

Akış

- Makine Öğrenmesi nedir?
- Günlük Hayatımızdaki Uygulamaları
- Verilerin Sayısallaştırılması
- Özellik Belirleme
 - Özellik Seçim Metotları
 - Bilgi Kazancı (Informaiton Gain-IG)
 - Sinyal-Gürültüye Oranı: (S2N ratio)
 - Alt küme seçiciler (Wrappers)
 - Yeni Özelliklerin Çıkarımı
 - Temel Bileşen Analizi (Principal Component Analysis)
 - Doğrusal Ayırteden Analizi (Linear Discriminant Analysis)
- Sınıflandırma Metotları
 - Doğrusal Regresyon
 - Karar Ağaçları (Decision Trees)
 - Yapay Sinir Ağları
 - En Yakın K Komşu Algoritması (k - Nearest Neighbor)
- Kümeleme Algoritmaları:
 - K-Ortalama (K-Means)
 - Kendi Kendini Düzenleyen Haritalar (Self Organizing Map -SOM)



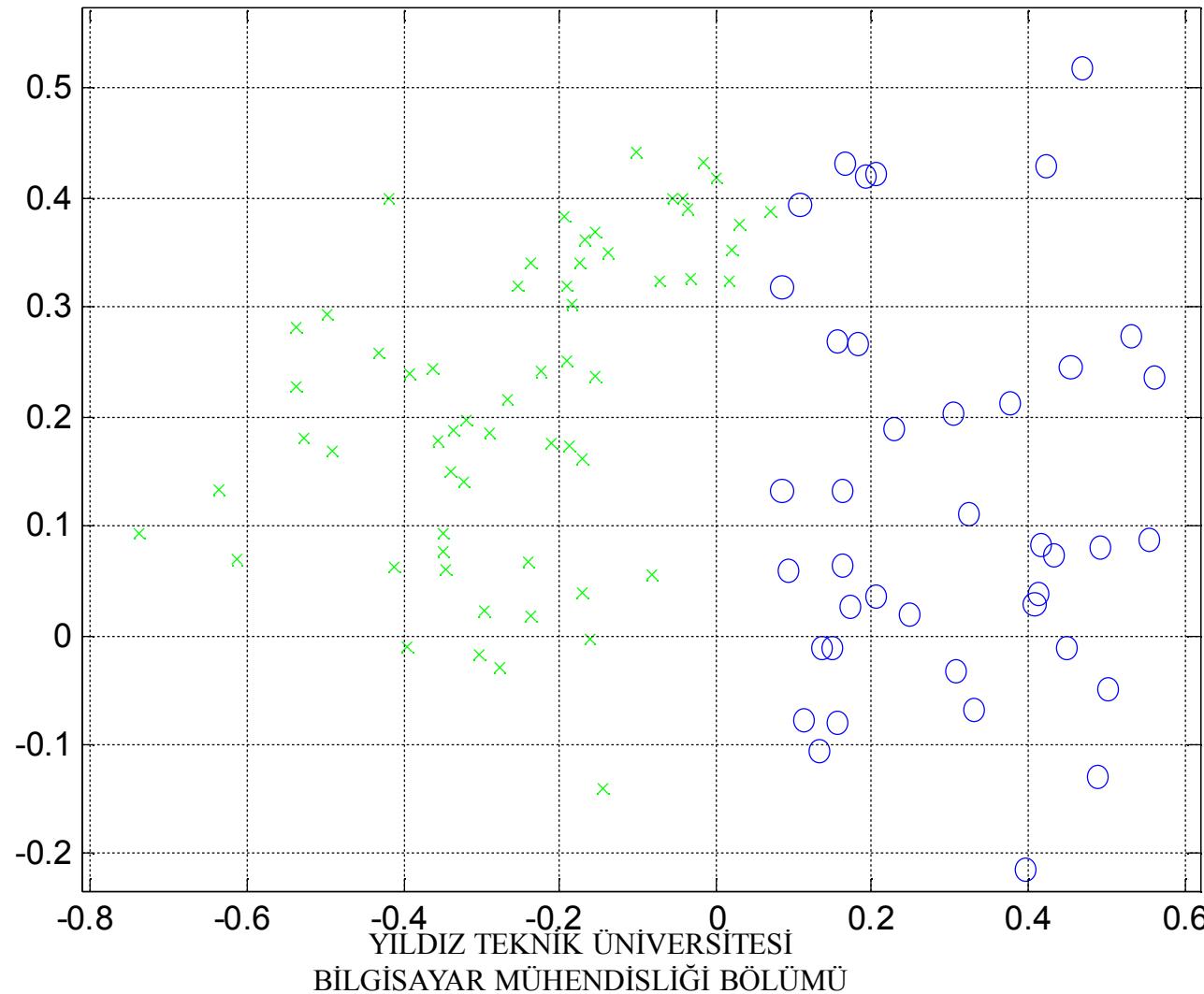
Özellik Belirleme

- Bir doktor
- Veri: Kişi bilgilerini içeren dosyalar
- Görev: Kimler hasta bul.
- Hangi bilgilere bakılır?
 - Ad soyad
 - Doğum yeri
 - Cinsiyet
 - Kan tahlili sonuçları
 - Röntgen sonuçları
 - VS.

1. Özellik	2. Özellik	Sınıf
1	3	A
2	3	B
1	4	A
2	3	B

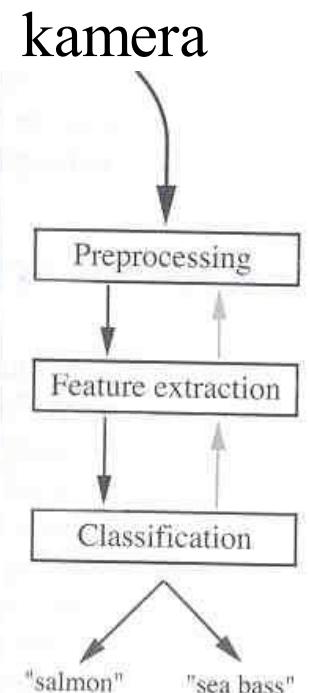
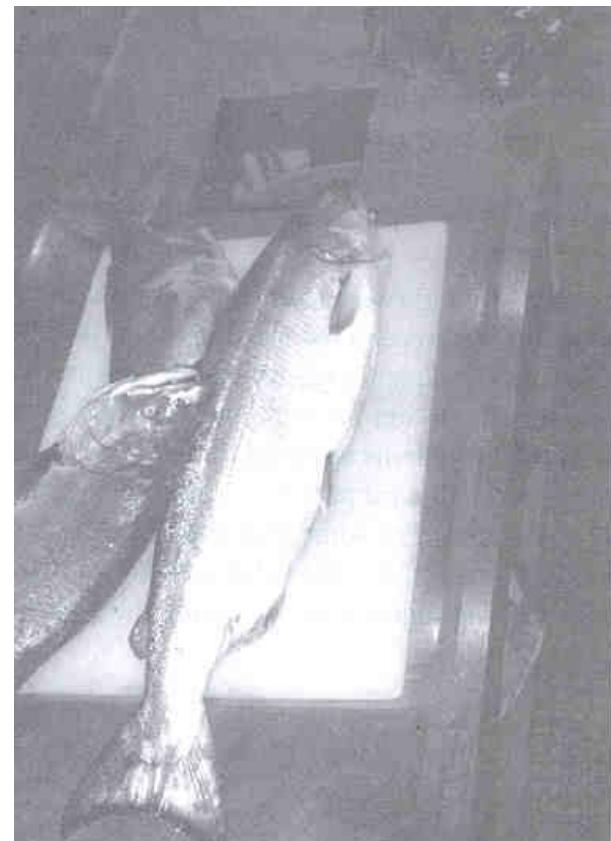


Hangi boyut?



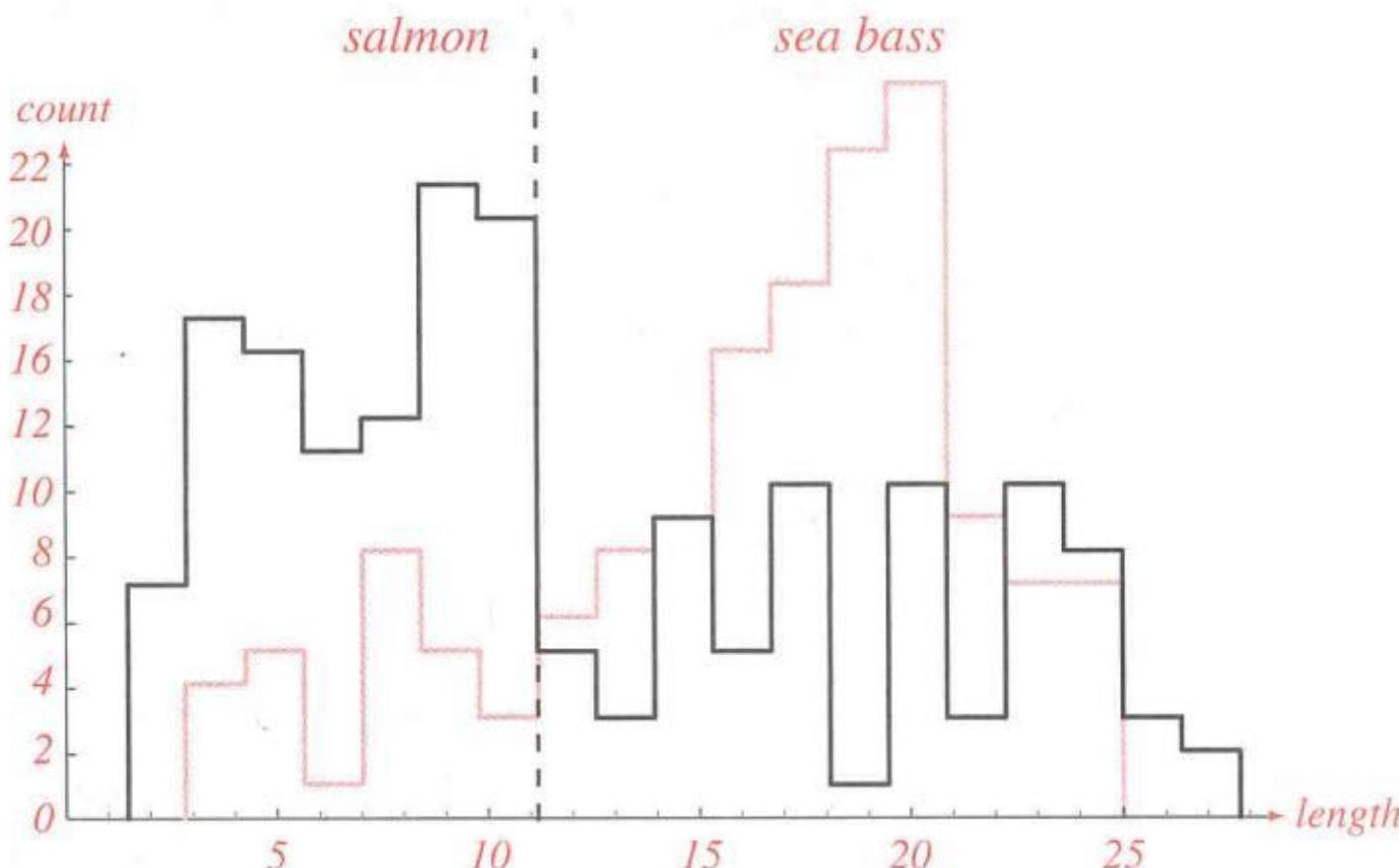
Balık Hali

- Kayan bant üzerindeki balığın türünü belirlemek(Salmon? Sea Bass?)



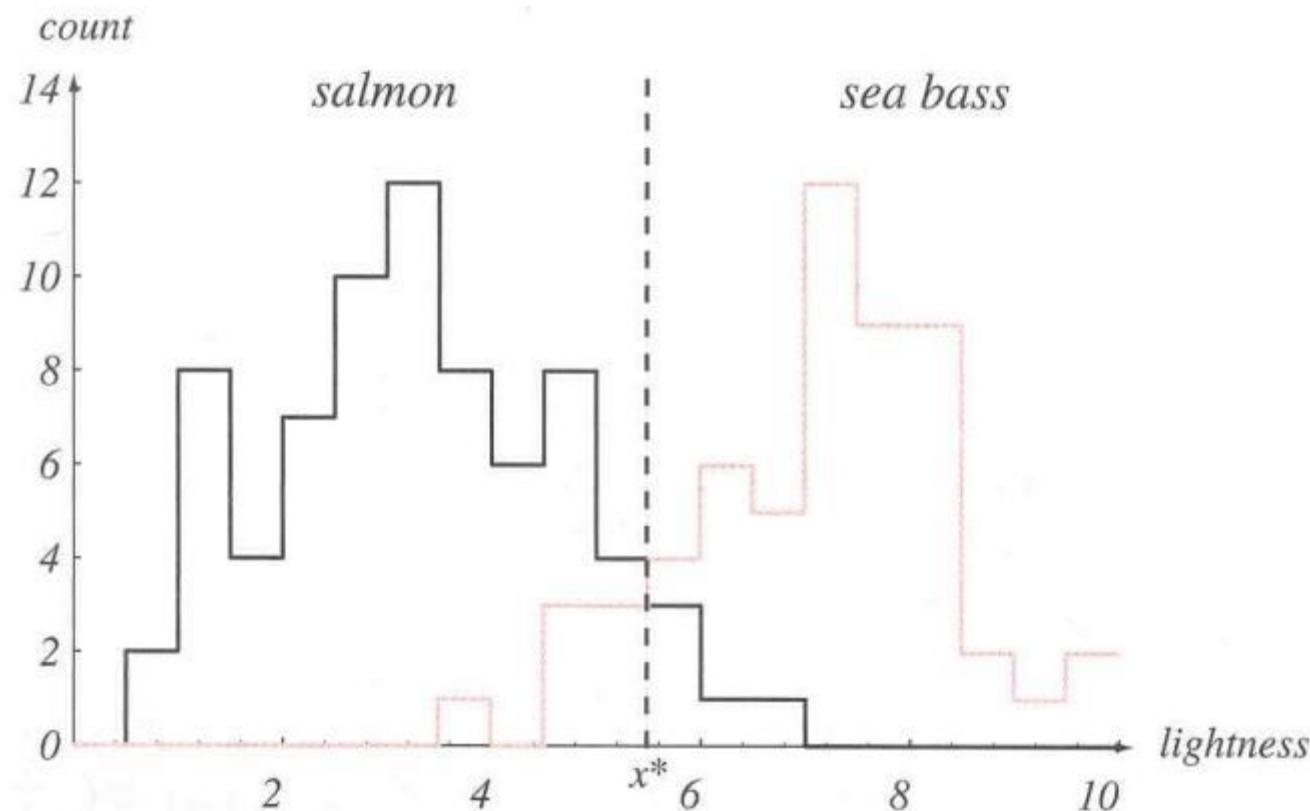
Balık Özellikleri. Uzunluk.

- Salmon lar genelde Sez Bass lardan daha kısalar.

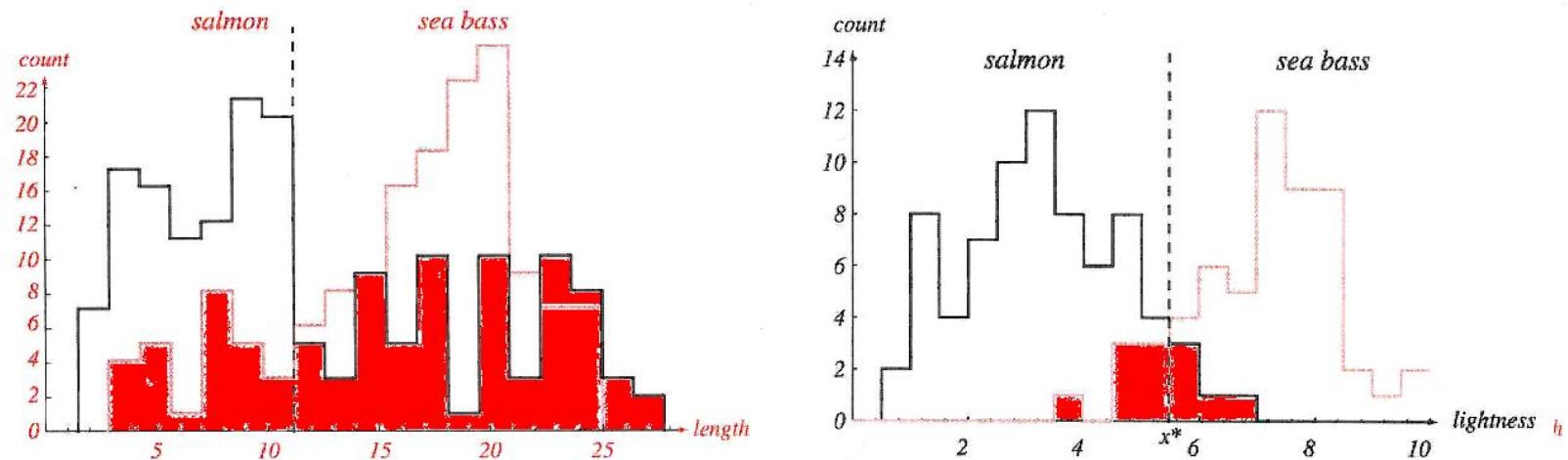


Balık Özellikleri. Parlaklık.

- Sea Bass genelde Salmon lardan daha parlaklar.



Hangi Özellik?



Kırmızı bölgeler yapılan hataları gösteriyor.

Doktoru yoralım 😊

- Hastalık dosyasında 5000 adet özellik olsaydı? Örneğin kişinin DNA dizisine bakarak hasta olup olmadığına karar verecek olsaydık ne yapardık? Nerelere bakacağımıza nasıl karar verirdik.
- Burada devreye makineleri sokmamız gerekiyor gibi gözükme.
- Bu olay bir insanın hesap yapma kabiliyetiyle, bir hesap makinesinininkini karşılaştırmaya benziyor.



Özellik seçimi

- Bu problem makinelerle iki farklı metotla çözülebilir.
 - Var olan özelliklerden bazılarını seçmek
 - Özellikleri tek tek değerlendirmek (Filter)
 - Özellik alt kümeleri oluşturup, sınıflandırıcılar kullanıp performanslarını ölçüp, bu alt kümeleri en iyilemek için değiştirerek (Wrapper)
 - Var olan özelliklerin lineer birleşimlerinden yeni özelliklerin çıkarımı



Özellikleri birer birer inceleme (Filters)

- Eğitim bilgilerindeki her bir özellik teker teker ele alınır.
- Örnek ile ilgili sadece o özellik elimizde olsaydı ne olurdu sorusunun cevabı bulunmaya çalışılır.
- Seçilen özellikle sınıf ya da sonucun birlikte değişimleri incelenir.
- Özelliğin değiştiğinde sınıf ya da sonuç ne kadar değişiyorsa o özelliğin sonuca o kadar etkisi vardır denilir.



Bilgi Kazancı - Information Gain

S eğitim seti içindeki A özelliğinin

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

N kavramının c farklı değeri varsa N'in entropisi,
N'in aldığı her değerin olasılıkları kullanılarak

$$\text{Entropy}(N) = \sum_{i=1}^c - p_i \log_2 p_i$$



daha önceki hava, nem, rüzgar, su sıcaklığını gibi değerlere göre pikniğe gidip gitmemeye kararları verilmiş 4 olay

Olay No	Hava	Nem	Rüzgar	Su sıcaklığı	Pikniğe gidildi mi?
1	güneşli	normal	güçlü	ılık	Evet
2	güneşli	yüksek	güçlü	ılık	Evet
3	yağmurlu	yüksek	güçlü	ılık	Hayır
4	güneşli	yüksek	güçlü	soguk	Evet



Her bir özelliğin piknik kavramı için bilgi kazancını bulalım

- Pikniğe gidildi mi? sorusunun iki cevabı vardır.
- Evet cevabının olasılığı $\frac{3}{4}$
- Hayır cevabının olasılığı $\frac{1}{4}$
- Dolayısıyla Pikniğin Entropi'si
- $E(\text{Piknik}) = -(3/4) \log_2(3/4) - (1/4) \log_2(1/4)$
 $= 0.811$ olarak bulunur.



- $\text{Gain}(\text{Piknik}, \text{Hava}) = 0.811 - \frac{(3/4)}{(1/4)} \left(\frac{-(3/3)}{0} \log_2 \frac{(3/3)}{0} - \frac{0}{1} \right) - \frac{(1/4)}{(1/4)} \left(\frac{0 - (1/1)}{1} \log_2 \frac{(1/1)}{1} \right) = 0.811$
- Hava özelliğinin IG'si hesaplanırken bulunan rakamların açıklamaları:

$0.811 \rightarrow$ Pikniğe gitme olayının Entropisi

$(3/4) \rightarrow$ havanın güneşli olma oranı

$(3/3) \rightarrow$ hava güneşli iken pikniğe gidilme oranı

$0 \rightarrow$ hava güneşli iken pikniğe gidilmeme oranı

$(1/4) \rightarrow$ havanın yağmurlu olma oranı

$0 \rightarrow$ hava yağmurlu iken pikniğe gidilme oranı

$(1/1) \rightarrow$ hava yağmurlu iken pikniğe gidilmeme oranı



- **Gain(Piknik,Nem)**= $0.811 - \left(\frac{1}{4} \right) \left(-(1/1) \log_2 (1/1) - 0 \right) - \left(\frac{3}{4} \right) \left(-(2/3) \log_2 (2/3) - (1/3) \log_2 (1/3) \right)$
 $= 0.811 - 0.688 = \mathbf{0.1225}$
- **Gain(Piknik,Rüzgar)**= $0.811 - \left(\frac{4}{4} \right) \left(-(3/4) \log_2 (3/4) - (1/4) \log_2 (1/4) \right)$
 $= 0.811 - 0.811 = \mathbf{0}$
- **Gain(Piknik,SuSıcaklığı)**= $0.811 - \left(\frac{3}{4} \right) \left(-(2/3) \log_2 (2/3) - (1/3) \log_2 (1/3) \right) - \left(\frac{1}{4} \right) \left(-(1/1) \log_2 (1/1) \right)$
 $= 0.811 - 0.688 = \mathbf{0.1225}$
- En büyük bilgi kazancına sahip özellik ‘Hava’dır.
- Gerçek uygulamalarda ise yüzlerce özelliğin bilgi kazançları hesaplanır ve en büyük olanları seçilerek kullanılır.

S2N

- Sınıflar arası ayrılıkların fazla sınıf içi ayrılıkların az olan özellikler seçilir.

$$S_i = \frac{m_1 - m_2}{d_1 - d_2}$$

$m_1 \rightarrow$ sınıf1'deki i. özelliklerin ortalaması

$m_2 \rightarrow$ sınıf2'deki i. özelliklerin ortalaması

$d_1 \rightarrow$ sınıf1'deki i. özelliklerin standart sapması

$d_2 \rightarrow$ sınıf2'deki i. özelliklerin standart sapması

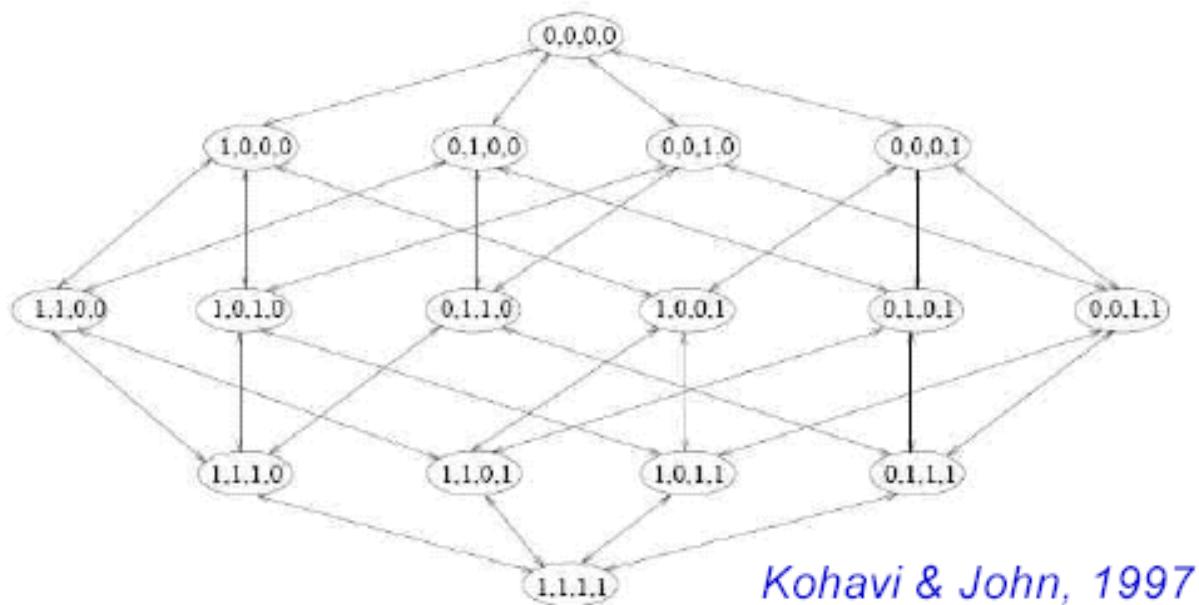
S değeri en yüksek olan özellikler

seçilerek sınıflandırmada kullanılırlar.



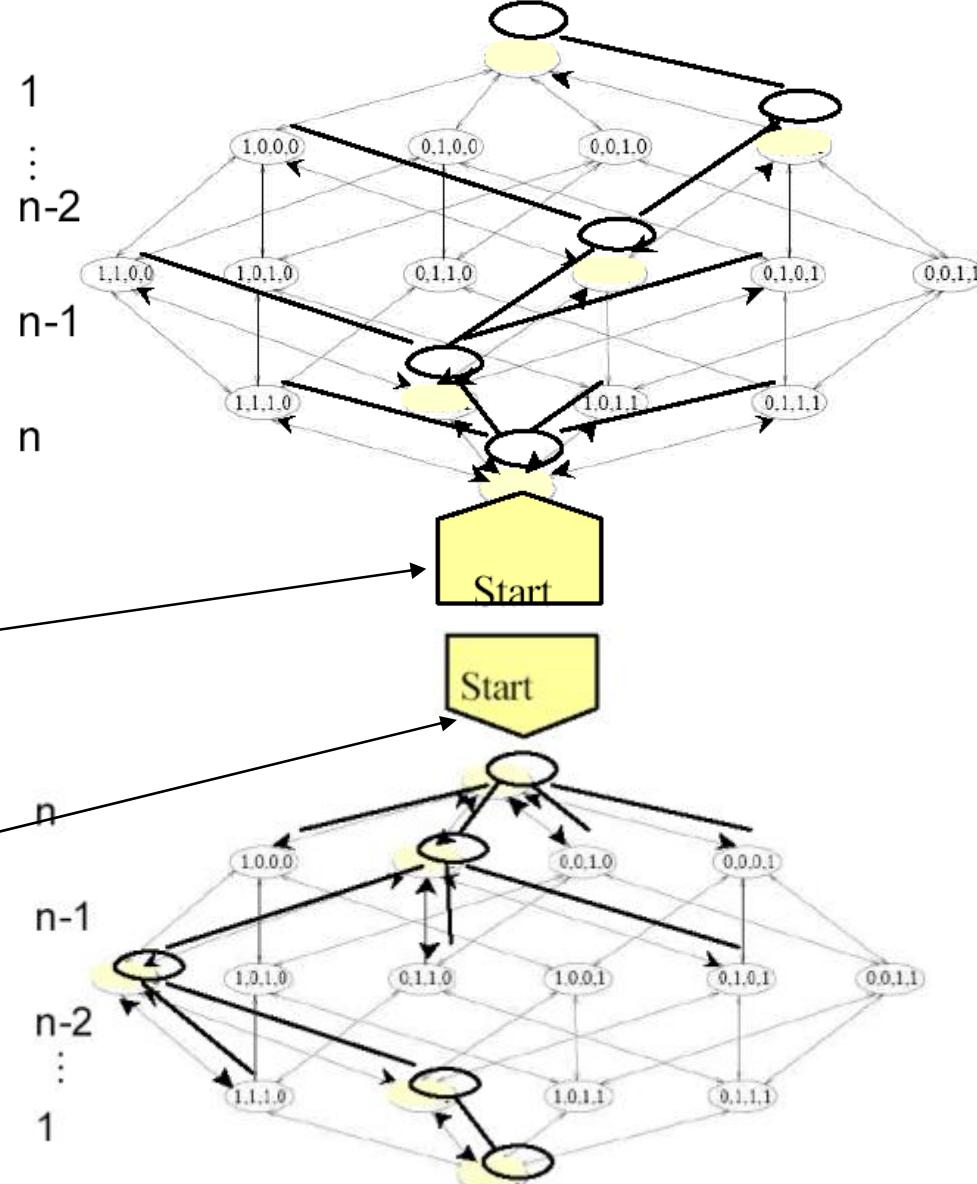
Özellik altkümesi seçiciler (Wrappers)

N özellik için olası 2^N özellik alt kümesi = 2^N eğitim



Özellik altkümesi seçiciler

- *Hızlandırmak için tüm olasılıkları denemek yerine*
 - Hepsiyle başlayıp her seferinde bir tane elemek
 - Tek özellikle başlayıp her seferinde bir tane eklemek



Hangi yoldan gidileceğine o özellik kümelerinin sınıflandırmadaki performansına bakılarak karar verilir.

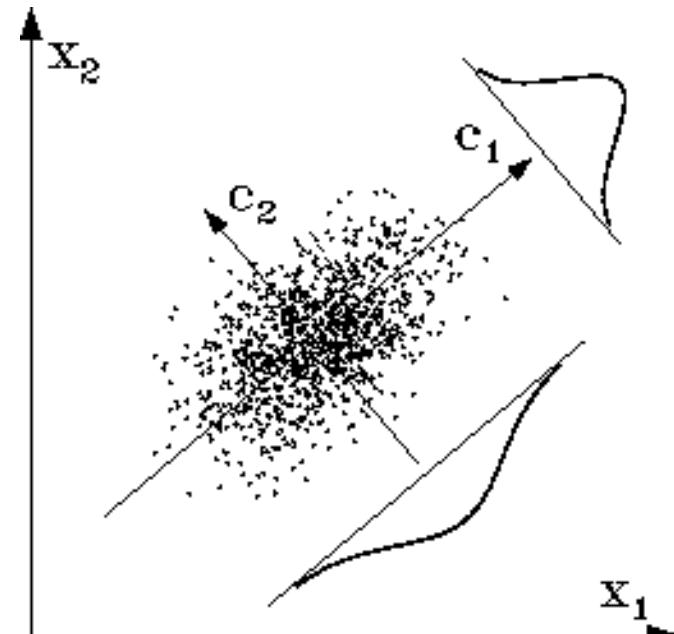
Yeni Özelliklerin Çıkarımı

- Var olan özelliklerin lineer birleşimlerinden yeni bir özellik uzayı oluşturular ve veriler bu uzayda ifade edilirler. Yaygın olarak kullanılan 2 metot vardır.
- PCA
- LDA



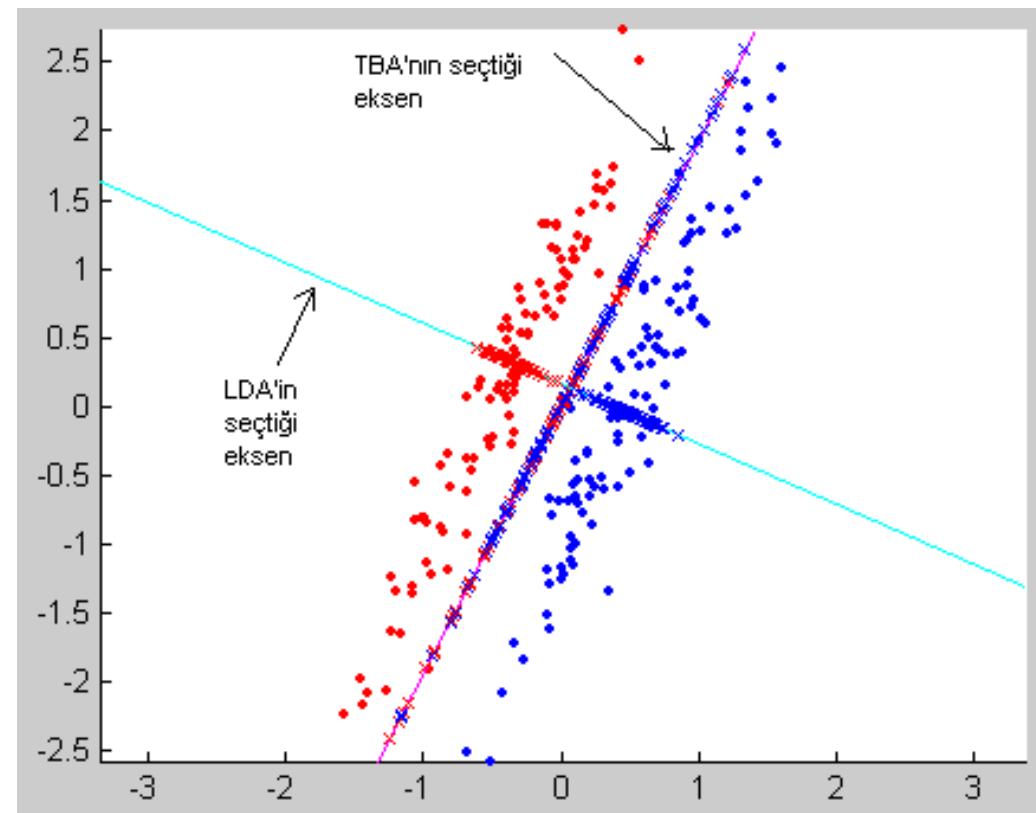
Temel Bileşen Analizi-TBA (Principle Component Analysis - PCA)

- Bu metotta örneklerin en fazla değişim gösterdikleri boyutlar bulunur. Yansa veriler c1 ve c2 eksenlerine izdüşümü yapıldığındaki dağılımları gösterilmiştir.
- C1 eksenindeki değişim daha büyüktür. Dolayısıyla veriler 2 boyuttan bir boyuta C1 eksenine iz düşürülerek indirgenmiş olur.



Doğrusal Ayırteden Analizi (Linear Discriminant Analysis - LDA)

Yandaki gibi durumlar için LDA önerilmiştir. LDA varyanslara ek olarak sınıf bilgisini de kullanarak boyut indirgene yapar. Sadece varyansa değil sınıflandırılmameye de bakar.



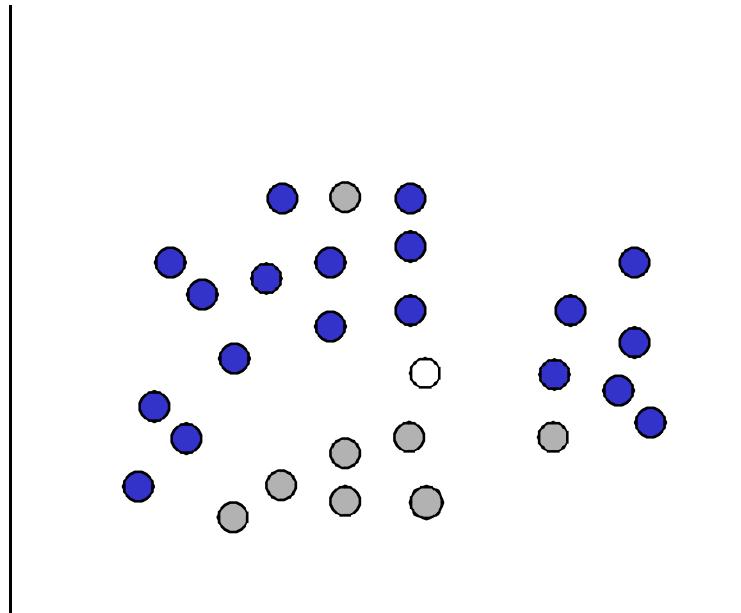
Hangisi

- Niye bu kadar çok metot var?
- Ne zaman hangisini kullanacağımız?



Sınıflandırma Metotları

Görev: Önceden etiketlenmiş örnekleri kullanarak
yeni örneklerin sınıflarını bulmak

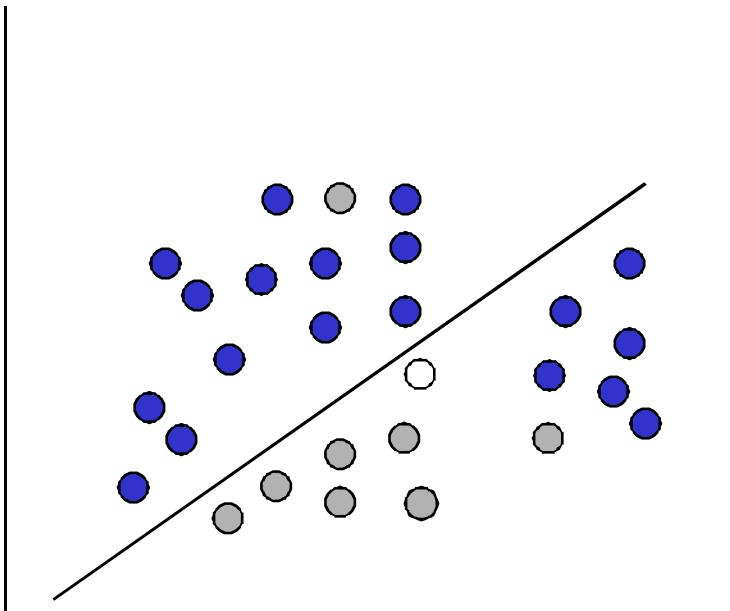


Metotlar:

Doğrusal Sınıflandırma,
Karar Ağaçları,
LVQ,
Yapay Sinir Ağları,
...

Mavi ve gri sınıfından örnekler • ○
Beyaz, mavi mi gri mi? ○

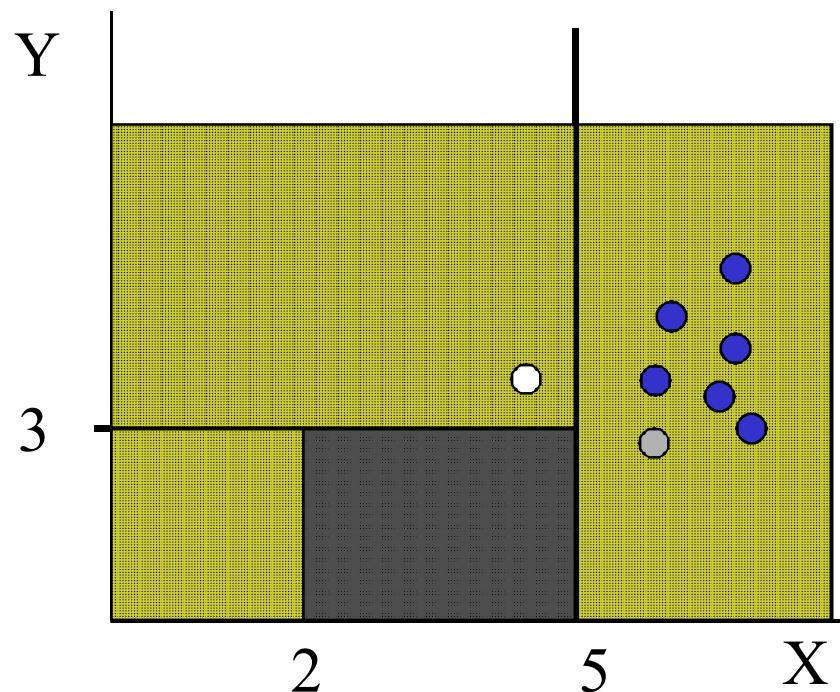
Doğrusal Sınıflandırma



- $w_0 + w_1 x + w_2 y \geq 0$
- En az hata yapan w_i leri bulmaya çalışır.
- Basit bir model
- Yeterince esnek değil

Karar Ağaçları

Böl ve yönet stratejisi



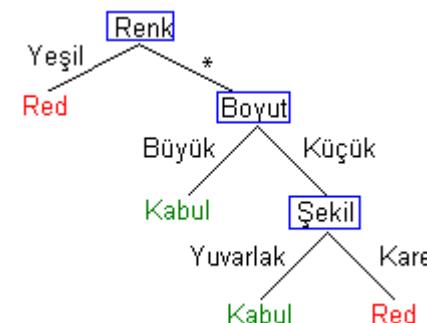
Nasıl böleceğiz?

```
if X > 5 then blue  
else if Y > 3 then blue  
else if X > 2 then green  
else blue
```

Karar Ağaçları

- Ürettikleri kurallar anlaşılır.
- Karar düğümleri ve yapraklardan oluşan hiyerarşik bir yapı.

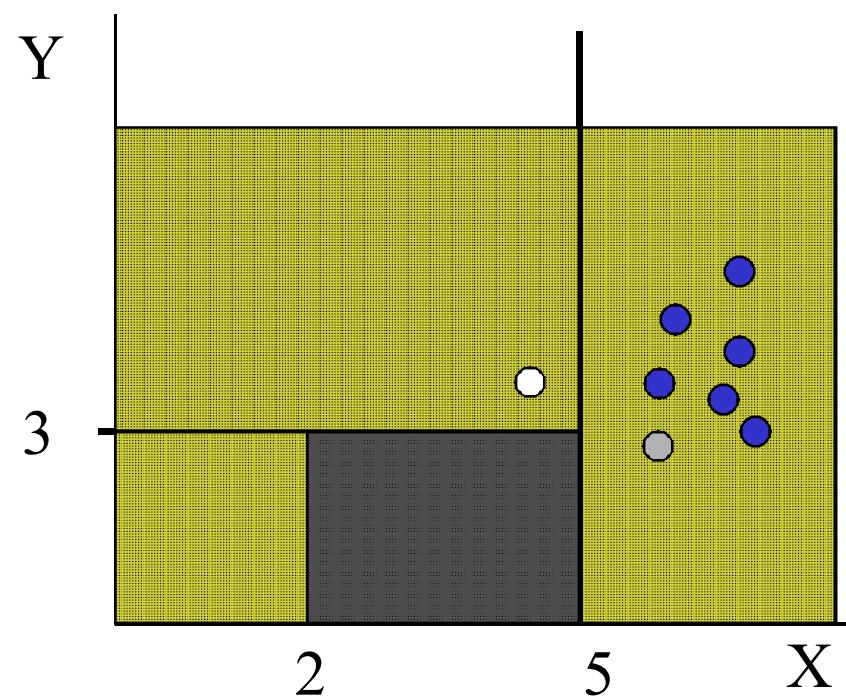
<u>Şekil</u>	<u>Renk</u>	<u>Boyun</u>	<u>Sınıf</u>
Yuvarlak	Yeşil	Küçük	Red
Kare	Siyah	Büyük	Kabul
Kare	Sarı	Büyük	Kabul
Yuvarlak	Sarı	Küçük	Red
Kare	Yeşil	Büyük	Red
Kare	Sarı	Küçük	Kabul



Karar Ağaçları Oluşturma

- Tüm veri kümesiyle başla.
- Bir özelliğin bir değerlerine göre veri kümesi iki alt kümeye böl. Bölmede kullanılan özellikler ve değerleri karar düğüme yerleştir.
- Her alt küme için aynı prosedür her alt kümede sadece tek bir sınıfı ait örnekler kalıncaya kadar uygula.





Karar Düğümleri Nasıl Bulunur?

- Karar düğümlerinde yer alan özelliğin ve eşik değerinin belirlenmesinde genel olarak **entropi** kavramı kullanılır.
- Eğitim verisi her bir özelliğin her bir değeri için ikiye bölünür. Oluşan iki alt kümenin entropileri toplanır. En düşük entropi toplamına sahip olan özellik ve değeri karar düğümüne yerleştirilir.

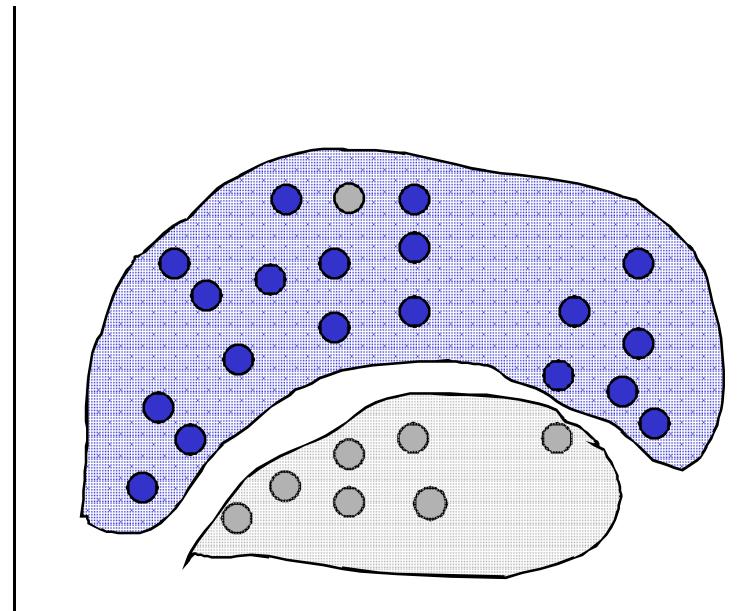


Karar Ağacılarıyla Sınıflandırma

- En tepedeki kök karar düğümünden başla.
- Bir yaprağa gelinceye kadar karar düğümlerindeki yönlendirmelere göre dallarda ilerle. (Karar düğümlerinde tek bir özelliğin adı ve bir eşik değeri yer alır. O düşüme gelen verinin hangi dala gideceğine verinin o düğümdeki özelliğinin eşik değerinden büyük ya da küçük olmasına göre karar verilir.)
- Verinin sınıfı, yaprağın temsil ettiği sınıf olarak belirle.

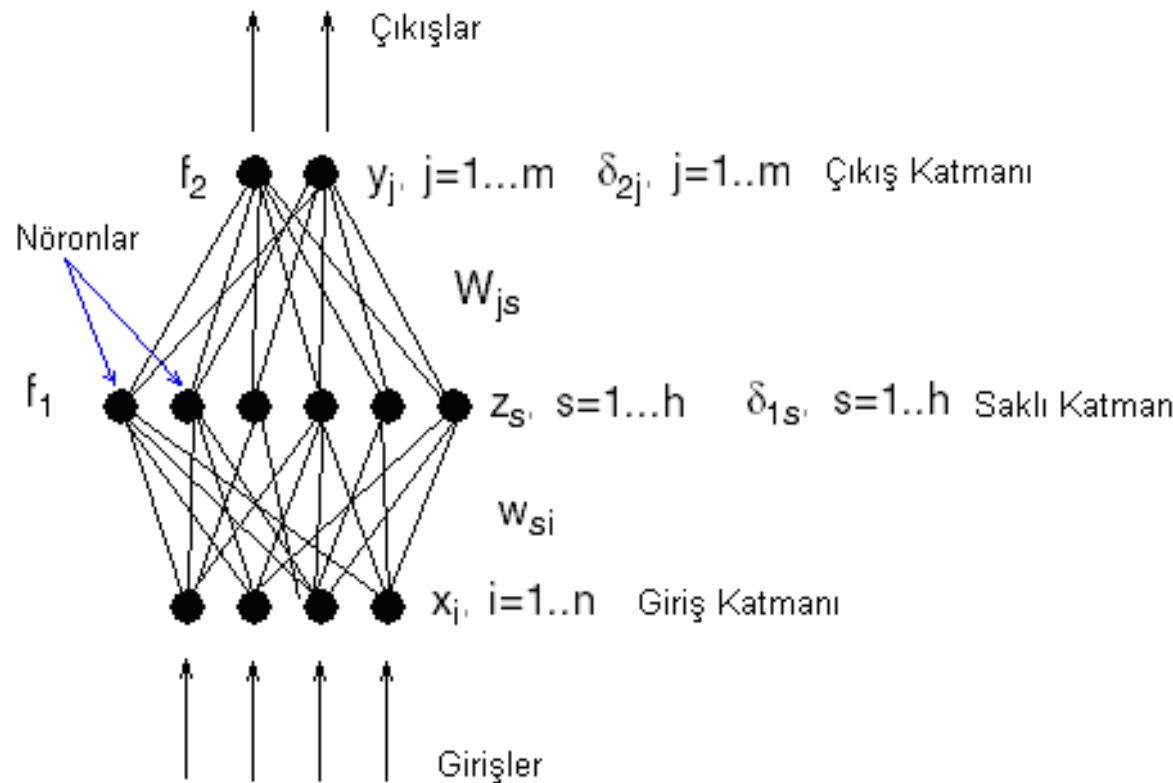


Yapay Sinir Ağları



- Daha kompleks karar sınırları üretebilirler.
- Can be more accurate
- Also can overfit the data – find patterns in random noise

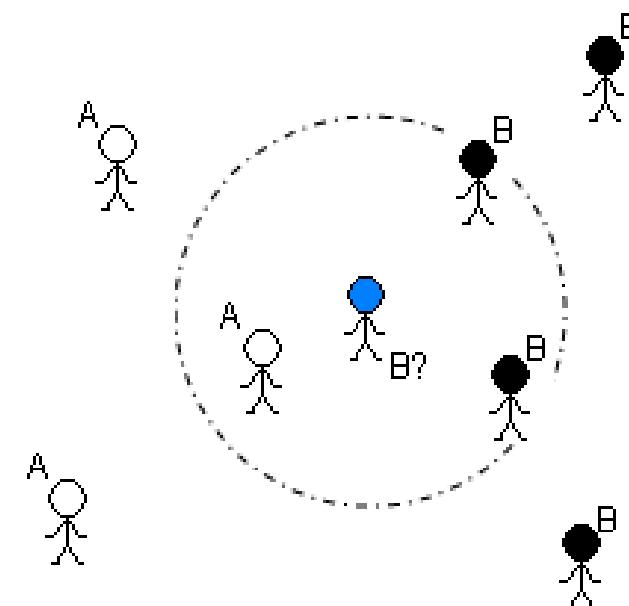
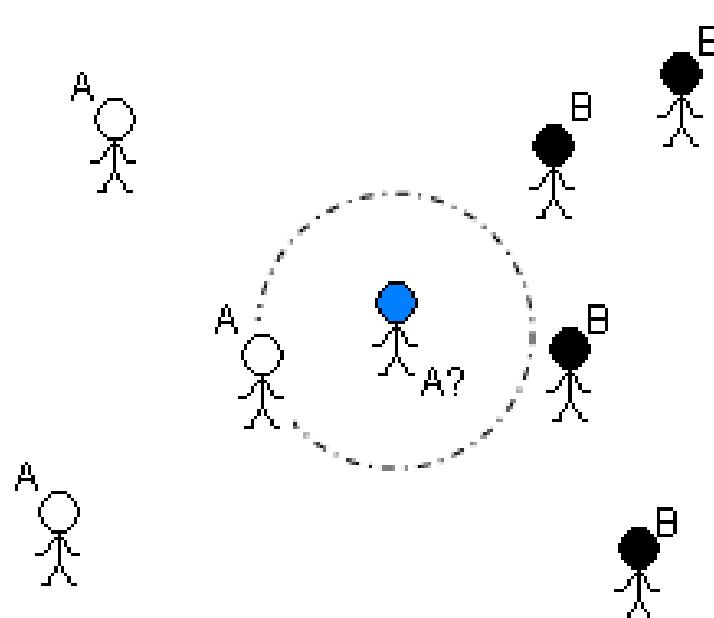
Yapay Sinir Ağları



Elimizdeki Eğitim seti Girişler ve Çıkışlar 1 içerir.
Bu girişler verildiğinde bu çıkışları verecek
Ağırlık değerlerini (W) bulmaya çalışır.

En Yakın K Komşu

- Bana Arkadaşını söyle, sana kim olduğunu söyleyeyim.



En yakın 1 komşu YILDIZ TEKNİK ÜNİVERSİTESİ En yakın 3 komşu
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



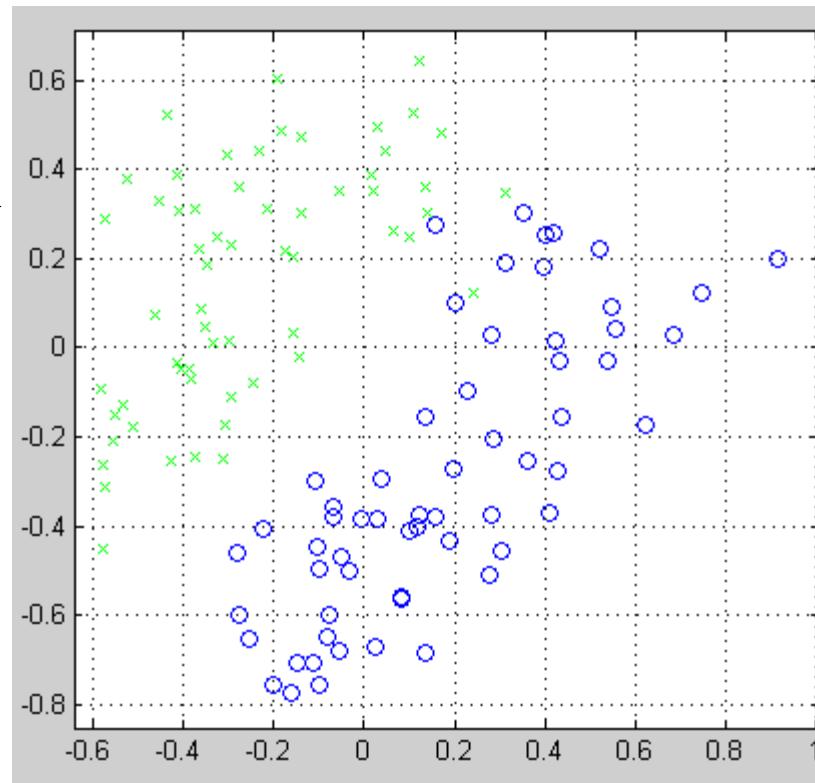
En Yakın K Komşu

- Eğitim yok.
- Test verileri en yakınlarındaki K adet komşularının sınıf değerlerine bakılarak sınıflandırılırlar.

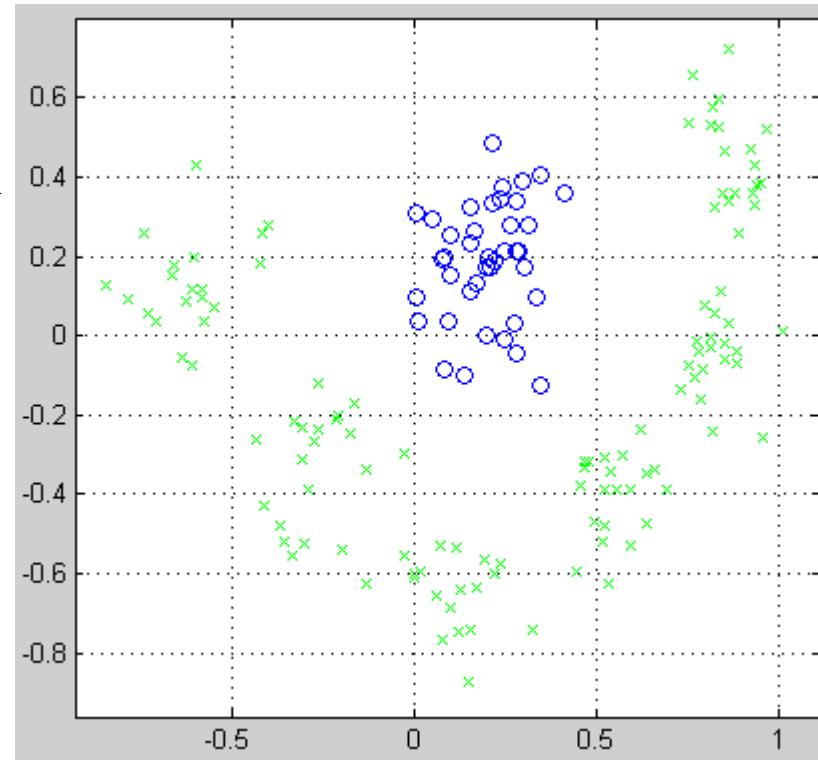


Hangisi?

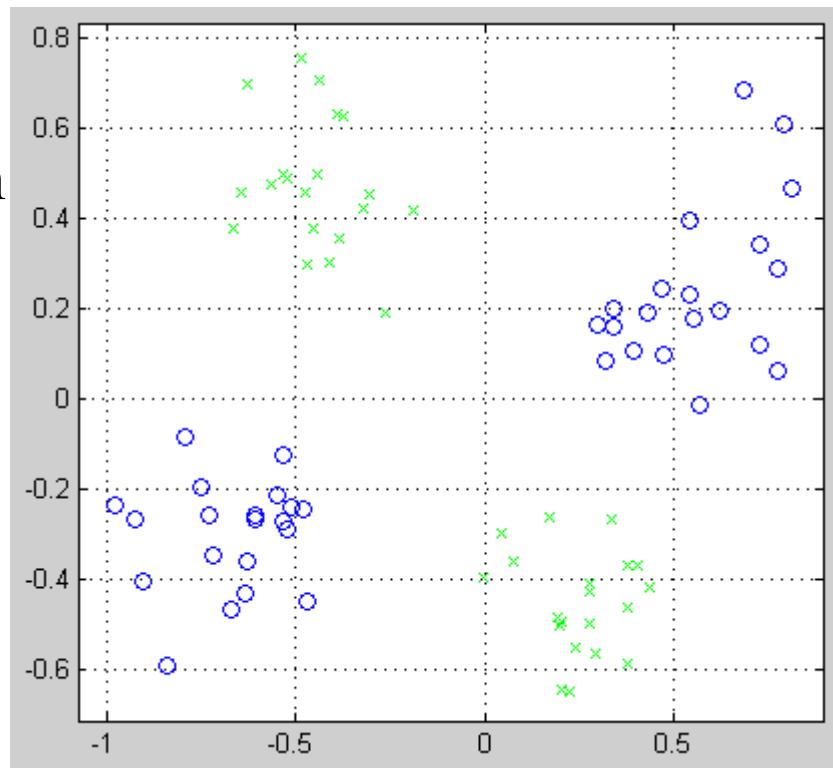
Lineer Regresyon
Karar Ağaçları
KNN
YSA
LVQ



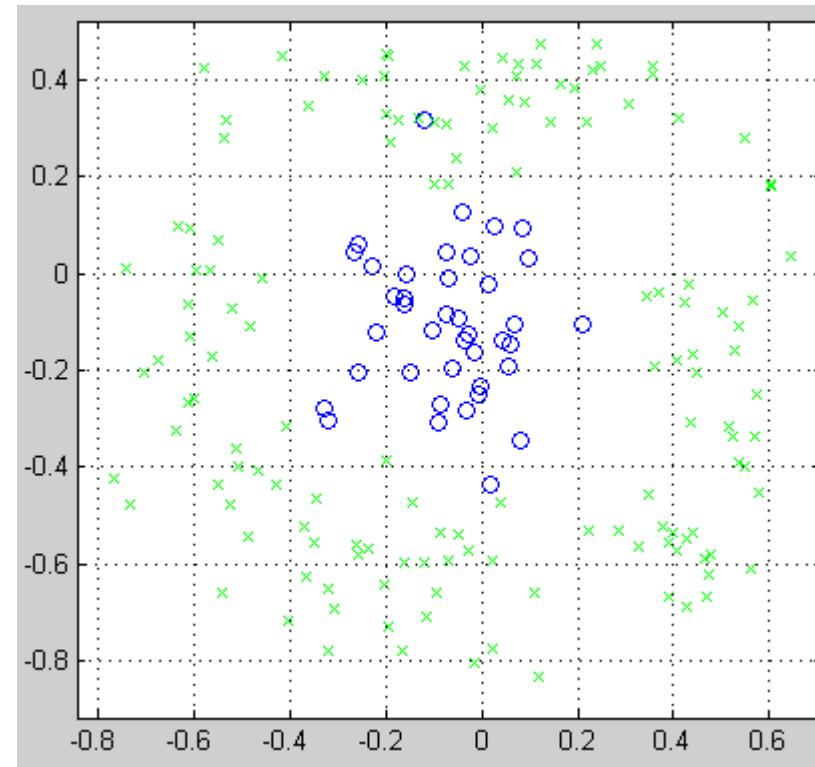
Lineer Regresyon
Karar Ağaçları
KNN
YSA
LVQ



Lineer Regresyon
Karar Ağaçları
KNN
YSA
LVQ



Başka bir model?



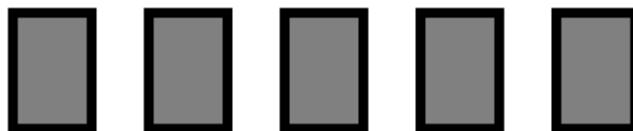
Eğitim kümesi üzerindeki performans neyi ifade eder?

- Borsa oynayan maymunlar*
- 1000 kişiye hisse senedi tahminleri göndermek*
- * Hatasız Düşünme Sanatı'ndan



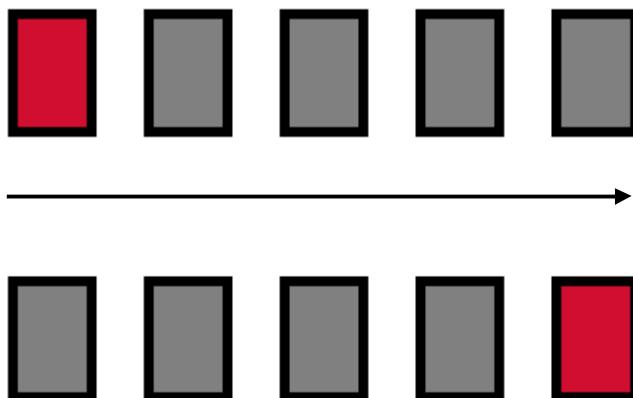
Çapraz Geçerleme

— Tüm dataseti eşit boyutlu N gruba böl



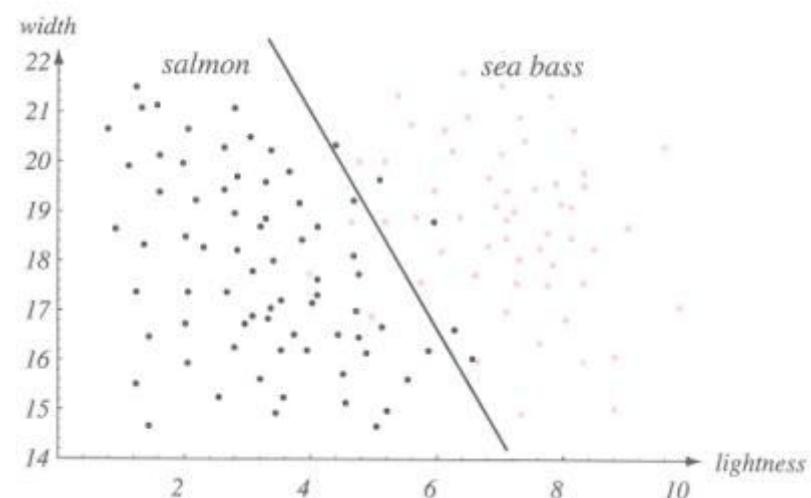
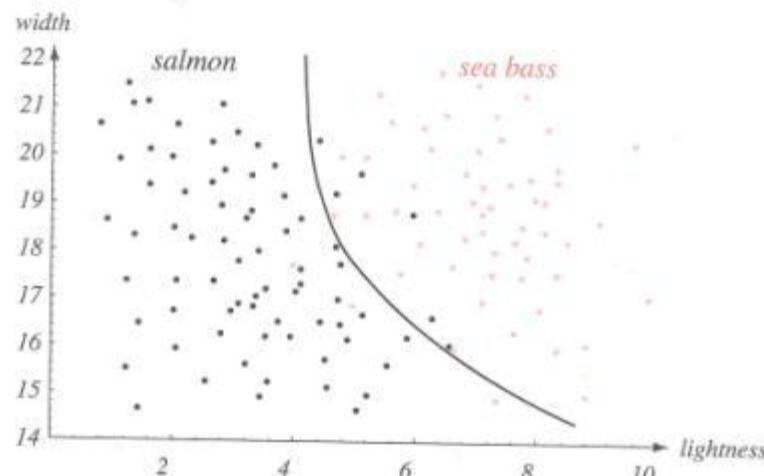
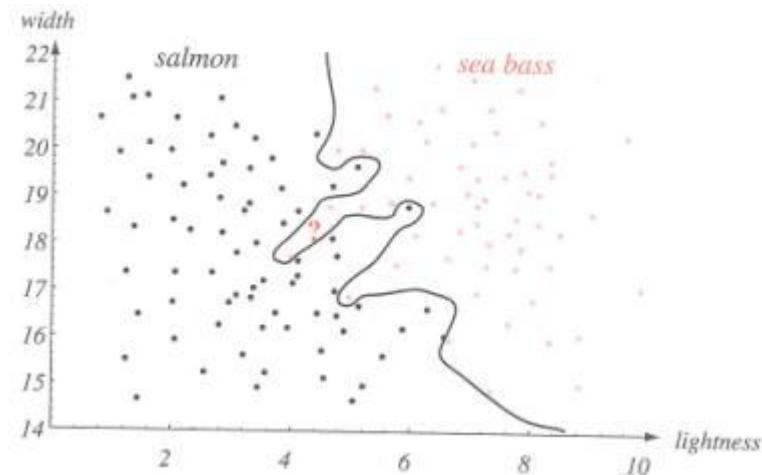
— Bir grubu test için geriye kalanların hepsini eğitim için kullan

— Test
— N defa tekrar et



Sınıflandırma Metotları- Sonuç

Farklı sınıflandırma metotlarının ürettikleri karar sınırları



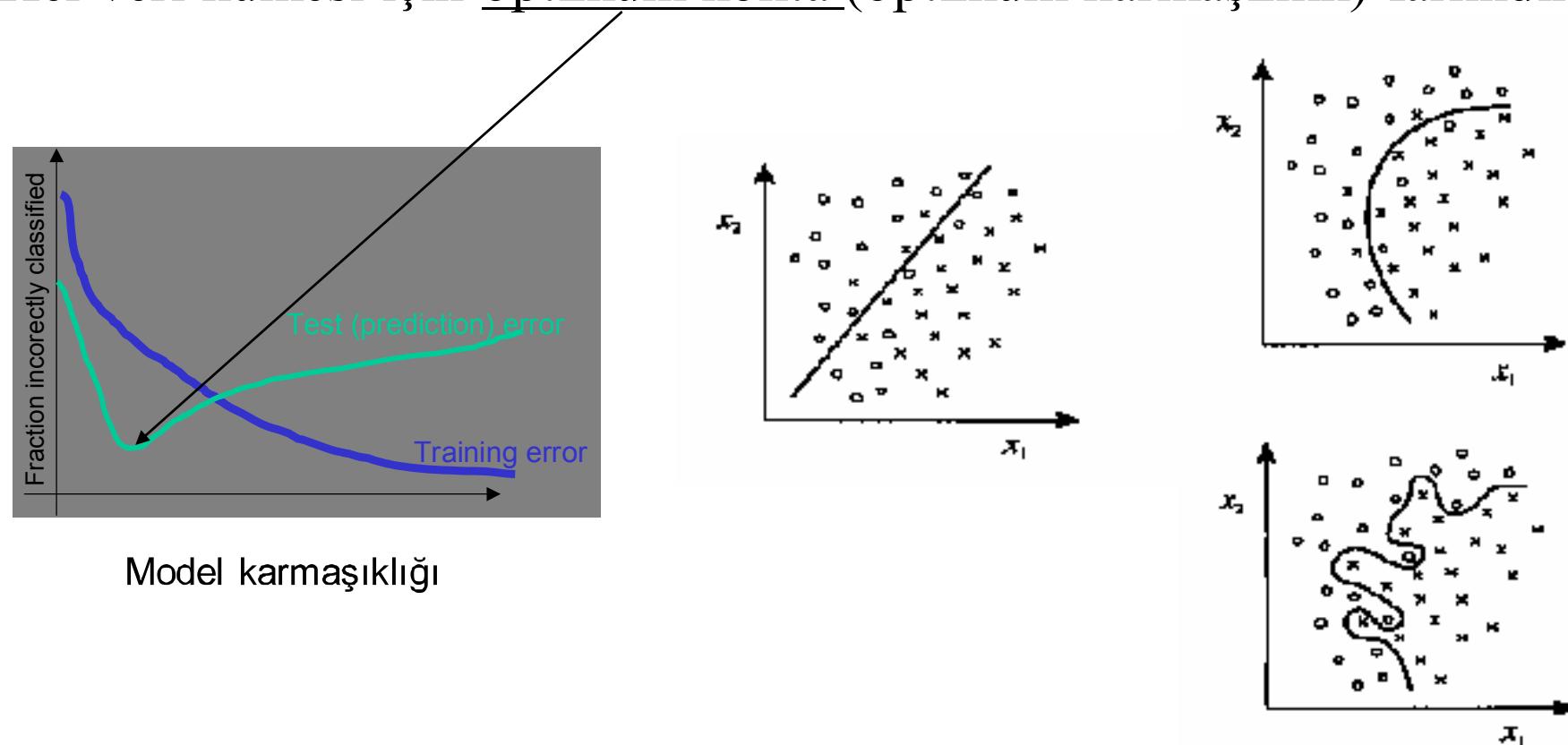
Sınıflandırma Metotları- Sonuç

- Neden bu kadar çok algoritma var?
- Ne zaman hangisini seçeceğiz?

dataset	amlall	ann	bi75ds3	derma	gkanser	Hava
Özellik sayısı	7129	21	470	34	30	34
Sınıf sayısı	2	3	9	6	2	2
Örnek sayısı	72	3772	315	286	456	281
NB	97,14	95,55	68,49	77,97	94,29	89,31
SVM	92,86	93,74	62,11	79,37	96,26	86,48
1NN	94,29	93,4	63,19	76,26	96,26	89,72
C45	83,39	99,58	65,01	75,2	93,62	91,82
RF	95,71	99,5	72	76,96	95,38	95,02

Modelim karmaşıklığı arttığında eğitim kümesindeki hata düşerken test kümesindeki hata yükselir.

Her veri kümesi için optimum nokta (optimum karmaşıklık) farklıdır.



Kümeleme Algoritmaları

- Kümeleme algoritmaları eğiticisiz öğrenme metotlarıdır.
- Örneklerde ait sınıf bilgisini kullanmazlar.
- Temelde verileri en iyi temsil edecek vektörleri bulmaya çalışırlar.
- Verileri temsil eden vektörler bulunduktan sonra artık tüm veriler bu yeni vektörlerle kodlanabilirler ve farklı bilgi sayısı azalır.
- Bu nedenle birçok sıkıştırma algoritmasının temelinde kümeleme algoritmaları yer almaktadır.



Kümeleme Algoritmaları

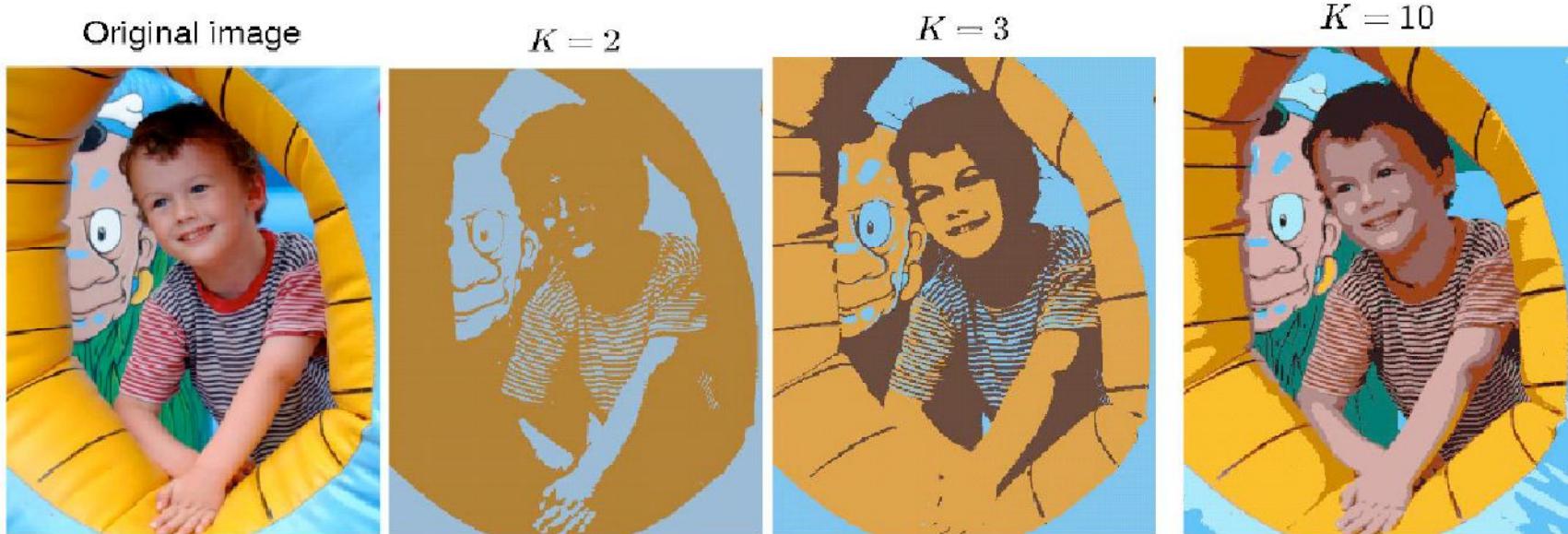
- Bir boyutlu (özellikli) 10 örnek içeren bir veri
12-15-13-87-4-5-9-67-1-2
- Bu 10 farklı veriyi 3 farklı veriyle temsil etmek istersek:
12-12-12-77-3-3-3-77-3-3
- şeklinde ifade edebiliriz.
- Kümeleme algoritmaları bu 3 farklı verinin değerlerini bulmakta kullanılırlar.
- Gerçek değerlerle temsil edilen değerler arasındaki farkları minimum yapmaya çalışırlar.

Yukarıdaki örnek için 3 küme oluşmuştur.

- 12-15-13 örnekleri 1. kümede
- 87-67 örnekleri 2. kümede
- 4-5-1-2-9 örnekleri 3. kümede yer almaktadır.



Renk Kümeleme



An Application of K-means: (Lossy) Data compression

- Original image has N pixels
- Each pixel $\rightarrow (R, G, B)$ values
- Each value is stored with 8 bits of precision
- Transmitting the whole image costs $24N$ bits

Compression achieved by K-means:

- Identify each pixel with the corresponding centroid
- We have K such centroids \rightarrow we need $\log_2 K$ bits per pixel
- For each centroid we need 24 bits
- Transmitting the whole image costs $24K + N \log_2 K$ bits

Original image = $240 \times 180 = 43,200$ pixels $\rightarrow 43,200 \times 24 = 1,036,800$ bits

Compressed images:

$K=2$: 43,248 bits

$K=3$: 86,472

$K=10$: 173,040 bits



Resim Kümelenme



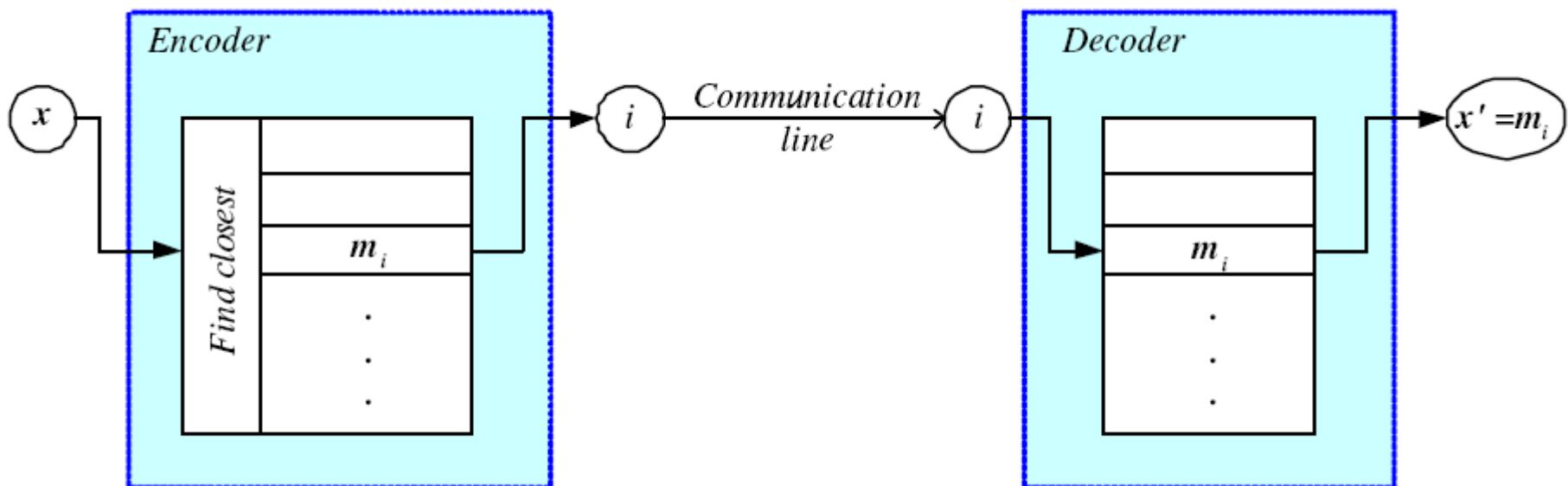
10*10 luk blokları ifade eden
vektörler kümelenmiş

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Nasıl Kullanılır?

Bulunan (renkleri yada blokları temsil eden) küme merkezlerinden bir kod kitabı oluşturulur. Bu kitap her iki merkeze verilir. Vektörlerin kendileri yerine sadece indisler kullanılır. İndisin maximum büyüklüğü kodlanması için gereken bit sayısını arttırır.
Bu yüzden farklı vektör sayısının az olması istenir.



ETHEM ALPAYDIN

© The MIT Press, 2004



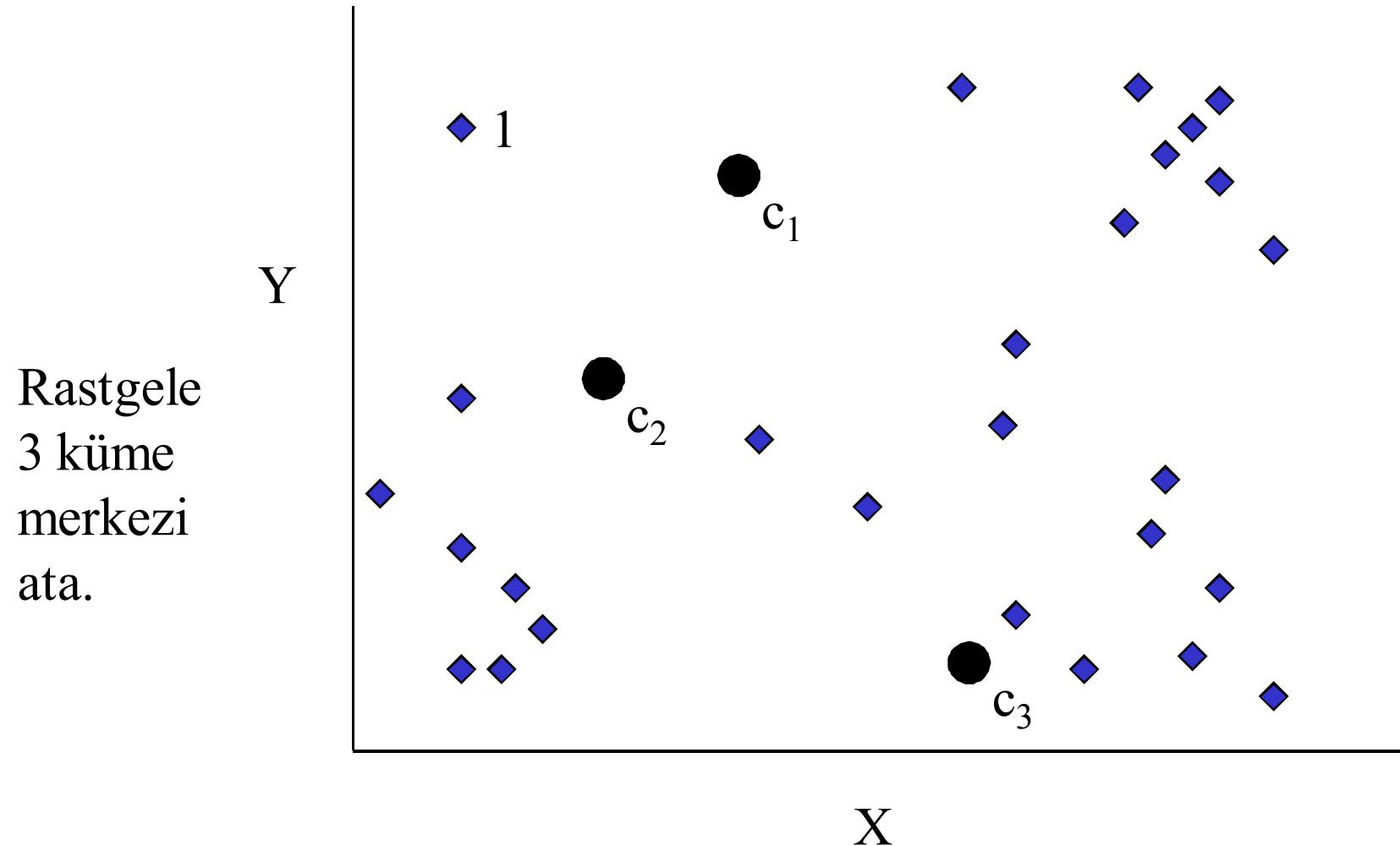
K-means

Works with numeric data only

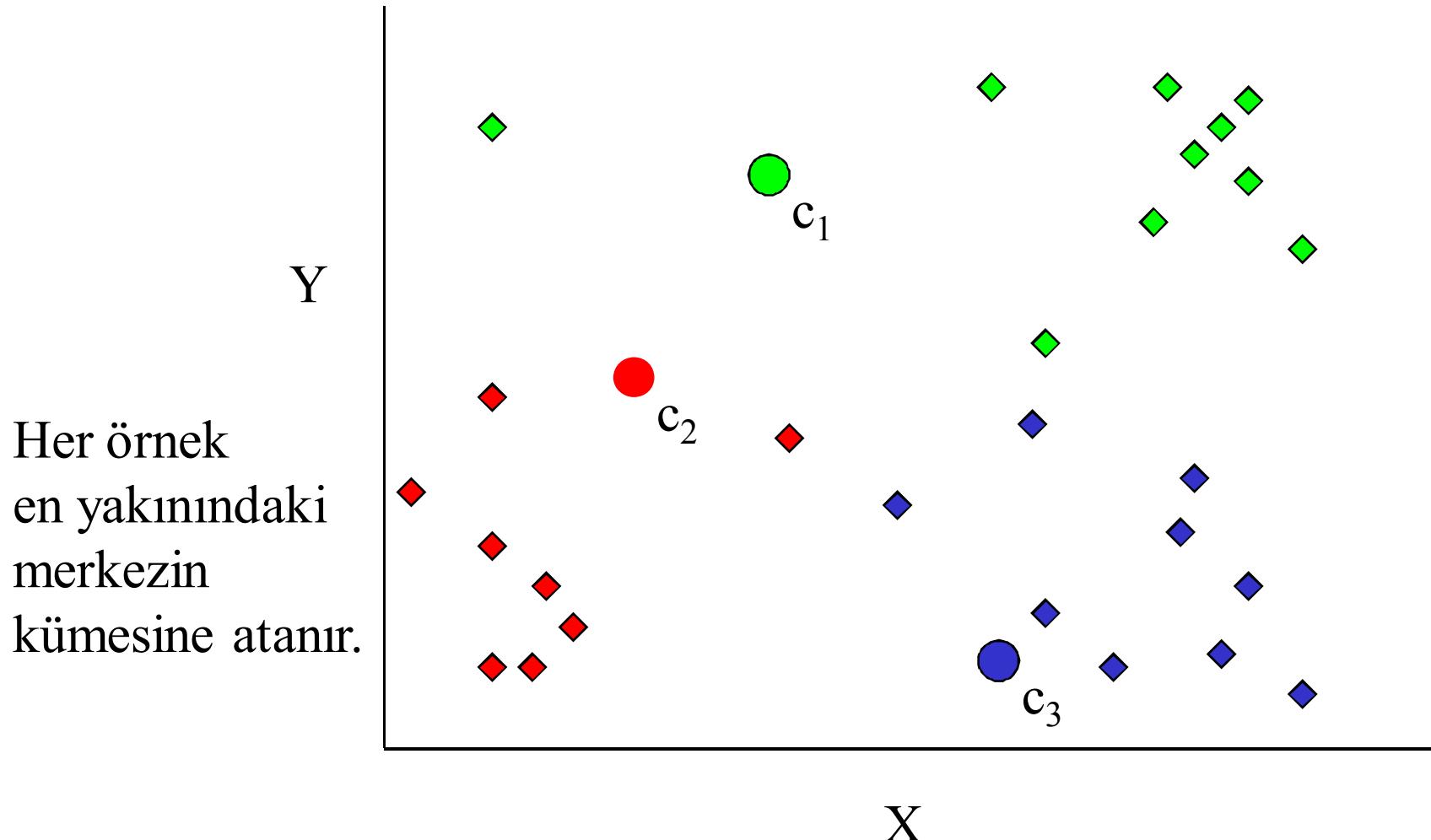
- 1) Rastgele K adet küme merkezi ata
- 2) Her örneği en yakınındaki merkezin kümese ata
- 3) Merkezleri kendi kümelerinin merkezine ata
- 4) 2. ve 3. adımları küme değiştiren örnek kalmayınca kadar tekrar et.



K-means örnek adım 1

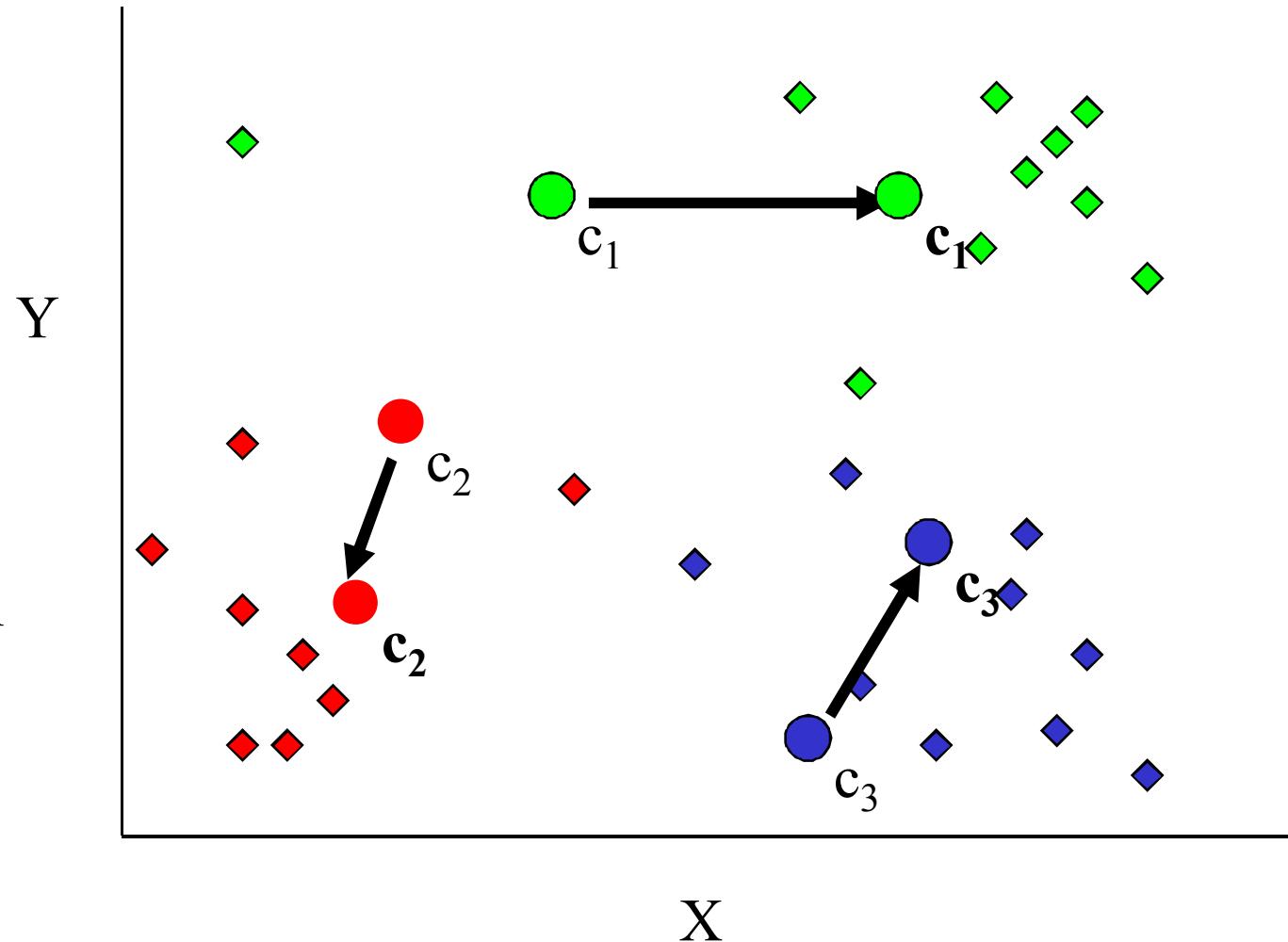


K-means örnek adım 2



K-means örnek adım 3

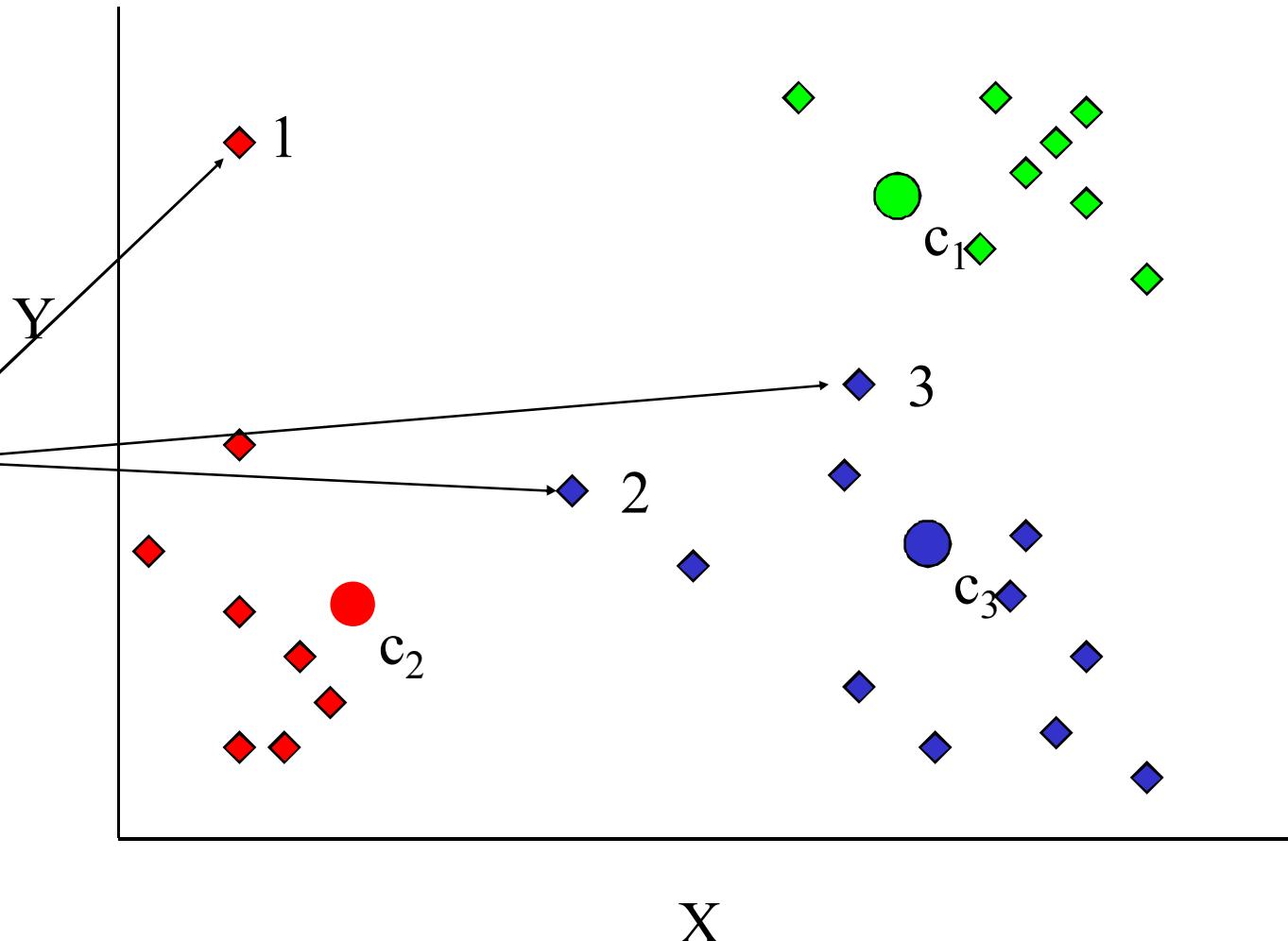
Merkezleri
kendi
kümelerinin
merkezine
götür.



K-means örnek adım 4

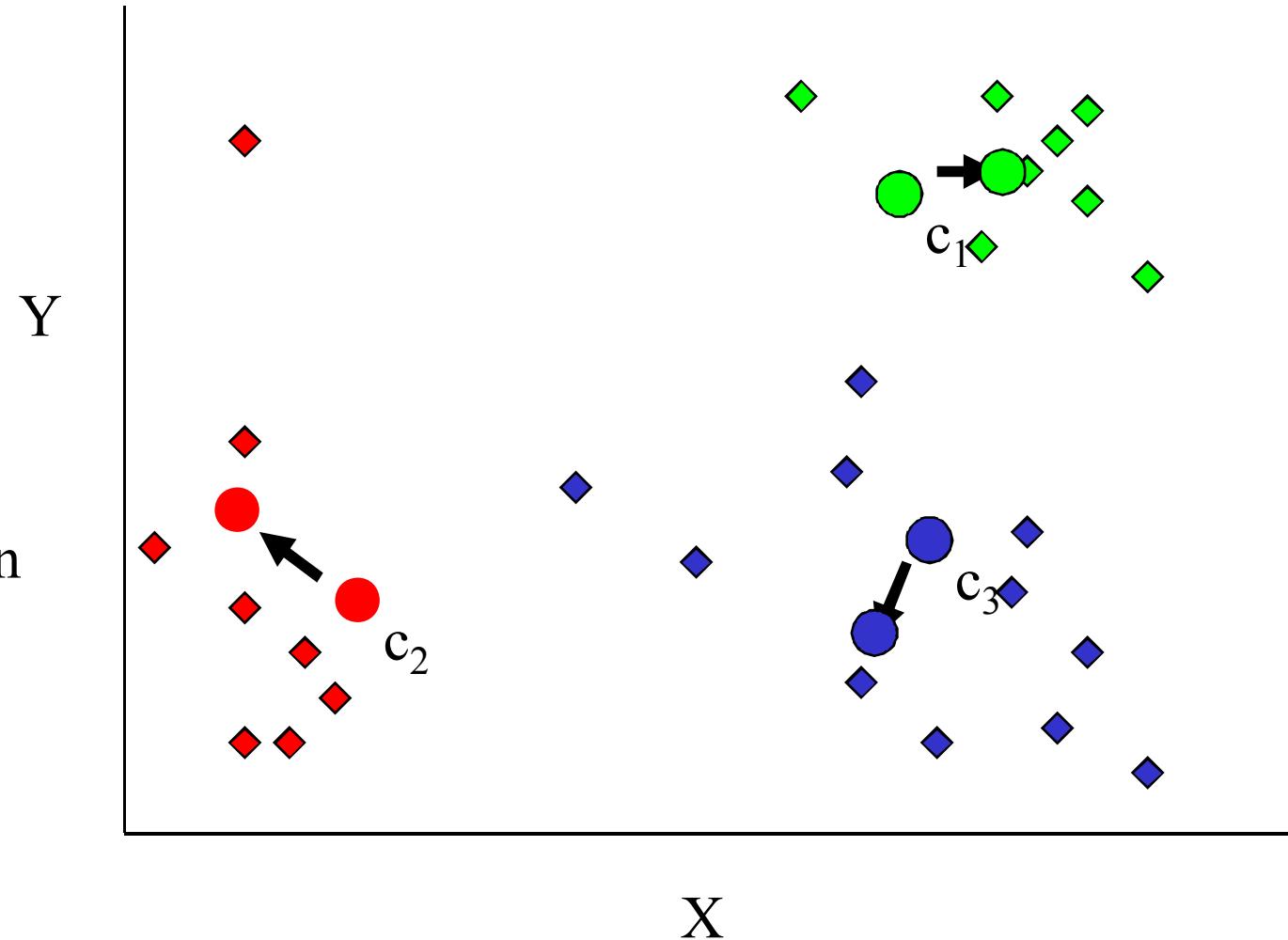
Her örneği
yeniden en
yakınındaki
merkezin
kümesine
ata.

*Q: Hangi
örneklerin
kümesi
değişti?*



K-means örnek adım 5

Merkezleri
kendi
kümelerinin
merkezine
götür.



Kendi Kendini Düzenleyen Haritalar

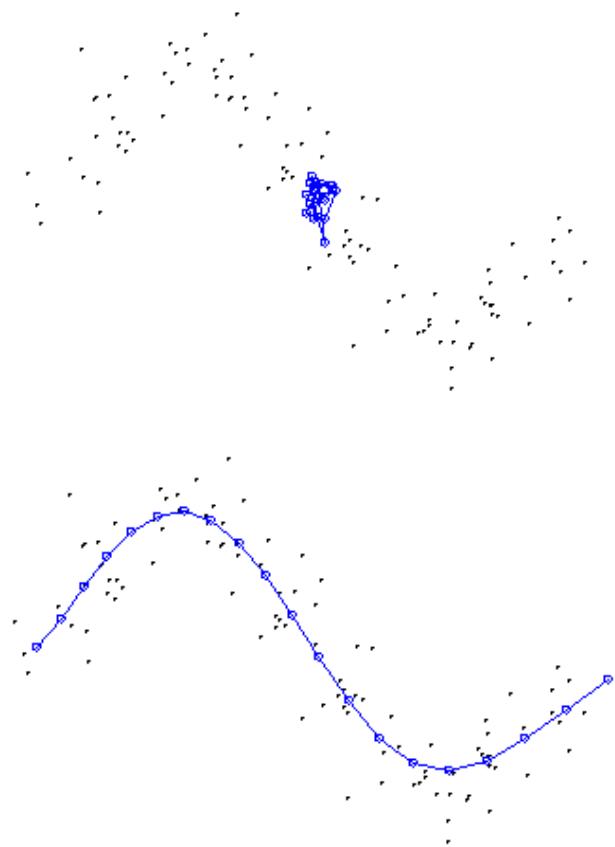
Self Organizing Maps

- Kmeans algoritmasında merkez noktalar arasında herhangi bir ilişki yoktur. SOM'da ise merkez noktalar 1 ya da 2 boyutlu bir dizi içinde yer alırlar. Buna göre birbirlerine 1 ya da 2 boyutlu uzayda komşudurlar.
- Kmeans algoritmasında sadece kazanan (en yakın) merkez güncellenirken SOM'da bütün merkezler kazanan nörona komşuluklarına göre güncellenir. Yakın komşular uzak komşulara göre daha fazla hareket ederler (güçellenirler).
- Merkezlerin birbirlerine bağlı oluşu verinin 1 ya da 2 boyutlu uzaydaki yansımاسının da elde edilmesini sağlar.



SOM

- SOM merkezleri 1 boyutlu bir dizide birbirlerine komşudurlar.
Başlangıçtaki durumları rasgele atıldığı için bir yumak şeklindedirler.
Eğitim tamamlandığında ise SOM merkezleri verinin şeklini almıştır.



Sonuç olarak

- Makineler insanlığın işgücüne sağladıkları katkıyı, makine öğrenmesi metotları sayesinde insanlığın beyin gücüne de sağlamaya başlamışlardır.



Kaynaklar

- Alpaydın E. (2004) “Introduction to Machine Learning”, The MIT Press, 3-6
- <http://www.autonlab.org/tutorials/infogain11.pdf>
- http://www.kdnuggets.com/dmcourse/data_mining_course/assignments/assignment-4.html
- http://pespmc1.vub.ac.be/asc/SENSIT_ANALY.html
- http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
- http://www.cavs.msstate.edu/hse/ies/publications/reports/isip_internal/1998/linear_discrim_analysis/lda_theory.pdf
- <http://www.kernel-machines.org>
- T.Kohonen," Self-Organization and associative Memory",3d ed, 1989, Berlin :Springer-Verlag.
- <http://www.willamette.edu/~gorr/classes/cs449/Classification/perceptron.html>
- O. T. Yıldız, E. Alpaydın, Univariate and Multivariate Decision Trees, Tainn 2000
- <http://www.ph.tn.tudelft.nl/PHDTheses/AHoekstra/html/node45.html>
- <http://mathworld.wolfram.com/K-MeansClusteringAlgorithm.html>
-



Weka



Copyright: Martin Kramer (mkramer@wxs.nl)

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

