

# ÖZELLİK BELİRLEME

Konu  
Sınıfı  
Dersi

## Özellik Seçim Yöntemleri



Sinyal  
Gürültü  
Oranı

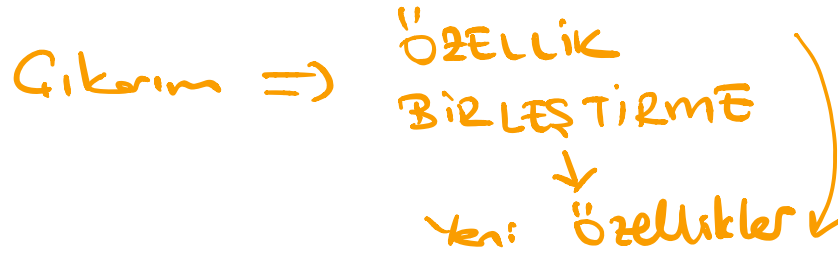
Alt  
Küme  
Seçimler



## Özellik Çıkarım Yöntemleri

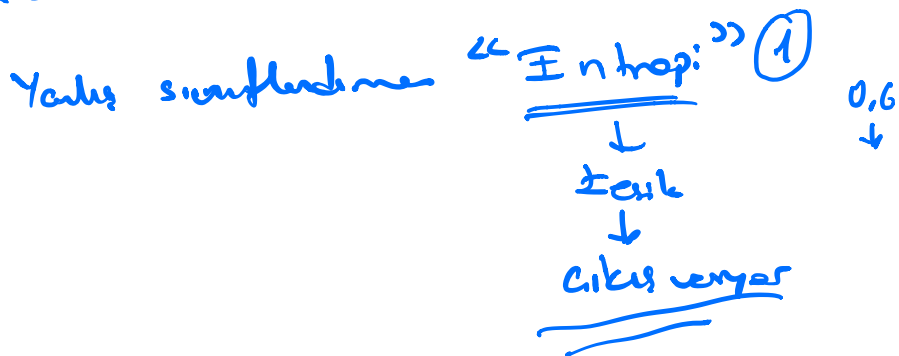
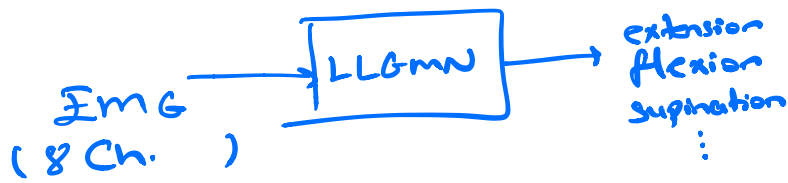
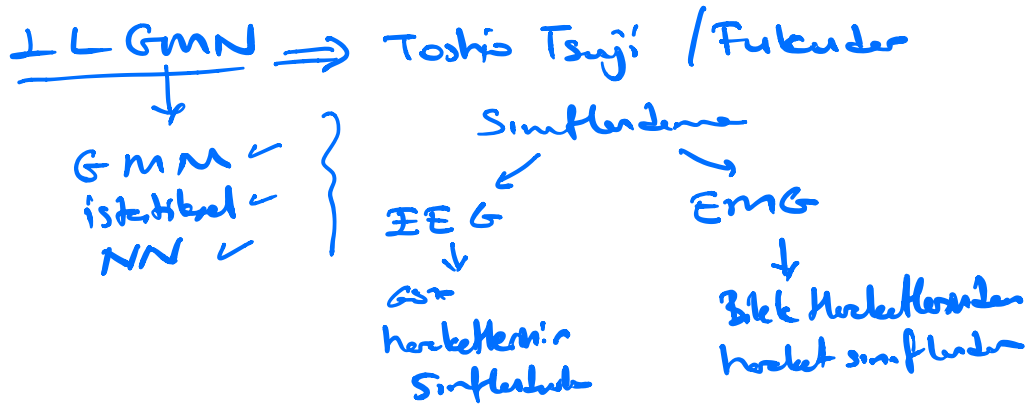
Temel Bileşen  
Analizi ✓

Deprusal Ayırt  
Eden Analizi ✓



Bilgi Kaynağı:  $0 \div 1$  aralığında

Entropi  $\Rightarrow$  Belirsizlik veya rastgelelik



Entropi  $\nearrow$   $\Rightarrow$  Kararsızlık / belirsizlik

S kaynağı  $\Rightarrow m_1, m_2, m_3, \dots, m_n$  n adet çıktı  
sırtlığını  
vorsejelen

$p_i \rightarrow m_i$  gerçekleşme olasılığı

$$P = \{ p_1, p_2, p_3, \dots, p_n \}$$

$H(S)$ : Entropi

$$H(S) = - \sum_{i=1}^n P_i \cdot \log_2(P_i)$$

Bilgi Kazancı  $\Rightarrow$   $IG(D) = H(D) - \sum_{i=1}^n P(D_i) H(D_i)$

D: Veri Kümesi      H: entropi

P: Olasılık

Örnek:

	Yaş	Mezuniyet	Sırt. Sahibi	Verilen Koor
1	Orta	Lise	Evet	iyi
2	Orta	Univ.	E	Fakir
3	Yaşlı	Lise	E	Zengin
4	Genç	Lise	Hayır	iyi
5	Genç	Univ.	E	Orta
6	Genç	Lise	H	iyi
7	Yaşlı	Univ.	H	iyi
8	Yaşlı	Univ.	E	Zengin
9	Yaşlı	Lise	E	iyi
10	Orta	Univ.	E	Fakir

5 örnek iyi, 2 örnek fakir, 2 örnek zengin  
1 örnek orta

$$\begin{aligned} H(S) &= - \left(\frac{5}{10}\right) \log_2 \left(\frac{5}{10}\right) - \left(\frac{2}{10}\right) \log_2 \left(\frac{2}{10}\right) \\ &\quad - \left(\frac{1}{10}\right) \log_2 \left(\frac{1}{10}\right) - \left(\frac{2}{10}\right) \log_2 \left(\frac{2}{10}\right) \\ &= - 1.7609 \end{aligned}$$

$G(S, \text{Yaş}) \Rightarrow 3 \text{ orta} \begin{cases} \nearrow 1 \text{ iyi} \\ \searrow 2 \text{ fakir} \end{cases}$

$$* H(S_{\text{orta}}) = -(1/3) \log_2(1/3) - (2/3) \log_2(2/3) = 0.918$$

4 Yaşlı durumu  $\begin{cases} \nearrow 2 \text{ zengin} \\ \searrow 2 \text{ iyi} \end{cases}$

$$* H(S_{\text{yaşlı}}) = -(2/4) \log_2(2/4) - (2/4) \log_2(2/4) = 1$$

3 Genç  $\begin{cases} \nearrow 2 \text{ iyi} \\ \searrow 1 \text{ orta} \end{cases}$

$$* H(S_{\text{genç}}) = -(2/3) \log_2(2/3) - (1/3) \log_2(1/3) = 0.918$$

$$G(S, \text{Yaş}) = H(S) - (P_{\text{orta}} H(S_{\text{orta}})) + P_{\text{yaşlı}} H(S_{\text{yaşlı}}) + P_{\text{genç}} H(S_{\text{genç}})$$

$$= 1.7609 - ((3/10 * 0.918) + (4/10 * 1) + ((3/10) * (0.918))) = 0.8101$$

mesureet / şirket sahibi - - -

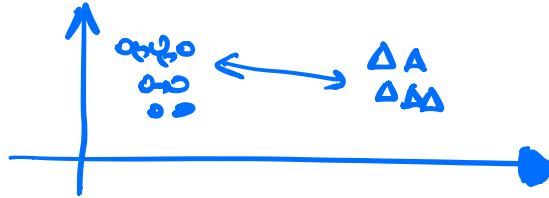
$$\text{mesureet} \rightarrow G(S, \text{mesureet}) = 0.4394$$

$$\text{şirket sahibi} \rightarrow G(S, \text{şirket sahibi}) = 0.3959$$

## Sinyal Güçlüğü Oranı (S2M Ratio)

Sınıflar arası özellikler farkları

sınıf içi                      "                      " z



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

$m_1$ : 1. sınıftaki özelliklerin ort.

$m_2$ : 2. " " "

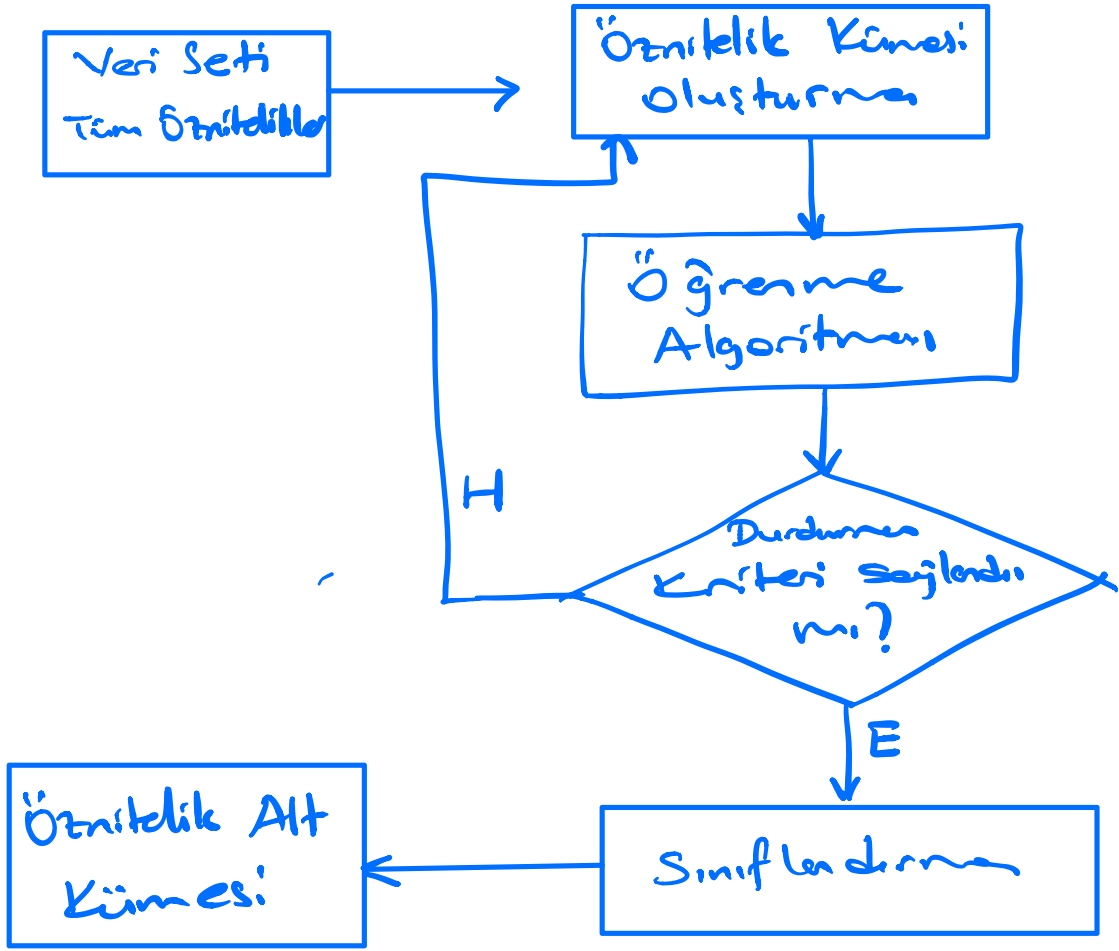
$d_1$ : 1. " " standart sapması

$d_2$ : 2. " " "

## Alt Küme Seçiciler (Wrappers)

Yavaş çalışırlar

Sınıflandırma başarısı yüksek



# Özellik Çıkarım Yöntemleri

## 1) Temel Bileşen Analizi:

1901 Karl Pearson

1933 Hotelling

Amaç boyut azaltma

Verinin temel yapısı

BOZULMAYACAK

VARİYANS

1. adım → Kovaryans

2. adım → özdeğerler → büyük  
ve  
özvektörleri

Eğer  $\lambda = 0$  ise

İlgili örnekler diğer vektörlerin bileşenleri olarak gösterilebilir.

- Gözaltıye karşı dayanıklı
- Kaparite / bellek ihtiyacı düşük
- Yüksek indekslere kabiliyeti

2) Doğrusal Ayırt Eden Analizi  
(Linear Discriminant Analysis)

- Ronald Aylmer Fisher

aynı sınıfa ait veriler birlikte analiz edilir.

---

Sınıf içi varyans  $\rightarrow$  en az

Sınıf dışı varyans  $\rightarrow$  en yüksek



1. adım: Veri kümesinin her bir elemanının ortalaması dır. Veri sayısı ile ortalamaya vektörü eklenir.

2. adım: Kovaryans matrisi hesaplanır.

3. adım: Özüektörleri, Özüdeğerler

4. adım: Özüdeğerlerin artan sıradan küçükten büyükten Özüektörleri sıralanır.

$k$  adet Özüektörü seçer  
 $d \times k$  boyutlu  $V$  ortak kovaryans matrisi oluşturulur

5. adım: Veri vektörlerini yeni alt uzayda denükleme için  $V$  matrisi denükleme matrisi olarak kullanılır.