

ANKARA ÜNİVERSİTESİ
BİYOTEKNOLOJİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ÇOKLU GEN HASTALIKLARININ İSTATİSTİKSEL
ÖĞRENME YÖNTEMLERİYLE ANALİZİ

Mehmet KORKMAZ

Danışman Öğretim Üyesi

Doç.Dr. H. Gökhan İlk

ANKARA 2007

Çoklu Gen Hastalıklarının İstatistiksel Öğrenme Yöntemleriyle Analizi

ÖZET

Kalp, yüksek tansiyon, tromboz ve ateroskleroz gibi, ortaya çıkmasında birden fazla genin aktif olarak rol oynadığı ve çevresel faktörlerle desteklenen hastalıklar çoklu gen hastalıkları olarak adlandırılmaktadır. Bu tip hastalıklar farklı genlerdeki mutasyonların (yada polimorfizmlerin) etkileşimiyle ortaya çıkar. Bu hastalıklarda birden fazla gen tetikleyici yada baskılayıcı özellik göstererek hastalığın ortaya çıkmasını yada engellenmesini sağlayabilir. Günümüzde bilim insanları genlerin bu tip hastalıkların üzerinde etkisini araştırırken çeşitli istatistik hesaplarıyla üzerinde çalıştıkları genlerin hastalık ile ilişkili olup olmadığını göstermeye çalışmaktadır.

İstatistiksel öğrenme; istatistik, veri madenciliği ve yapay zeka gibi alanlarda önemli rol oynamaktadır. İstatistiksel öğrenme kabaca “veriden öğrenme”dir. Eldeki ölçülmüş veriyi (niceliksel yada kategorik) kullanır. Verideki bazı öznitelikleri kullanarak tahminde bulunmaya çalışır..

Bu tez çalışmasında; çoklu gen hastalıklarında, ilgili genlerin hastalık ile olan ilişkisi araştırılırken, elle hesaplanamayan durumlarda istatistiksel öğrenme yöntemleri kullanılmaya çalışılmıştır. Hasta ve kontrol gruplarından oluşan veri, eğitim verisi olarak kullanılmıştır. Her ne kadar çalışmanın hedefi, hasta ve kontrol grupları arasındaki ayrışmayı gerçekleştirmek olsa da yapılan hesaplamalar sonucu bunun mümkün olmadığı, verinin istatistiksel öğrenme yöntemleri ile analiz yapılmaya uygun bir veri olmadığı ortaya konulmuştur.

Anahtar Kelimeler: İstatistiksel Öğrenme, Çoklu Gen Hastalıkları, Topaklandırma, PCA

Analysis of Polygenic Diseases Using Statistical Learning Methods

ABSTRACT

Polygenic diseases as atherosclerosis, thrombosis and hypertension have multigenic basis and also have interactions with environmental factors.. Multiple genes can play important role to suppress or to trigger the occurrence of the disease. Scientists use statistical tools and methods to find evidence whether suspected genes have relation with the diseases they study on.

Statistical learning play important roles in areas such as statistics, data mining and artificial intelligence. Statistical learning can be explained briefly as “learning from data” and it uses measured qualitative or quantitative data. In statistical learning, some features that exist in data are used to make a prediction.

In this thesis, it is intended that statistical learning methods can help to identify the roles of the genes in polygenic diseases. Dataset that contains case and control group data is used as training data. Even if the aim of this thesis is to classify the case and control groups of polygenic diseases, the analysis of the data used in this study showed that this dataset is not a proper one to conduct a statistical learning analysis.

Key words: Statistical Learning, Polygenic Diseases, Clustering, PCA

TEŞEKKÜR

Yüksek Lisans çalışmalarım boyunca, eğitimini gördüğüm konularda beni yönlendiren ve ufkumu açan, yaptığım hatalardan sonra bile sakin bir şekilde benimle ilgilenen danışmanım sayın Doç. Dr. H. Gökhan İLK'e,

Yüksek lisans yapmam için her türlü maddi ve manevi desteğini benden esirgemeyen ve inancını kaybetmeyen değerli hocam Prof. Dr. Nejat AKAR'a,

Genetikle ilgili konularda yılmadan bana yardımcı olan ve teşvik eden sevgili hocam Doç. Dr. Hilal ÖZDAĞ'a,

2003 yılından bu yana en iyi ve en kötü günlerimde yanımda olan, kendisini sıkıntıya sokma pahasına maddi ve manevi hiçbir desteğini ve sevgisini benden esirgemeyen, sevgili dostum Devrim AYDIN'a,

Biyoteknoloji Enstitüsü'nde bulunduğum süre boyunca bana bir baba ve anne şevkiyle yaklaşıp bana her türlü desteği gösterme kararlılığında olan sevgili büyüklerim Ali GÖKÇE ve Birgül YILDIRIM'a,

Beni kendilerin ayrı tutmayan tüm Biyoteknoloji Enstitüsü çalışanlarına,

Bana her konuda yardımcı olan ve bu çalışmamda beni destekleyen sevgili arkadaşım Dilay DÜNGÜL'e,

Ankara'da geçirdiğim yılları anlamlı kılan kadim dostum Ömer YURTSEVER'e,

Tez çalışmam süresince manevi desteğini esirgemeyen ve varlığıyla geleceğimi aydınlatan kız arkadaşım Güneş Özlem ÖZTÜRK'e,

Her zaman hislerime tercüman olabilcek sözler yazmış olan, sesiyle beni dünyaya bağlayan ve beni de Tugboat Kaptanları kervanına ekleyen Dean WAREHAM'a,

Birgün, Balkan Dağlarında beraber koyun yetiştirmeyi ümit ettiğim, sıradışı insan ve gönül adamı Jeff MANGUM'a

Teşekkürü bir borç bilirim.

Bu tez çalışmasını başta babam Sefer KORKMAZ olmak üzere bana destek olan ve inancını kaybetmeyen aileme ithaf ediyorum.

İÇİNDEKİLER

ÖZET	i
TEŞEKKÜR	iii
İÇİNDEKİLER.....	v
ŞEKİLLER DİZİNİ.....	vii
ÇİZELGELER DİZİNİ	1
1. GİRİŞ	2
1.1 Genel Bilgiler	3
1.1.1 Genetik Hastalıklar.....	4
1.1.2 Çoklu Gen Hastalıkları.....	5
1.1.3 Pediatrik İnme	7
1.1.4 İstatistiksel Öğrenme.....	7
2. MATERYAL VE YÖNTEM	11
2.1. Materyal	11
2.2. Yöntem	12
2.2.1. Eğitim Verisinin Matematiksel Olarak Kullanılabilir Hale Getirilmesi	12
2.2.2 Verinin İstatistiksel Olarak Kullanılabilirliğinin Testi.....	20
2.2.3 Verinin PCA ile analizi	22
2.2.4 K-means Clustering.....	22
3. ARAŞTIRMA BULGULARI	23
4. SONUÇ, TARTIŞMA ve ÖNERİLER	28

KAYNAKÇA.....	30
EKLER.....	32
Ek-1. Hasta Verisi	32
Ek-2. Kontrol Verisi.....	37
Ek-3. Matlab Kodları.....	43
Ek-3.1 T-test ve Kolmogorov-Smirnov Testleri	43
Ek-3.1 PCA ve K-means Clustering	45

ŞEKİLLER DİZİNİ

Şekil 1.1 (a) Merkezi Dogma. (b) Tek Gen Hastalıkları (Davison 2005).....	5
Şekil 1.2 Çoklu Gen Hastalıkları (Davison 2005).....	5
Şekil 2.1. Pediatrik inme verisinin histogramı	14
Şekil 2.2. Hasta grubuna ait verinin histogramı	18
Şekil 2.3. Kontrol grubuna ait verinin histogramı.....	19
Şekil 2.4. Hasta ve kontrol grubuna ait verinin histogramı.....	19
Şekil 2.5. Hasta ve kontrol grubu histogramlarının birlikte gösterimi.....	20

ÇİZELGELER DİZİNİ

Çizelge 2.1 Kullanılan verinin hasta ve kontrol grubuna ait sayısal bilgileri.	11
Çizelge 2.2 Vektörler	12
Çizelge 2.3 Vektör boyutları için atanan nümerik değerler	15
Çizelge 2.4 Atama sonrası oluşan yeni veri düzeni	17
Çizelge 2.5 Pediatrik inme verisine ait istatistiksel parametreler	17
Çizelge 3.1 K-means clusterin sonrası analiz sonuçları	25

1. GİRİŞ

Kalp, yüksek tansiyon, tromboz ve ateroskleroz gibi ortaya çıkmasında birden fazla genin aktif olarak rol oynadığı ve çevresel faktörlerle desteklenen hastalıklar çoklu gen hastalıkları olarak adlandırılmaktadır (Jameson 1998). Bu tip hastalıklar farklı genlerdeki mutasyonların (yada polimorfizmlerin) etkileşimiyle ortaya çıkar. (Wink 2006). Bu hastalıklarda birden fazla gen tetikleyici yada baskılayıcı özellik göstererek hastalığın görünmesini yada engellenmesini sağlayabilir. Günümüzde bilim insanları genlerin bu tip hastalıkların üzerinde etkisini araştırırken çeşitli istatistik hesaplarıyla üzerinde çalıştıkları genlerin hastalık ile ilişkili olup olmadığını göstermeye çalışmaktadır. Manuel olarak yapılan bu hesaplamalarda bilim insanları, en fazla 3-4 genin etkili olduğu durumlarda hesaplamalarını gerçekleştirebilmektedir. İlgilenilen gen sayısı arttıkça bilim insanları bu hesaplamaları elle yapamamaktadır.

İstatistiksel öğrenme (statistical learning); istatistik, veri madenciliği ve yapay zeka gibi alanlarda önemli rol oynamaktadır. İstatistiksel öğrenme kabaca “veriden öğrenme”dir. Eldeki ölçülmüş veriyi (niceliksel yada kategorik) kullanır. Verideki bazı özellikleri (feature) kullanarak tahminde bulunmaya çalışır (Hastie 2001).

Pubmed’te, BMC Bioinformatics ve Oxford Bioinformatics dergilerinin elektronik ortamdaki veritabanlarında yapılan anahtar kelime sorguları sonucu şunlar görülmüştür: Tez konusunun belirlendiği tarihte bu konuyla ilgili doğrudan ilişkili bir çalışmanın bildirilmemiş olduğu, fakat Ocak 2007’de yayınlanan bir çalışmanın benzer bir konu üzerinde çalıştığı tesbit edilmiştir (Gaulton 2007). Bir başka çalışma da ise 7 tane çoklu gen hastalığında etkili olan genlerin saptanması amacı ile DNA mikrodizinin teknolojisi kullanılarak 1400 vaka ve 3000 ortak kontrol kullanılarak bu hastalıklarla ilişkili genler belirlenmiştir. (Weir 2002, The Wellcome Trust Consortium 2007) tarafından gerçekleştirilen bir çalışmada da genom boyunca incelemeler yapılarak çoklu gen hastalıkları tanımlanmak istenmiştir.

Bu tez çalışmasında; çoklu gen hastalıklarında, ilgili genlerin hastalık ile olan ilişkisi araştırılırken, elle hesaplanamayan durumlarda istatistiksel öğrenme yöntemleri kullanılması amaçlanmıştır. Böylelikle hasta ve kontrol gruplarından oluşan veri eğitim verisi olarak kullanılacaktır. Bu çalışmanın hedefi, hasta ve kontrol grupları arasındaki ayrışmayı gerçekleştirmek ve bu ayrımı gerçekleştirmeyi sağlayan bir bilgisayar yazılımı

ortaya çıkarmak hedeflenmiştir. Ayırtırmayı “öğrenmiş” sistem ile hangi gruba ait olduğunu bilmediğimiz bir veri ile sorgulandığında bu verinin ait olabileceği grubu belirleyebilmesi amaçlanmıştır.

1.1 Genel Bilgiler

Yaşayan tüm canlıların soylarının devamı onların genomları tarafından belirlenir. Genom nükleik asit serilerinden oluşan DNA içinde şifrelenmiş tir. DNA canlıyı tek başına oluşturabilen bir yapı değildir. DNA bir organizmanın tüm hücre ve dokularının oluşturulması için gerekli bilgiyi taşır (Lewin 2004, Lodish 2003).

Tüm canlılar aynı temel mekanizmayı kullanarak genleri okur ve ifade eder. Buna Merkezi dogma denir. Merkezi dogmaya göre bilgi DNA’dan mRNA (messenger RNA/ haberci RNA)’ya geçer. mRNA, proteini oluşturan amino asit zinciri bilgisini taşır (Raven 2001). RNA’daki sıralı her üç amino asit kodon olarak adlandırılır. Her kodon bir amino asit bilgisini taşır. Bazı karmaşık etkileşimler (yer, zaman ve diğer fiziksel ve kimyasal etkileşimler) sonucu bu mekanizmayla proteinler oluşturulur. Proteinler organizmanın yapısında yada bu yapıların oluşturulmasında yer aldığı gibi işlevsel anlamda organizmanın metabolizmasının ana dinamlarından birini teşkil eder (Lewin 2004).

Özdeş ikizler dışında birbirinin aynı DNA dizilimine sahip iki insan yoktur. Genel olarak popülasyondaki iki birey arasında ise DNA dizilimi düzeyinde % 0.1’lik bir farktan söz etmek mümkündür. Bu binde bir oranındaki farklılık genel anlamda bireylerin dış görünüşlerini, metabolizmal farklılıklarını belirlemede etkindir (Meyers 1995).

DNA’nın farklılığı iki şekildedir: mutasyon ve polimorfizm. Mutasyon, genin bir allelinin başka bir allele dönüşmesidir. DNA’daki kayıp (delesyon), eklenme (insersiyon) ve değişim (tranzisyon, transversiyon) geçirme gibi değişiklikler bu farklılıkların yapısal çeşitleri arasında sayılabilir. DNA seviyesindeki mutasyonlar iki çeşittir. Pürin başka bir pürin ile primidin başka bir primidin ile yerdeğiştirir yada pürin yerine primidin, primidin yerine pürin gelebilir. Protein seviyesinde mutasyon ş u üç başlıkta gruplanır. Sessiz mutasyon: DNA seviyesindeki değişim sonucu amino asit bilgisini taşıyan kodon yine aynı amino asit bilgisini taşıyan başka bir kodona dönüşebilir. Yanlış Mutasyon: Değişen kodon başka bir amino asit şifresini taşır. Anlamsız Mutasyon: Değişen kodon protein sentezini

bitirme şifresini taşır (Griffiths 1999). Mutasyon, protein sentezinde amino asit diziliminin belirlenmesi aşamasında farklı bir amino asitin proteinin yapısına girmesine neden olabileceği gibi, ilgili proteine yeni bir veya birkaç aminoasitin katılması veya bu proteinin yapısındaki aminoasitlerin eksilmesine de yol açabilir. Bu şekilde ortaya çıkan bir mutasyonun sonucunda oluşan protein artık asıl işlevini yerine getiremeyecektir (Nei 1975).

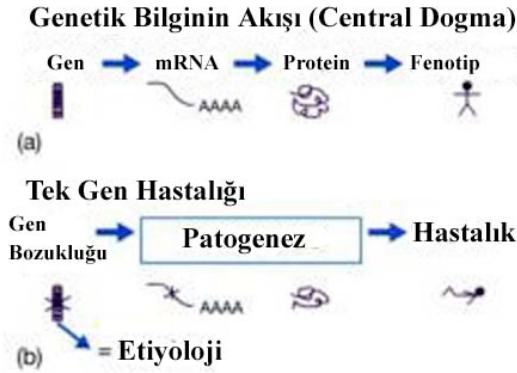
Genetik çeşitlilik birçok organizmanın doğal ortamlarında bulunur. Bireyler arasındaki genetik farklılıklar DNA polimorfizmi olarak adlandırılır. Polimorfizm “birden çok formda” anlamına gelir (Hartl 2005). Polimorfizmlerde genelde tek nükleotit değişikliği gözlemlenir. Bu değişim yukarıda bahsedildiği üzere genel olarak fenotipte bir değişiklik meydana getirir, çevre, diyet benzeri etkenlerle olan etkileşimde farklılıklar (bazı alerjik reaksiyonlar, alkole aşırı duyarlılık, bazı hastalıklara yatkınlık gibi) açığa çıkabilir.

1.1.1 Genetik Hastalıklar

Genetik hastalıklar, ilk olarak dış etkilerle alakası olmadığı belirlenen hastalıkların aile içinde tekrar ettiğinin farkedilmesiyle tartışılmaya başlandı. Gregor Mendel’in kalıtımın kurallarını belirlemesinden sonra genetik hastalıkların önemli bir kısmının mendelyen kalıtım gösterdiği bir kısmının ise mendelyen kalıttan saptığı tespit edildi (Friedmann 1994).

DNA’daki mutasyonlar sonucu mutasyonu içeren mRNA hatalı aminoasit kodunu taşıması nedeniyle yanlış proteinin sentezlenmesine neden olur. Farklı amino asit kodu proteinin sentezinin son aşamasında işlevselliğini yitirmiş ya da farklı işlevlere sahip protein üretilmesine neden olur. Bunun sonucunda ilgili genin kodladığı proteinin görev aldığı metabolik yolda aksamalar olacağı için birey rahatsızlık yaşar. Kalıtsal hastalıklarda bu mutasyonlar aile içinde nesiller arası geçiş göstermeye devam eder (Meyers 1995). Kalıtım profilleri rahatlıkla belirlenen tek gen hastalıkları mendelyen kalıtım kuralları ile uyum gösterir. Bunlara örnek olarak hemofili, sistik fibrosis gibi iyi bilinenler verilebilir. Tek gen hastalıklarının dışında birden çok gen ile açıklanabilen bazı kanser formları, bazı erken kalp hastalıkları, obezite, bazı körlük ve sağırılık vakaları da genetik hastalıklar arasında sayılabilmektedir. Bugüne dek bildirilmiş binlerce hastalık vardır (Gormley 1989; OMIM-Online Mendelian Inheritance in Man, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM>).

Genetik kaynaklı hastalıklar daha çok iki başlık altında incelenir. Tek gen hastalıkları ve çoklu gen hastalıkları. Tek bir gendeki mutasyon nedeniyle açığa çıkan bu grup hastalıklara tipik örnek olarak sistik fibrosis verilebilir. Bu durum CFTR (Cystic fibrosis transmembrane conductance regulator) geni tarafından kodlanan klorid kanalında işlevsellik kaybından kaynaklanır (Wink 2006).

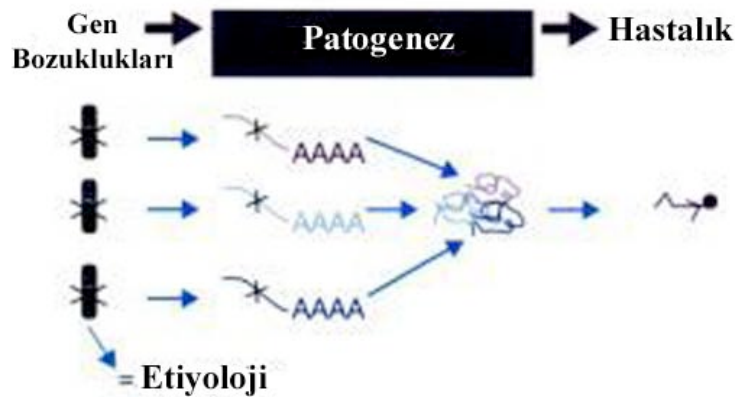


Şekil 1.1 (a) Merkezi Dogma. (b) Tek Gen Hastalıkları (Davison 2005)

Tek gen hastalıklarının en önemli özelliği, gendeki mutasyonun kendisi hastalık bilgisini (birincil nedenini) vermesidir (Davison 2005).

1.1.2 Çoklu Gen Hastalıkları

Çoklu gen hastalıkları, birkaç gendeki değişimlerin etkileşimi sonucu ortaya çıkar. Tek başlarına etkisi az olan bu değişimler hepsi bir arada gerçekleştiğinde ciddi sonuçlar ortaya çıkarabilir. (Wink 2006).



Şekil 1.2 Çoklu Gen Hastalıkları (Davison 2005)

Çoklu gen hastalıkları genelde, insan popülasyonunun büyük bir yüzdesini etkilemektedir (Miller 2000). Çoklu gen hastalıkları temel hipertansiyon gibi her 5 kişiden birinde görülen tiplerinden, her 1000 kişiden birinde rastlananlarına dek değişik sıklıkta görülen rahatsızlıklardır (Davison 2005).

Büyük popülasyonlara dahil küçük popülasyonlar, şeker, yüksek kan basıncı, astım, şizofreni vb hastalıkların çok yada az görülmelerine göre tanımlanabilmektedir. Çoklu gen hastalıkları bu ayırım sonrasında küçük popülasyonda daha yüksek oranlarda görünür. Fenotiplerin açıklığa kavuşturulmasında ve fenotip ifadesinde çevresel faktörlerin etkisi çalışmalarında küçük popülasyonlar üzerinde çalışılmaktadır (Miller 2000).

Çoklu gen hastalıkları dış faktörlerle birlikte kompleks bir yapıda olduklarından, hastalığa neden olan faktörleri ve genleri tesbit etmek tek gen hastalıklarındaki gibi kolay değildir. Hastalığın ortaya çıkmasına neden olan yolağın çok açık olmamasından dolayı, riskli genleri tanımlamak için pedigrî analizi gibi aileyi analiz eden klasik yöntemler yeterli olamamaktadır (Wink 2006). Hastalık prognozu hastalığa neden olan metabolik arıza veya arızalarla ilgili olarak bazı bilgiler verebilir. Bu bilgilerin ışığında bazı genler hastalıkla ilintili aday genler olarak belirlenip popülasyon bazında incelenmeye alınabilir. Bu tip hastalıklarla ilintilendirilen genlerin üzerinde tespit edilen tekli nükleotid polimorfizmleri (TNP-Single Nucleotide Polymorphism/SNP) üzerindeki araştırmalar son 10 yıldır giderek artan bir ilgiyle devam etmektedir (Walliker 1996, Camm 2006).

Tanım gereği SNP aynı tür organizmalardaki nükleotid dizisindeki tek bir nükleotiddeki değişikliktir. İnsan genomundaki çeşitliliğin %90'ını SNP'ler oluşturur (Ziegler,2006). SNP'ler genelde sessizdirler (sessiz polimorfizmdirler) fakat işlevsel olarak aktif olan çeşitleri de bulunmaktadır. Bu tipte olanlar gen ifadesini ya da proteinin işlevini etkileyebilir. Bunlar genelde genin regülatör ya da kod içeren kısımlarında bulunurlar. İşlevsel SNP'lerin çevresel baskılarla etkileşimi sonucu doğal, faydalı ya da zararlı etkileri olabilir. Yukarıda da bahsedilen konulardan dolayı SNP'ler ile kompleks rahatsızlıklar arasındaki ilişkiyi bulabilmek için genel popülasyon temelli yaklaşım geliştirilmiştir. (Camm 2006).

1.1.3 Pediatrik İnme

İnme, vasküler tıkanma ya da damar yırtılması ile bunu izleyen kanama ve nörolojik aksaklıklar sonucu ortaya çıkan fokal beyin hasarı olarak tanımlanabilir. Bu tıkanma sonucu beynin bir kısmına kan sağlamaması sonucu oksijen yetersizliğine neden olur. Bundan dolayı bu bölgede bulunan sinir hücreleri çalışamaz ve kısa sürede ölür. Sinir hücrelerinin ölümü nedeniyle, bu hücrelerin ilgili olduğu vücut bölgeleri iş görmez olur. İnme yetişkenler ve pediatrik yaş grupları arasında bazı farklılıklar gösterir. Yetişkinlerde kolaylıkla teşhis edilebilir ve erken tedavi uygulanabilmektedir (Demir 2004).

Çocuklarda görülen inmenin potansiyel risk yaratan kalıtsal sebeplerden bazıları; faktör V Leiden mutasyonu, PT 20210A değişimi, faktör VIII yüksekliği gibi faktörlerdir. Bu faktörlerin yanısıra TFPI (Doku Faktörü Yolu İnhibitörü) ve Seritonin (5HT)2A Reseptör genlerinde görülen polimorfizmlerin de etkili olduğu düşünülmektedir.

1.1.4 İstatistiksel Öğrenme

1.1.4.1. Giriş

İnsanlar başka kişilerin yüzlerini, seslerini, söyledikleri sözcükleri, el yazılarını ayırt edebilirler. Ceplerindeki anahtarı dokunarak bulabilirler yada bir meyveyi kokusundan tahmin edebilirler. Ses ve yüz tanıma, parmak izinden kimlik belirleme, optik karakter çözümlene, DNA dizi kimliği belirleme vb işleri yapabilecek makinelerin olması insanı günlük yaşamda birçok alanda rahatlatmıştır (Duda 2000). Bilgisayarların gelişmesiyle makinenin sağlanan birtakım veriler ile bunları belirleyebilmesi mümkün kılınmıştır. Bunun yapılması için bazı matematiksel ve istatistiksel yöntemler kullanılmaktadır (Mitchell 1998).

İstatistiksel öğrenme; istatistik, veri madenciliği ve yapay zeka gibi alanlarda önemli rol oynamaktadır. İstatistiksel öğrenme kabaca “veriden öğrenme”dir. Eldeki ölçülmüş veriyi (niceliksel yada kategorik) kullanır. Verideki bazı özellikleri (öznitelik) kullanarak tahminde bulunmaya çalışır (Hastie 2001).

1.1.4.2 Tipik İstatistiksel Öğrenme Analizinde İşlemler

İstatistiksel öğrenme ile problem çözümünde ve sonucun güvenilir olması için sırasıyla şu işlemler yapılır:

- a. Verinin toplanması: İstatistiksel öğrenmenin en önemli ve zor kısımlarından birisi veri toplamadır. Eğitilecek sistemin daha doğru ve güvenilir cevap verebilmesi için mümkün olduğunca çok veri toplanmalıdır. Toplanan veride, karşılaştırılacak öğelerin istatistiksel olarak birbirinden bağımsız olması gerekmektedir.(Duda 2000).
- b. Öznitelik (feature) seçimi: Gözlenen bir olayda başlıbaşına ayırt edici olabilen ve ölçülebilen özelliklere öznitelik denir. Öznitelik seçimi, istatistiksel öğrenme yöntemlerinin en kritik aşamasıdır. Problemin karakteristiğine bağlıdır. Öznitelik seçilirken dikkat edilmesi gereken hususlar ş unlardır: Kolay ayrıştırılabilir olmalıdır. Dönüşümler yapıldığında değişmemeli ve gürültüye karşı hassas olmamalıdır (Duda 2000). Öznitelik sayısının çok olması fayda sağlar. Öznitelik sayısı gereksiz yere arttırılmamalıdır, bu durum “curse of dimensionality” olarak tanımlanan ve verinin gereğinden fazla ve sınıflandırmayı zorlaştıracak şekilde betimlenmesine neden olabilir.
- c. Modelin oluşturulması: Problemin ve özniteliklerin yapısına göre uygun yöntem seçilmelidir. Analiz sonucu niceliksel olacak ise regresyon, niteliksel olacak ise sınıflandırma tahmin edici yöntemler olarak kullanılır (Hastie 2001).
- d. Eğitim: Seçilen yönteme göre eldeki veri (train data) kullanılarak “sistem” eğitilir.
- e. Değerlendirme: Seçilen yöntem sonucunda ortaya çıkacak ayrılmış gruplar, bu gruplardan olduğu bilinen değerlendirme verisi (test data) ile değerlendirilmeye alınır. Sistem doğru eğitilmiş ise değerlendirme verisi ait olduğu gruba bağlı görünecektir.

1.1.4.3 Principal Component Analysis (PCA)

“Curse of dimensionality” ile karşılaşmamak için kullanılan yöntemlerden biri de veriyi daha az sayıda boyuta indirgemektir (Wasserman 2006). PCA (Karhunen-Loève transform), güdümsüz olarak boyut indirgeme için yaygın kullanılan parametrik olmayan bir istatistiksel tekniktir (Marques de Sa 2001, Ding 2004). Temelde sadece önemli eigen vektörleri çıkarmak için ortonormal transformasyon yapma fikrine dayanır, yani N boyutlu veri sadece en önemli boyutları içeren k boyuta indirgenmiş olur ($N > k$) (Marques de Sa 2001).

1.1.4.4 Clustering (Kümelendirme) ve K-means Clustering

Kümelendirme basit olarak bir objeyi, elemanları bir şekilde benzer olan gruplar şeklinde düzenleme işlemidir. Birden fazla kümelendirme yöntemi vardır. Analiz sonucunda oluşacak küme sayısı tam olarak belirtilmek istenirse k-means kümelendirme yöntemi kullanılır.

K-means, güdümsüz öğrenme konularının da en yaygın kullanılan itereatif kümelendirme yöntemlerinden biridir (Hastie 2001, Ding 2004). Bu yöntemde k adet küme merkezi (cluster centroids) bulunur. K-means kümelendirme algoritmasının temel fikri, örneklerin küme içindeki merkeze uzaklıklarının toplamını minimize etmektir (Marques de Sa 2001). “Squared Euclidean Distance” benzemezlilik(dissimilarity) ölçüsü olarak kullanılır (Hastie 2001). K-means kümelendirme algoritması şu şekildedir(Duda 2000, Hastie 2001, Marques de Sa 2001):

1. Küme sayısı c belirlenir. Başlangıç değeri için $k=1$ olarak verinin ortalama değeri bulunur.
2. Bulunan ortalama değerin çok küçük farklarla sağ ve sol tarafında iki yeni merkez seçilir.
3. Tüm veriler uzaklığı en küçük olan merkeze göre bağlanarak kümelendirilir.
4. Kümelere bağlanan verilerin yeniden ortalama değeri bulunur ve 2. Ve 3. basamak tekrarlanır.
5. Bu işlem daha önce belirlenen maksimum iterasyon sayısı kadar yada her seferinde bulunan ortalama değerin bir öncekiyle aynı olduğu yada farkın çok küçük olduğu aşamaya kadar devam ettirilir.
6. 5. Basamak sonunda belirlenen kümeler için merkezler belirlenmiş olur.
7. İstenilen küme sayısı ikiden fazla ise 6. basamaktan sonra 2. basamağa geçilip aynı yöntem ile küme sayısı artırılabilir.

Bu tez çalışmasında; çoklu gen hastalıklarında, ilgili genlerin hastalık ile olan ilişkisi araştırılırken, elle hesaplanamayan durumlarda istatistiksel öğrenme yöntemleri kullanılması amaçlanmıştır. Böylelikle hasta ve kontrol gruplarından oluşan veri eğitim

verisi olarak kullanılacaktır. Bu çalışmanın hedefi, hasta ve kontrol grupları arasındaki ayrışmayı gerçekleştirmek ve bu ayrımı gerçekleştirmeyi sağlayan bir bilgisayar yazılımı ortaya çıkarmak hedeflenmiştir. Ayrıştırmayı “öğrenmiş” sistem ile hangi gruba ait olduğunu bilmediğimiz bir veri ile sorgulandığında bu verinin ait olabileceği grubu belirleyebilmesi amaçlanmıştır.

Bu bağlamda tez çalışmasında kullanılmak üzere temin edilen veri öznitelikleri belirlenerek (V Leiden mutasyonu, PT 20210A değişimi, faktör VIII, TFPI, (5HT)2A polimorfizmleri) k-means algoritması kullanılarak kümelendirme analizi yapılacaktır. Bu analiz sonucunda eğitilen sistem sınama verileri ile sınanacaktır. Böylelikle eğitilmiş sistemin doğru çalışıp çalışmadığı gösterilecektir. K-means algoritması kullanılmadan önce PCA kullanılarak k-means uygulaması sonucunda nasıl bir sonuc alınabileceği öngörüsü yapılacaktır.

2. MATERYAL VE YÖNTEM

2.1. Materyal

Bu tez çalışmasında Ankara Üniversitesi, Tıp Fakültesi, Dahili Tıp Bilimleri Bölümü, Çocuk Sağlığı ve Hastalıkları Anabilim Dalı, Pediatrik Moleküler Genetik Bölümü'nden gelen pediatrik inme hastalığı ile ilgili yapılan çalışmalar sonucu elde edilen veri kullanılmıştır.

Bütün inme hastaları, klinik olarak teşhis edilmiştir ve beyindeki infarkt beynin kranyal görüntülenmesi ile belirlenmiştir. Çalışmaya dahil edilen tüm bireyler/aileler bilgilendirilmiş ve tüm bireylerden/ailelerden imzalı aydınlatılmış onam formu alınmıştır.

Bu veri; hasta adı ve hastadaki FVL1691, PT20210, FXIII, TFPI, 5-HT2A polimorfizmlerini içeren excel dosyası olarak temin edilmiştir. Bu veri hasta ve kontrol gruplarından oluşmaktadır. Hasta grubu verisi Çizelge Ek-1.1'de, kontrol grubu verisi Çizelge Ek-2.1.'de verilmiştir.

Temin edilen bu veri, hasta adı, gen polimorfizmleri, verinin ait olduğu grup bilgilerini içerecek şekilde MYSQL veritabanına aktarılmıştır. (Bilgi içermeyen alanlar XX olarak işaretlendi). Böylelikle SQL altyapısı ile veri üzerinde her türlü sıralama ya da çeşitli kombinasyonlarda veriyi kullanabilme olanağı sağlanmıştır.

Kullanılan veride bazı hasta yada kontrol bilgilerinde tüm 5 gen polimorfizm bilgileri bulunmamaktadır. Bundan dolayı temin edilen veride tüm 5 gen polimorfizmini içermeyen veriler temizlenerek elde edilen veri kullanılmıştır. Kullanılan verinin hasta ve kontrol grubuna ait sayısal bilgileri Çizelge 2.1'te verilmiştir.

	Temin edilen veri sayısı	Kullanılabilir veri sayısı
Hasta Grubu	121	105
Kontrol Grubu	119	58

Çizelge 2.1. Kullanılan verinin hasta ve kontrol grubuna ait sayısal bilgileri.

2.2. Yöntem

2.2.1. Eğitim Verisinin Matematiksel Olarak Kullanılabilir Hale Getirilmesi

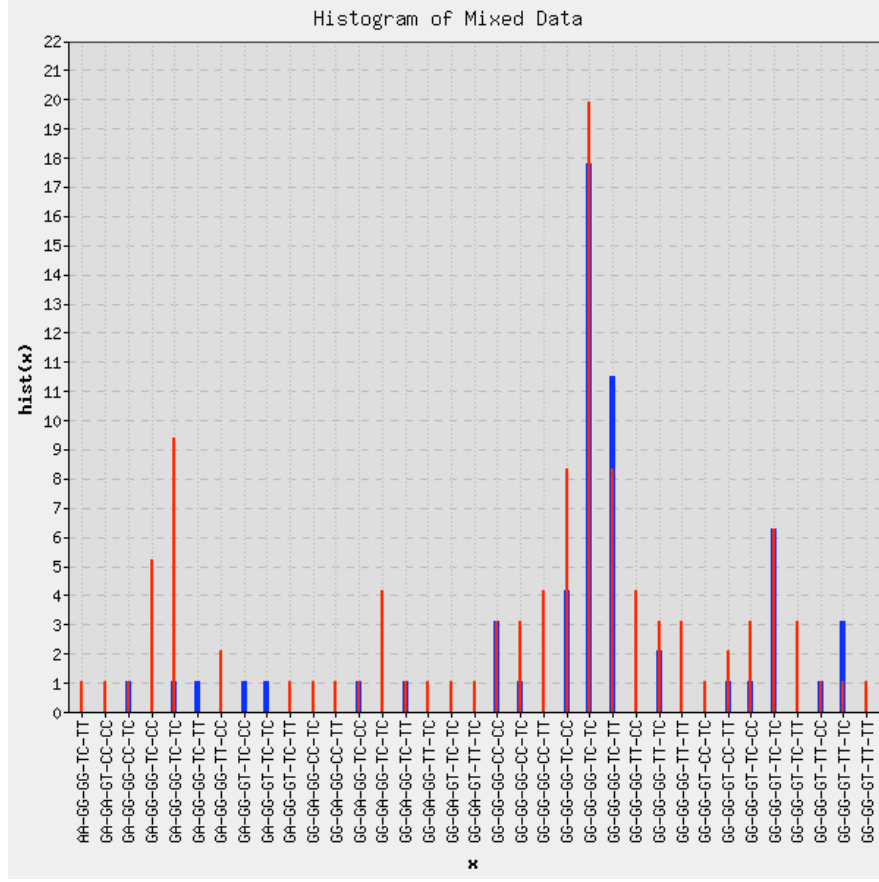
5 ayrı gen polimorfizminden oluşan verinin, çeşitli istatistiksel yöntemlerle analiz edilebilmesi için verinin matematiksel olarak ifade edilebilir bir şekilde getirilmesi gerekmektedir. Bu çalışmada gen polimorfizmlerinin her biri bir boyut olarak düşünülerek 5 boyutlu uzayda dağılmış bir veri yığını haline getirilmiştir. Dolayısıyla her bir hasta yada kontrol grubu verisi uzayda bir nokta (yada vektör) olacak şekilde ayarlanmıştır. Bu çalışma sonrasında hasta ve kontrol gruplarından oluşan 35 farklı vektör elde edilmiştir. Bu değerlerin listesi Çizelge 2.2’te, histogram gösterimi ise Şekil 2.1’de verilmiştir.

	Kontrol	Hasta	Toplam
AA-GG-GG-TC-TT	0	1	1
GA-GA-GT-CC-CC	0	1	1
GA-GG-GG-CC-TC	1	1	2
GA-GG-GG-TC-CC	0	5	5
GA-GG-GG-TC-TC	1	9	10
GA-GG-GG-TC-TT	1	0	1
GA-GG-GG-TT-CC	0	2	2
GA-GG-GT-TC-CC	1	0	1
GA-GG-GT-TC-TC	1	0	1
GA-GG-GT-TC-TT	0	1	1
GG-GA-GG-CC-TC	0	1	1
GG-GA-GG-CC-TT	0	1	1
GG-GA-GG-TC-CC	1	1	2

GG-GA-GG-TC-TC	0	4	4
GG-GA-GG-TC-TT	1	1	2
GG-GA-GG-TT-TC	0	1	1
GG-GA-GT-TC-TC	0	1	1
GG-GA-GT-TT-TC	0	1	1
GG-GG-GG-CC-CC	3	3	6
GG-GG-GG-CC-TC	1	3	4
GG-GG-GG-CC-TT	0	4	4
GG-GG-GG-TC-CC	4	8	12
GG-GG-GG-TC-TC	18	20	38
GG-GG-GG-TC-TT	11	8	19
GG-GG-GG-TT-CC	0	4	4
GG-GG-GG-TT-TC	2	3	5
GG-GG-GG-TT-TT	0	3	3
GG-GG-GT-CC-TC	0	1	1
GG-GG-GT-CC-TT	1	2	3
GG-GG-GT-TC-CC	1	3	4
GG-GG-GT-TC-TC	6	6	12
GG-GG-GT-TC-TT	0	3	3
GG-GG-GT-TT-CC	1	1	2
GG-GG-GT-TT-TC	3	1	4

GG-GG-GT-TT-TT	0	1	1
----------------	---	---	---

Çizelge 2.2. Vektörler.



Şekil 2.1. Pediatrik inme verisinin histogramı

Elde edilen bu vektörler, kontrol, hasta, toplam ve fark değerleri veritabanında bir tablo oluşturularak sorgulanabilir şekilde veritabanına aktarıldı. Yeni bir forma sokula sokulan veritabanından, kontrol ve hasta grubu verisinin toplamaları en büyükten küçüğe doğru olacak şekilde sorgulanarak yeni bir düzen elde edildi.

İstatistiksel öğrenme yöntemleriyle analiz edilmek istenen bir veri, istatistiksel olarak farklı kaynaklardan geldiğinin gösterilmesi gerekmektedir. Bunun için bazı hipotez testlerinden geçirilmesi gerekmektedir.

Bu noktada AA-GG-GG-TC-TT vb. şekilde temsil edilen verinin nümerik olarak ifade edilmesi gerekmektedir. Bu değerlerin nümerik olarak ifade edilmesinde verinin normal

dağıldığı varsayımıyla ve kontrol ile hasta gruplarını birbirinden ayırabilmek amacıyla şu yöntem uygulanmıştır;

- a) Vektörler kontrol sıklığı öncelikli olmak üzere toplamda görülme sıklığına göre sıralanmıştır.
- b) En sık görünen vektöre 0 (sıfır) değeri atanmıştır.
- c) Toplamlarına göre sık görünme sırasına göre sıralanan vektörlerde, aralarındaki farka bakılarak; fark 0 yada pozitif bir sayı ise bir önce atanan pozitif değer bir artırıldı, fark negatif bir sayı ise bir önce atanan negatif değer bir azaltıldı.

Vektörün herbir boyutu için de nümerik atama gerçekleştirilmiştir. Atamada kullanılan nümerik değerler Çizelge 2.3'te verilmiştir.

Olası SNP	Nümerik Değer
GG	50
GC-CG	40
GA-AG	60
GT-TG	20
CC	30
CA-AC	10
CT-TC	70
AA	80
AT-TA	90
TT	100

Çizelge 2.3 Vektör boyutları için atanan nümerik değerler

Atama sonrası elde edilen yeni veri düzeni Çizelge 2.4'te verilmiştir.

Vektör – Nümerik Değeri	Kontrol	Hasta	Toplam	Nümerik Değer
GG-GG-GG-TC-TC - 50 50 50 70 70	18	20	38	0
GG-GG-GG-TC-TT - 50 50 50 70 100	11	8	19	1
GG-GG-GT-TC-TC - 50 50 20 70 70	6	6	12	2
GG-GG-GG-TC-CC - 50 50 50 70 30	4	8	12	-1
GA-GG-GG-TC-TC - 60 50 50 70 70	1	9	10	-2
GG-GG-GG-CC-CC - 50 50 50 30 30	3	3	6	3
GG-GG-GG-TT-TC - 50 50 50 100 70	2	3	5	-3
GA-GG-GG-TC-CC - 60 50 50 70 30	0	5	5	-4
GG-GG-GT-TT-TC - 50 50 20 100 70	3	1	4	4
GG-GG-GG-CC-TC - 50 50 50 30 70	1	3	4	-5
GG-GG-GT-TC-CC - 50 50 20 70 30	1	3	4	-6
GG-GG-GG-CC-TT - 50 50 50 30 100	0	4	4	-7
GG-GG-GG-TT-CC - 50 50 50 100 30	0	4	4	-8
GG-GA-GG-TC-TC - 50 60 50 70 70	0	4	4	-9
GG-GG-GT-CC-TT - 50 50 20 30 100	1	2	3	-10
GG-GG-GG-TT-TT - 50 50 50 100 100	0	3	3	-11
GG-GG-GT-TC-TT - 50 50 20 70 100	0	3	3	-12
GG-GA-GG-TC-TT - 50 60 50 70 100	1	1	2	5
GA-GG-GG-CC-TC - 60 50 50 30 70	1	1	2	6
GG-GA-GG-TC-CC - 50 60 50 70 30	1	1	2	7
GG-GG-GT-TT-CC - 50 50 20 100 30	1	1	2	8

GA-GG-GG-TT-CC - 60 50 50 100 30	0	2	2	-13
GA-GG-GG-TC-TT - 60 50 50 70 100	1	0	1	9
GA-GG-GT-TC-CC - 60 50 20 70 30	1	0	1	10
GA-GG-GT-TC-TC - 60 50 20 70 70	1	0	1	11
GG-GG-GT-CC-TC - 50 50 20 30 70	0	1	1	-14
AA-GG-GG-TC-TT - 80 50 50 70 100	0	1	1	-15
GG-GA-GT-TT-TC - 50 60 20 100 70	0	1	1	-16
GG-GA-GT-TC-TC - 50 60 20 70 70	0	1	1	-17
GG-GA-GG-TT-TC - 50 60 50 100 70	0	1	1	-18
GG-GA-GG-CC-TT - 50 60 50 30 100	0	1	1	-19
GG-GA-GG-CC-TC - 50 60 50 30 70	0	1	1	-20
GA-GG-GT-TC-TT - 60 50 20 70 100	0	1	1	-21
GA-GA-GT-CC-CC - 60 60 20 30 30	0	1	1	-22
GG-GG-GT-TT-TT - 50 50 20 100 100	0	1	1	-23

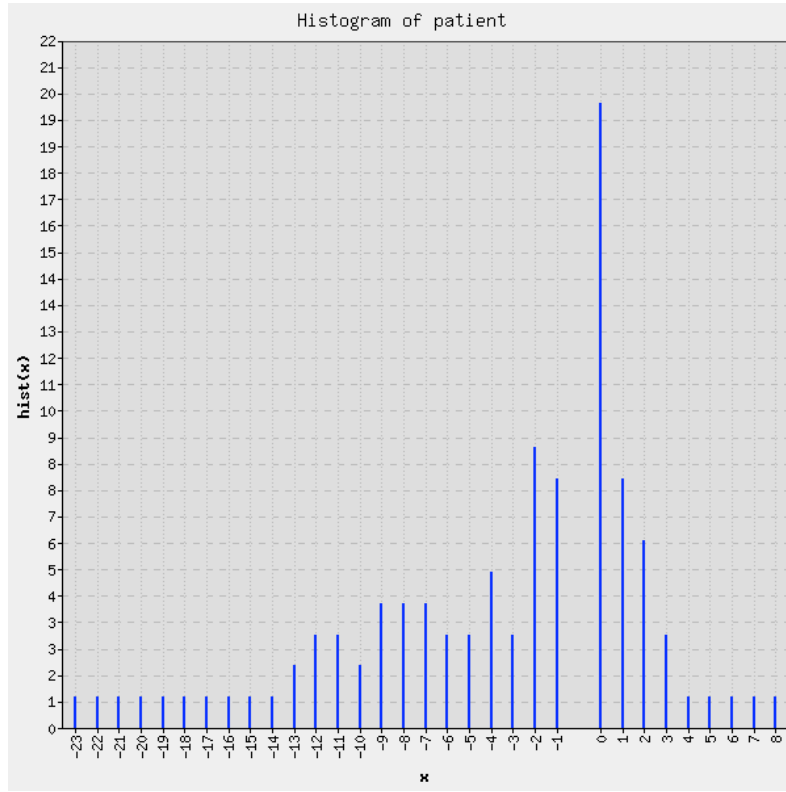
Çizelge 2.4. Atama sonrası oluşan yeni veri düzeni.

Atama sonrası oluşan yeni verinin kontrol, hasta ve her iki grup birlikte olarak gözönünde bulundurularak oluşan verinin istatistiksel parametreleri Çizelge 2.5 verildiği üzere hesaplanmıştır.

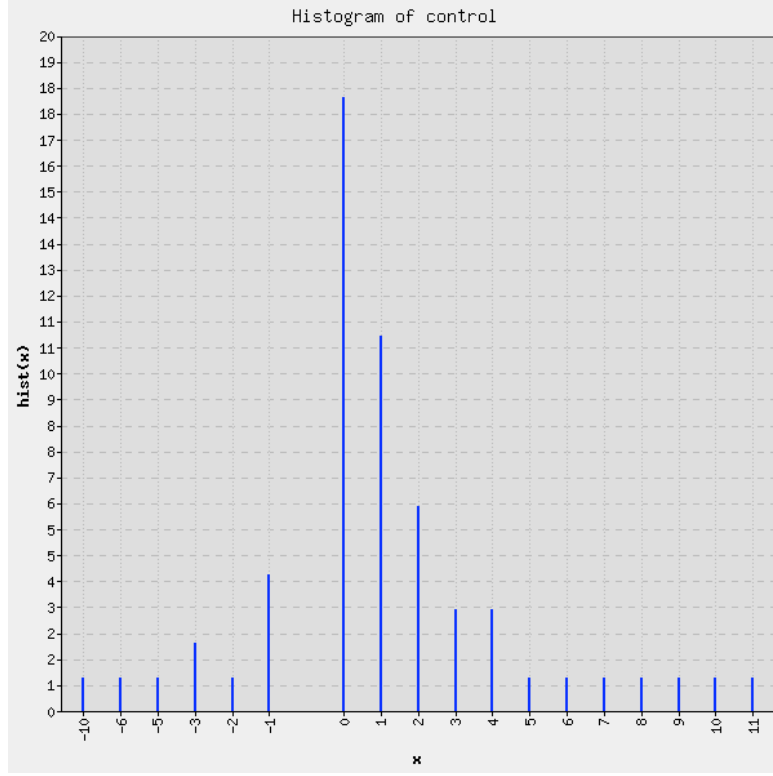
	Ortalama Değer	Varyans	Standart Sapma	N
Hasta	-4.047619047619	43.045787545788	6.5609288630336	105
Kontrol	1.1551724137931	12.203569267998	3.4933607411771	58
Heriki grup	-2.196319018405	38.171097477846	6.1782762545751	163

Çizelge 2.5 Pediatrik inme verisine ait istatistiksel parametreler

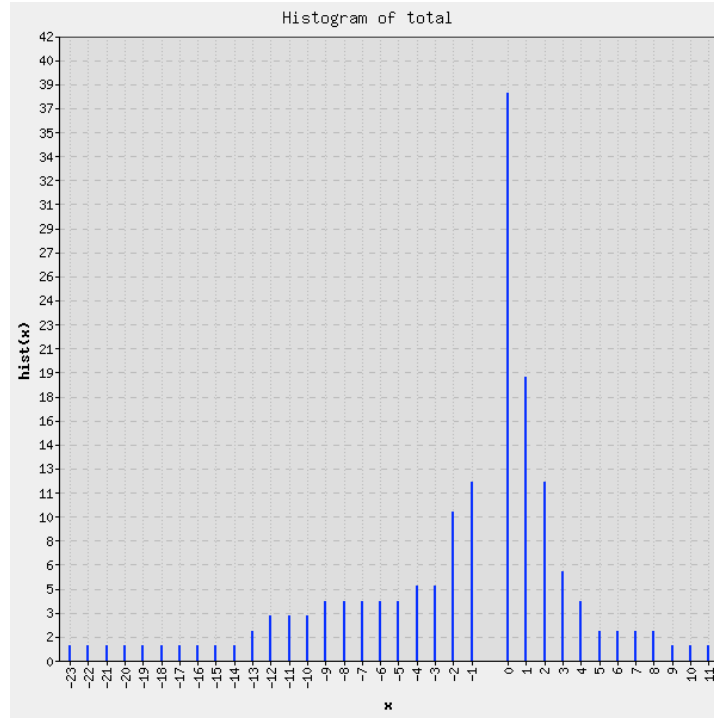
Bu verinin histogramları Şekil 2.2, Şekil 2.3 ve Şekil 2.4'te verilmiştir.



Şekil 2.2. Hasta grubuna ait verinin histogramı

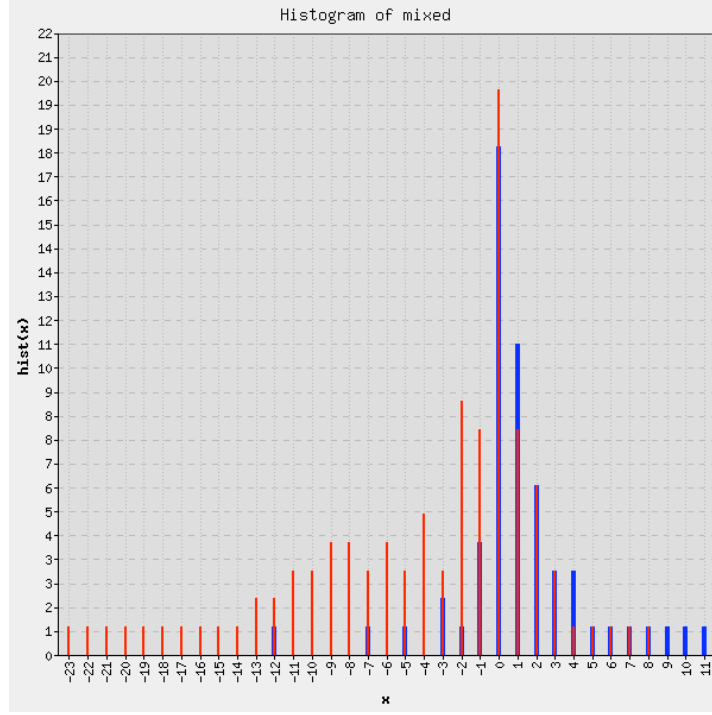


Şekil 2.3. Kontrol grubuna ait verinin histogramı



Şekil 2.4. Hasta ve kontrol grubuna ait verinin histogramı.

Kontrol ve Hasta gruplarının verilerinin ayrışmalarını göstermek için oluşturulmuş üstüste bindirilmiş histogramlar Şekil2.5'te verilmiştir.



Şekil 2.5. Hasta ve kontrol grubu histogramlarının birlikte gösterimi

2.2.2 Verinin İstatistiksel Olarak Kullanılabilirliğinin Testi

Nümerik olarak da ifade edilen hasta ve kontrol grubu verilerinin herhangi bir istatistiksel model içinde kullanılmadan önce her iki grup verilerinin istatistiksel olarak birbirinden bağımsız olduğunun, diğer bir deyişle farklı kaynaklardan geldiğinin gösterilmesi zorunludur.

Her iki grup verilerinin birbirinden bağımsız olduğunun gösterilmesi için

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Hipotez testleri ile t-test testinden geçirilmelidir. Nümerik olarak ifade edilen data t-test için kullanılabilir hale getirildi. Bunun için görünme sıklığına göre veri yeniden

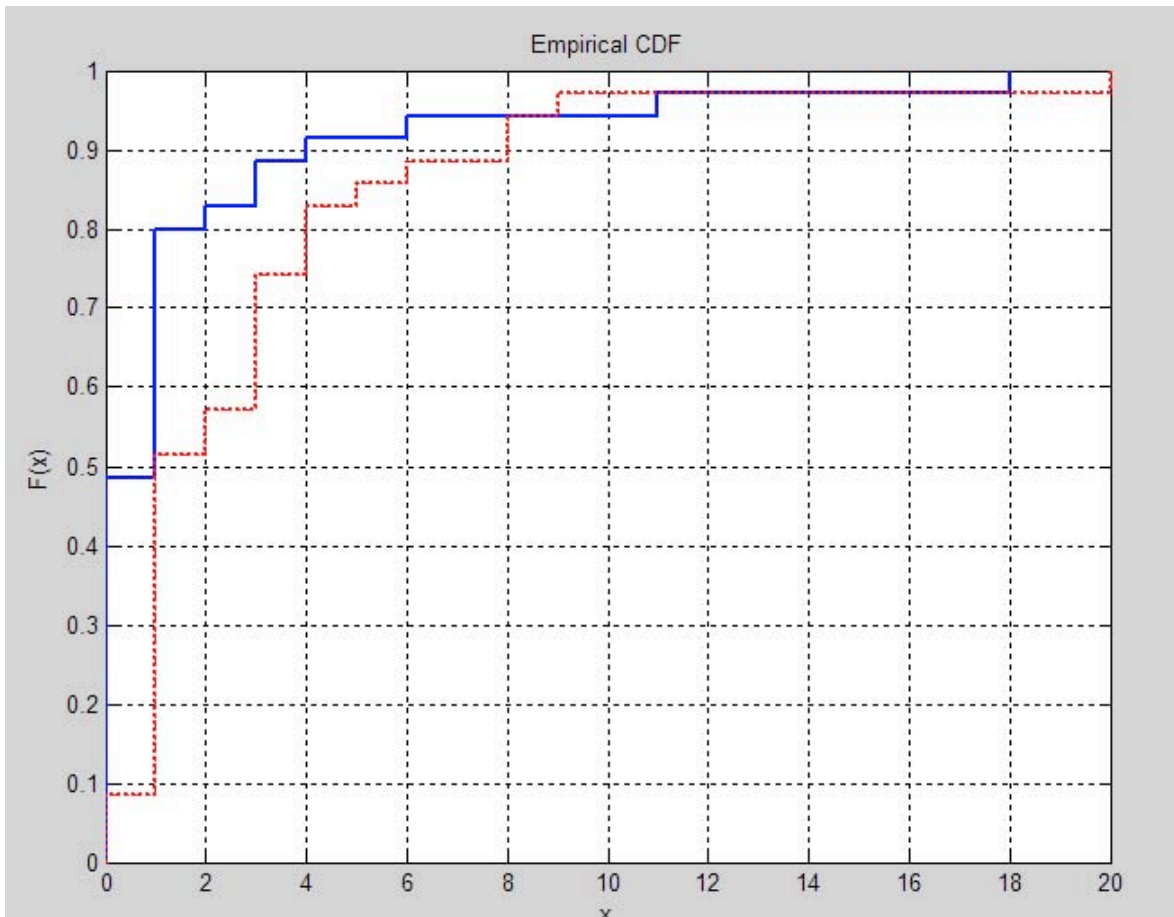
oluřturuldu. Oluřturulan yeni veri Matlab kullanılarak t-test ve Kolmogorov-Smirnov testleri uygulanmıřtır.(EK 3.1)

2.2.2.1 Pediatrik inme verisine ait t-test sonuları

h:1 (alpha:0.01, tstat: 6.6056, df: 160.8444, significance: 0.0000000006) ıkmıřtır. Alpha 0.01 de yani %99 gven aralıęında bu null hipotezinin reddedilmesi demektir. Daęılımların ortalama deęerleri birbirinden farklıdır.

2.2.2.1 Pediatrik inme verisine ait Kolmogorov-Smirnov test sonucu :

h:1 (KSSTAT: 0.4000) ıkmıřtır. Daęılımların aynı olduęu hipotezi reddedilmiřtir. Kolmogorov-Smirnov test sonucu elde edilen fark ařaęıdaki Őekil 2.6'daki grafikte verilmiřtir.



Őekil 2.6 Kolmogorov-Smirnov test sonucu iki daęılım arasındaki farkın grafiksel gsterimi

2.2.3 Verinin PCA ile analizi

Boyutları nümerik olarak ifade edilerek yeniden oluşturulan veri Matlab'da PCA uygulanarak kontrol grubu ile hasta grubunun birbirinden ayrışıp ayrışmadığı gösterilmiştir. (EK 3.2)

Varyansların kümülatif toplamları tek bir principal component ile %52.0046, iki principal component ile %84.5991, üç principal component ile %97.8439, dört principal component ile %99.3451, beş principal component %100.0000 olarak hesaplanmıştır. Buradan hareketle verilen ilk iki boyuta bakarak veri hakkında bir tahminde bulunulması iki principal component ile %84.59 güvenilirlikle mümkündür.

2.2.4 K-means Clustering

Yeniden oluşturulan veri sırasıyla şu işlemlerden geçirilmiştir.

1. Hasta ve kontrol gruplarından rasgele 10'ar veri test verisi olarak seçilmiştir.
2. 1. basamaktaki işlem 10 kez tekrar edilmiştir..
3. Sistem, Matlab kullanılarak, her bir eğitim verisi kullanılarak k-means clustering algoritmasıyla eğitilmiştir. (EK 3.2)
4. Herbir eğitim verisi, Matlab kullanılarak, kendisinden rasgele ayrılmış olan test verisi ile sınanmıştır. (EK 3.2)

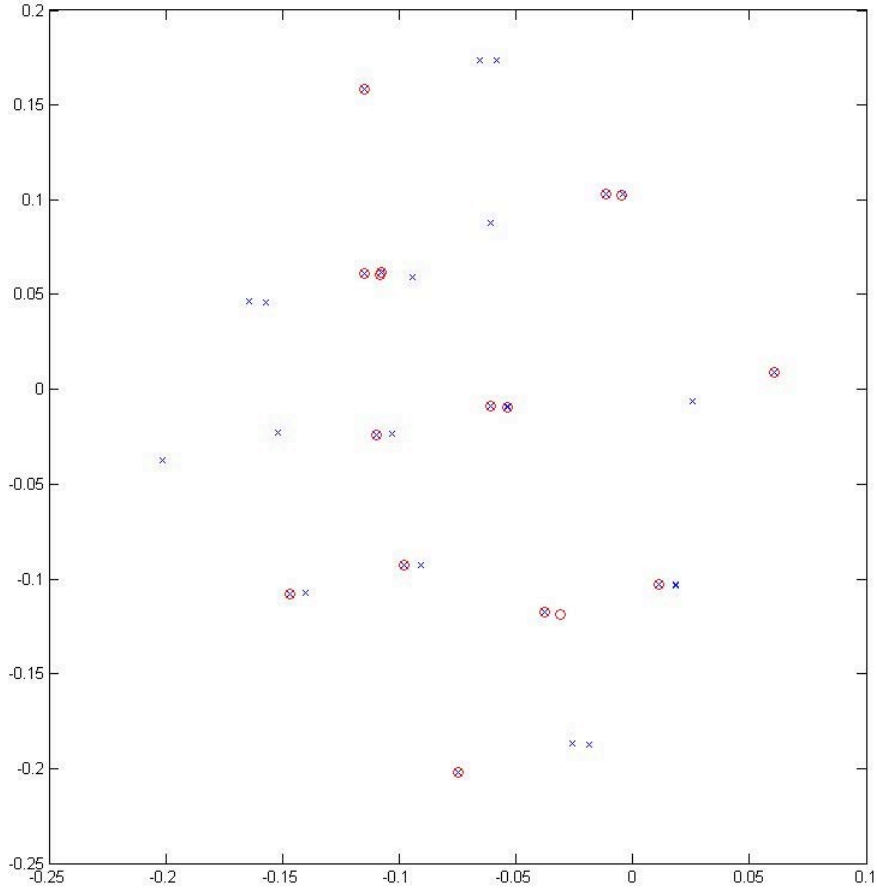
3. ARAŞTIRMA BULGULARI

Üst üste bindirilmiş histogramlarda (Şekil 2.5) her iki grubunda yoğunlaşması birbirine yakın noktalarda olmaktadır. Bazı noktalarda üst üste binme gerçekleşse de bu noktalarda genelde tek bir vaka görülmektedir.

Kullanılan verinin istatistiksel olarak kullanılabilirliğinin testi sonucunda, t-test'e göre hasta ve kontrol dağılımlarının ortalama değerlerinin birbirinden farklı olduğu görülmüştür. Dolayısıyla hasta ve kontrol gruplarının birbirinden farklı iki kaynaktan geldiği görülmüştür.

Kolmogorov-Smirnov testi sonucunda her iki dağılımın aynı olduğu hipotezi reddilmiş ve dağılımların farklı olduğu sonucunu ortaya koymuştur. Ancak pediatrik inme verisinde görülen polimorfizmler bir çok değer için hasta ve kontrol grubunda tıpatıp aynı gelmiştir.

Varyansların kümülatif toplamları tek bir principal component ile %52.0046, iki principal component ile %84.5991, üç principal component ile %97.8439, dört principal component ile %99.3451, beş principal component %100.0000 olarak hesaplanmıştır. Buradan hareketle verilen ilk iki boyuta bakarak veri hakkında bir tahminde bulunulması iki principal component ile %84.59 güvenilirlikle mümkündür. Bu iki boyut kullanılarak hasta ve kontrol verilerinin iki boyutlu gösterimi Şekil 3.1'deki grafikteki gibidir. Şekil 3.1'deki grafikte de görüldüğü gibi "o" ile gösterilen kontrol verisi "x" gösterilen hasta verisiyle örtüşmektedir. Bölüm 2.2.4 de bahsedilen rasgele oluşturulan 10 adet eğitim datasıyla da yapılan tüm PCA uygulamalarının sonuçları benzer sonuçlar vermiştir.



Şekil 3.1Verinin PCA analizi sonucu 2 boyutlu gösterimi

Bölüm 2.2.4 de açıklanan ve rasgele oluşturulan 10 adet eğitim datasıyla K-means clustering algoritması kullanılarak sistemler eğitilmiştir. Eğitilen sistemler, daha önce rasgele seçilen, kendilerine ait test verisiyle sınanmıştır. Bu sınamanın sonuçları Çizelge 3.1 de görülmektedir.

Çizelge 3.1 de Veri kümesi rasgele oluşturulan eğitim verilerinden oluşturulan kümeleri göstermektedir. Kontrol test ve hasta test sütunları test verisi olarak rasgele seçilen verinin eğitilmiş kümede doğru yerleşme oranlarını vermektedir. Centroid Hasta ve Centroid Kontrol sütunlarında ise eğitim sonrası oluşan kontrol ve hasta grubuna ait kümelerin merkezlerini göstermektedir.

Eđitim Kumesi	Kontrol Test	Hasta Test	Centroid Hasta	Centroid Kontrol
#1	1/10	2/10	{51.3084, 50.9346, 42.9907, 68.9720, 79.2523}	{52.5000, 50.5556, 44.1667, 68.8889, 30.0000}
#2	2/10	8/10	{51.3889, 51.0185, 43.3333, 68.5185, 79.4444}	{52.5714, 50.8571, 44.0000, 68.8571, 30.0000}
#3	0/10	7/10	{51.3084, 51.0280, 43.2710, 67.7570, 79.8131}	{52.5000, 50.8333, 43.3333, 71.1111, 30.0000}
#4	4/10	7/10	{51.6216, 50.8108, 42.9730, 68.7387, 79.7297}	{52.5000, 50.6250, 43.4375, 70.3125, 30.0000}
#5	2/10	8/10	{51.4815, 50.8333, 43.3333, 70.1852, 80.0000}	{52.2857, 50.8571, 44.0000, 70.0000, 30.0000}
#6	5/10	9/10	{51.7647, 50.7843, 42.9412,	{51.7073, 50.7317, 43.4146,

			72.9412, 55.8824}	57.3171, 94.1463}
#7	3/10	7/10	{51.4545, 51.0000, 43.4545, 68.7273, 79.0000}	{52.4242, 50.9091, 44.5455, 68.1818, 30.0000}
#8	1/10	7/10	{51.3889, 50.9259, 43.0556, 69.5370, 79.4444}	{52.5714, 50.8571, 43.1429, 68.0000, 30.0000}
#9	3/10	6/10	{51.7476, 50.6796, 43.0097, 73.5922, 56.4078}	{51.7500, 51.0000, 43.2500, 56.5000, 94.7500}
#10	3/10	9/10	{51.5741, 50.8333, 43.6111, 68.0556, 79.7222}	{52.2857, 50.5714, 44.8571, 68.0000, 30.0000}
Ortalama	2.4/10	7/10		

Çizelge 3.1 K-means clustering sonrası analiz sonuçları

Çizelge 3.1 incelendiğinde ortaya şöyle bir sonuç çıkmaktadır: Eldeki veri kullanılarak eğitilen sistem sırandığında verdiği cevaplar beklenileni karşılamamıştır. Kümelendirme işlemi sonucunda eğitilmiş sistemler kontrol test verilerini %24, hasta test verilerini %70 bir ortalamayla doğru cevap vermektedir. Kontrol test verilerininin %24'lük bir ortalamayla

dođru cevap vermesi PCA uygulaması sonucu elde edilen grafikte (Şekil 3.1) görünen içiçe geçmiş görüntüyü dođrulamaktadır.

Çizelge 3.1 de ayrıca, hasta ve kontrol kümelenme merkezleri incelendiğinde, tamamıyla rasgele seçilen eğitim ve test verilerinin %80 oranında benzer merkezler de yoğunlaştığı görülmektedir.

4. SONUÇ, TARTIŞMA ve ÖNERİLER

Ankara Üniversitesi, Tıp Fakültesi, Dahili Tıp Bilimleri Bölümü, Çocuk Sağlığı ve Hastalıkları Anabilim Dalı, Pediatrik Moleküler Genetik Bölümü'nden sağlanan pediatrik inme hastalığına ait verinin istatistiksel olarak anlamlılığının testi sonucunda, hasta ve kontrol gruplarının birbirinden farklı iki kaynaktan geldiği gösterilmiştir. Dolayısıyla bu verinin istatistiksel öğrenme yöntemleriyle analizi mümkündür.

İstatistiksel öğrenme yöntemleriyle yapılan analizlerde görülmüştür ki, kullanılan hasta ve kontrol grupları veriler, her ne kadar ayrı kaynaklardan gelseler bile, içiçe geçmiş bir yapıdadırlar. Hasta ve kontrol grubuna ait ayrıştırılmak istenen verinin gerçek yaşamdaki durumu modelleyebiliyor olması gerekmektedir. Bir verinin, hastalık söz konusu olduğunda, sağlıklı bir yığın içinde hasta olan durumları gösterebilmesi için kontrol sayısının hasta sayısından daha fazla olması beklenmektedir. Eldeki veride 105 hasta verisine karşılık 58 kontrol verisinin olması verinin sağlıklı sonuçlar vereceğine işaret olarak görülebilmektedir.

Bir istatistiksel analizin güvenilirliği, analizde kullanılan verinin çokluğu ile doğru orantılıdır. Eldeki veri ne kadar çok ise istatistiksel analiz sonrası elde edilen sonucun gerçek hayatta gerçekleşen durumu modellemesi konusunda, daha güvenilir olacağı söylenebilir.

Verinin derlenmesi sırasında ortaya çıkan tabloya bakıldığında şunlar görülmektedir:

- Hasta ve kontrol verisi aynı noktalarda yoğunluk göstermektedir.
- Hasta ve kontrol verisine ait histogram aynı noktada en yüksek değeri almaktadır.
- Hasta ve kontrol verisinin tıpatıp aynı olduğu polimorfizmler bulunmaktadır.
- Hasta ve kontrol verisinin birbirinden ayrıldığı noktalar bulunmaktadır fakat bu noktalarda genelde tek bir vaka gözlenmiştir.

Buradan hareketle eldeki verinin toplanması aşamasında seçilen gen polimorfizmlerinin hasta ve kontrol grubunu ayırtmakta yeterli olmadığı görülmektedir. Yoğunlaşmanın ve en yüksek değer noktasının aynı olması şu durumlardan dolayı gerçekleşmiş olabilir:

- Bu noktalarda, hasta olan bireylerde, hastalığı tetikleyen başka bir faktör (başka bir gen yada çevresel faktörler) rol oynamış olabilir
- Bu noktalarda, kontrol grubundaki bireylerde, hastalığı baskılayan başka bir faktör (başka bir gen yada çevresel faktörler) rol oynamış olabilir.

Bu noktalar açık değildir ve araştırılması gerekir.

Tek bir vaka gözlenen noktalarda, her ne kadar hasta ve kontrol grubu ayrışmış gibi görünse de, yeni bir vaka geldiğinde, o noktalar da hasta ve kontrol grubunun birbiriyle aynı oldukları noktalardan biri olma ihtimali vardır.

Çoklu gen hastalıklarının kompleks yapıda hastalıklar olduğunu bilinmektedir. Bu çalışmada kullanılan verinin bu kompleks yapıyı iyi temsil edemediği görülmüştür. Analizde kullanılan polimorfizmlerden biri baskılayıcı etken olarak bu kompleks yapıda görev alıyor olabilir. Bunun yanında başka bir gen yada çevresel etken bu kompleks yapıda tetikleyici unsur olarak bulunuyor olabilir. Gerek Şekil 2.5'deki üstüste bindirilmiş histogram grafiği incelendiğinde, gerek Şekil 3.1'deki PCA uygulaması sonucu verinin 2 boyutlu gösterimi incelendiğinde bu durum daha net bir şekilde görülmektedir. Bu kompleks yapının aydınlatılabilmesi için analize başka bir gen yada genlerdeki polimorfizmler yada çevresel faktörler de dahil edilmelidir.

Bu tip çalışmalarda, çalışmanın sağlığını ve güvenilirliğini sağlamak için toplanan verinin mümkün olduğunca çok olması sağlanmalı ve kontrol grubu verisinin hasta grubu verisinden daha çok olması gözden kaçırılmamalıdır. Bu sağlandığında, hasta ve kontrol grubunun birbirinden bağımsız kaynaklardan geldiği kesinleştirildiği takdirde güdümlü öğrenme yöntemiyle analizine geçilebilektir.

Bu konuda çalışma yapan araştırmacılar, istatistiksel öğrenme açısından önemli olan train verisini mümkün olduğunca çok toplamaları gerekmektedir. Veri toplandıktan sonra özniteliklerin özenle seçilip, bunların kullanılıp kullanılmayacağı PCA gibi uygulamalarla önceden öngörülme çalışılmalıdır.

KAYNAKÇA

Davison, A. M. 2005. Oxford Textbook of Clinical Nephrology, Oxford University Press, 2648.

Ding, C., He, X.2004. K-means Clustering via Principal Component Analysis. Proc. of Int'l Conf. Machine Learning (ICML 2004), 225-232

Demir, B. N. 2004 Pediatrik İnmelerde Koruyucu Olduđu Düşünölen Gen Deđişimlerinin Etkisi, Yüksek Lisans Tezi (yayınlanmamış), Ankara Üniversitesi, 66, Ankara

Duda, R.O., Hart, P. E., Stork, D.G. 2000. Pattern Classification, Wiley-Interscience, 680

Friedmann, T. 1994. Gene Therapy: fact and fiction in biology's new approaches to disease, CSHL Press, 124.

Gaulton, K. J., Mohlke, K. L. ve Vision, T. J. 2007. A computational system to select candidate genes for complex human traits. Bioinformatics, 23 (9 2007); 1132-1140.

Gormley, M. V. 1989. Family Diseases: Are You at Risk?, Genealogical Publishing, 166

Griffiths, A. J. F., Gelbart, W. M., Miller, J. H., Lewontin, R. C. 1999. Modern Genetic Analysis, W. H. Freeman, 576

Hartl, D. L., Jones, E. W. 2005, Genetics: Analysis of Genes and Genomes, Jones & Bartlett, 854

Hastie, T., Tibshirani, R., Friedman, J. 2001, The Elements of Statistical Learning - Data Mining, Inference and Prediction, Springer, 533.

Jameson, J. L., Collins, F. S. 1998. Principles of Molecular Medicine, Humana Press, 1123.

Kwok, P.2002. Single Nucleotide Polymorphisms: Methods and Protocols, Humana Press, 288

Le, C.T. 2003. Introductory Biostatistics, Wiley-Interscience, 552

Lewin, B. 2004. Genes VIII, Pearson Prentice Hall, 1056

- Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A., Krieger, M., Scott, M. P., Zipursky, L., Darnell, J. 2003. *Molecular Cell Biology*, W. H. Freeman, 1152
- Marques de Sa, J.P. 2001. *Pattern Recognition: Concepts, Methods And Applications* Springer, 318
- Meyers, R. A. 1995. *Molecular Biology and Biotechnology: A Comprehensive Desk Reference*, Wiley-VCH, 1072.
- Miller, M. S., Cronin, M. T. 2000. *Genetic Polymorphisms and Susceptibility to Disease*, CRC Press, 1109
- Mitchell, T.M. 1997, *Machine Learning*, McGraw-Hill, 342
- Nei, M. 1975. *Molecular Population Genetics and Evolution*, North-Holland Publishing, 228
- Pence, G. E. 1998. *Who's Afraid of Human Cloning?*, Rowman & Littlefield, 192
- Raven, P. H., Johnson, G. B. 2001. *Biology*, McGraw-Hill, 1344
- Stork, D.G., Yom-Tov E. 2004. *Computer Manual in MATLAB to accompany Pattern Classification*, Wiley-Interscience, 134
- Vapnik, 1998.V.N. *Statistical Learning Theory*, Wiley-Interscience, 732
- Wasserman, L. 2006. *All of Nonparametric Statistics*, Springer, 268
- Wink, M. 2006. *An Introduction to Molecular Biotechnology: Molecular Fundamentals, Methods and Applications in Modern Biotechnology*. Wiley-VCH, 828
- Ziegler A., König, I.R. 2006. *A Statistical Approach to Genetic Epidemiology: Concepts and Applications*, Wiley-VCH, 361

EKLER

Ek-1. Hasta Verisi

Hasta Adı	FVL 1691	PT20210	FXIII	TFPI	5-HT2A
AF	GG	GG	GG	CC	XX
ABT	GG	GG	GG	TT	CC
AD	GG	GG	GG	CC	TC
AS	GG	GG	XX	CC	TC
AK	GG	GG	GT	TC	TT
AT	GG	GG	GT	XX	TC
AA	GG	GG	GG	TC	CC
AB	GG	GG	GT	TC	TC
AP	GG	GG	GG	TC	TC
AU	GA	GG	GG	TC	TC
A	GG	GG	XX	XX	TT
BC	GG	GG	GG	TT	CC
BL	GG	GG	XX	XX	TC
BA	GA	GG	GG	TT	CC
B	GA	GG	GG	TC	TC
BK	GG	GA	GG	TC	TC
BA	GG	GG	GG	TC	TC
B	GG	GG	GG	TT	TC
BZS	GG	GG	GG	TC	TC
B	GG	GG	GG	TC	TT

BK	GA	GG	GG	TC	CC
BI	GG	GG	GT	TC	TC
BB	GG	GG	GG	TC	TC
CE	GG	GG	GG	TC	TC
C	GG	GG	GG	TC	TC
CA	GG	GG	GT	TT	TC
D	GG	GA	GT	TT	TC
DA	GG	GG	GG	TC	TT
DT	GG	GG	GT	TC	TC
DO	GG	GG	GG	CC	CC
EB	GA	GA	GT	CC	CC
E	GG	GG	GG	TC	TC
E	GG	GA	GG	CC	TC
ED	GG	GG	GG	TC	TT
ENK	GG	GG	GG	TC	CC
ES	GG	GG	GT	TT	TT
ED	GG	GG	GG	CC	TT
ET	GG	GG	GG	TC	TC
EY	GG	GG	GG	XX	XX
FT	GG	GG	GG	TT	XX
FA	GG	GG	GG	TC	CC
F	GG	GG	GG	TC	TC
FB	GG	GA	GG	TC	TC

FBG	GG	GG	GG	TC	TC
FD	GG	GG	GG	TC	TC
FG	GG	GG	GG	TC	TC
GY	GG	GG	GG	CC	TT
HK	GG	GG	GG	TT	CC
H	GA	GG	GG	TC	TC
HZ	GG	GG	GG	TC	TC
HK	GA	GG	GG	TC	XX
HB	AA	GG	GG	TC	TT
H	GG	GA	GG	TC	TT
HD	GG	GG	GT	CC	TC
HE	GG	GG	GG	TC	TC
HA	GG	GG	GG	TC	TC
HS	GG	GG	GG	TC	TC
HK	GG	GG	GG	TC	CC
IA	GG	GG	GT	CC	TT
IS	GG	GG	GG	TC	TT
IM	GG	GG	GG	TC	TT
I	GG	GG	GG	CC	TC
KK	GA	GG	GG	TC	TC
KK	GA	GG	XX	XX	TC
KA	GG	GG	GG	TC	TT
KD	GG	GA	GG	TC	TC

KA	GG	GA	GG	TT	TC
KK	GG	GG	GT	TC	TC
LS	GA	GG	GG	TC	CC
L	GA	GG	GG	TC	TC
MAB	GG	GG	GT	TC	CC
MD	GG	GG	GG	CC	TT
MY	GG	GG	GG	CC	CC
MB	GG	GG	GT	TC	XX
MK	GA	GG	GG	TC	CC
MS	GG	GG	GT	TC	TT
MD	GG	GG	GG	TC	CC
MC	GA	GG	GG	TC	TC
NT	GG	GG	GG	XX	XX
NB	GG	GG	GG	TC	TC
NG	GG	GG	GG	CC	XX
NB	GA	GG	GG	TC	TC
NP	GG	GG	GG	TC	TT
NG	GA	GG	GT	TC	TT
NA	GG	GG	GG	TC	CC
OC	GA	GG	GG	TC	CC
OG	GG	GG	GT	TC	CC
OB	GG	GA	GT	TC	TC
OD	GG	GG	GT	TT	CC

OS	GG	GA	GG	TC	TC
ÖY	GG	GG	GG	TT	TT
Ö	GA	GG	GG	CC	TC
RG	GG	GG	GG	CC	TC
RD	GG	GG	GG	CC	CC
RK	GA	GG	GG	TC	TC
S	GG	GG	GG	TT	CC
STY	GA	GG	GG	TC	CC
SK	GA	GG	GG	TT	CC
SS	GG	GG	GT	TC	TT
SS	GG	GG	GG	TC	TC
SC	GG	GG	GT	TC	TC
SK	GG	GG	XX	XX	TC
S	GG	GG	GG	TC	TT
S	GG	GG	GT	CC	TT
ŞK	GG	GG	GG	CC	TT
ŞS	GG	GG	GG	TT	TT
ŞA	GG	GA	GG	TC	CC
TG	GG	GG	GG	TC	TC
TA	GG	GA	GG	CC	TT
T	GG	GG	GG	TT	TC
UY	GG	GG	GG	TT	TC
UBO	GG	GG	GT	TC	TC

U	GG	GG	GG	TT	TT
UK	GG	GG	GG	TC	TC
ÜS	GG	GG	GG	TC	XX
VS	GG	GG	GG	TC	CC
Y	GG	GG	GT	TC	CC
YD	GA	GG	GT	TC	XX
YA	GG	GG	XX	XX	TC
Y	GG	GG	GG	TC	CC
YR	GA	GG	GG	TC	TC

Ek-2. Kontrol Verisi

Hasta Adı	FVL 1691	PT20210	FXIII	TFPI	5-HT2A
AH	GG	GG	GT	TT	CC
AK	GG	GG	GG	TC	CC
AK	GG	GG	GT	TC	XX
AU	GG	GG	GG	TC	TT
A	GG	GG	GT	CC	TT
A	GG	GG	GG	TC	TC
A	GG	GG	XX	TC	XX
AU	GA	GG	GT	TC	TC
A	GG	GG	GT	TT	TC
AA	GG	GG	XX	XX	CC

AB	GA	GG	GG	CC	TC
AK	GG	GG	GG	CC	XX
AT	GG	GG	GG	TC	TT
AD	GG	GG	GG	TC	XX
A	GG	GG	GT	TC	XX
BY	GG	GG	GG	TC	XX
BA	GG	GG	GG	TC	CC
BD	GG	GG	GT	TC	TC
BK	GG	GG	GG	TC	TT
CK	GG	GG	GG	TC	TC
CK	GG	GA	GG	TC	CC
C	GG	GG	GG	TC	TT
ÇE	GG	GG	GG	TC	XX
ÇU	GG	GG	GG	XX	TC
DT	GA	GG	GG	TC	TC
DA	GG	GG	GG	TC	TT
DS	GG	GG	GT	CC	XX
DB	GG	GG	GG	TT	TC
D	GG	GG	GG	TC	TT
DG	GG	GG	GG	TC	TC
DK	GG	GG	GG	TC	TT

DA	GA	GG	GG	TC	XX
DH	GG	GG	GT	TC	XX
D	GG	GG	GG	TC	XX
EA	GG	GG	GT	TC	XX
EU	GG	GG	XX	TC	XX
E	GG	GG	XX	TC	XX
EB	GG	GG	GG	TC	CC
E	GG	GG	GT	TC	CC
EK	GG	GG	GG	TC	TT
EY	GG	GG	GG	TT	TC
FP	GA	GG	GG	TC	XX
FA	GG	GG	GG	TC	TC
FV	GG	GG	GG	TC	XX
FA	GG	GG	GT	TT	TC
FE	GG	GG	GG	XX	XX
F	GG	GG	GG	TC	XX
GD	GG	GG	GT	TT	TC
GT	GA	GG	XX	XX	CC
GB	GG	GG	GG	TC	TC
GD	GG	GG	GG	CC	CC
G	GG	GG	GG	CC	CC

HG	GG	GG	GG	TC	XX
HA	GG	GG	GT	TC	TC
H	GA	GG	XX	XX	TC
HT	GG	GG	GG	TC	TC
HB	GG	GG	GG	XX	TT
HG	GG	GG	XX	TC	XX
HA	GG	GG	GG	TC	XX
HE	GA	GG	GG	XX	TC
Ī	GG	GG	XX	XX	TC
ĪP	GG	GG	GG	TC	TC
KU	GA	GG	GG	TC	XX
MS	GG	GA	TT	XX	TT
MH	GG	GG	GG	XX	TC
M	GG	GG	GT	XX	CC
MB	GG	GA	XX	TC	XX
MT	GG	GG	GG	TC	XX
MM	GG	GG	GG	TC	XX
NK	GG	GG	GG	TC	CC
NA	GG	GG	GG	TC	XX
N	GG	GG	GG	TC	TT
NP	GG	GG	GG	TC	TC

OK	GG	GG	XX	TC	XX
ÖU	GG	GG	GT	TC	XX
ÖB	GG	GG	GG	TC	TT
PA	GG	GG	GG	TC	XX
PH	GG	GG	GG	TC	TC
P	GG	GG	GT	TC	TC
PAY	GG	GG	GG	TC	TC
RT	GG	GG	GT	TC	XX
RA	GG	GG	GG	CC	TC
RC	GG	GG	GG	TC	XX
STK	GG	GG	GT	TC	XX
SM	GG	GG	GG	TC	XX
SM	GG	GG	GT	CC	XX
SB	GG	GG	GG	TC	XX
SB	GG	GG	XX	TC	XX
SK	GG	GG	GG	TC	TC
SU	GG	GG	GG	XX	TC
SE	GG	GG	XX	XX	TT
SG	GG	GG	GG	TC	TC
S	GG	GG	GG	TC	TC
SS	GA	GG	GG	TC	TT

ŞE	GG	GG	GT	TC	TC
ŞK	GG	GG	GG	TC	TC
ŞA	GG	GG	GG	TC	TC
ŞG	AA	GG	GG	TC	XX
TK	GG	GG	GG	TC	XX
TA	GG	GG	XX	XX	TT
TG	GG	GG	XX	XX	TC
T	GG	GG	GG	TC	TT
TS	GG	GG	GG	CC	CC
TA	GG	GG	GG	TC	TC
TD	GG	GG	GG	TC	TC
U	GA	GG	GT	TC	CC
VE	GG	GG	GT	TC	TC
VK	GG	GG	TT	XX	TC
VA	GG	GG	GG	XX	TC
VC	GG	GG	GG	TC	XX
Y	GG	GG	GT	TC	XX
Y	GA	GG	GT	XX	XX
YE	GA	GA	GG	TC	XX
Y	GG	GG	XX	TC	XX
YS	GG	GG	XX	XX	TC

ZK	GG	GG	XX	TC	XX
ZM	GG	GA	GG	TC	TT
Z	GG	GG	GT	TC	TC
ZY	GG	GG	GG	TC	TC

Ek-3. Matlab Kodları

Ek-3.1 T-test ve Kolmogorov-Smirnov Testleri

ttest_and_kol-smi.m

```

% Clear all variables and functions from the memory %
clc
clear

% Define control, patient and all data as arrays %
cData=[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, -1, -1, -1, -1, -2, 3, 3, 3, -3, -
3, 4, 4, 4, -5, -6, -10, 5, 6, 7, 8, 9, 10, 11];
pData=[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, -1, -1, -1, -1, -1, -1, -1, -1, -2, -
2, -2, -2, -2, -2, -2, -2, -2, 3, 3, 3, -3, -3, -3, -4, -4, -4, -4, -4,
4, -5, -5, -5, -6, -6, -6, -7, -7, -7, -7, -8, -8, -8, -8, -9, -9, -9, -
9, -10, -10, -11, -11, -11, -12, -12, -12, 5, 6, 7, 8, -13, -13, -14, -
15, -16, -17, -18, -19, -20, -21, -22, -23];
allData=[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, -1, -1, -1, -1, -2, 3, 3, 3, -
3, -3, 4, 4, 4, -5, -6, -10, 5, 6, 7, 8, 9, 10, 11, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2,
2, 2, 2, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -2, -2, -2, -2, -2, -2, -2, -2,
-2, 3, 3, 3, -3, -3, -3, -4, -4, -4, -4, -4, -4, 4, -5, -5, -5, -6, -6, -6, -
7, -7, -7, -7, -8, -8, -8, -8, -9, -9, -9, -9, -10, -10, -11, -11, -11, -
12, -12, -12, 5, 6, 7, 8, -13, -13, -14, -15, -16, -17, -18, -19, -20, -
21, -22, -23];

histValsC=[18, 11, 6, 4, 1, 3, 2, 0, 3, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1,
1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0];

```

```
histValsP=[20, 8, 6, 8, 9, 3, 3, 5, 1, 3, 3, 4, 4, 4, 2, 3, 3, 1, 1, 1,
1, 2, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1];
```

```
% alpha %
```

```
alpha=.01;
```

```
tail='both';
```

```
% Hasta ve kontrol gruplarını karşılaştırmak için ttest uyguluyoruz
```

```
%
```

```
[h,significance,ci,stats] = ttest2(cData,pData,alpha,tail,'unequal');
```

```
if((h ~= 1))
```

```
    disp('Mean1 != Mean2')
```

```
else
```

```
    outputt=sprintf('alpha:\t\t\t%.2f',alpha);
```

```
    disp(outputt)
```

```
    outputt=sprintf('T-test result: \th:%d The result h = 1 means that  
you can reject the null hypothesis. ',h);
```

```
    disp(outputt)
```

```
    outputt=sprintf('significance:\t%.10f',significance);
```

```
    disp(outputt)
```

```
    ci;
```

```
    stats
```

```
%
```

```
%
```

```
% Kolmogorov-Smirnov test to compare the distribution of two  
samples %
```

```
%
```

```
%
```

```

%
%

tail='unequal';

[H,P,KSSTAT] = kstest2(histValsC,histValsP,alpha,tail);

if(H==1)
    KSSTAT
    disp('Kolmogorov-Smirnov test result: the hypothesis that the
distributions are the same is rejected.')
else
    disp('Kolmogorov-Smirnov test result: the hypothesis that the
distributions are the same is not rejected')
end

cdfplot(histValsC)
hold on
cdfplot(histValsP)
h = findobj(gca,'type','line');
set(h(1),'linestyle',':','color','r')

end % end of ELSE:1

```

Ek-3.1 PCA ve K-means Clustering

pca_and_kmeans.m

```
clear all
```

```
clc
```

```
trainData_0=[50,50,20,100,30; 50,50,50,70,30; 50,50,20,30,100;
50,50,50,70,70; 50,50,20,100,70; 60,50,50,30,70; 50,50,50,70,30;
50,50,20,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
60,50,50,70,70; 50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100;
50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70; 50,50,50,70,70;
60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70;50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70; 50,50,50,100,30;
60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,20,100,70; 50,60,20,100,70; 50,50,50,70,100; 50,50,20,70,70;
50,50,50,30,30; 60,60,20,30,30; 50,50,50,70,70; 50,60,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30;
60,50,50,70,70; 50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100;
50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,30; 50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70;
50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30;
50,50,20,70,30; 50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30;
50,50,20,70,100; 50,50,50,70,30; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30; 50,60,50,70,70;
50,50,50,100,100; 50,50,50,30,70; 50,50,50,30,30; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30; 50,50,20,70,100;
50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100;
50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100;
50,50,50,100,70; 50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,30; 60,50,50,70,70];
```

```
trainDataC_0=[50,50,20,100,30; 50,50,50,70,30; 50,50,20,30,100;
50,50,50,70,70; 50,50,20,100,70; 60,50,50,30,70; 50,50,50,70,30;
50,50,20,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
60,50,50,70,70; 50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,100;
```

50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100;
50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70; 50,50,50,70,70;
60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70];

trainDataP_0=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70; 50,50,50,100,30;
60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,20,100,70; 50,60,20,100,70; 50,50,50,70,100; 50,50,20,70,70;
50,50,50,30,30; 60,60,20,30,30; 50,50,50,70,70; 50,60,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30;
60,50,50,70,70; 50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100;
50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,30; 50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70;
50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30;
50,50,20,70,30; 50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30;
50,50,20,70,100; 50,50,50,70,30; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30; 50,60,50,70,70;
50,50,50,100,100; 50,50,50,30,70; 50,50,50,30,30; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30; 50,50,20,70,100;
50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100;
50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100;
50,50,50,100,70; 50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,30; 60,50,50,70,70];

testDataC_0=[50,50,50,70,100; 60,50,20,70,70; 50,50,50,70,100;
50,60,50,70,30; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70];

```
testDataP_0=[50,50,50,70,30; 50,60,50,70,70; 50,50,20,30,100;  
50,50,50,70,100; 50,50,50,70,100; 60,50,50,70,70; 60,50,50,70,70;  
60,50,50,30,70; 50,50,50,70,70; 50,50,20,70,30];
```

```
trainData_1=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;  
50,50,20,30,100; 50,50,50,70,70; 60,50,20,70,70; 60,50,50,30,70;  
50,50,50,70,100; 50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70;  
50,60,50,70,30; 50,50,50,70,100; 50,50,50,70,100; 50,50,50,70,100;  
50,50,50,70,70; 50,50,50,70,100; 50,50,20,70,30; 50,50,50,70,100;  
50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,50,70,70;  
50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70; 50,50,50,70,70;  
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;  
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70;  
50,50,50,70,70; 50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70;  
50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70;  
60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;  
50,50,50,70,70; 50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;  
50,50,50,70,30; 50,50,50,70,70; 60,50,50,70,70; 50,50,50,100,30;  
60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70; 50,50,50,100,70;  
50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70;  
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70;  
50,60,20,100,70; 50,50,20,70,70; 50,50,50,30,30; 60,60,20,30,30;  
50,50,50,70,70; 50,60,50,30,70; 50,50,50,70,100; 50,50,50,70,30;  
50,50,20,100,100; 50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,70;  
50,60,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,100,30;  
60,50,50,70,70; 50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100;  
50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;  
50,50,20,30,100; 50,50,50,70,100; 50,50,50,30,70; 50,50,50,70,100;  
50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30;  
60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,100; 50,50,50,30,30;  
60,50,50,70,30; 50,50,20,70,100; 50,50,50,70,30; 60,50,50,70,70;  
50,50,50,70,70; 60,50,50,70,70; 50,50,50,70,100; 60,50,20,70,100;  
50,50,50,70,30; 60,50,50,70,30; 50,60,20,70,70; 50,50,20,100,30;  
50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70;  
50,50,50,30,30; 60,50,50,70,70; 50,50,50,100,30; 60,50,50,70,30;  
60,50,50,100,30; 50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70;  
50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100; 50,50,50,100,100;  
50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100; 50,50,50,100,70;  
50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100; 50,50,50,70,70;  
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];
```



```
trainDataC_1=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;
50,50,20,30,100; 50,50,50,70,70; 60,50,20,70,70; 60,50,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70;
50,60,50,70,30; 50,50,50,70,100; 50,50,50,70,100; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 50,50,20,70,30; 50,50,50,70,100;
50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,50,70,70;
50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70;
60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70];
```

```
trainDataP_1=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,50,70,70; 60,50,50,70,70; 50,50,50,100,30;
60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70; 50,50,50,100,70;
50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70;
50,60,20,100,70; 50,50,20,70,70; 50,50,50,30,30; 60,60,20,30,30;
50,50,50,70,70; 50,60,50,30,70; 50,50,50,70,100; 50,50,50,70,30;
50,50,20,100,100; 50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,70;
50,60,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,100,30;
60,50,50,70,70; 50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100;
50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,20,30,100; 50,50,50,70,100; 50,50,50,30,70; 50,50,50,70,100;
50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30;
60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,100; 50,50,50,30,30;
60,50,50,70,30; 50,50,20,70,100; 50,50,50,70,30; 60,50,50,70,70;
50,50,50,70,70; 60,50,50,70,70; 50,50,50,70,100; 60,50,20,70,100;
50,50,50,70,30; 60,50,50,70,30; 50,60,20,70,70; 50,50,20,100,30;
50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70;
50,50,50,30,30; 60,50,50,70,70; 50,50,50,100,30; 60,50,50,70,30;
60,50,50,100,30; 50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70;
50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100; 50,50,50,100,100;
50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100; 50,50,50,100,70;
50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];
```

```
testDataC_1=[50,50,20,100,70; 50,50,20,70,70; 60,50,50,70,70;  
50,50,50,100,70; 50,50,50,70,30; 50,50,20,100,70; 50,50,50,70,30;  
50,50,50,30,70; 60,50,50,70,100; 50,50,50,70,70];
```

```
testDataP_1=[50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,100;  
50,50,50,30,100; 50,50,50,70,70; 50,50,50,30,100; 50,50,50,70,30;  
50,50,50,70,100; 60,50,50,70,70; 50,50,20,70,30];
```

```
trainData_2=[50,50,20,100,30; 50,50,50,70,30; 50,50,20,30,100;  
50,50,50,70,70; 50,50,20,100,70; 60,50,50,30,70; 50,50,50,70,100;  
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,100; 50,50,50,70,70;  
50,60,50,70,30; 50,50,50,70,100; 60,50,50,70,70; 50,50,50,70,100;  
50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30;  
50,50,20,70,30; 50,50,50,70,70; 50,50,20,100,70; 50,50,20,100,70;  
50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70;  
50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;  
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70;  
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100;  
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100;  
50,50,50,30,30; 50,50,50,70,70; 60,50,20,70,30; 50,60,50,70,100;  
50,50,50,70,70; 50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;  
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;  
50,50,50,100,30; 60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70;  
50,50,50,70,70; 50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,100;  
60,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70;  
50,50,50,70,70; 50,50,20,100,70; 50,60,20,100,70; 50,50,50,70,100;  
50,50,20,70,70; 60,60,20,30,30; 50,50,50,70,70; 50,60,50,30,70;  
50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100;  
50,50,50,70,70; 50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70;  
50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30;  
50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100; 50,50,20,30,70;  
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30;  
50,50,20,30,100; 50,50,50,70,100; 50,50,50,70,100; 60,50,50,70,70;  
50,50,50,70,100; 50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70;  
60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,100;  
60,50,50,70,30; 50,50,20,70,100; 50,50,50,70,30; 60,50,50,70,70;  
50,50,50,70,70; 60,50,50,70,70; 60,50,20,70,100; 50,50,50,70,30;  
60,50,50,70,30; 50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30;  
50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70;  
50,50,50,30,30; 50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30;
```

50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100;
50,50,50,30,100; 50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70;
50,60,50,30,100; 50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30];

trainDataC_2=[50,50,20,100,30; 50,50,50,70,30; 50,50,20,30,100;
50,50,50,70,70; 50,50,20,100,70; 60,50,50,30,70; 50,50,50,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,100; 50,50,50,70,70;
50,60,50,70,30; 50,50,50,70,100; 60,50,50,70,70; 50,50,50,70,100;
50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30;
50,50,20,70,30; 50,50,50,70,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70;
50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100;
50,50,50,30,30; 50,50,50,70,70; 60,50,20,70,30; 50,60,50,70,100;
50,50,50,70,70];

trainDataP_2=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70;
50,50,50,70,70; 50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,100;
60,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,20,100,70; 50,60,20,100,70; 50,50,50,70,100;
50,50,20,70,70; 60,60,20,30,30; 50,50,50,70,70; 50,60,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100;
50,50,50,70,70; 50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30;
50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100; 50,50,20,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30;
50,50,20,30,100; 50,50,50,70,100; 50,50,50,70,100; 60,50,50,70,70;
50,50,50,70,100; 50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70;
60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,100;
60,50,50,70,30; 50,50,20,70,100; 50,50,50,70,30; 60,50,50,70,70;
50,50,50,70,70; 60,50,50,70,70; 60,50,20,70,100; 50,50,50,70,30;
60,50,50,70,30; 50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30;
50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70;
50,50,50,30,30; 50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100;
50,50,50,30,100; 50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70;

50,60,50,30,100; 50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30];

testDataC_2=[50,50,50,70,100; 60,50,20,70,70; 50,50,50,100,70;
50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,20,70,70; 50,50,20,70,70];

testDataP_2=[50,50,50,30,30; 50,50,50,70,30; 60,50,50,70,70;
50,50,50,30,70; 50,50,50,30,30; 50,50,50,70,100; 60,50,50,70,70;
50,50,20,70,100; 50,50,50,100,70; 60,50,50,70,70];

trainData_3=[50,50,20,100,30; 50,50,50,70,30; 50,50,20,30,100;
50,50,50,70,70; 60,50,20,70,70; 50,50,20,100,70; 60,50,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,70; 50,50,50,70,100;
50,50,50,100,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,100; 50,50,50,100,70;
50,50,20,100,70; 50,50,20,100,70; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,70;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70;
50,50,50,70,70; 60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100;
50,50,50,70,70; 50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70;
50,50,50,70,100; 50,50,20,70,70; 50,50,50,30,30; 60,60,20,30,30;
50,50,50,70,70; 50,60,50,30,70; 50,50,50,70,100; 50,50,50,70,30;
50,50,20,100,100; 50,50,50,30,100; 50,50,50,70,70; 50,50,50,70,30;
50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30; 60,50,50,70,70;
50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100; 50,50,20,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,30,100; 50,50,50,70,100;
50,50,50,70,100; 60,50,50,70,70; 50,50,50,70,100; 50,60,50,70,70;
50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30; 60,50,50,70,70;
50,50,20,70,30; 50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30;
50,50,20,70,100; 50,50,50,70,30; 60,50,50,70,70; 60,50,50,70,70;

50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,60,20,70,70; 50,50,20,100,30; 50,60,50,70,70; 50,50,50,100,100;
60,50,50,30,70; 50,50,50,30,70; 50,50,50,30,30; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30; 50,50,20,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100;
50,50,50,30,100; 50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70;
50,50,50,100,70; 50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];

trainDataC_3=[50,50,20,100,30; 50,50,50,70,30; 50,50,20,30,100;
50,50,50,70,70; 60,50,20,70,70; 50,50,20,100,70; 60,50,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,70; 50,50,50,70,100;
50,50,50,100,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,100; 50,50,50,100,70;
50,50,20,100,70; 50,50,20,100,70; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,70;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70;
50,50,50,70,70; 60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100;
50,50,50,70,70];

trainDataP_3=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70;
50,50,50,70,100; 50,50,20,70,70; 50,50,50,30,30; 60,60,20,30,30;
50,50,50,70,70; 50,60,50,30,70; 50,50,50,70,100; 50,50,50,70,30;
50,50,20,100,100; 50,50,50,30,100; 50,50,50,70,70; 50,50,50,70,30;
50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30; 60,50,50,70,70;
50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100; 50,50,20,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,30,100; 50,50,50,70,100;
50,50,50,70,100; 60,50,50,70,70; 50,50,50,70,100; 50,60,50,70,70;
50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30; 60,50,50,70,70;
50,50,20,70,30; 50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30;
50,50,20,70,100; 50,50,50,70,30; 60,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,60,20,70,70; 50,50,20,100,30; 50,60,50,70,70; 50,50,50,100,100;

60,50,50,30,70; 50,50,50,30,70; 50,50,50,30,30; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30; 50,50,20,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100;
50,50,50,30,100; 50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70;
50,50,50,100,70; 50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];

testDataC_3=[50,50,50,70,100; 50,60,50,70,30; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30; 50,50,50,70,30;
50,50,50,70,100; 50,50,50,70,70; 50,50,20,70,70];

testDataP_3=[60,50,50,100,30; 50,50,50,70,70; 50,60,20,100,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,30,70; 50,50,50,70,70;
50,50,20,70,30; 50,60,50,30,100; 50,50,50,70,70];

trainData_4=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;
50,50,50,70,70; 60,50,20,70,70; 50,50,50,70,100; 50,50,20,70,70;
50,50,50,70,100; 50,50,50,70,70; 50,60,50,70,30; 50,50,50,70,100;
60,50,50,70,70; 50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,30,30; 50,50,50,30,30; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,70;
60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,50,100,30; 60,50,50,100,30; 60,50,50,70,70;
50,60,50,70,70; 50,50,50,70,70; 50,50,50,100,70; 50,50,50,70,70;
50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70; 50,60,20,100,70;
50,50,50,70,100; 50,50,20,70,70; 50,50,50,30,30; 60,60,20,30,30;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100;
50,50,50,30,100; 50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,100;
50,50,50,100,30; 60,50,50,70,70; 50,50,50,70,70; 80,50,50,70,100;
50,60,50,70,100; 50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,30,100; 50,50,50,70,100;

50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70; 50,50,50,70,100;
50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30;
60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,30; 50,50,20,70,100;
50,50,50,70,30; 60,50,50,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30; 50,60,50,70,70;
50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70; 50,50,50,30,30;
60,50,50,70,70; 50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30;
50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100;
50,50,20,30,100; 50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70;
50,60,50,30,100; 50,50,50,100,70; 50,50,50,100,70; 50,50,20,70,70;
50,50,50,100,100; 50,50,50,70,30; 50,50,50,70,30; 60,50,50,70,70];

trainDataC_4=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;
50,50,50,70,70; 60,50,20,70,70; 50,50,50,70,100; 50,50,20,70,70;
50,50,50,70,100; 50,50,50,70,70; 50,60,50,70,30; 50,50,50,70,100;
60,50,50,70,70; 50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,30,30; 50,50,50,30,30; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,70;
60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70];

trainDataP_4=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,50,100,30; 60,50,50,100,30; 60,50,50,70,70;
50,60,50,70,70; 50,50,50,70,70; 50,50,50,100,70; 50,50,50,70,70;
50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70; 50,60,20,100,70;
50,50,50,70,100; 50,50,20,70,70; 50,50,50,30,30; 60,60,20,30,30;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100;
50,50,50,30,100; 50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,100;
50,50,50,100,30; 60,50,50,70,70; 50,50,50,70,70; 80,50,50,70,100;
50,60,50,70,100; 50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,30,100; 50,50,50,70,100;
50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70; 50,50,50,70,100;
50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30;

60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,30; 50,50,20,70,100;
50,50,50,70,30; 60,50,50,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30; 50,60,50,70,70;
50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70; 50,50,50,30,30;
60,50,50,70,70; 50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30;
50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100;
50,50,20,30,100; 50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70;
50,60,50,30,100; 50,50,50,100,70; 50,50,50,100,70; 50,50,20,70,70;
50,50,50,100,100; 50,50,50,70,30; 50,50,50,70,30; 60,50,50,70,70];

testDataC_4=[50,50,20,30,100; 50,50,20,100,70; 60,50,50,30,70;
50,50,50,70,30; 50,50,50,70,70; 50,50,50,70,70; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,30,70; 50,50,50,30,30];

testDataP_4=[50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,60,50,30,70; 50,60,50,70,70; 50,50,50,30,100; 60,50,50,70,30;
50,50,50,30,100; 50,50,50,70,70; 50,50,20,70,30];

trainData_5=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;
50,50,20,30,100; 50,50,50,70,70; 60,50,20,70,70; 50,50,20,100,70;
60,50,50,30,70; 50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70;
50,60,50,70,30; 60,50,50,70,70; 50,50,50,70,100; 50,50,50,100,70;
50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,20,70,70;
50,50,50,30,70; 50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30;
50,50,50,70,70; 60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100;
50,50,20,70,70; 50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,100,30; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,100,70; 50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,50,70,100;
50,50,20,70,70; 50,50,50,30,30; 50,60,50,30,70; 50,50,50,70,100;
50,50,20,100,100; 50,50,50,30,100; 50,50,50,70,70; 50,50,50,70,30;
50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70; 50,50,50,70,70;

50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30; 60,50,50,70,70;
50,50,50,70,70; 80,50,50,70,100; 50,50,20,30,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,30,100; 50,50,50,70,100;
50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70; 50,50,50,70,100;
50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30;
60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,100; 50,50,50,30,30;
60,50,50,70,30; 50,50,20,70,100; 50,50,50,70,30; 60,50,50,70,70;
50,50,50,70,70; 60,50,50,70,70; 50,50,50,70,100; 60,50,20,70,100;
60,50,50,70,30; 50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30;
50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70;
50,50,50,30,30; 60,50,50,70,70; 50,50,50,100,30; 60,50,50,70,30;
60,50,50,100,30; 50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70;
50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100; 50,50,50,100,100;
50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100; 50,50,50,100,70;
50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];

trainDataC_5=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;
50,50,20,30,100; 50,50,50,70,70; 60,50,20,70,70; 50,50,20,100,70;
60,50,50,30,70; 50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70;
50,60,50,70,30; 60,50,50,70,70; 50,50,50,70,100; 50,50,50,100,70;
50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70; 50,50,20,70,70;
50,50,50,30,70; 50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30;
50,50,50,70,70; 60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100;
50,50,20,70,70];

trainDataP_5=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,100,30; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,100,70; 50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,50,70,100;
50,50,20,70,70; 50,50,50,30,30; 50,60,50,30,70; 50,50,50,70,100;
50,50,20,100,100; 50,50,50,30,100; 50,50,50,70,70; 50,50,50,70,30;
50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30; 60,50,50,70,70;
50,50,50,70,70; 80,50,50,70,100; 50,50,20,30,70; 50,50,50,70,70;

50,50,50,70,70; 50,50,50,70,30; 50,50,20,30,100; 50,50,50,70,100;
50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70; 50,50,50,70,100;
50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70; 60,50,50,70,30;
60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,100; 50,50,50,30,30;
60,50,50,70,30; 50,50,20,70,100; 50,50,50,70,30; 60,50,50,70,70;
50,50,50,70,70; 60,50,50,70,70; 50,50,50,70,100; 60,50,20,70,100;
60,50,50,70,30; 50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30;
50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70;
50,50,50,30,30; 60,50,50,70,70; 50,50,50,100,30; 60,50,50,70,30;
60,50,50,100,30; 50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70;
50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100; 50,50,50,100,100;
50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100; 50,50,50,100,70;
50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];

testDataC_5=[50,50,50,70,100; 50,50,50,70,100; 50,50,50,70,100;
50,50,50,70,100; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,70;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70];

testDataP_5=[60,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,60,20,100,70; 60,60,20,30,30; 50,50,50,70,70; 50,50,50,70,30;
50,60,50,70,100; 50,50,50,70,70; 50,50,50,70,30];

trainData_6=[50,50,50,70,100; 50,50,20,30,100; 50,50,50,70,70;
60,50,20,70,70; 50,50,20,100,70; 50,50,50,70,30; 50,50,20,70,70;
50,50,50,70,100; 50,50,50,70,70; 50,60,50,70,30; 50,50,50,70,100;
60,50,50,70,70; 50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,100;
50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,70,70; 50,50,50,30,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70;
50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30;
50,50,50,70,70; 60,50,20,70,30; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70;
50,50,50,70,70; 50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,100;

60,50,50,70,30; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,20,100,70; 50,60,20,100,70; 50,50,50,70,100; 50,50,20,70,70;
50,50,50,30,30; 60,60,20,30,30; 50,50,50,70,70; 50,60,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100;
50,50,50,70,70; 50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,100,30; 60,50,50,70,70;
50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100; 50,50,20,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30;
50,50,20,30,100; 50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70;
50,50,50,70,100; 50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70;
60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,100;
50,50,50,30,30; 60,50,50,70,30; 50,50,20,70,100; 50,50,50,70,30;
60,50,50,70,70; 60,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30;
60,50,50,70,30; 50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30;
50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70;
50,50,50,30,30; 60,50,50,70,70; 50,50,50,100,30; 60,50,50,70,30;
50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100;
50,50,20,30,100; 50,50,50,30,100; 50,60,50,70,30; 50,50,50,70,70;
50,60,50,30,100; 50,50,50,100,70; 50,50,50,100,70; 50,50,50,100,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,30; 60,50,50,70,70];

trainDataC_6=[50,50,50,70,100; 50,50,20,30,100; 50,50,50,70,70;
60,50,20,70,70; 50,50,20,100,70; 50,50,50,70,30; 50,50,20,70,70;
50,50,50,70,100; 50,50,50,70,70; 50,60,50,70,30; 50,50,50,70,100;
60,50,50,70,70; 50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,100;
50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,70,70; 50,50,50,30,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70;
50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30;
50,50,50,70,70; 60,50,20,70,30; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70];

trainDataP_6=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70;
50,50,50,70,70; 50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,100;
60,50,50,70,30; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,20,100,70; 50,60,20,100,70; 50,50,50,70,100; 50,50,20,70,70;

50,50,50,30,30; 60,60,20,30,30; 50,50,50,70,70; 50,60,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100;
50,50,50,70,70; 50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,100,30; 60,50,50,70,70;
50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100; 50,50,20,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30;
50,50,20,30,100; 50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70;
50,50,50,70,100; 50,60,50,70,70; 50,60,50,100,70; 50,50,20,70,70;
60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30; 50,50,50,30,100;
50,50,50,30,30; 60,50,50,70,30; 50,50,20,70,100; 50,50,50,70,30;
60,50,50,70,70; 60,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30;
60,50,50,70,30; 50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30;
50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,70;
50,50,50,30,30; 60,50,50,70,70; 50,50,50,100,30; 60,50,50,70,30;
50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100;
50,50,20,30,100; 50,50,50,30,100; 50,60,50,70,30; 50,50,50,70,70;
50,60,50,30,100; 50,50,50,100,70; 50,50,50,100,70; 50,50,50,100,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,30; 60,50,50,70,70];

testDataC_6=[50,50,20,100,30; 50,50,50,70,30; 60,50,50,30,70;
50,50,50,70,100; 50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,70,70];

testDataP_6=[50,50,20,70,70; 50,50,50,70,30; 50,50,50,30,100;
50,50,50,70,100; 50,50,50,70,70; 60,50,20,70,100; 60,50,50,100,30;
50,50,50,100,100; 50,50,20,70,70; 50,50,20,70,30];

trainData_7=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;
50,50,50,70,70; 50,50,20,100,70; 60,50,50,30,70; 50,50,50,70,100;
50,50,20,70,70; 50,50,50,70,70; 50,60,50,70,30; 50,50,50,70,100;
60,50,50,70,70; 50,50,50,70,100; 50,50,50,70,100; 50,50,50,70,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,100;
50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,20,100,70;
50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70; 50,50,50,70,70;
60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;

50,50,50,70,70;50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
60,50,50,100,30; 50,60,50,70,70; 50,50,50,100,70; 50,50,50,70,70;
50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,50,70,100;
50,50,20,70,70; 50,50,50,30,30; 60,60,20,30,30; 50,50,50,70,70;
50,60,50,30,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100;
50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30; 60,50,50,70,70;
50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100; 50,50,20,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30;
50,50,20,30,100; 50,50,50,70,100; 50,50,50,70,100; 50,50,50,30,70;
60,50,50,70,70; 50,50,50,70,100; 50,60,50,70,70; 50,60,50,100,70;
50,50,20,70,70; 60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30;
50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30; 50,50,20,70,100;
50,50,50,70,30; 60,50,50,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30; 50,60,50,70,70;
50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,30; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30; 50,50,20,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100;
50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100;
50,50,50,100,70; 50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,70,30; 60,50,50,70,70];

```
trainDataC_7=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;  
50,50,50,70,70; 50,50,20,100,70; 60,50,50,30,70; 50,50,50,70,100;  
50,50,20,70,70; 50,50,50,70,70; 50,60,50,70,30; 50,50,50,70,100;  
60,50,50,70,70; 50,50,50,70,100; 50,50,50,70,100; 50,50,50,70,70;  
50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,100;  
50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,20,100,70;  
50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70; 50,50,50,70,70;  
50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;  
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70;  
50,50,50,70,70; 50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,70;  
50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70; 50,50,50,70,70;  
60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;  
50,50,50,70,70];
```

```
trainDataP_7=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;  
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70; 60,50,50,70,70;
```

60,50,50,100,30; 50,60,50,70,70; 50,50,50,100,70; 50,50,50,70,70;
50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70; 50,50,50,70,100;
50,50,20,70,70; 50,50,50,30,30; 60,60,20,30,30; 50,50,50,70,70;
50,60,50,30,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100;
50,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,30,100; 50,50,50,100,30; 60,50,50,70,70;
50,50,50,70,70; 80,50,50,70,100; 50,60,50,70,100; 50,50,20,30,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30;
50,50,20,30,100; 50,50,50,70,100; 50,50,50,70,100; 50,50,50,30,70;
60,50,50,70,70; 50,50,50,70,100; 50,60,50,70,70; 50,60,50,100,70;
50,50,20,70,70; 60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30;
50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30; 50,50,20,70,100;
50,50,50,70,30; 60,50,50,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30; 50,60,50,70,70;
50,50,50,100,100; 60,50,50,30,70; 50,50,50,30,30; 60,50,50,70,70;
50,50,50,100,30; 60,50,50,70,30; 60,50,50,100,30; 50,50,20,70,100;
50,50,50,70,70; 50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100;
50,50,50,100,100; 50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100;
50,50,50,100,70; 50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,70,30; 60,50,50,70,70];

testDataC_7=[50,50,20,30,100; 60,50,20,70,70; 50,50,50,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,30,70; 60,50,50,70,100; 50,50,50,70,70];

testDataP_7=[50,50,50,100,30; 60,50,50,70,70; 50,50,50,70,70;
50,60,20,100,70; 50,50,50,30,100; 50,50,50,70,70; 50,50,50,70,30;
50,50,50,30,70; 50,50,50,30,100; 50,50,50,70,30];

trainData_8=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;
50,50,20,30,100; 60,50,20,70,70; 50,50,20,100,70; 60,50,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,100;
50,50,50,100,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70;
50,50,20,100,70; 50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30;
50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70;

50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30;
50,50,50,70,70; 60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100;
50,50,50,70,70; 50,50,50,100,30; 50,50,20,70,100; 50,50,20,70,70;
50,50,50,70,70; 60,50,50,70,70; 50,50,50,100,30; 60,50,50,100,30;
60,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70; 50,50,50,100,70;
50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70; 50,60,20,100,70;
50,50,50,70,100; 50,50,20,70,70; 50,50,50,70,70; 50,60,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,70; 50,60,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,100;
50,50,50,100,30; 60,50,50,70,70; 50,50,50,70,70; 80,50,50,70,100;
50,60,50,70,100; 50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,30,100; 50,50,50,70,100;
50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70; 50,60,50,70,70;
50,60,50,100,70; 60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30;
50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30; 50,50,20,70,100;
50,50,50,70,30; 60,50,50,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30; 60,50,50,30,70;
50,50,50,30,70; 50,50,50,30,30; 60,50,50,70,70; 50,50,50,100,30;
60,50,50,70,30; 60,50,50,100,30; 50,50,20,70,100; 50,50,50,70,70;
50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100;
50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100; 50,50,50,100,70;
50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];

trainDataC_8=[50,50,20,100,30; 50,50,50,70,30; 50,50,50,70,100;
50,50,20,30,100; 60,50,20,70,70; 50,50,20,100,70; 60,50,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,100;
50,50,50,100,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,70; 50,50,20,100,70;
50,50,20,100,70; 50,50,50,70,70; 50,50,50,30,30; 50,50,50,30,30;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,30;
50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70;
50,50,20,70,70; 50,50,50,70,70; 50,50,50,30,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30;

50,50,50,70,70; 60,50,20,70,30; 50,50,20,70,70; 50,60,50,70,100;
50,50,50,70,70];

trainDataP_8=[50,50,50,100,30; 50,50,20,70,100; 50,50,20,70,70;
50,50,50,70,70; 60,50,50,70,70; 50,50,50,100,30; 60,50,50,100,30;
60,50,50,70,70; 50,60,50,70,70; 50,50,50,70,70; 50,50,50,100,70;
50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70; 50,60,20,100,70;
50,50,50,70,100; 50,50,20,70,70; 50,50,50,70,70; 50,60,50,30,70;
50,50,50,70,100; 50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,70; 50,60,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,100;
50,50,50,100,30; 60,50,50,70,70; 50,50,50,70,70; 80,50,50,70,100;
50,60,50,70,100; 50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,20,30,100; 50,50,50,70,100;
50,50,50,70,100; 50,50,50,30,70; 60,50,50,70,70; 50,60,50,70,70;
50,60,50,100,70; 60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30;
50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30; 50,50,20,70,100;
50,50,50,70,30; 60,50,50,70,70; 50,50,50,70,70; 60,50,50,70,70;
50,50,50,70,100; 60,50,20,70,100; 50,50,50,70,30; 60,50,50,70,30;
50,50,20,70,30; 50,60,20,70,70; 50,50,20,100,30; 60,50,50,30,70;
50,50,50,30,70; 50,50,50,30,30; 60,50,50,70,70; 50,50,50,100,30;
60,50,50,70,30; 60,50,50,100,30; 50,50,20,70,100; 50,50,50,70,70;
50,50,20,70,70; 50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100;
50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100; 50,50,50,100,70;
50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];

testDataC_8=[50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,70;
50,60,50,70,30; 60,50,50,70,70; 50,50,50,70,100; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,70,70];

testDataP_8=[50,50,50,30,70; 50,50,50,70,30; 50,50,50,70,70;
50,50,50,30,30; 60,60,20,30,30; 50,50,50,70,100; 50,50,20,70,70;
50,60,50,70,70; 50,50,50,100,100; 50,50,50,100,100];

trainData_9=[50,50,50,70,30; 50,50,50,70,100; 50,50,20,30,100;
50,50,50,70,70; 60,50,20,70,70; 50,50,20,100,70; 60,50,50,30,70;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,100; 50,50,50,70,70;

50,50,50,70,100; 60,50,50,70,70; 50,50,50,100,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,70; 50,50,50,70,70;
50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70;
50,50,50,70,70; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,50,70,70; 60,50,50,70,70; 50,50,50,100,30;
60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70; 50,50,50,100,70;
50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70;
50,60,20,100,70; 50,50,50,70,100; 50,50,20,70,70; 50,50,50,30,30;
60,60,20,30,30; 50,50,50,70,70; 50,60,50,30,70; 50,50,50,70,100;
50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,30,100; 50,50,50,100,30; 60,50,50,70,70; 80,50,50,70,100;
50,60,50,70,100; 50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,100;
50,50,50,30,70; 60,50,50,70,70; 50,50,50,70,100; 50,60,50,100,70;
50,50,20,70,70; 60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30;
50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30; 50,50,20,70,100;
50,50,50,70,30; 60,50,50,70,70; 50,50,50,70,70; 60,50,50,70,70;
60,50,20,70,100; 60,50,50,70,30; 50,50,20,70,30; 50,60,20,70,70;
50,50,20,100,30; 50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70;
50,50,50,30,70; 50,50,50,30,30; 50,50,50,100,30; 60,50,50,70,30;
60,50,50,100,30; 50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70;
50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100; 50,50,50,100,100;
50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100; 50,50,50,100,70;
50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];

trainDataC_9=[50,50,50,70,30; 50,50,50,70,100; 50,50,20,30,100;
50,50,50,70,70; 60,50,20,70,70; 50,50,20,100,70; 60,50,50,30,70;
50,50,50,70,30; 50,50,20,70,70; 50,50,50,70,100; 50,50,50,70,70;
50,50,50,70,100; 60,50,50,70,70; 50,50,50,100,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,70,30; 50,50,20,70,30;
50,50,50,70,100; 50,50,50,100,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,30,30; 50,50,50,30,30; 50,50,20,70,70; 50,50,50,70,70;

50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,70; 50,50,50,70,100;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,30,70; 50,50,50,70,70;
50,50,50,70,70; 60,50,50,70,100; 50,50,20,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,100; 50,50,50,30,30; 50,50,50,70,70;
50,50,50,70,70; 50,50,20,70,70; 50,60,50,70,100; 50,50,20,70,70;
50,50,50,70,70];

trainDataP_9=[50,50,50,100,30; 50,50,50,30,70; 50,50,20,70,100;
50,50,50,70,30; 50,50,50,70,70; 60,50,50,70,70; 50,50,50,100,30;
60,50,50,100,30; 60,50,50,70,70; 50,60,50,70,70; 50,50,50,100,70;
50,50,50,70,70; 50,50,50,70,100; 60,50,50,70,30; 50,50,20,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70; 50,50,20,100,70;
50,60,20,100,70; 50,50,50,70,100; 50,50,20,70,70; 50,50,50,30,30;
60,60,20,30,30; 50,50,50,70,70; 50,60,50,30,70; 50,50,50,70,100;
50,50,50,70,30; 50,50,20,100,100; 50,50,50,30,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,50,70,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,30,100; 50,50,50,100,30; 60,50,50,70,70; 80,50,50,70,100;
50,60,50,70,100; 50,50,20,30,70; 50,50,50,70,70; 50,50,50,70,70;
50,50,50,70,70; 50,50,50,70,30; 50,50,50,70,100; 50,50,50,70,100;
50,50,50,30,70; 60,50,50,70,70; 50,50,50,70,100; 50,60,50,100,70;
50,50,20,70,70; 60,50,50,70,30; 60,50,50,70,70; 50,50,20,70,30;
50,50,50,30,100; 50,50,50,30,30; 60,50,50,70,30; 50,50,20,70,100;
50,50,50,70,30; 60,50,50,70,70; 50,50,50,70,70; 60,50,50,70,70;
60,50,20,70,100; 60,50,50,70,30; 50,50,20,70,30; 50,60,20,70,70;
50,50,20,100,30; 50,60,50,70,70; 50,50,50,100,100; 60,50,50,30,70;
50,50,50,30,70; 50,50,50,30,30; 50,50,50,100,30; 60,50,50,70,30;
60,50,50,100,30; 50,50,20,70,100; 50,50,50,70,70; 50,50,20,70,70;
50,50,50,70,100; 50,50,20,30,100; 50,50,50,30,100; 50,50,50,100,100;
50,60,50,70,30; 50,50,50,70,70; 50,60,50,30,100; 50,50,50,100,70;
50,50,50,100,70; 50,50,20,70,70; 50,50,50,100,100; 50,50,50,70,70;
50,50,50,70,30; 50,50,20,70,30; 50,50,50,70,30; 60,50,50,70,70];

testDataC_9=[50,50,20,100,30; 50,50,50,70,100; 50,60,50,70,30;
50,50,50,70,100; 50,50,20,100,70; 50,50,20,100,70; 50,50,50,70,70;
50,50,20,70,70; 50,50,50,70,70; 60,50,20,70,30];

testDataP_9=[50,50,20,70,70; 50,50,50,70,70; 50,60,50,70,70;
50,50,50,70,70; 50,50,50,70,70; 50,50,20,30,100; 50,60,50,70,70;
50,50,50,70,100; 50,50,50,70,30; 60,50,50,70,70];

```

[pc, zscores, pcvars] = princomp(trainData_0);
cumsum(pcvars./sum(pcvars) * 100)

res=pca(trainData_0,2);
res
figure
plot(res.W(1:48,1),res.W(1:48,2), 'ro')
hold;
plot(res.W(48:143,1),res.W(48:143,2), 'x')

%{

%figure
%plot3(res.W(1:50,1),res.W(1:50,2),res.W(1:50,3), 'o')
%hold;
%plot3(res.W(51:162,1),res.W(51:162,2),res.W(51:162,3), 'x')
%}

disp('C1: Centroid Of Group Patients')
disp('C2: Centroid Of Group Controls')
disp('-----')
disp(' ')

disp('-----Train Data 0-----')

disp(' ')

[C] = kmeansx(trainData_0,2);

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----+-----C2-----')

```

```

for i=1:1:10

    tst_X=testDataC_0(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+')
    end
    if(ED1>ED2)

        disp('-----+-----X----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X----')
    end

end

disp(' ')

for i=1:1:10

    tst_X=testDataP_0(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+')
    end
    if(ED1>ED2)

        disp('-----+-----X----')
    end
    if(ED1==ED2)

```

```

        disp('-----X-----+-----X-----')
    end

end

C

disp('-----End of Test-----')

disp(' ')

%%-----
--
%%-----
--

disp('-----Train Data 1-----')

disp(' ')

[C] = kmeansx(trainData_1,2);

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----+-----C2-----')
disp(' ')

for i=1:1:10

    tst_X=testDataC_1(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+-----')

```

```

end
if(ED1>ED2)

    disp('-----+-----X----')
end
if(ED1==ED2)

    disp('-----X-----+-----X----')
end

end

disp(' ')

for i=1:1:10

    tst_X=testDataP_1(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+-----')
    end
    if(ED1>ED2)

        disp('-----+-----X----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X----')
    end

end

end

C

disp('-----End of Test-----')

disp(' ')

```

```

%%-----
--
%%-----
--

disp('-----Train Data 2-----')

disp(' ')

[C] = kmeansx(trainData_2,2);

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----+-----C2-----')
disp(' ')

for i=1:1:10

    tst_X=testDataC_2(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+-----')
    end
    if(ED1>ED2)

        disp('-----+-----X-----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X-----')
    end

end

end

```

```

disp(' ')

for i=1:1:10

    tst_X=testDataP_2(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+')
    end
    if(ED1>ED2)

        disp('-----+-----X----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X----')
    end

end

C

disp('-----End of Test-----')

disp(' ')
%%-----
--
%%-----
--

disp('-----Train Data 3-----')

disp(' ')

[C] = kmeansx(trainData_3,2);

```



```

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----+-----C2-----')
disp(' ')

for i=1:1:10

    tst_X=testDataC_3(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if (ED1<ED2)

        disp('-----X-----+-----')
    end
    if (ED1>ED2)

        disp('-----+-----X-----')
    end
    if (ED1==ED2)

        disp('-----X-----+-----X-----')
    end
end

disp(' ')

for i=1:1:10

    tst_X=testDataP_3(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if (ED1<ED2)

        disp('-----X-----+-----')

```

```

end
if(ED1>ED2)

    disp('-----+-----X----')
end
if(ED1==ED2)

    disp('-----X-----+-----X----')
end

end

C

disp('-----End of Test-----')

disp(' ')
%%-----
--
%%-----
--

disp('-----Train Data 4-----')

disp(' ')

[C] = kmeansx(trainData_4,2);

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----+-----C2-----')
disp(' ')

for i=1:1:10

    tst_X=testDataC_4(i,:);

    ED1 = norm(x1 - tst_X);

```

```

ED2 = norm(x2 - tst_X);
if(ED1<ED2)

    disp('-----X-----+')
end
if(ED1>ED2)

    disp('-----+-----X----')
end
if(ED1==ED2)

    disp('-----X-----+-----X----')
end

end

disp(' ')

for i=1:1:10

    tst_X=testDataP_4(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+')
    end
    if(ED1>ED2)

        disp('-----+-----X----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X----')
    end

end

```

C

```

disp('-----End of Test-----')

disp(' ')
%%-----
--
%%-----
--

disp('-----Train Data 5-----')

disp(' ')

[C] = kmeansx(trainData_5,2);

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----+-----C2-----')
disp(' ')

for i=1:1:10

    tst_X=testDataC_5(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+-----')
    end
    if(ED1>ED2)

        disp('-----+-----X-----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X-----')
    end
end

```

```

end

disp(' ')

for i=1:1:10

    tst_X=testDataP_5(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+')
    end
    if(ED1>ED2)

        disp('-----+-----X----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X----')
    end

end

end

C

disp('-----End of Test-----')

disp(' ')
%%-----
--
%%-----
--

disp('-----Train Data 6-----')

disp(' ')

```

```

[C] = kmeansx(trainData_6,2);

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----+-----C2-----')
disp(' ')

for i=1:1:10

    tst_X=testDataC_6(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+-----')
    end
    if(ED1>ED2)

        disp('-----+-----X-----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X-----')
    end
end

disp(' ')

for i=1:1:10

    tst_X=testDataP_6(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

```

```

        disp('-----X-----+')
    end
    if(ED1>ED2)

        disp('-----+-----X-----')
    end
    if(ED1==ED2)

        disp('-----X-----X-----')
    end

end

C

disp('-----End of Test-----')

disp(' ')
%%-----
--
%%-----
--

disp('-----Train Data 7-----')

disp(' ')

[C] = kmeansx(trainData_7,2);

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----C2-----')
disp(' ')

for i=1:1:10

    tst_X=testDataC_7(i,:);

```

```

ED1 = norm(x1 - tst_X);

ED2 = norm(x2 - tst_X);
if (ED1<ED2)

    disp('-----X-----+')
end
if (ED1>ED2)

    disp('-----+-----X----')
end
if (ED1==ED2)

    disp('-----X-----+-----X----')
end

end

disp(' ')

for i=1:1:10

    tst_X=testDataP_7(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
if (ED1<ED2)

    disp('-----X-----+')
end
if (ED1>ED2)

    disp('-----+-----X----')
end
if (ED1==ED2)

    disp('-----X-----+-----X----')
end

end

```


C

```
disp('-----End of Test-----')
```

```
disp(' ')
```

```
%%-----  
--  
%%-----  
--
```

```
disp('-----Train Data 8-----')
```

```
disp(' ')
```

```
[C] = kmeansx(trainData_8,2);
```

```
x1=C(1,:);
```

```
x2=C(2,:);
```

```
disp('-----C1-----+-----C2-----')
```

```
disp(' ')
```

```
for i=1:1:10
```

```
    tst_X=testDataC_8(i,:);
```

```
    ED1 = norm(x1 - tst_X);
```

```
    ED2 = norm(x2 - tst_X);
```

```
    if(ED1<ED2)
```

```
        disp('-----X-----+-----')
```

```
    end
```

```
    if(ED1>ED2)
```

```
        disp('-----+-----X-----')
```

```
    end
```

```
    if(ED1==ED2)
```

```

        disp('-----X-----+-----X-----')
    end

end

disp(' ')

for i=1:1:10

    tst_X=testDataP_8(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if (ED1<ED2)

        disp('-----X-----+-----')
        end
        if (ED1>ED2)

            disp('-----+-----X-----')
            end
            if (ED1==ED2)

                disp('-----X-----+-----X-----')
                end
            end

end

C

disp('-----End of Test-----')

disp(' ')
%%-----
--
%%-----
--

disp('-----Train Data 9-----')

```

```

disp(' ')

[C] = kmeansx(trainData_9,2);

x1=C(1,:);
x2=C(2,:);

disp('-----C1-----+-----C2-----')
disp(' ')

for i=1:1:10

    tst_X=testDataC_9(i,:);

    ED1 = norm(x1 - tst_X);

    ED2 = norm(x2 - tst_X);
    if(ED1<ED2)

        disp('-----X-----+-----')
    end
    if(ED1>ED2)

        disp('-----+-----X-----')
    end
    if(ED1==ED2)

        disp('-----X-----+-----X-----')
    end

end

disp(' ')

for i=1:1:10

    tst_X=testDataP_9(i,:);

    ED1 = norm(x1 - tst_X);

```

```

ED2 = norm(x2 - tst_X);
if(ED1<ED2)

    disp('-----X-----+')
end
if(ED1>ED2)

    disp('-----+-----X----')
end
if(ED1==ED2)

    disp('-----X-----X----')
end

end

C

disp('-----End of Test-----')

disp(' ')
%%-----
--
%%-----
--

```