{46} değerlerinin birbirine eşit olmasıdır. r_{15} ve r_{16} değerlerinin biribirine eşit olması ilginç bir başka gözlemdir. Ayrıca r_{12} değerinin 1'e yakın olması x=1.00 değerini bu kez de iyi temsil edildiğini gösterir.
<u>Sonuç</u>: Y.N.=8 YSA nın (x,y,z) yapısını iyi denilebilecek bir ölçüde olarak temsil eder.

Tablo 5.11 n_v=(2,2,1), Y.N.=9 için final ağırlık vektörü bileşenlerinin hesaplanmış korelasyonları.

r	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{23}	r_{24}	r_{25}	r_{26}	r_{34}	r_{35}	r_{36}	r_{45}	r_{46}	r_{56}
iw=1	0.99	0.70	0.79	-0.55	-0.56	0.67	-0.62	-0.62	0.97	0.11	0.11	-0.09	-0.09	-0.09	1.0
iw=2	0.98	0.40	0.39	0.18	0.18	0.25	0.34	0.35	0.99	-0.66	-0.66	-0.60	0.60	0.60	1.0
iw=3	0.88	-0.01	0.58	-0.62	-0.62	0.79	-0.77	-0.77	0.67	0.09	0.09	-0.56	0.56	0.60	1.0

Burada görsel veri olarak sunulmasa da n_v=(2,3,1) ve n_v=(2,4,1) YSA nın da daha düşük kompleksiteli YSA ındaki eğilimlere benzer eğilimler gösterdiği gözlendi. Özellikle x=1.00 değerinin bulunduğu yapılar için (4,6,9) için X_1 ve X_2 inputlarını biribirine bağlayan ağırlık bileşenlerinin korelasyonlarının 1'e çok önemli ölçüde yakın olduğunu gösterdi. Bu eğilim ise daha önce de vurgulandığı gibi x=1.0 durumunda X_1 inputunun X_2 cinsinden (ya da tersi) yazılabileceğini ve bu input bileşenlerinden birisinin elenebileceği anlamına gelir.

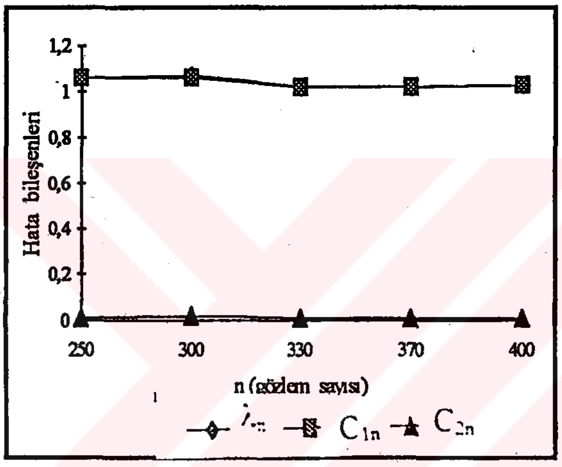
n_v=(2,n_h,1) (n_h=1,2,3,4) YSA nın final ağırlık vektörü bileşenleinin asimptotik korelasyonları için genel sonuç: X_1 ve X_2 inputları mükemmel korelasyonlu iken (x=1.00) bu iki inputu biribirine bir gizli nöronu ile bağlayan ağırlık bileşenlerinin asimptotik korelasyonları da 1'e oldukça yakındır. x in küçük olduğu durumlarda (≤0.4) da ise sözü edilen ağırlık bileşenleri düşük korelasyonludur. Yani x'in yüksek ya da düşük olduğu durumlarda veri yapısının korelasyonları (x,y,z) final ağırlık vektörlerinin korelasyonlarına oldukça benzer eğilimler gösterir. Ara durumlar için ise kesin bir ilişki bulunduğu incelenen örnekler çerçevesinde söylenemez. Ayrıca biribirleriyle doğrudan bağlantılı olmayan bazı ağırlık bileşenlerinin bütün iw ler için aynı korelasyon değerlerini sergilemesi ilginç bir nokta olarak not edilmeye değerdir.

5.5.3 YSA nın genelleştirme yeteneği

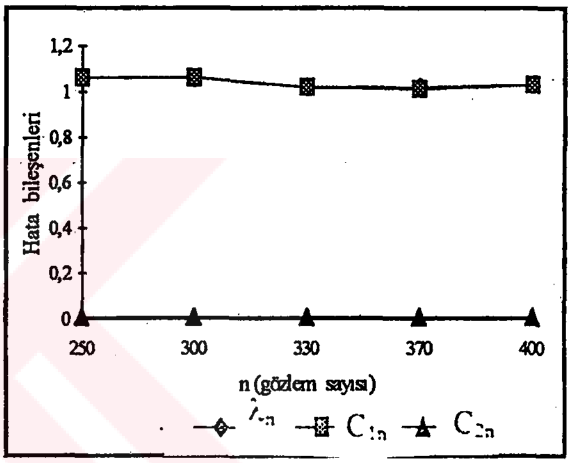
Bu kesimde Kesim 5.5.2.1 de elde edilen optimum final ağırlık vektörleri $\mathbf{w}_n$ (n=250) başlangıç vektörleri olarak alınarak YSA nın test kümesi üzerindeki performansı (genelleştirme yeteneği) üzerine bazı deneysel sonuçlar sağlanacaktır. Bu yetenek ölçülürken kuşkusuz ağa herhangi bir eğitme yapılmayacaktır. (Test kümesindeki n değerleri 250, 290, 330, 360, 380 ve 400 olarak alınacaktır).

<u>Yapı No: 1</u> (x,y,z)=(0.0,0.0,0.2)

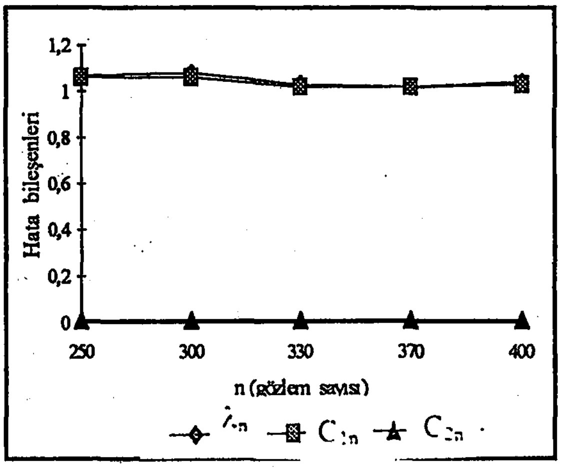
Şekil 5.27 den görüleceği gibi n=250 eğitme kümesiyle elde edilen final ağırlık vektörüyle n=400'e kadar yapılan genelleştirmeler oldukça iyidir. Bir başka not şudur. n=400'e yaklaşıldıkça zaten çok küçük olan kestirim hatası daha da azalmaktadır. Ayrıca n_h=1 ve n_h=3 için elde edilen değerler biribirlerine oldukça yakındırlar. Yani YSA daki bileşen sayısı azaltılabilir.



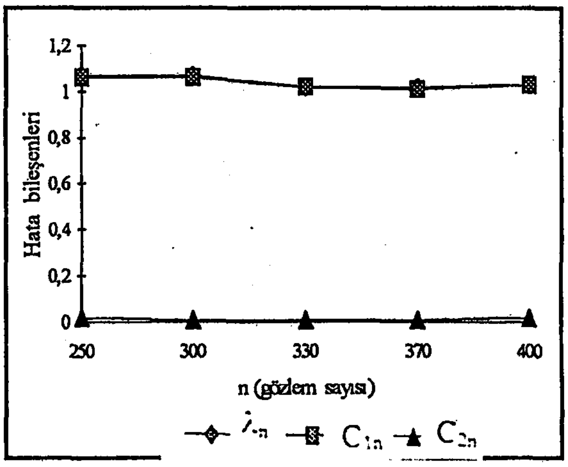
Şekil 5.27a Y.N.=1 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=1.00)



Şekil 5.27b Y.N.=1 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=3, λ_{th}=1.00)



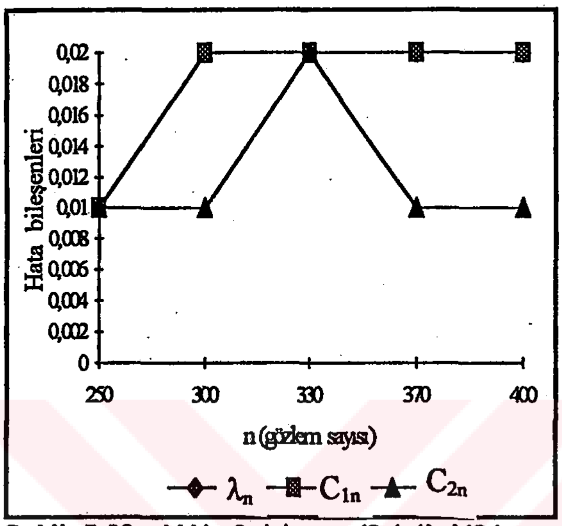
Şekil 5.27c n_v=(2,3,1) YSA da Y.N.=1 için genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=1.00)



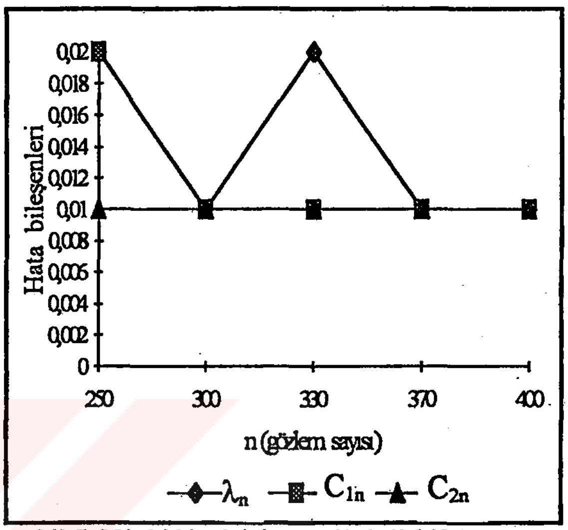
Şekil 5.27d n_v=(2,3,1) YSA da Y.N.=1 için genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=3, λ_{th}=1.00)

Yapı No: **3** (x,y,z)=(0.0,0.0,1.0)

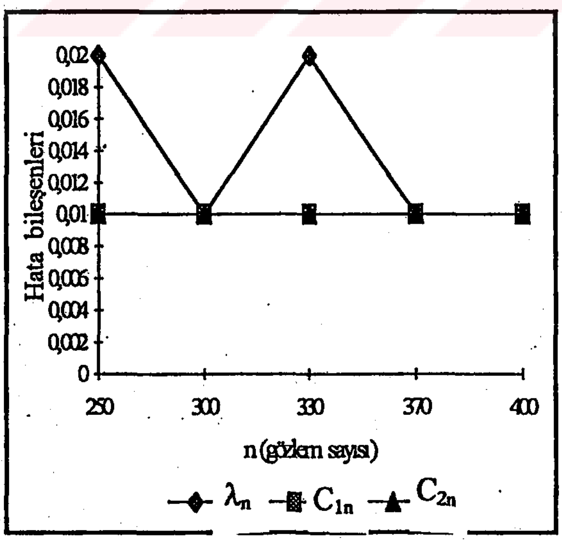
Şekil 5.28a-d Y.N.=3 için YSA nın genelleştirme yeteneği üzerine sonuçlar sunar. n=400 için bile kestirim oldukça iyidir. Ayrıca (2,1,1) YSA nın kestirim yeteneği ile (2,3,1) YSA nın yeteneği biribirine oldukça yakındır. Şekillerdeki değerlerin çok küçük olması nedeniyle hata bileşenlerinin λ_n ile C_{1n} biribirinden aslında oldukça farklıymış izlenimi verebilir oysa bu doğru değildir, çünkü şekiller geniş ölçekte çizilmiştir.



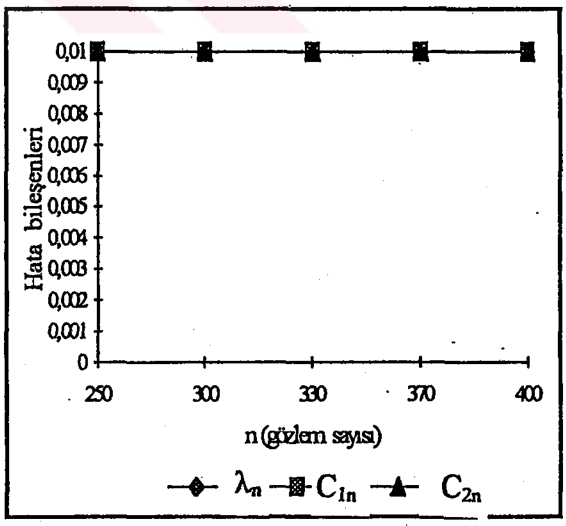
Şekil 5.28a Y.N.=3 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=0.00)



Şekil 5.28b Y.N.=3 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=3, λ_{th}=0.00)



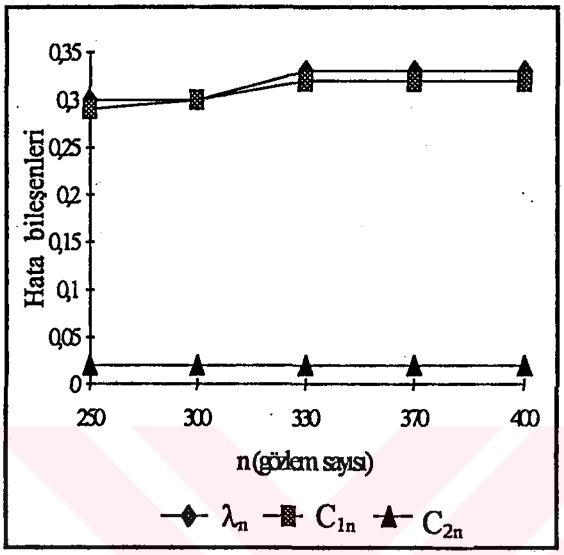
Şekil 5.28c Y.N.=3 için n_v=(2,3,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=0.00)



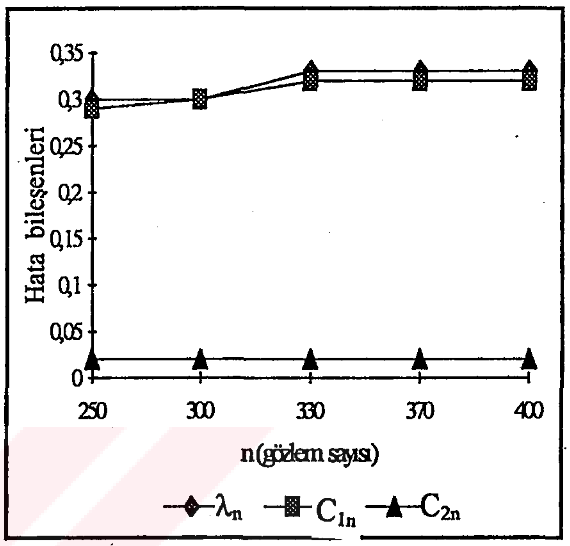
Şekil 5.28d Y.N.=3 için n_v=(2,3,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=3, λ_{th}=0.00)

Yapı No: 7 (x,y,z)=(0.8,0.8,0.8)

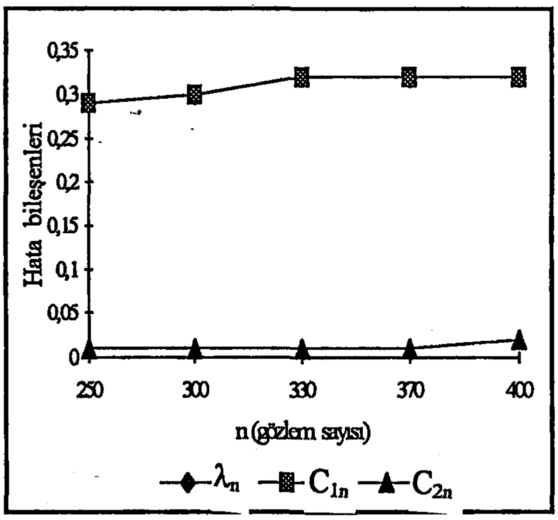
Şekil 5.29 den görüleceği gibi n=250 ile elde edilen final ağırlık vektörüyle n=400'e kadar yapılan genelleştirmeler oldukça iyidir. Bir başka not şudur. n=400'e yaklaşıldıkça zaten çok küçük olan kestirim hatası daha da azalmaktadır. n_h=1 ve n_h=3 için elde edilen değerlerde n_h=3 için elde edillenlerin λ_{th}=0.289 teorik hata değerine biraz daha yakın olduğu gözlendi. Yani kompleksitenin artması bu kez çok azda olsa genelleştirme yeteneğini iyileştirdi.



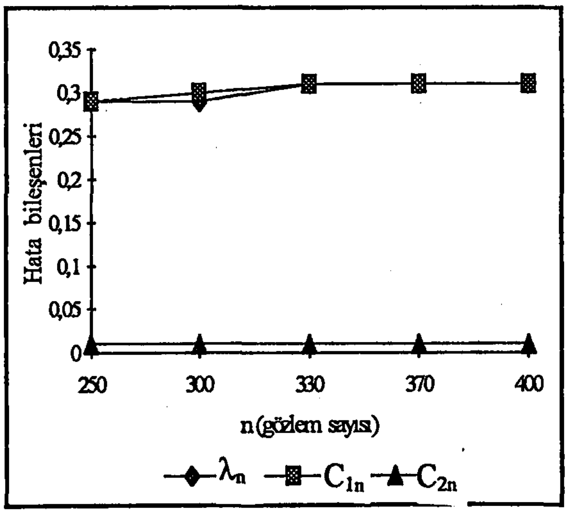
Şekil 5.29a Y.N.=7 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=0.289)



Şekil 5.29b Y.N.=7 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=3, λ_{th}=0.289)



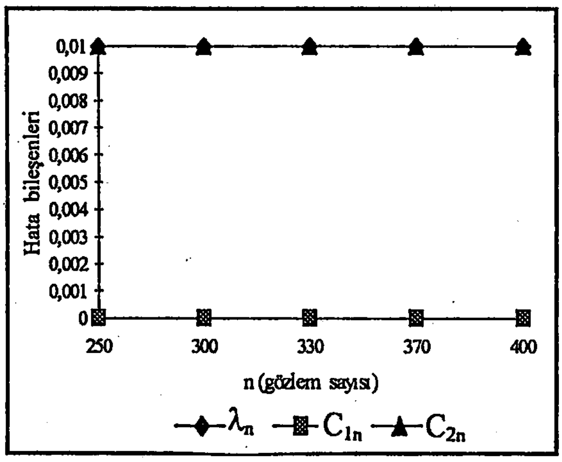
Şekil 5.29a Y.N.=7 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=0.289)



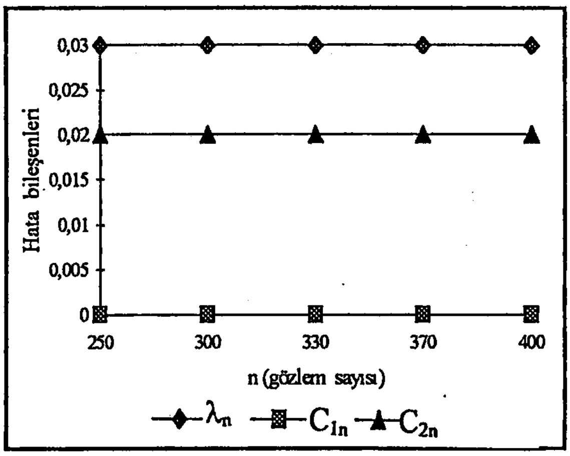
Şekil 5.29d Y.N.=7 için n_v=(2,3,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=0.289)

Yapı No: 9 (x,y,z)=(1.0,1.0,1.0)

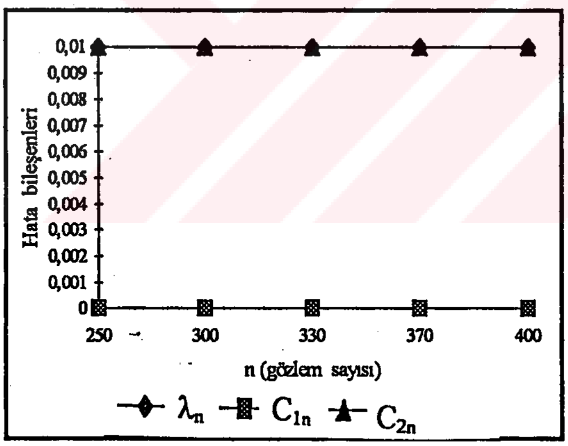
Şekil 5.30a-d Y.N.=9 için de ağın genelleştirme yeteneğinin oldukça iyi olduğunu gösterir. Şekil 5.30a ve Şekil 5.30b karşılaştırıldığında iw lere bağlı olarak bazı sapmaların olabileceğini gösterir. Ancak bu sapmalar önemsenmeyecek ölçüde küçüktür.



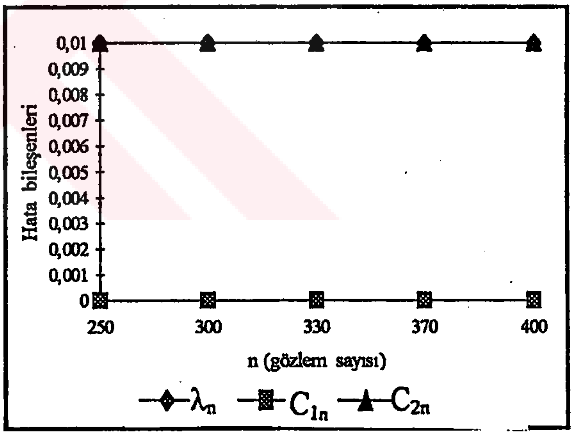
Şekil 5.30a Y.N.=9 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=0.00)



Şekil 5.30b Y.N.=9 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=3, λ_{th}=0.00)



Şekil 5.30c Y.N.=9 için n_v=(2,1,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=0.00)



Şekil 5.30d Y.N.=9 için n_v=(2,3,1) YSA nın genelleştirme yeteneği. n=250 eğitme kümesi. n=300-400 test kümesi. (iw=1, λ_{th}=0.00)

Genelleştirme yeteneği için genel sonuç: Gerek görsel verili olarak sunulan yapılar ve gerekse görsel verili sunulmayan öteki yapılar da dikkate alındı ve bütün iw ve n_h değerleri için n=250 elemanlı eğitme kümesinden elde edilen optimum final vektörleri başlangıç vektörleri olarak alınıp ağa eğitme yaptırılmadan n=400 e kadar olan veriyi kestirmesi istendi. Ağın kestirim (genelleştirme) yeteneğinin bütün yapılar için oldukça iyi (% 0.5 hatayı geçmeksizin) olduğu gözlendi. Ayrıca kompleksitenin artırılmasının genelleştirme yeteneğini çok azda olsa artırdığını gösterdi. Bir başka sonuç ise n=250 değerinin asimptotik olarak yeterli bir değer olabileceği bu örneklerle bir kez daha kanıtlandı.

5.6 ÇBND den Farklı Diğer Bazı Dağılımlar İçin Asimptotik Sonuçlar

Daha önce Kesim 5.2.1 de söz edildiği gibi bu tezde stokastik verinin YSA tarafından dönüştürülmesinde belirli bir düzeyde genelliğe ulaşmak için , ÇBND den farklı diğer bazı dağılımlardan üretilen veriye karşı YSA nın performansının araştırılması uygun olacaktır. ÇBND dışında çok boyutlu dağılımları üretebilecek etkili algoritmalar yok denecek kadar azdır. Bu nedenle üretilen ÇBND verisi kaynak olarak alınıp uygun dönüşümlerle yeni dağılımlar elde edildi. Bu verilerin işlendiği İUB1 ve İUB2 li YSA nın (n_h=1,2,3 parametreli) performansı araştırıldı.

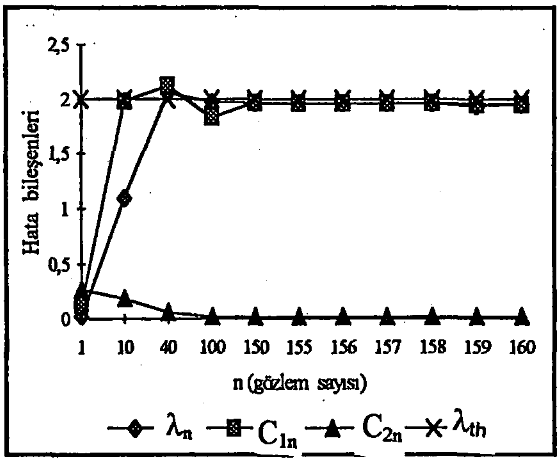
Dağılım Tipi	*Betimleme*
1	X_1 ve X_2 standart normal ve X_1 ve X_2 istatitistik bağımsız olmak üzere , yeni değişkenler $Y_1=X_1+X_2$ ve $Y_2=X_1-X_2$ olarak tanımlanır. Y_1 ve Y_2 de istatistik bağımsızdırlar.

$g(Y_1)=E(Y_2|\ Y_1)=0$ ve $C_1=E(Y_2-g(Y_1))^2=2$ olarak bulunur. Burada bu türetmeler yapılmayacaktır. Bu konu için örn. bkz. (Mood ve ark.1974 s198). Bununla birlikte yalnızca bu dağılıma ilişkin olmak üzere C_1 bileşenin nasıl elde edildiğini göstermek yararlı olacaktır.

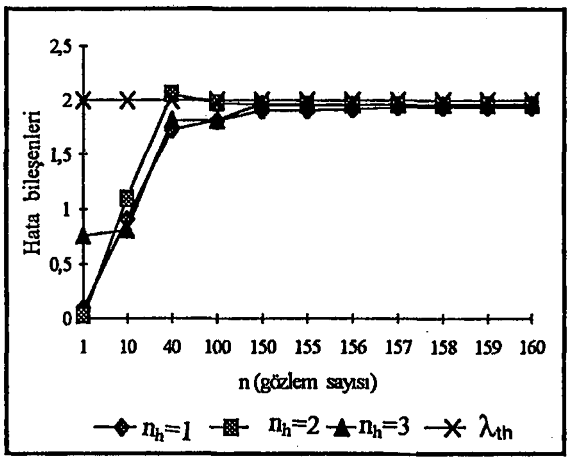
$g(Y_1)= E(Y_2|Y_1)=E(X_1 -X_2|X_1+X_2)=E(X_1|X_1)+E(X_1|X_2)-E(X_1|X_2)+E(X_2|X_2)=E(X_1)+E(X_2)=0+0=0$

O halde $C_1=E(Y_2)^2=E(X_1^2-2X_1X_2+X_2^2)=E(X_1^2)-2E(X_1X_2)+E(X_2^2)=1$(Standart normal olduğu için $VarX_1=1$)+ 0 (X_1 ve X_2 istatistiksel bağımsız oldukları için)+ 1 (Standart normal olduğu için $VarX_2=1$)=1+0+1=<u>2</u> (Aranılan sonuç).

Özet olarak : Dağılım Tipi 1 için $\lambda_{th}=C_1+C_2=2+0$ (Burada YSA nın C_2 yi sıfıra yakın ölçüde bulduğu varsayılmıştır. Şekil 5.31a-b Dağılım Tipi 1 (D.T.=1) için n=160 için YSA nın performansını (çeşitli hata bileşenlerinin teorik değerlerine yakınlığı cinsinden) sergiler. {Burada n için 250 yerine 160 alınmasının biricik nedeni yapılan ön simülasyon deneylerinde hemen hemen bütün dönüşüm tipleri için n=160 ya da 200 den sonra n=250 sonuçlarına oldukça yakın sonuçlar elde edilmiş olmasıdır}. Şekil 5.31a den sabit ağ kompleksitesi n_h=2 için çeşitli hata bileşenlerinin asimptotik davranışının teorik beklentilere uygun olduğu görülür. Şekil 5.31b de ise hata metriği λ_n in artan ağ kompleksitesi için büyük *n* davranışının uyumlu olduğu sonucu çıkar. Genel olarak n_h' ın artması performansı biraz iyileştirir (Şekil 5.31b) ancak n_h=1 bile verinin temsili için yeterlidir.



Şekil 5.31a D.T.=1 için çeşitli hata bileşenlerinin büyük n için evrimi. (n_h=2 ve λ_{th}=2.00 teorik hata değeri, yani asimptotik limit)

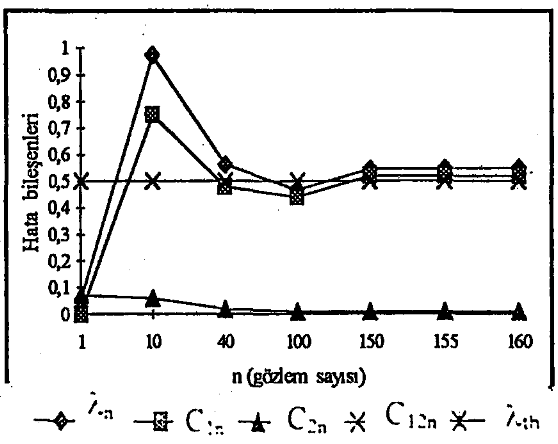


Şekil 5.31b D.T.=1 için λ_n hata metriğinin artan ağ kompleksitesi altında büyük n için evrimi.

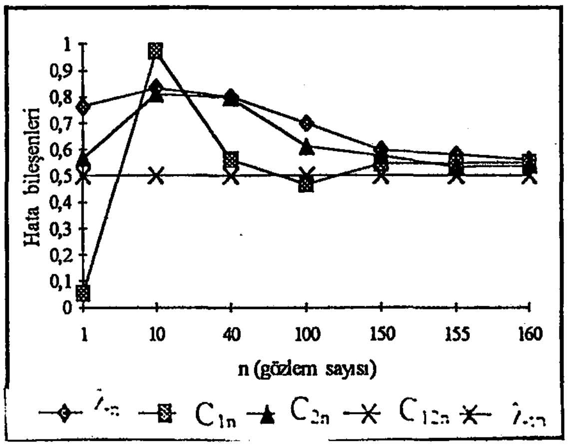
Dağılım Tipi	*Betimleme*
2	X_1, X_2 standart normal ve istatistiksel bağımsız. $Y_1=\sum_{i=1}^{2}(X_i-\overline{X}_n)^2$ ve $Y_2=(1/2)\sum_{i=1}^{2}X_i=\overline{X}_n$. $Y_1\approx$ (1 serbestlik dereceli chi (ki) kare ($\chi^2(1)$)) ve $Y_2\approx N(0,1/2)$. Y_1 ve Y_2 istatistiksel bağımsız .

$g(Y_1)=E(Y_2|Y_1)=E(Y_2)=0 \Rightarrow C_1=E(Y_2-g(Y_1))^2=\mathrm{Var}Y_2=1/2=\underline{\mathbf{0.5}}$

Şekil 5.32a-b D.T.=2 için hata bileşenlerinin büyük n için evrimini sergiler. Gözlemler teorik beklentilere oldukça uygundur. D.T.=2 için de ağın performansının artan ağ kompleksitesi ile bir miktar arttığını gösterir. Bununla birlikte n_h=1 li YSA yeterli olabilir.



Şekil 5.32a D.T.=2 için çeşitli hata bileşenlerinin büyük n için evrimi. (n_h=2 ve λ_{th}=0.50 teorik hata değeri, yani asimptotik limit)



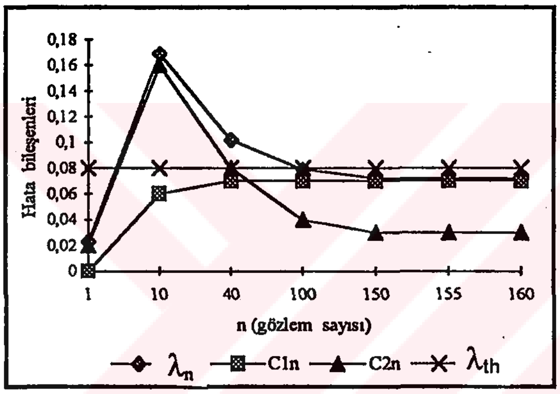
Şekil 5.32b D.T.=2 için λ_n hata metriğinin artan ağ kompleksitesi altında büyük n için evrimi (λ_{th}=0.50)

Dağılım Tipi	Betimleme
3	X_1, X_2,X_3 standart normal ve istatistiksel bağımsız. $Y_1=\alpha X_1^2$, $Y_2=e^{\beta X2}$, $Y_3=\gamma X_3+1$ (α,β,γ pozitif sabitler). Bu durumda Y_1,Y_2,Y_3 istatistiksel bağımsızdır.

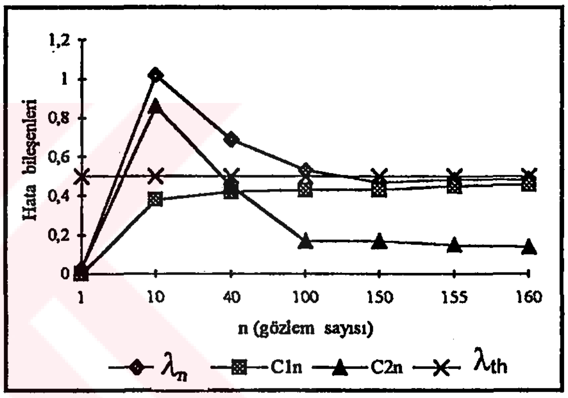
Bu dağılım tipi Y_1, Y_2 ya da Y_3 bileşenlerinden birinin output öteki ikisinin ise input bileşenleri olmasına göre durumlandırılacaktır.

D.T.=3 <u>Durum 1</u>: Y_1 (output) ve Y_2,Y_3 (input bileşenleri)

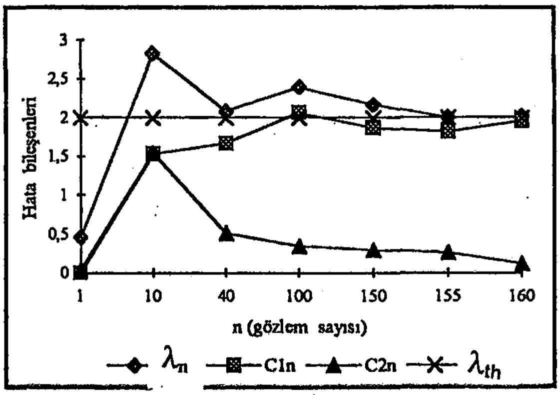
$E(Y_1|Y_2,Y_3)= g(Y_2,Y_3)=\underline{2\alpha^2}$ (Bu sonuç β ve γ dan bağımsızdır). Şekil 5.33a-d çeşitli α,β,γ için n_h=2 li YSA nın performansını sergiler. Sonuçlar teori ile iyi uyumludur.



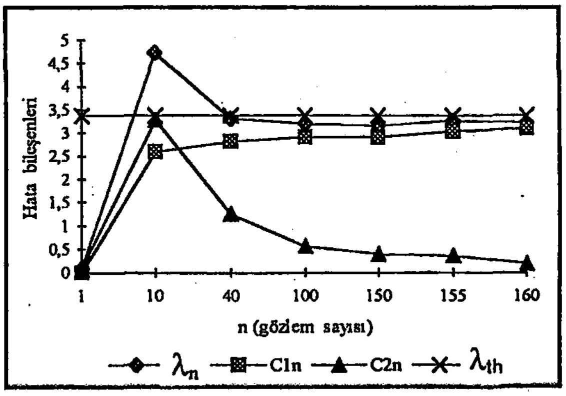
Şekil 5.33a D.T.=3 ve Durum 1 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=0.2, β=0.5, γ=1.0) λ_{th}=0.08.



Şekil 5.33b D.T.=3 ve Durum 1 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=0.5, β=1.2, γ=0.5) λ_{th}=0.50.



Şekil 5.33c D.T.=3 ve Durum 1 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=1.0, β=0.3, γ=0.7) λ_{th}=2.00.

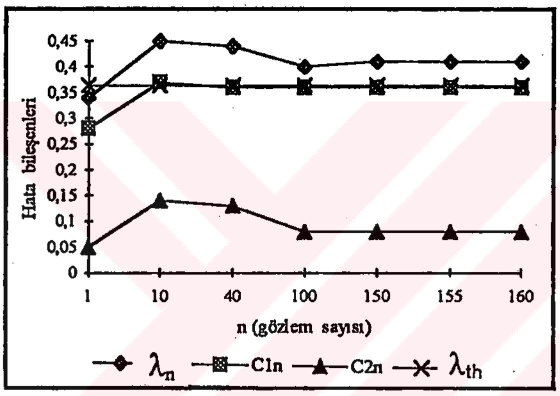


Şekil 5.33d D.T.=3 ve Durum 1 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=1.3, β=0.8, γ=2.2) λ_{th}=3.38.

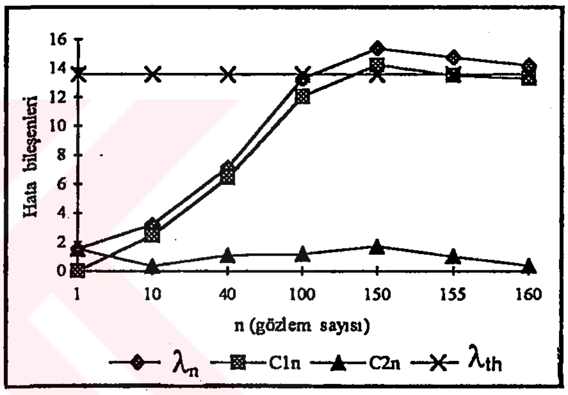
D.T.=3 <u>*Durum 2*</u>: $\mathbf{Y_2}$ (output) ve $\mathbf{Y_1, Y_3}$(input bileşenleri)

$$\boxed{E(Y_2|Y_1,Y_3)=g(Y_1,Y_3)=e^{\frac{1}{2}\beta^2} \text{ ve } E(Y_2-g(Y_1,Y_3))^2=E(Y_2-E(Y_2))^2=\mathrm{Var}Y_2=e^{2\beta^2}-e^{\beta^2}}$$

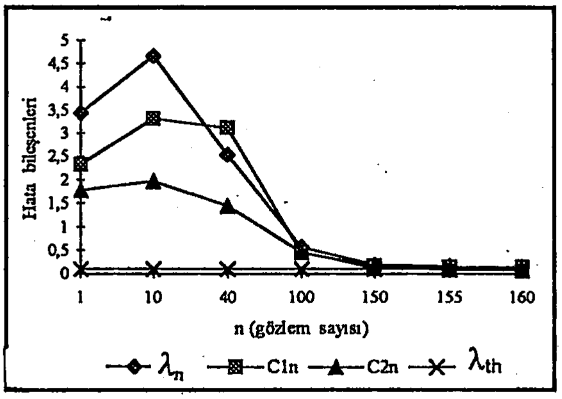
Burada $\mathrm{Var}\mathbf{Y_2}=C_1$ dir. (Yukarıda kutu içine alınan değer $\mathbf{Y_2}$ nin lognormal bir dağılım olduğu göz önüne alınarak hesaplanmıştır). Şekil 5.34a-d çeşitli α,β,γ için $n_h=2$ li YSA nın performansını sergiler. Sonuçlar teorik öngörüleri doğrular niteliktedir. Burada özellikle Şekil 5.34b deki teorik değerin yakınsamak için oldukça büyük bir değer (13.59) olmasına karşın YSA nın görevini yerine getirip yakınsama işini başarıyla gerçekleştirdiğine özellikle dikkat çekilmelidir. Not: $n_h=1$ ve 3 için görsel veri sunulmadı, ancak her ikisi de için sonuçlar $n_h=2$ dekine yakındır.



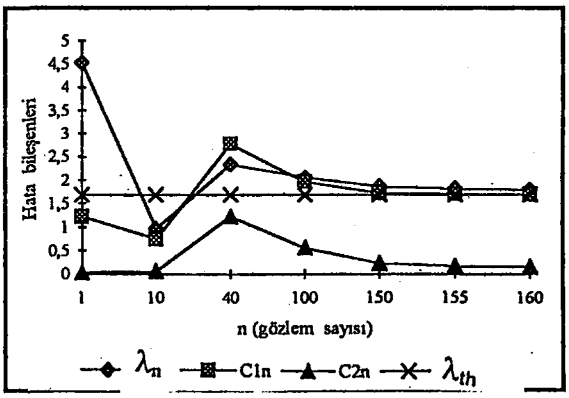
Şekil 5.34a D.T.=3 ve Durum 2 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=0.2, β=0.5, γ=1.0) λ_{th}=0.364.



Şekil 5.34b D.T.=3 ve Durum 2 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=0.5, β=1.2, γ=0.5) λ_{th}=13.59.



Şekil 5.34c D.T.=3 ve Durum 2 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=1.0, β=0.3, γ=0.7) λ_{th}=0.102.



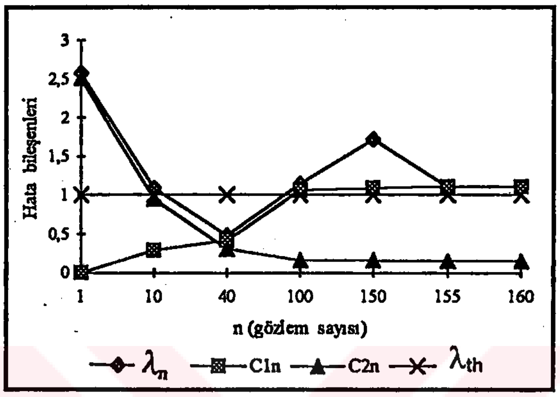
Şekil 5.34d D.T.=3 ve Durum 2 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=1.3, β=0.8, γ=2.2) λ_{th}=1.69.

D.T.=3 *Durum 3*: Y_3 (output), Y_1, Y_2 (input)

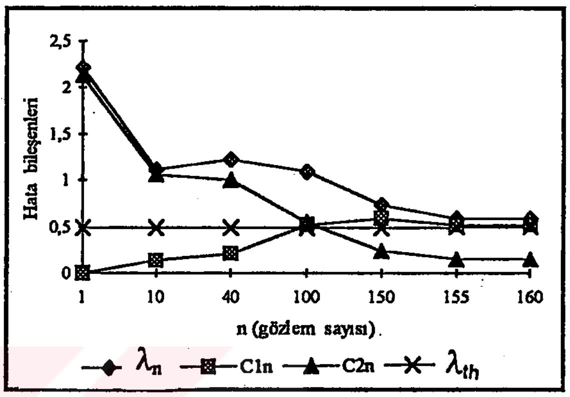
$E(Y_3|Y_1,Y_2)=g(Y_1,Y_2)=1$ ve $C_1 = \mathrm{Var}Y_3 = \gamma^2$

Şekil 5.35a-b $\gamma = 1$ ve 0.49 için $n_h=3$ li bir ağda çeşitli hata bileşenlerinin evrimini gösterir Sonuçlar teori ile uyumludur.

Sonuç: D.T.=3 için YSA teorinin öngördüğü performansı sergiler.



Şekil 5.35a D.T.=3 ve Durum 2 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=0.2, β=0.5, γ=1.0) λ_{th}=1.0.



Şekil 5.35b D.T.=3 ve Durum 2 için çeşitli hata bileşenlerinin asimptotik evrimi. (α=1.0, β=0.3, γ=0.7) λ_{th}=0.49.

Dağılım Tipi	*Betimleme*
4	X_1,X_2,X_3 standart normal ve istatistiksel bağımsız. Yeni dönüşüm: $Y_1=X_1$, $Y_2=(X_1+X_2)/2$, $Y_3=(X_1+X_2+X_3)/3$. Burada Y_1, Y_2, Y_3 biribirlerine *istatistiksel bağımlıdırlar*. Bunu görmek için bkz. (Mood ve ark.1974 s212). Birleşik olasılık dağılımı fonksiyonu marjinal olasılık dağılımı fonksiyonları cinsinden yazılamaz.

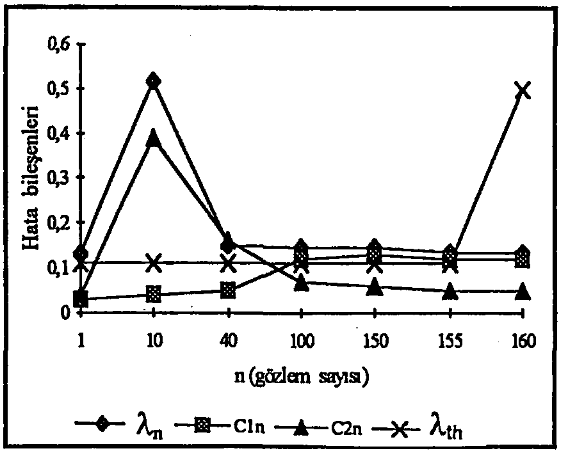
Durum 1: Y_3 (output) ve Y_1,Y_2 (input bileşenleri)

Yapılan analitik hesaplar sonucunda $E(Y_3|(Y_1,Y_2)=g(Y_1,Y_2)=2/3\ Y_2$ ve $C_1=E(Y_3-g)^2=1/9=\underline{0.111}$ elde edildi.

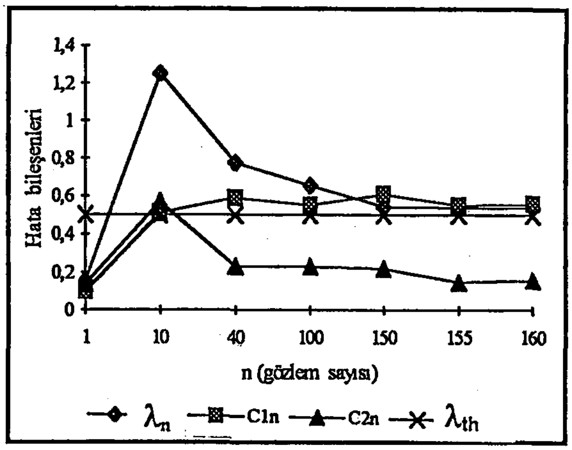
Durum 2 : Y_1 (output), Y_2,Y_3 (input bileşenleri) $E(Y_1|(Y_2,Y_3)=g(Y_2,Y_3)=Y_2$ ve $C_1=E(Y_1-g)^2=E(Y_1-Y_2)^2=.E((X_1-X_2)/2)^2=(1/4)E(X_1^2-2X_1X_2+X_2^2)=(1/4)(1-0+1)=\underline{0.5}$ elde edilir.

Şekil 5.36a-b D.T:=4 için n_h=2 li YSA nın performansını sergiler. Görüldüğü gibi sonuçlar teorinin öngördüklerine oldukça iyi uyumludurlar.

Sonuç: *İstatistiksel bağımlı* değişkenlerin de YSA tarafından işlenmesi teorik beklentilerle aynı doğrultudadır.



Şekil 5.36a D.T.=4 ve Durum 1 için çeşitli hata bileşenlerinin asimptotik evrimi. λ_{th}=0.111.

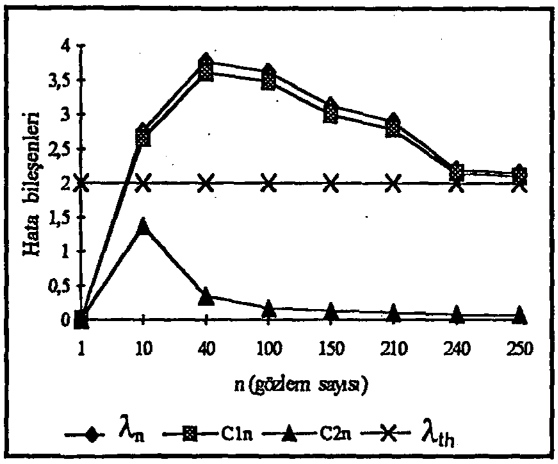


Şekil 5.36b D.T.=4 ve Durum 2 için çeşitli hata bileşenlerinin asimptotik evrimi. λ_{th}=0.50.

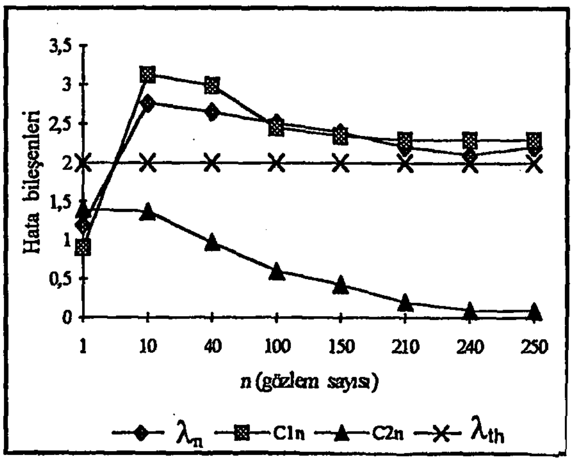
Dağılım Tipi	*Betimleme*
5	$\mathbf{X_1}$, $\mathbf{X_2}$ standart normal ve istatistiksel bağımsızdırlar. $\mathbf{Y_1=X_1+X_2}$, $\mathbf{Y_2}=(\mathbf{X_1-X_2})^2/2$. $\mathbf{Y_1,Y_2}$ istatistiksel bağımsızdırlar.

D:T.=5 *Durum 1*: $E(\mathbf{Y_2}|\mathbf{Y_1})=1$ ve $C_1=E(\mathbf{Y_2}-1)^2=E\mathbf{Y_2}^2-2E\mathbf{Y_2}+1$. Bu son ifade $\mathbf{Y_2}$ nin marjinal olasılık dağılımıyla hesaplanabilir. $f_{Y2}\,(y_2)=\frac{1}{\sqrt{2\pi}}y_2^{\frac{-1}{2}}e^{\frac{-y}{2}}\;(y_2\geq 0)$ (Bu ifade r=1/2 ve λ=1/2 parametreli gama dağılımıdır:) Bu dağılım kullanılarak $E\mathbf{Y_2}^2=\int y_2^2\frac{1}{\sqrt{2\pi}}y_2^{\frac{-1}{2}}e^{\frac{-y}{2}}\;(y_2\geq 0)=$ $(2^{1/2}2^2\int_0^\infty u^{3/2}e^{-u}du)/2^{1/2}\sqrt{\pi}=\frac{4}{\sqrt{\pi}}\frac{3\sqrt{\pi}}{4}=\mathbf{3}$ olur. $\therefore C_1=3-2.1+1=\underline{2}$

D.T.=5 *Durum 2* : $E(\mathbf{Y_1}|\mathbf{Y_2})=E(\mathbf{Y_1})=g(\mathbf{Y_2})=0+0=0$ ve $C_1=E(\mathbf{Y_1}-g)^2=E\mathbf{Y_1}^2=E(\mathbf{X_1}^2+2\mathbf{X_1X_2}+\mathbf{X_2}^2)=1+0+1=\underline{2}$. Görüldüğü gibi her iki durumda da C_1=2 dir. Şekil 5.37a-b D.T:=5 için n_h=3 li YSA nın performansını sergiler. Sonuçlar bu kez de teorinin öngördüğü gibidir. Hata bileşenleri λ_{th}=2 , C_1=2 ve C_2=0 asimptotik değerlerine yaklaşırlar. Bu örnekte yeniden n=250 alındığına dikkat edilmelidir.



Şekil 5.37a D.T.=5 ve Durum 1 için çeşitli hata bileşenlerinin asimptotik evrimi. λ_{th}=2.00.



Şekil 5.37b D.T.=5 ve Durum 2 için çeşitli hata bileşenlerinin asimptotik evrimi. λ_{th}=2.00.

Dağılım Tipi	*Betimleme*
6	X_1, X_2 standart normal ve istatistiksel olarak bağımsız. $Y_1=X_1+X_2$ ve $Y_2=X_1 / X_2$. Y_1 ve Y_2 ,aşağıda analitik yazımı verilen $f(y_1,y_2)$ birleşik olasılık fonksiyonundan görüleceği gibi, *istatistiksel olarak bağımlıdırlar.*

$f(y_1,y_2)=\frac{1}{2\pi}\frac{|y_1|}{(1+y_2)^2}\exp\left[-\frac{1}{2}\frac{(1+y_2^2)y_1}{(1+y_2)^2}^2\right]$ $y_1,y_2 \in(-\infty,\infty)$. Bu fonksiyon $h(y_1)$ $t(y_2)$ gibi ik

fonksiyonun çarpımı olarak yazılamaz. O nedenle Y_1, ve Y_2 istatistiksel olarak bağımlıdırlar.

$$E(Y_2|Y_1)=g(Y_1)=I=\frac{\int_{-\infty}^{\infty}\frac{y_2}{(1+y_2)^2}\exp(-\frac{1}{2}\frac{(1+y_2^2)}{(1+y_2)^2}y_1^2)dy_2}{\int_{-\infty}^{\infty}\frac{1}{(1+y_2)^2}\exp(-\frac{1}{2}\frac{(1+y_2^2)}{(1+y_2)^2}y_1^2)dy_2}$$

I integrali $1+y_2=y$ dönüşümü ile kolayca alınabilir. $I=\frac{\int_{-\infty}^{\infty}\frac{y-1}{y^2}\exp(-\frac{1}{2}\frac{(1+(y-1)^2)}{y^2}y_1^2)dy}{\int_{-\infty}^{\infty}\frac{1}{y^2}\exp(-\frac{1}{2}\frac{(1+(y-1)^2)}{y^2}y_1^2)dy}$

I integralinin payındaki integrand $(\frac{1}{y}-\frac{1}{y^2})\exp(\frac{-1}{2}\frac{(1+(y-1)^2)}{y^2}y_1^2)$ olarak yeniden yazılır. E

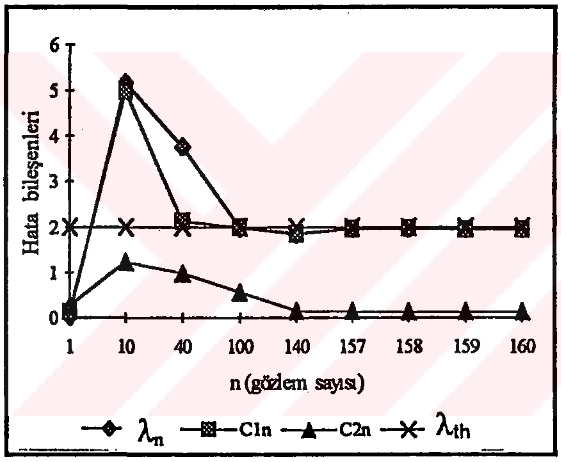
yazım kullanıldığında I integralinin son durumu

$$I = \frac{\int_{-\infty}^{\infty} \frac{1}{y}\exp(-\frac{1}{2}\frac{(1+(y-1)^2)}{y^2})y_1^2 dy - \int_{-\infty}^{\infty} \frac{1}{y^2}\exp(-\frac{1}{2}\frac{(1+(y-1)^2)}{y^2}y_1^2)dy}{\int_{-\infty}^{\infty} \frac{1}{y^2}\exp(-\frac{1}{2}\frac{(1+(y-1)^2)}{y^2}y_1^2)dy} = \frac{I_1 - I_2}{I_3}$$

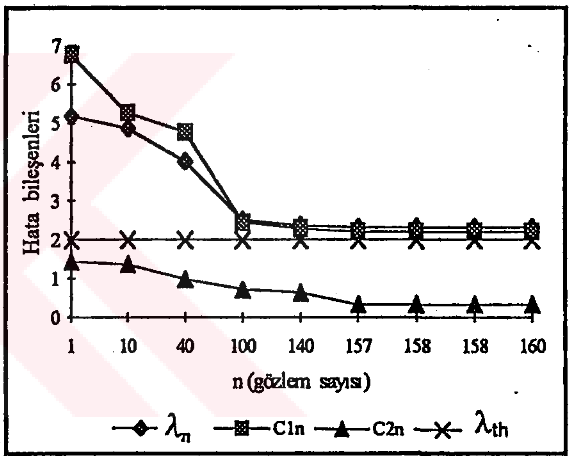
Burada $I_2=I_3$ dir ve ayrıca I_1 integralinin integrandı $y=-y$ dönüşümü altında tek fonksiyonudur ve böylece $I_1=0$ olur. I integrali ise -1 dir. $g(Y_1)=\underline{-1}$ olarak elde edilir.

$C_1=E(Y_1-g)^2=E(Y_1+1)^2=E(Y_1^2)+2E(Y_1)+1=1+0+1=\underline{2}$ bulunur.

Şekil 5.38 a-b den görüleceği gibi n=100 ya da 140 dan sonra λ_n ve C_{1n} değerleri terorik limit değerine (=2.00) doğru yakınsamaktadırlar. n_h=3 li YSA ile n_h=1 li YSA nın performansları arasında bu dağılım tipi için göze çarapan bir fark yoktur.



Şekil 5.38a D.T.=6 için n_h=1 li YSA da çeşitli hata bileşenlerinin asimptotik evrimi. λ_{th}=2.00.



Şekil 5.38b D.T.=6 ve n_h=3 çeşitli hata bileşenlerinin asimptotik evrimi. λ_{th}=2.00.

ÇBND' den farklı dağılımlar için genel sonuç: İncelenen dağılım tipleri çerçevesinde YSA nın ÇBND'den farklı dağılımlar için stokastik veriyi işlemesinin sonuçları teorik beklentilere uygundur. Bu uygunluğun göstergesi analitik olarak hesaplanan g koşullu beklenen değere YSA tarafından görece düşük hatalarla yaklaşım yapılmış olmasıdır. Ayrıca bu dağılım tipleri içinde n=100 ya da 140 dan sonra asimptotik limite oldukça yaklaşılmıştır. C_2 bileşenenine olan yakınsama ise daha hızlıdır (n=40 ya da 100 den sonra).

5.7 GENEL SONUÇ

(i) **Farklı metotlarla veriyi dönüştürme yeteneği:** Çok sayıda stokastik veri matrislerinin Geriye Yaymalı Yapay Sinir Ağları (GYYSA) ile işlenmesi söz konusu olduğunda standart olmayan GYYSA tipindeki Genişletimiş Delta Bar Delta Metodu (GDBDM) standart GYYSA'ya oranla daha efektif olarak kullanılabilir. Özellikle standart GYYSA nın minimizasyon sürecinde çok düşük yakınsama hızına sahip olması nedeniyle GDBDM görece çok daha yüksek yakınsama hızı sağlayabilir. Ancak bu çalışmada bu iki metodun karşılaştırılması için yalnızca Çok Boyutlu Noramal Dağılım (ÇBND) kullanıldığı için GDBDM'in olası üstünlüğüne karar verebilmek için ÇBND dışındaki dağılımlar da denenmelidir.

(ii) **Asimptot altı bölgede performans:** Stokastik veri (iki boyutlu normal dağılımdan çekilen) YSA tarafindan asimptot altı bölgede işlenirken sabit n_h (gizli tabaka nöronları sayısı) verildiğinde r_{XY} (**X**:input vektörü, **Y**:hedef vektörü) korelasyon katsayısı sıfirdan bire arttığında YSA nın dönüştürme yeteneği kesinlikle artar. Bununla birlikte **X** ve **Y** arasındaki ilişki deterministik olmadığında ($r_{XY}\neq\pm1$) ne n_h 'ı artırmak ne de minimizasyon için gerekli iterasyon sayısını artırmak YSA nın dönüştürme yeteneğini artırır. Bunun temel nedeni, C_1 ile gösterilen koşullu variyans değerinin alttan sınırlı olmasıdır. Bu çalışmada asimptot altı bölgede bile olsa elde edilen sonuçlar White (1989a) tarafından ve Hornik ve ark.(1989) tarafından geliştirilen teorik değerlendirmeleri destekler niteliktedir. White (1989a) gösterdi ki, $r_{XY}\neq\pm1$ durumunda YSA nın yapabileceği en iyi iş **X** verildiğinde **Y** nin koşullu beklenen değerini yaklaşımla bulmaktır. Bu ise $C_2\approx0$ ve $C_1\neq0$ durumuna karşılık gelir. Oysa $r_{XY}\neq\pm1$ iken **X** ile **Y** arasında ölçüsü sıfır olan kümenin dışındaki bölgede Y= aX+b biçiminde afin bir ilişki (deterministik) vardır ve C_1=0 ve C_2=0 dır. Bu ise Hornik ve ark.(1989) ulaştıkları sonuçları destekler. Yani deterministik ilişki olduğunda (**X** ve **Y** arasında) YSA nın yaklaşım yapma yeteneği istenilen derecede iyidir.

(iii) **Asimptotik bölgede ÇBND için dönüştürme yeteneği:** YSA nın asimptotik davranışında (3 boyulu ÇBND veri matrisleri için) ise yine White (1989a) tarafından geliştirilen teorik değerlendirmelere uygun sonuçlar elde edildi. YSA nın ÇBND verisini dönüştürmedeki başarısı λ_n hata fonksiyonlarının (n: input ve hedef için gözlem sayısı) bileşenleri cinsinden ölçüldüğünde, YSA nın limit hata değerlerine başarıyla ulaştığı görüldü. Ayrıca veri matrislerinin büyüklüğü sabit iken, n_h değeri arttıkça YSA nın performansının bir miktar arttığı gözlendi. Bununla birlikte, teorik limitler nedeniyle hiçbir zaman belirli korelasyon yapıları için sıfır hata düzeyine erişilmediği görüldü.

(iv) Veri matrisinin korelasyon yapısı için final ağırlık vektör bileşenlerinin korelasyonları arasında bazı benzerlikler vardır: Veri matrisinin korelasyon yapısıyla YSA nın final ağırlık vektörleri bileşenlerinin korelasyon yapılarıyla bir benzerlik olabileceğine ilişkin bazı gözlemler yapıldı. Özellikle r_{XY}=1 ve 0 durumlarında (not: burada r_{XY} notasyonu jeneriktir. Yani iki input durumunda r_{X1Y} ya da r_{X2Y} değerlerinden herhangi birini gösterebilir) bu benzerlikler daha çarpıcı olarak gözlendi. Ancak ara durumdaki r_{XY} ler için daha fazla araştırma yapılmaya değerdir.

(v) Genelleştirme yeteneği: YSA nın genelleştirme yeteneğinin ÇBND verisini kullanılarak oldukça iyi olduğu çok sayıda korelasyon yapısı için gözlendi.

(vi) ÇBND 'den farklı dağılımlar: Son olarak, ÇBND den farklı diğer bazı dağılımlardan üretilen veri matrislerinin işlenmesiyle elde edilen sonuçlarda White (1989a) tarafından ortaya konulan teorinin geçerli olduğu gözlendi. Yani elde edilen sonuçlar denk.(2.2.) ile tutarlıdır ve onu önemli ölçüde destekler. Bütun bunları göz önüne alarak incelenen örnekler çerçevesinde belirli bir güven düzeyiyle YSA nın stokasrtik veriyi işlemede oldukça başarılı olduğu söylenebilir.

(vii) Çeşitli dağılım grupları için yakınsama hızları: ÇBND ve ÇBND'den farklı diğer bazı dağılımlar için YSA nın yakınsama hızları arasında önemli ölçüde bir fark gözlenmedi. Ya da daha açıkçası iki ana dağılım grubu (gerek ÇBND ve gerekse ÇBND olmayan dağılımlar) için YSA nın yakınsama hızı biribirine oldukça yakındır. Örneğin λ_n ve C_{1n} bileşenleri n her iki dağılım grubu için de yaklaşık olarak n= 160 dan sonra (bazan n=100 den sonra) asimptotik limit değerlerine yakınsamaktadırlar. C_{2n} bileşeni ise her iki dağılım grubu için genel olarak n= 40 ya da 100 den sonra asimptotik limit değerine yakınsamaktadır. Kısaca YSA nın stokastik veriyi işlemesinde asimptotik limit değerlerine erişilmesini dağılım tiplerine bağımlı olmadığı (en azından burada incelenen dağılım grupları için) söylenebilir.

Gelecekteki Çalışmalar İçin Öneriler

Bu çalışma stokastik verilerinin işlenmesinde YSA literatüründe var olan teorilerin sistematik deneylerle denenmesi için bir ilk adımdır. Bundan sonra White(1989a) tarafından önerilen öteki konulara da eğilmek son derece yararlı olacaktır. Örneğin, fazlalık input hipotezi ve fazlalık gizli tabaka nöronu hipotezinin deneylerle araştırılması gerekir. Bundan başka bir olanak ise input tabakası nöronları sayısının adım adım artırılarak YSA nın performansı araştırılmalıdır. Hornik ve ark.(1989) tarafından önemle vurgulandığı gibi input uzayı boyutunun belirli bir ölçüde artırıldığında n_h sayısının optimum olarak ne olması gerektiği ise ilginin odaklanacağı bir bir başka konu olmalıdır.

KAYNAKLAR

Amit, D.J. **(1986)**. Neural networks-achievements, prospects difficulties, The Physics of structure formation int. symposium, Tubingen.

Anderson, T.W. **(1958)**. An introduction to multivariate statistical analysis. John Wiley and Sons, Inc., London.

Aoyama, T., and Ichikawa, H.**(1991)**. Chem. Pharm.Bull.,**39**,358-366.

Ash ,R. **(1972)**. Basic Probability Theory. John Wiley and sons, Inc., New York.

Barron , A. **(1989)**. Statistical properties of artificial neural networks. In 'Proceedings of 28 th IEEE Conf. on Decision and Control. Tampa, Florida, 280-285, New York: IEEE press.

Cerny, V. **(1985)**. Thermodynamical approach to the traveling salesman problem: An sufficient algorithm. J.Opt. Theory Appl., **45**, 41-51

Davis, R.B. **(1977)**. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, **64**, 247-254.

Davis, R.B. **(1987)**. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, **64**, 33-43.

Gallant, A.R., and White, H. **(1988)**. A unified theory of estimation and inference for nonlinear dynamics models. Oxford: Basil Blackwell.

Golden, R. (1988). A unified framework for connectionist systems. Biological Cybernetics, **59**, 109-120.

Haaser, N.B., and Sullivan, J.A. **(1971)**. Real Analysis. Van Nostrand Reinhold Company, New York.

Hansen, B.E. **(1992)**. Inference when a nuisance parameters is not identified under the null hypothesis. Revised manuscript. (Bu makale H.White tarafindan yazara gönderilmiştir).

Harth, E. **(1982)**. Windows on the mind. Pelican.

Hopfield , J.J.**(1982)**. Natl. Acad. Sci. USA., **79**, 2554.

Holland, J. (1975). Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor.

Hornik, K., Stincombe, M.,and White, H. **(1989)**. Multilayer feedforward networks are universal approximators. Neural Networks, **2**, 359-368.

Jakobs, R.A. **(1988)**. Increased rates of convergence through learning rate adaptation. Neural Networks, Vol.1, 295-307.

Kirkpatrick, S. Gellatt, C.D. Jr., and Vecchi, M.P.(**1983**). Optimization by simulated annealing. Science, **220**, 671-680.

Lee , T.H., White, H., and Granger, C.W.J. (**1989**). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. UCSC Department of Economics Discussion Paper.

Loeve, M. (**1977**). Probability theory. Vol.1. New York: Springer-Verlag.

Mielnizcuk, J., and Tyrcha J. (**1993**). Consistency of multilayer perceptron regression estimators. Neural Networks, **I**, 425-464.

Minai, A.A., and Williams, R.D. (**1990**). Acceleration of back-propagation through learning rate and momentum adaptation. Proceedings of int. joint conference on neural networks., **1**, 676-679.

Minsky, M. (**1987**). The society of mind. MIT Press. Cambridge (Mass.)

Mood, M.M., Graybill, F.A., and Boes, D.C. (**1974**). Introduction to the theory of statistics. 3 rd. Ed. McGraw-Hill book company.

NAG Fortran Library. Numerical Algorithms Group Limited. (**1990**). Volume 7. Oxford, England.

Ronning, G. (**1977**). A simple scheme for generating multivariate gamma distributions with non-negative covariance matrix, Techno., **19**, 179-183.

Rumelhart, D.E., and McClelland, J.L. (**1986**). Parallel distributing processing, Vol. **1**. MIT press. Cambridge (Mass.).

Salt, D.W., Yıldız, N., Livingstone, D.J., and Tinsley C.J. (**1992**). The use of Artificial Neural Networks in QSAR. Pestic. Sci. 36, 161-170.

Sietsma, J., and Dow, R.J.F. (**1991**). Creating artificial neural networks that generalize. Neural Networks, **4**, 67-79.

Spect, D.F. (1991). A general regression neural network. IEEE transactions on neural networks. Vol. **2**. No:6. 568-576.

Stillings, N.A. (**1987**). Cognitive science: An introduction. MIT Press. Cambridge (Mass.)

Sungur, M. (**1995**). Learning in feedforward random neural networks. The fourth Turkish symposium on artificial intelligence and neural networks, 133-140, Gebze-Turkey.

Tisby, N., Levin, E., and Solla, S. **(1989)** Consistent inference of probabilities in layered networks: prediction and generalization. Proceedings of the int. joint conference on neural networks, Washington D.C. II, 403-408 IEEE, New York.

White, H.**(1984)**. Asimptotic theory for neconometricians. Academic Press .New York.

White, H. **(1989a)**. Learning in artificial neural networks: A statistical perspective, Neural Computation, **1**, 425-464.

White, H. **(1989b)**. Some asymptotic results for learning in a single hidden layer feed-forward network models., Journal of American Statistical Association, **84**, 1003-1013.

White, H. **(1989c)**. An additional hidden unit test for neglected non-linearity . Proceedings of the International Joint Conf. on Neural Networks, Washington D.C., New York: IEEE Press, **II**, 451-455.

White, H. **(1990)**. Connectionist nonparametric regression: Multilayer feedforward networks can learn mappings. Neural Networks, **3**, 535-549.

White, H. **(1992)**. Estimation, Inference and Specification analysis. New York. Cambridge University Press. (in press). (Bu kitap H.White tarafindan yazara kişisel görüşme yoluyla gönderilmiştir).

Yıldız N. **(1987)**. Yapay sinir ağlarının spin camı modeliyle incelenmesi. Yayınlanmamış M.Sc.tezi. Cumhuriyet Üniversitesi, Sivas, Türkiye.

Yıldız, N. **(1995a)**. A comparison of the performances of two different type of feedforward neural networks in a stochastic environment. Turkish Journal of Engineering and Environmental Sciences, **19**, 199-203.

Yıldız, N. **(1995b)**. The correlation structure of training data and the fitting of Back Propagation network. The fourth Turkish symposium on artificial intelligence and neural networks. 233-242.

ÖZGEÇMİŞ

1961 yılında Sivas'ta doğdu. 1984 yılında Hacettepe Üniversitesi (Ankara) Müdendislik Fakültesi Fizik Mühendisliği bölümünden mezun oldu. 1984 Ekim ayında Cumhuriyet Üniversitesi (Sivas) Fen-Edebiyat Fakültesi Fizik Bölümünde Araştıma Görevlisi olarak akademik yaşamına başladı. 1987 Ekim de Cumhuriyet Üniversitesi Fen Bilimleri Enstitüsünden Fizikte M.Sc. derecesi (Yüksek Lisans derecesi) aldı. 1990-1993 yılları arasında İngiltere'nin Porstmouth şehrinde University of Portsmouth'ta lisans üstü araştırmalar yaptı ve aynı üniversiteden 1993 de M.Phil diploması aldı. Halen Cumhuriyet Üniversitesi Fen-Edebiyat Fakültesi Fizik Bölümünde Araştırma Görevlisi olarak görevini sürdürmektedir.

EK:

```
C
C  PROGRAM BACKPROPAGATION
C   Yazan: Nihat  YILDIZ
C  Bu program belirli korelasyon yapılarına göre Çok Boyutlu Noramal Dağılım (ÇBND) ve ÇBND
C  den farklı diğer bazı dağılmlar için Bölüm 5 de analizi  yapılan  λn= C1n+ C2n  hata bileşenlerini
C  hesaplar. Bunun için Geriye Yaymalı (Back Propagation) bir Yapay Sinir Ağı (YSA) kullanır.
C  Sonuç  ağırlık vektörünü bir dosyaya yazar. İstenirse input- hedef  ve predikte edilmiş hedef
C  (output)  vektörünü de ayrı bir dosyaya yazar. Programda λn  E ile C1n  BL1 ile ve C2n ise BL2 ile
C  gösterilir. X(input vektörü), Y(hedef  vektörü), YP (predikte edilmiş outputu) gösterir.
C  1) Declerations
C  1.1) Character  Constants
C  FM  karakter bşilgileri için FORMAT  bilgisini gösterir.
C  İnput  ve hedef için
     CHARACTER* 20  FMXY, FMXYYP
C  Ağırlıklar  için
      CHARACTER *20  FMTWALL
C   1.2)  Parametreler
      INTEGER FLIPTG1, FLINIW, FLFWS, FLPMTS, FLERRO, FLİPTG2
C   FLIPTG1: Veri  matrislerini içeren  input dosyası
C   FLINIW: İlk ağırlık dosyasını saklayan input dosyası
C   FLFWS: Son  ağırlık dosyasını saklayan output dosyası
C   FLPMTS: Gözlem sayısı (NPT), iterasyon sayısı ( niter) vb. gibi parametreleri saklayan input
C   dosyası
C   FLERRO:  Hatanın evrimini istenirse saklayan output dosyası
C   FLİPTG2: İnput-hedef ve output vektörlerini  saklayan output dosyası
C   Kullanılan  dosyaların birim numaraları
      PARAMETER(fliptg1=1,fliniw=2,flfws=3,flpmts=4,flerro=5,fliptg2=6)
C   Tabaka  nöronlarının  boyutları
C   N1: İnput , N2: Gizli , N3: Output tabakası nöronları sayısı, Nimp: İnput + output nöronu sayısı
      INTEGER  N1,N2,N3, Nimp
C    sz: Bir veri matrisinde toplam gözlem sayısı, NSML: Herhangi bir andaki toplam gözlem sayısı
C    pt:  Toplam hata hesaplanırken kullanılacak  geçici bir değişken
      INTEGER  sz, NSMPL,pt
      PARAMETER(sz=250, N1=2,N2=3, N3=1, Nimp=N1+1)
      REAL   X(sz,Nimp), Y(sz,N3), YP(sz,N3)
C   Ağırlık Vektörleri-  W1: İnput-gizli tabaka arası,  W2: Gizli tabaka-output arası
     REAL  W1(N2,N1), W2(N3,N2)
C   Wvect: (W1,W2) toplam vektörü
      REAL  Wvect(sz)
C    Nethid(.,.): Gizli tabakadaki toplam input, Hid(.,.): Gizli tabakadaki toplam output,
C    Netout(.,.): Output tabakasındaki toplam input
       REAL*8  Nethid(sz,N2), Hid(sz,N2), Netout(sz, N3)
C    Hata Bileşenleri
C    C12=C1+C2
      REAL*8 E, ERSUM, C1,C2, C12
C    Xywrite: 1 ise  XYYP  bileşenleri dosyaya yazılır, 0 ise yazılmaz. Wvecwrt: 1 ise ağırlıklar yazılır
C    0  ise yazılmaz.
       INTEGER  XYWRITE, WVECWRT
C    E  hatasının hoşgörü  sınırı Epsilon
       REAL  EPSİLON
C    Lrate: Eğitme adımı
       REAL  LRATE
```

```
C     Ercall:  Hata yazılacaksa ilgili rutin çağrılır.
      REAL  ERCALL
C     TIMES: Itearsyon  sayısısnın  katsayısı,
      INTEGER  TİMES
C     İnput-hedef  korelasyon katsayıları
      REAL   RHOXY,rX1X2,rYX1,rYX2
C     Epoch: İterasyon sürecinin döngü değişkeni
      INTEGER  Epoch

C     2)  İlgili  dosyaların açılıp  okunması  (Alt rutin 1)
      CALL  FILEOPENER (FLINIW, FLFWS, FLPMTS, FLERRO, FLİPTG)
C     *****************************************************************
C     3) İnput uzayının boyutunu oku. Bu aşağıdaki koşullu beklenen değeri 1 ya da 2 boyut
C     olduğuna göre  hesaplayacaktır.
      WRITE(*,*)' INDIM/ FMT I1'
      READ(*,'(I1)') INDIM
C     INDIM=1 ise rXY= X ile Y arasındaki korelasyon katsayısı, INDIM=2 ise rX1X2, rX1Y,, rX2Y
C     korelasyon  katasayıları okunacaktır. Burada  C1 (x,y,z) (Bölüm 5) fonksiyonunda paydanın
C     sıfır olmasını önlemek için korelasyon katsayıları 1 yerine 0.998 olarak alınacaklardır.
      IF(INDIM.EQ.1) THEN
      WRITE(*,*)' Read  RHOXY  F3.2'
      READ(*,'(F3.2)') RHOXY
      ELSEIF (INDIM.EQ.2)  THEN
      WRITE (*,*)' Read rX1X2, rYX1,rYX2'
      READ(*,10)rX1X2,rYX1,rYX2
10    FORMAT(3(F3.2,1X))
      IF (rX1X2.EQ.1.00) rX1X2=0.998
      IF (rYX1.EQ.1.00) rX1X2=0.998
      IF (rYX2.EQ.1.00) rX1X2=0.998
      IF (rX1X2.EQ.0.00) rX1X2=0.020
      IF (rYX1.EQ.1.00) rX1X2=0.020
      IF (rYX2.EQ.1.00) rX1X2=0.020
C     ***************************************
C
C     4) Çeşitli  parametreleri  oku  (Alt rutin 2)
      CALL  FMTPMREADER (flpmts, NPT,fmtwall,FMXY,FMXYYP,Lrate,
    + Epsilon,IER, XYWRITE)
C     NS:  Ağırlık  vektörünün  boyutu
      NS=N1*N2+N2*N3
C     ****************************************
C     5)  Times:?  ve   hata evrimi yazılacak mı?
      WRITE(*,*)'TIMES ?  NSMPL*TIMES,  FORMAT=I3'
      READ(*,2-'(I3,1X,I1)')TIMES,ERCALL
C     ****************************************
C     6) Örneklem  matrislerini  okuyarak   toplam hatayı  hesapla
C     NPT: İnput ya da hedef değişkenine ilişkin kaç gözlem yapılacağını  bildirir
      DO  20  NSMPL=1,NPT
C     6.1)  Veriyi  oku (Alt rutin 3)
      CALL  XYREADER (sz,FMXY,X,Y,Nimp,N1,N2,N3,
    + NSMPL, fliptg1)

C     6.2)  Başlangıç  ağırlık  vektörünü  oku (Alt  rutin 4)
      CALL  WEIGHTREADER (sz,W1,W2,N1,N2,N3,fmtwall,fliniw)
```

```
        Rewind(fliniw)
C       ******************************************
        pt=0
        Ersum=0.0
        NITER=NSMPL*TIMES
C       7) İterasyon süreci
        DO 30 Epoch=1,NİTER
        pt=pt+1
C       7.1) İnput verisine bağlı olarak outputu hesapla (Alt rutin 5)
        CALL NODEACTIVATOR (sz,N1,N2,N3,NSMPL,Nimp,pt,X,Y,E,
     +  Nethid,Hid,Netout,,YP,W1,W2,Epoch)
        IF(E.GT.Epsilon) THEN
C       7.2) Hatayı geriye yaymak için ağırlıkları modifiye et (Alt rutin 6)
        CALL WMODIF (sz, N1,N2,N3,NSMPL,Nimp,pt,W1,W2,X,
     +  D1,D2,Hid,Y,YP,Epoch,Lrate)
        ELSE
        ENDIF

C       7.3) Toplam hatayı hesapla
        IF(pt.EQ.NSMPL) THEN
        DO 100 ip=1,NSMPL
        DO 110 j=1,N3
        Ersum=Ersum+(Y(ip,j)-YP(ip,j))**2
110     CONTINUE
100     CONTINUE
        Ersum=Ersum/NSMPL
C       7.4) Eğer hata epsilondan küçükse yaz değilse ağırlık modifikasyonuna devam et
        IF((pt.EQ.NSMPL).and.(Ersum.LE.Epsilon)) THEN
        GO TO 500
        ELSEIF(((pt.EQ.NSMPL).and.(Ersum.GT.Epsilon)) THEN
C       7.5) Hata yazıcıyı çağır (Alt rutin 7)
        CALL EWRITER(flerro, Ier, Epoch, Ersum,sz,Epsilon,NITER,NSMPL,ERCALL)
        pt=0
        Ersum=0.0
        ELSE
        ENDIF

30      CONTINUE
C       ******************************************
500     WRITE(*,*)
C       8) Koşullu beklenen değerin hesaplanması
        IF(INDIM.EQ.1) THEN
        C1=(C1+(Y(NSMPL,1)-RHOXY*X(NSMPL,1))**2)/NSMPL
        C2= (C2+(Netout(NSMPL,1)-RXOXY*X(NSMPL,1))**2/NSMPL
        ELSEIF(INDIM:EQ:2) THEN
C       ****
C       A1,A2,A3 Bölüm 5 de C1(x,y,z) nin analitik ifadesindeki terimler
        A1=(1/(1-(rX1X2)**2))
        A2=(rYX1-rYX2*rX1X2)
        A3=(rYX2-rYX1*rX1X2)
C       ****
        C1=(C1+(Y(NSMPL,1)*A1*(X(NSMPL,1)*A2+
     +  X(NSMPL,2)*A3))/NSMPL
```

```
      C2=(C2+(Netout(NSMPL,1)-A1*(X(NSMPL,1)*A2+
     +  X(NSMPL,2)*A3))/NSMPL
C     ****
      ELSE
      ENDIF
      C12=C1+C2
      WVecwrt=1
C     9) Final ağırlık vektörünü ve hatayı yaz (Alt rutin 8)
      CALL WVECTWRITER(NSMPL,Epoch,sz,N1,N2,N3,
     + NS,W1,W2,Wvect,Wvecwrt,fmtwall, FLFWS)

      C1=C1/NSMPL
      C2=C2/NSMPL
20    CONTINUE
      STOP
      END
C     **** ANA PROGRAMIN SONU ! *******


C     ***********ALT PROGRAMLAR************
C     Not: Alt rutin 1,2,3,4,7,8 kullanıcı tarafından istenildiği gibi değiştirilebileceğinden
     + buraya alınmamışlardır. Buradaki alt programlar toplam hatanın hesaplanabilmesi
     + için asıl olarak gereken rutinlerdir.

C     **********ALT RUTİN 5: İLERİ DOĞRU YAYILMA*******
      SUBROUTINE NODEACTIVATOR (sz,N1,N2,N3,NSMPL,Nimp,pt,X,Y,E,
     + Nethid,Hid,Netout,,YP,W1,W2,Epoch)
      INTEGER sz,N1,N2,N3,NSMPL,Nimp,Epoch
      INTEGER p,k
      REAL X(sz,Nimp),Y(sz,N3)
      REAL*8 Nethid(sz,N2),Hid(sz,N2),Netout(sz,N3),YP(sz,N3)
      REAL W1(N2,N1),W2(N3,N2)
C     # Net'leri ve out'ları sıfırla
      DO 1 p=1,sz
      DO 2 i=1,N2
      Nethid(p,i)=0.0
      Hid(p,i)=0.0
2     CONTINUE
1     CONTINUE
C     # Nöronların üzerindeki toplam inputu hesapla & Toplam hatayı hesapla
      DO 10 p=1,NSMPL
      DO 20 i=1,N2
      DO 30 j=1,N1
      Nethid(p,i)=Nethid(p,i)+X(p,j)*W1(i,j)
30    CONTINUE
      Hid(p,i)=1/((1+EXP(-Nethid(p,i))))
      DO 40 k=1,N3
      DO 50 i=1,N2
      Netout(p,k)=Netout(p,k)+Hid(p,i)*W2(k,i)
50    CONTINUE
C     # Aktivasyon fonksiyonu afindir (AF2). O nedenle output tabakasında üstel
C     fonksiyon kullanılmadı.
      YP(p,k)=Netout(p,k)
```

```
        E=E+(YP(p,k)- Y(p,k))**2
40      CONTINUE
10      CONTINUE
        RETURN
        END
C       **********Alt  program sonu*******


C        *****ALT RUTİN 6: GERİYE DOĞRU YAYILMA: Ağırlık Modifikasyonu****
C       Bu program standart geriye yayma(SBPM) ile ağırlıkları  adapte eder
        SUBROUTINE WMODIF (sz, N1,N2,N3,NSMPL,Nimp,pt,W1,W2,X,
    +   D1,D2,Hid,Y,YP,Epoch,Lrate)
        INTEGER sz,N1,N2,N3,NSMPL,Epoch,p,j,k
        REAL T1, Lrate,
        REAL  W1(N2,N1),W2(N3,N2)
        REAL X(sz,Nimp),Y(sz,N3)
         REAL*8 Hid(sz,N2),YP(sz,N3)
        REAL*8 D1(sz,N2),D2(sz,N3)

C       # D1 ve D2 lerin sıfırlanması
        DO 1 p=1,NSMPL
        DO 2 j=1,N2
        DO 3 k=1,N3
        D1(p,j)=0
        D2(p,k)=0
3       CONTINUE
2       CONTINUE
1       CONTINUE
C        #  Output nöronları için D2 lerin hesaplanması
        DO 10 p=1,NSMPL
        DO  20 j=1,N3
        D2(p,j)=(Y(p,j)-YP(p,j))
20      CONTINUE
10      CONTINUE
C       T1 bu program boyunca jenerik toplayıcı gibi işlev görür.
        T1=0
        DO 30 p=1,NSMPL
        DO 40 j=1,N2
        DO 50 k=1,N3
C       D1pj=k üzerinden toplam (D2pk*W2kj)
        T1=T1+D2(p,k)*W2(k,j)
50      CONTINUE
        D1(p,j)=T1
        T1=0
40      CONTINUE
        T1=0
30      CONTINUE
        T1=0
         DO 60 j=1,N2
         DO 70 p=1,NSMPL
         T1=D1(p,j)*(1-Hid(p,j))*(Hid(p,j))
70       CONTINUE
         D1(p,j)=T1
```

```
        T1=0
60      CONTINUE
       T1=0
C      #  W2  ağırlıklarını değiştir (modifiye et)
        DO 80 j=1,N3
        DO 90 i=1,N2
        DO 100 p=1,NSMPL
C       D2pj bütün p ler üzerinden topla
        T1=T1+D2(p,j)*Hid(p,j)
100     CONTINUE
       W2(j,i)=W2(j,i)+Lrate*T1
        T1=0
90      CONTINUE
        T1=0
80      CONTINUE
        T1=0
C       # Bu kısım istenirse bias modifikasyonu için kullanılacaktır
C        DO 110  j=1,N3
C        DO 120 p=1,NSMPL
C        T1=T1+D2(p,j)*1
C        120          CONTINUE
C        B2(j)=B2(j)+T1
110      CONTINUE
C        T1=0

C        # W1 ve B1(Bias1) ağırlıklarını  modifiye et
         DO 130  k=1,N1
         DO 140  j=1,N2
         DO  150 p=1,NSMPL
         T1=T1+D1(p,j)*X(p,k)
150       CONTINUE
         T1=0
140      CONTINUE
         T1=0
C         DO 160 j=1,N2
C         DO 170 p=1,NSMPL
C         T1=T1+D1(p,j)*1
C         B1(j)=B1(j)+Lrate*T1
C         T1=0
C         160          CONTINUE
C         T1=0
          RETURN
          END
C         ********Alt program sonu*******
```