

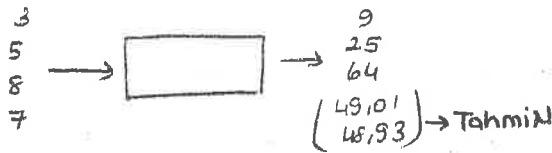
makine öğrenmesi bir problemi probleme ait ortanolan edinilen veriye göre modelleyen bilgisayar algoritmalarının genel adıdır. Birçok yaklaşım ve algoritma mevcuttur. Yaklaşımlar

- * Tahmin (prediction)
- * Kestirim (estimation)
- * Sınıflandırma (classification)

yeteneğine sahiptir.

1. **Tahmin:** Veriden öğrenen modelde sistem çıkışının nicel olması durumunda kullanılan yöntemlerin drettliği değeridir.

Rasgele değişkenin bilinmeyen değerinin belirlenmesi tahmindir.



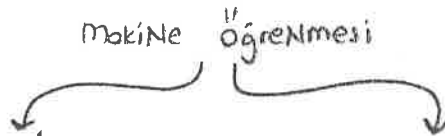
2. **Sınıflandırma:** Giriş verisine ait çıkışların nitel olduğu durumlarda kullanılan yöntemlerin her veri örneğinin hangi sınıfa ait olduğunu belirlemesidir.

Sınıflandırma tahminin daha özel halidir.

3. **KESTİRİM:** Giriş ve çıkış arasındaki parametre denklemdir.

X	Y
3	9
4	12
5	15

$$f(x) = 3x$$



1. **Deneyimsiz öğrenme**

Girdiye karşılık gelen çıktıların

öğrenilmesi.

(Tabloların ve örneklerin tamamıyla öğrenilmesi)

	X ₁	X ₂	Y (Sınıf)
1.			
2.			
3.			
4.			

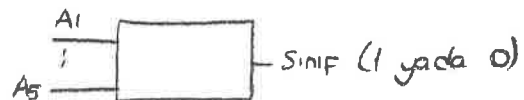
2. **Deneyimli öğrenme**

Tabloda sınıf kolonu dahil

edilmemiş olursa, örneklerin

birbirine benzerliği gösterilir.

	A ₁	A ₂	A ₃	A ₄	Sınıf
1					Dişabet
2					Değil
3					Dişabet



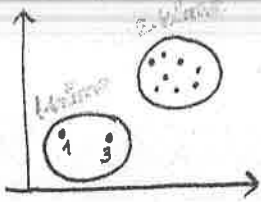
Takviyeli Öğrenme: Sisteme geri beslemenin yapılması (Çıktının doğru yada yanlış olması)
Boltzmann makinesi, LVA, genetik algoritma... vb

Sınıflandırma, tahmin \Rightarrow Danışmanlı (Eğitici) Öğrenme

Kümeleme \Rightarrow Danışmansız Öğrenme

Kümelemede sınıf kumu yoktur. Eğitici öğrenmeden bahsedilemez. Sadece girdiler vardır.

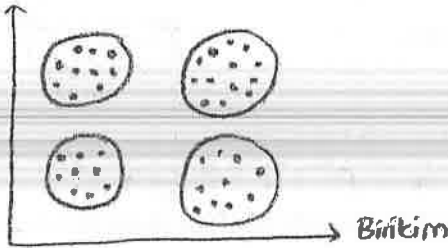
1 ve 3 kayıtlarının değerleri birbirine yakın:



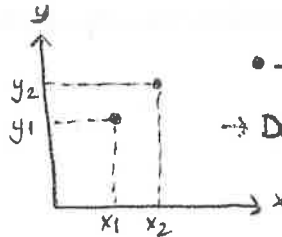
Değerleri birbirine yakın olan kayıtlar kendi doğrultusunda küme oluşturulur.

Küme sayısı dışarıdan verilir.

Örnek: Yıllık Kazanç



Yıllık kazanç ve yaptıkları birikimlere göre insanlar kümelenecek kredi verirken, faiz oranlarında banka bir mekanizma oluşturabilir.



• \rightarrow Müşteri

\rightarrow Değerleri yakın aynı kümede bulunan iki müşteri

NOTASYONLAR

\rightarrow Gerçek Sınıf

\rightarrow Attribütler (Özellikler... A_1, A_2, A_3)

	A_1	A_2	A_3	A_4	D
X_1	X_{11}	X_{12}			
X_2	X_{21}	X_{22}			
X_3					
\vdots					
X_n					

$X \rightarrow$ Örnek

$Y \rightarrow$ Tahmin edilen çıktı

Öğrenme

Online

Offline

Gerçek zamanlı
sistemde öğrenmenin
çalışma anında olması

Sistem çalışmazken eğitilir,
Eğitilen yöntem sisteme yüklenir
ve çalıştırılır.

Veriye dayalı olarak eğitim yapılmasının temel amacı eğitilen sistemin benzer konuda hiç bilinmeyen bir örneğe mantıklı cevap üretebilmesidir.

Çapraz Geçerlilik: Verilen veri kümesi ikiye bölünür. Bir kısmı eğitim verisi bir kısmı test verisi olur. Eğitim algoritması ile sistem eğitim kümesini öğrenir. Test verisi ile sistemin öğrenmesi test edilir.

	A ₁	A ₂	A ₃	A ₄	A ₅	Sınıf
1						
2						
...						
100						
101						
...						
150						

→ 1-100 arası

→ 101-150 arası

50 tane veriden 40 tane doğru tahmin edilmişse başarı %80'dir.

Hiçbir öğrenmede başarı %100 olmaz.

1. Basamaklı Örneklem: En yüksek başarıyı sağlayan yöntemdir. 300 veri olsa 10 parçaya ayrılırsa bazıları test, bazıları eğitimde kullanılır.

$\frac{2}{3} \rightarrow$ eğitim $\frac{1}{3} \rightarrow$ test $200 \rightarrow$ eğitim verisi
 $100 \rightarrow$ test verisi

2. K parçaya: K parçaya ayır. K=30 olursa her parçada 10 örnek var. Bir parça K'yı test yap, sonra diğerlerini eğit. 30 defa tekrarlanır. Ortalama alınması başarı oranını verir.

3. Bütünsel test: 300 tane örnekten 1'ini test için ayır 299 tanesini eğit.

? Herhangi bir özellik (attribute) için kabul edilebilir değer.
 ∅ Hiçbir değer kabul edilemez

$\langle ?, ?, ?, ?, ?, ? \rangle \rightarrow$ En general hipotesis

$3 \times 2 \times 2 \times 2 \times 2 \times 2 = 96$ farklı örnek

$$5 \times 4 \times 4 \times 4 \times 4 \times 4 = 5120 \text{ farklı hipotez (Teorik olarak)}$$
$$1 + (4 \times 3 \times 3 \times 3 \times 3 \times 3) = 973 \text{ Faktli hipotez (protikite)}$$

↳ ?'den dolayı:

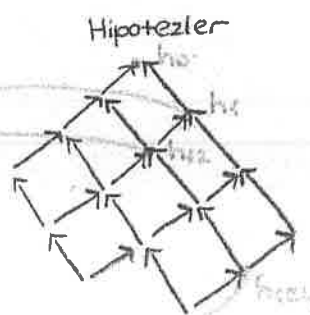
Yukarıdaki verilerde evet çoğunlukta olduğu için evete göre bakılır.

$h_1: \langle \text{Güneşli, Sıcak, Normal, Güçlü, Sıcak, Aynı} \rangle$

$h_2: \langle \text{Güneşli, Sıcak, ?}, \text{Güçlü, Sıcak, Aynı} \rangle$

h₄: < Güneşli, Sıcak, ?, Güçlü, ?, ? >

6 parametreden 3 parametre bizim için önemlidir.



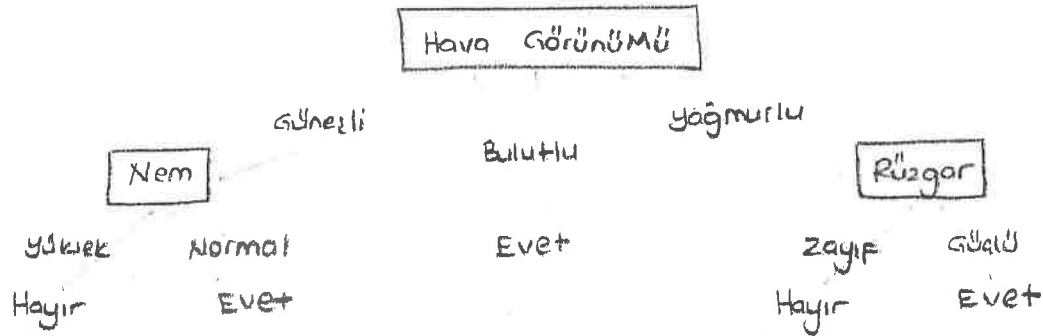
En Spesifik Durum

— Bir birim altta olması ? geldiğini gösteriyor—

En genel Durum

NOT: Düşümlerin sağda ya da solda olması değil, seviyesi önemli

Temel fikir; giriş verisinin bir kümeleme algoritması yardımıyla tekrar tekrar gruplara bölünmesine dayanır. Grubun tüm elemanları aynı sınıf etiketine sahip olana kadar kümeleme işlemi devam eder.



$(\text{Havanın Görünümü} = \text{Güneşli} \wedge \text{Nem} = \text{Normal}) \vee (\text{Havanın Görünümü} = \text{Bulutlu}) \vee$

$(\text{Havanın Görünümü} = \text{Yağmurlu} \wedge \text{Rüzgar} = \text{Güçlü})$

GÜN	Havanın Görünümü	Sıcaklık	Nem	Rüzgar	TENİS
1	Güneşli	Sıcak	Y	Z	H
2	Güneşli	Sıcak	Y	G	H
3	Bulutlu	Sıcak	Y	Z	E
4	Yağmurlu	Soğuk	Y	Z	E
5	Yağmurlu	Ilık	N	Z	E
6	Yağmurlu	Ilık	N	G	H
7	Bulutlu	Ilık	N	G	E
8	Güneşli	Soğuk	Y	Z	H
9	Güneşli	Ilık	N	Z	E
10	Yağmurlu	Soğuk	N	Z	E
11	Güneşli	Soğuk	N	G	E
12	Bulutlu	Soğuk	Y	G	E
13	Bulutlu	Sıcak	N	Z	E
14	Yağmurlu	Soğuk	Y	G	H

Düğümle parametreyi, yapraklar sınıfı gösterir.

Karar Ağacı için Uygun Problemler

- Örnekler $\langle \text{parametre, değer} \rangle$ çiftleriyle temsil ediliyorsa
- Hedef fonksiyon (sınıf) ayrık çıkış değerlerine sahipse
- Veya'lı tanımlanalar varsa

2. Eğitim verisi hatalar içerebilir.

3. Eğitim verisi eksik değer içerebilir.

* Hangi parametre en iyi sınıflandırıcıdır?

Bilgi kazancı, hangi parametrenin daha büyüğe en iyi sınıflandırıcıdır.

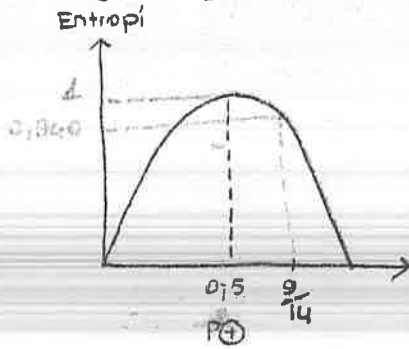
Bilgi kazancı hesaplamak için entropi kavramı vardır.

P_{\oplus} → Pozitif örnekleri oranı ($9/14$)

P_{\ominus} → Negatif örneklerin oranı ($5/14$)

$$\text{Entropi}(S) = -P_{\oplus} \log_2 P_{\oplus} - P_{\ominus} \log_2 P_{\ominus}$$

Entropi max elde etmek için örneklerin yarısının evet, yarısının hayır olmasıyla elde edilir. Böylece uygun sınıflandırıcı elde edilir.



2 sınıftan fazla sınıf olduğunda;

$$\text{Entropi}(S) = \sum_{i=1}^c -P_i \log_2 P_i$$

$$\text{Bilgi Kazancı } (S, A) = \text{Entropi}(S) - \sum \frac{|S_v|}{|S|} \cdot \text{Entropi}(S_v) \quad \underbrace{V \in \text{Values}(A)}_{\text{Rüzgarın alacağı değerler}}$$

Eğitim verileri Parametre

A Rüzgar

$S = (9+, 5-) \rightarrow$ (Toplamda 9 → evet 5 → hayır sınıfı mevcut)

$S_{2\text{ayrık}} = (6+, 2-) \rightarrow$ (2ayrık parametresinin sınıflarında kaç evet kaç hayır var)

$S_{güçlü} = (3+, 3-)$

$$\text{Bilgi Kazancı } (S, \text{Rüzgar}) = \text{Entropi}(S) - \sum \frac{|S_v|}{|S|} \cdot \text{Entropi}(S_v) \quad V \in (2\text{ayrık}, \text{güçlü})$$

$$= \underbrace{\text{Entropi}(S)}_{0.940} - \frac{8}{14} \cdot \text{Entropi}(S_{2\text{ayrık}}) - \frac{6}{14} \cdot \text{Entropi}(S_{güçlü})$$

1 → (3+, 3-) oranı +1 oluyor

$$= 0.048$$

Bilgi kazancı (S , Hava görünümü) = 0,246

Bilgi kazancı (S , nem) = 0,151

Bilgi kazancı (S , sıcaklık) = 0,029

Bulutlu ($D3, D7, D12, D13$) ($4+, 0-$) → evete 4 tane gidiyor (Yapraktır)

Güneşli ($D1, D2, D8, D9, D10$) ($2+, 3-$) → Yaprak Değil

Bilgi kazancı ($S_{\text{güneşli}}, \text{Nem}$)

Bilgi kazancı ($S_{\text{güneşli}}, \text{Sıcaklık}$)

Bilgi kazancı ($S_{\text{güneşli}}, \text{Rüzgar}$)

YAPAY SINIR AĞLARI

Biyojik sinir 4 bölüme ayrılır; * Dendrit

* Aksan

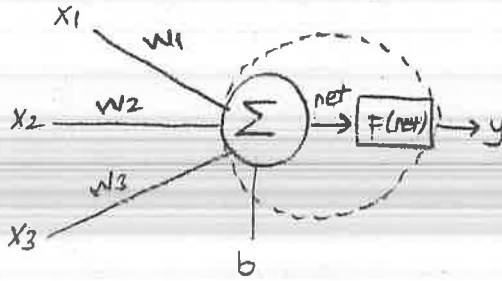
* Çekirdek

* Bağlantılar

- Yapay sinir ağı makine öğrenmesini gerçekleştirir.
- Yapay sinir ağı örnekleri kullanarak öğrenirler.
- Görünmemiş belgeler hakkında bilgi üretebilirler.
- Eksik bilgi ile çalışabilirler
- Hata toleransına sahip. (sınıflandırmadaki hatalara uygun çözüm üretebiliyor)
- Dereceli bozulma gösterirler.

Dağıtık belleğe sahiptir. (Paralel olarak çalışabilirler)

Yapay sinir Hücresi



$$net = \sum_{i=1}^M x_i w_i + b$$

$F(net) \rightarrow$ aktivasyon fonksiyonu

$$F(net) = \frac{1}{1 + e^{-net}}$$

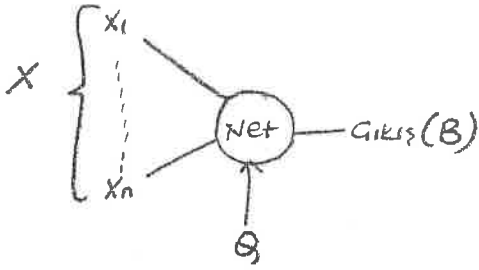
Sigmoid Fonk

Yapay sinir hücresi çalışma prensibi; Ağırlığın skaler değerine bakılır.

- Girdi katmanı; işlem yapılmaz.
- Ara katmandaki hücreler için girdi ağırlık alınıp toplanarak net hesaplanır.
- Girdi ve çıktı sayısı teorik olarak az olmalı iyidir. Fakat problemde girdi, çıktı sayısı sabittir.
- Ara katmandaki hücre sayısı artırılabilir.
- Çıktının aktivasyon fonk. ara katmanın aktivasyonundan farklı olabilir.

Basit Algılayıcı Öğrenme Kuralı

Algılayıcı tek nöronla oluşur



(X, B)

$$Net = \sum_{i=1}^n w_i x_i$$

Eşik değeri

* $NET > \theta$ ise $gikis \rightarrow 1$

$B \rightarrow$ Beklenen çıktı

* $NET < \theta$ ise $gikis \rightarrow 0$

Gerçeklenen çıktı ile beklenen çıktı aynıysa ağırlıkta değişiklik olmaz

Beklenen değer $\rightarrow 0$ } Ağırlıklar azaltılmalıdır.

Gerçeklenen değer $\rightarrow 1$ }

NOT: Yapay sinir ağlarında öğrenme ağırlıklarla aynıdır. Ağırlık değiştirerek girdiye uygun çıktı üretilmeye çalışılır.

Ağırlıkların değişim oranı girdi oranı kadardır.

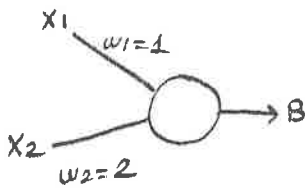
$$w_n = w_0 - \lambda \cdot x$$

Beklenen $\rightarrow 1$

$$w_n = w_0 + \lambda \cdot x$$

Gerçeklenen $\rightarrow 0$

Soru:



x_1	x_2	B
1	0	1
0	1	0

Eşik değeri $= \theta = -1$

Öğrenme katsayısı $= \lambda = 0,5$

$$NET = x_1 \cdot w_1 + x_2 \cdot w_2 = 1 \cdot 1 + 0 \cdot 2 = 1$$

$NET > \theta \rightarrow gikis \rightarrow 1$

2. örnek için:

$$NET = 0 \cdot 1 + 1 \cdot 2 = 2$$

$NET > \theta \rightarrow gikis \rightarrow 1$ (Ağırlıklar azaltılmalıdır)

Beklenen ve gerçekleşen aynı değil.

Azaltmak için; $w_n = w_0 - \lambda \cdot x$

$$w_1 = 1 - 0,5 \cdot 0 = 1$$

$$w_2 = 2 - 0,5 \cdot 1 = 1,5$$

w değerleri 2. değere (örneğin)

göre yapılır.

3. iterasyon

1. örnek değişme yok.

4. iterasyon (2. örnek)

$$NET = 1 \cdot 0 + 1 \cdot 1,5 = 1,5$$

\rightarrow Ağırlık azaltılmaya devam edilmektedir.

NOT!

İterasyon değeri $NET < \theta$ yani çıkış 0 olana kadar azaltma yapılmaya devam edilir.

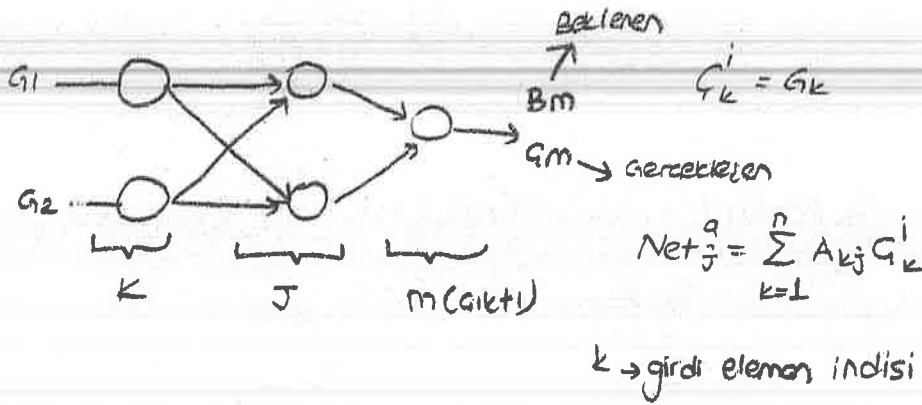
Perseptum = Hücre

ÇOK KATMANLI Algılayıcı

En basit problemi ex-or'dur.

* 1. hafta ; ileri doğru çıktı hesaplama

* 2. hafta ; ağırlıkları değiştirme



$$G_j^a = \frac{1}{1 + e^{-(Net_j^a + B_j^a)}}$$

* sigmoid fonk. kullanılmasının nedeni türevi kolay olan bir fonksiyon olmasıdır.

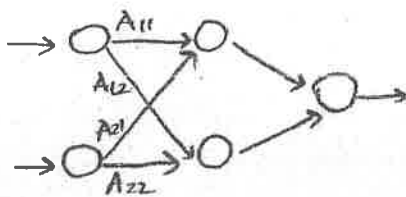
$$E_m = B_m - G_m$$

Hata

$$T_H = \frac{1}{2} \sum_m E_m^2$$

Ara Katman / Çıktı Katman arasında Ağırlık Değişimi

Soru:	Örnek 1	Girdi 1	Girdi 2	Çıkış
	1	0	0	0
	2	0	1	1
	3	1	0	1
	4	1	1	1



$$\lambda = 0,5$$

$$\alpha = 0,8$$

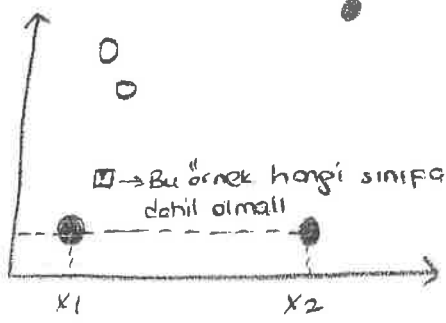
Toplam Hata = 0,03 (0,03'den küçük olduğunda mavi)

makine öğrenmesi

Uzaklığa Dayalı Sınıflandırma

İki örnek aynı sınıftaysa yüksek benzerlik / düşük uzaklık demektir.

• K en yakın komşu (K-NN)



K-NN

1-NN → Kendisine en yakın örnek hangi sınıftaysa

0 sınıfa dahil değildir. (siyah sınıfa ait olur)

3-NN (Beyaz sınıfa ait olur)

Uzaklık azaldıkça benzerlik artar.

Sınıfı bilinmeyen noktaya sınıfı olan uzaklıklar hesaplanır. Hesaplanan uzaklıklar sıralanır. (Genelde öklid uzunluğu kullanılır). En küçük uzaklığa sahip sınıfa dahil edilir.

Dezavantajı: * Bellek gerektir.

* Veri büyüdüğünde hesaplama süresi artar.

* Bütün atributlar katıldığı için sınıflandırmayı olumsuz yönde etkiler.

Benzerlik ve Farklılık;

İki nesnenin benzerliği;

Benzerlik 0 ile 1

arasında değer alır.

	A ₁	A ₂	A ₃	A ₄	A _k
1.					
2.					

1 → Bu iki nesne aynı

0 → Bu iki nesne farklı

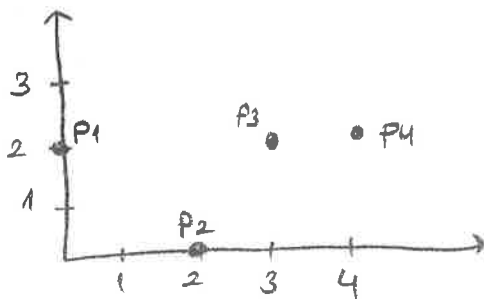
$$\text{Benzerlik} = 1 - \text{Farklılık}$$

Yukarıdaki matris uzaklığı

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \end{bmatrix}$$

p adet niteliği olan i ve j nesneleri arasındaki uzaklık;

$$d(i,j) = \sqrt{(x_1 - x_2)^2 + (x_1' - x_2')^2 + \dots} \rightarrow \text{öklid uzaklığı}$$



$$d(P_3, P_4) = 1$$

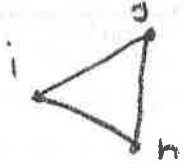
öklid uzaklığı hipotenüs bulmadır.

manhattan uzaklığı \rightarrow üst 1 olursa

minkowski uzaklığı \rightarrow euklidin genelleştirilmiş hali.

$$d(i, j) = d(j, i)$$

$$d(i, i) = 0$$



$$d(i, j) \leq d(i, h) + d(h, j)$$

Benzerlik özellikleri

$$\text{sim}(i, j) > 0$$

İkili Değişken Arası Benzerlik;

		Neve j	
Neve i	0	M ₀₀	M ₀₁
	1	M ₁₀	M ₁₁

$$p = 1000000000$$

$$q = 0000001001$$

$$M_{01} = 2$$

$$M_{10} = 1$$

$$M_{00} = 7$$

$$M_{11} = 0$$

$$\frac{7}{10} = 0,7 \rightarrow \text{Benzerlik}$$

$$\text{Jaccard katrasyonu} = 0$$

Kosinüs Benzerliği

$$d_1 = 3205000200$$

$$d_2 = 1000000102$$

$$3 \times 1 + 2 \times 1 = 5 \quad (d_1 \cdot d_2)$$

$$|d_1| = 0,451$$

$$|d_2| = 2,249$$

$$\cos(d_1, d_2) = 0,3150$$

Uzaklığa Dayalı Gruplandırma:

Uzaklık için en çok kullanılan öklit uzaklığıdır.

$$S_{ab} = \frac{1}{1+|a-b|} \rightarrow \text{Benzerlik}$$

Benzerlikle uzaklık ters orantılıdır.

Kümeleme; sınıf bilgisi içermeyen benzer örneklerin gruplanmasıdır.

K değeri kaç kümeye ayrılacak

Veriden rasgele seçilen K adet noktaya ($K=2$) küme merkezi olarak görülüyor.

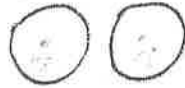
Daha sonra merkezlere noktaların yakınlığına göre gruplanıyor.

Ağırlık merkezi bulunduğundan sonra tekrar uzaklıklara bakılır ve yeni küme oluşturulur. Değişiklik olmayınca kadar bu işlem devam eder.

(Değişik olmaması demek örneklerin grubu değiştiği anlamına gelmez)

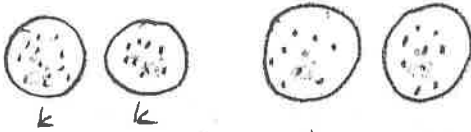
$$J = \sum_{i=1}^K \left(\sum_k \|x_k - c_i\|^2 \right)$$

↳ Hatayı verir



$x_k \rightarrow$ örnekler
 $c_i \rightarrow$ merkez

$(x_k - c_i) \rightarrow$ bir noktanın merkeze uzaklığı



Her 2 sinde 10 örnek var.

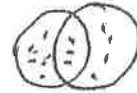
Hangisinin J değeri küçükse daha iyi bir kümelemedir.

İterasyon değiştikçe J değeri fazla olmuyorsa iterasyon bitirilebilir.

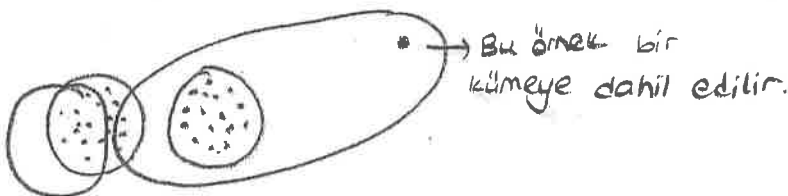
Dezavantajlar

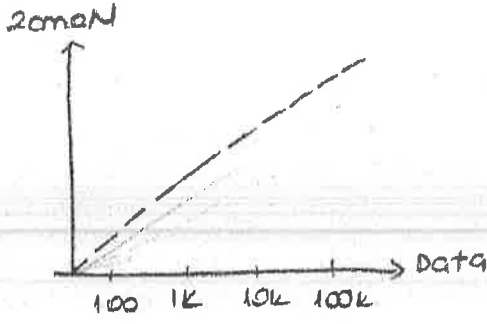
↳ Rasgele noktalarla dolayı çok karışık değildir.

↳ Kümeleme keskin yapar.



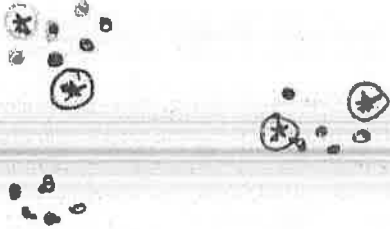
Overlapping





Veri arttikça lineerlik dilina çıkılmıyorsa bu ölçeklenebilir denir.

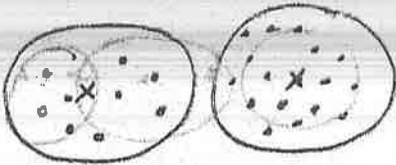
K-means örneği



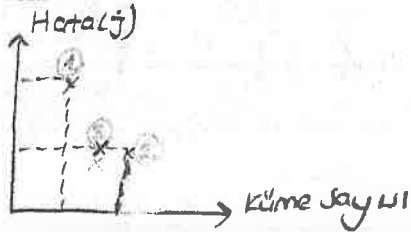
En son bulduğn küme her merkez etrafında örnekler vardır

Örnek sayısı kadar sıfır olursa hata minimum olur.

Soru: Bir kümeyi önce 2'ye sonra 3'e bölüyoruz. Hangi kümelemelerde hata az olur?



Küme sayısı ile hata oranı arasında doğrudan bir ilişki yoktur.



Küme sayısı olabildiğince az olsun isteriz.

Yukardaki 3 noktayı analiz edelim. birbirine üstünlüğünü yoktur. 3-2'ye baskındır. Çünkü aynı hata oranı daha az küme sayısı ile elde edilmektedir.

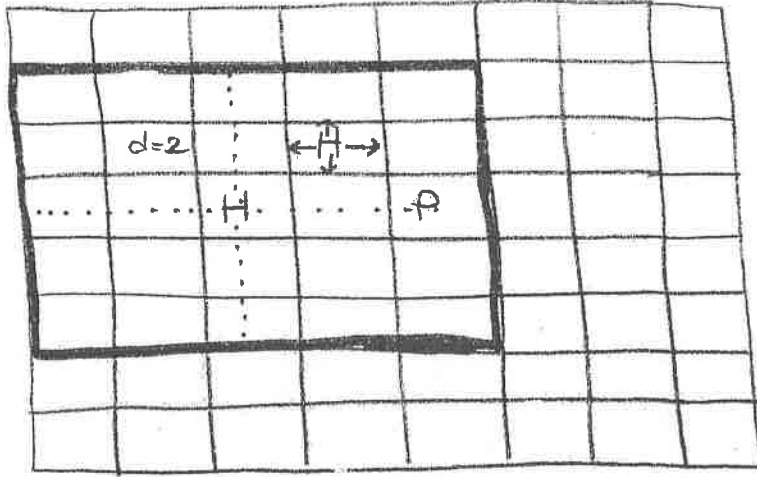
Hata oranı max olduğu zaman tek kümede bulunur bütün örnekler.

10/04/2014

TAKVİYELİ ÖĞRENME (Pekistirek öğrenme)



Eğitici öğrenmede sınıf ve eğitim verisi vardır.



$H \rightarrow Avcı$ $P \rightarrow Av$

Sadece 5 hareket var

H'nin görme derinliği olsun. (2)

H	P	Durum Sayısı
0	0	1
0	1	25
1	0	25
1	1	$+25 \times 25$
		676

$\rightarrow H$ bir kareye

konulup P 25 kare dolabilir

H ile P aynı karede bulunuyorsa P yakındır.

Not: Görme derinliği 2 olduğundan matris 5×5 'dir. Yani P bu 5×5 'lik matriste 25 birimdede bulunabilir.

$$v = (2d+1)^2 + 1$$

$n_P \rightarrow P$ 'nin sayısı (Av sayısı) $n_H \rightarrow Avcı$ sayısı

$$\begin{pmatrix} v \\ n_P \end{pmatrix} \times \begin{pmatrix} v \\ n_H - 1 \end{pmatrix}$$

$$\frac{v!}{(v-n_P)! \cdot n_P!}$$

$d \rightarrow$ görme derinliği

$v \rightarrow$ sonal derinlik

Bu problem için 676 durum gözlemleyebiliyoruz.

$Q(s,a)$	↑	↓	←	→	•	→ Hareketler(a)
s_1	0,1	0,1	0,1	0,1	0,1	
s_2						
s_3						
s_4						
s_5						
⋮						
s_{100}						
s_{101}						
⋮						
s_{676}						

"Öğrenme Algoritması:

1. Şu anki S durumunu gözlemle
2. a hareketini seç ve uygula
3. Yeni S durumunu gözlemle
4. Ortamdan bir r ödülünü al
5. Aşağıdaki formüle göre Q değerini güncelle

$$Q(s,a) = (1-\beta) Q(s,a) + \beta (r + \gamma \max_{a'} Q(s',a'))$$

$$Q_{t+1}(s,a) = Q_t(s,a) + \beta (r + \gamma \max_{a'} Q(s',a') - Q_t(s,a))$$

Tablodaki H, şu an S durumunu gözlemledik. 5 hareketten birini yaptık. Yeni S durumu gözlemledik. (Her 3 üde simultane hareketi yapar)

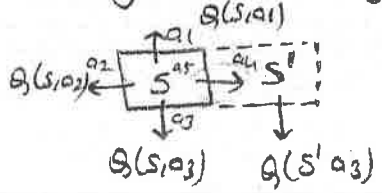
Her biri aynı hareketi yaparsa yine başlangıçtaki durum gözlenir. Tek fark görme derinliği kayar.

* Bu problemde av ile avcı aynı karede bulunursa ödül verilir.

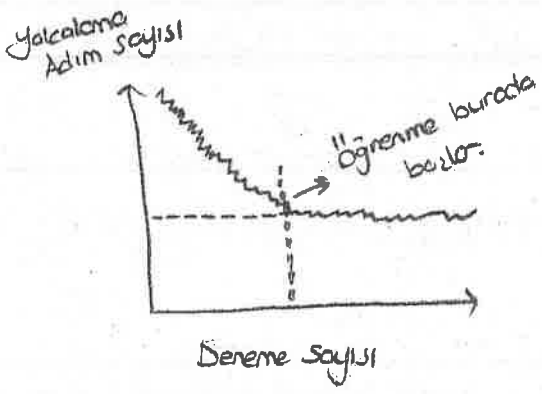
$\beta \rightarrow$ Öğrenme Oranı
 $r \rightarrow$ Ödül

S_1-S_{100} 'e geçme durumunda yakabildiyse S_1 'de sağa git hareketi bir anda diğer değerlere göre sızer. Bu aynı bu durumda olduğunda sağa git der.

Tablo güncellenince hangi değer yüksekse o hareket yapılmış olur.



"Ödül sadece av ile avcı aynı durumdayken verilir."



	<div> <div>Av</div> <div>AVCI</div> <div>1/1</div> </div>	<div> <div>Av</div> <div>AVCI</div> <div>1/2</div> </div>
Rasgele	∞	75000
Öğrenen Avcular	∞	100

