

# Veri Madenciliği

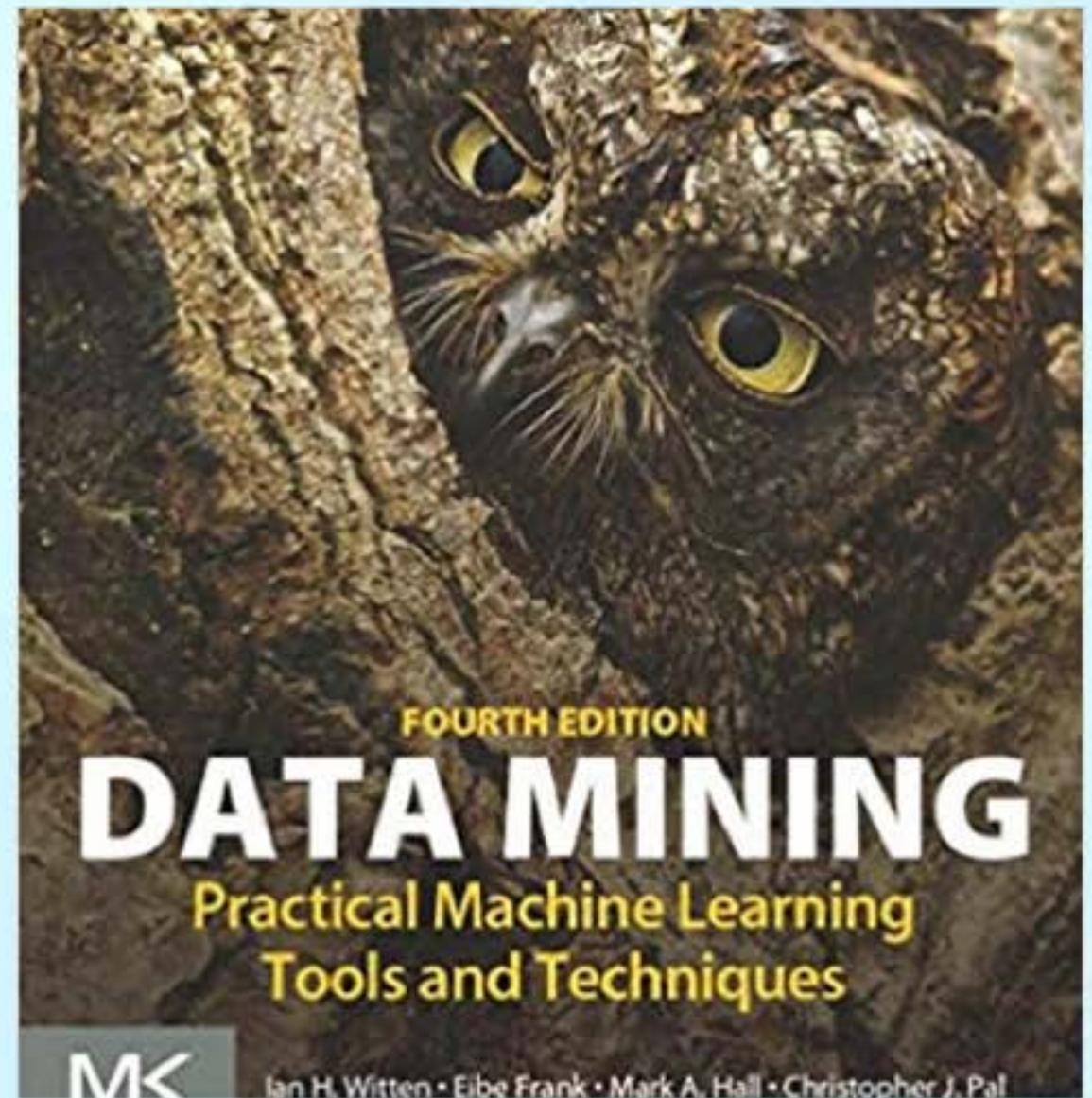
Güz 2023

Ders 11

Çıktı: Kurallar

# Dersin Kitabı

- Data Mining: Practical Machine Learning Tools and Techniques, 4th Ed., by Ian Witten, Eibe Frank, Mark Hall, and Christopher Pal (Morgan Kaufmann Publishers, 2017. ISBN: 978-0-12-804291-5)





# Örnek Tabanlı Gösterim

- Öğrenmenin en basit şekli: **ezberci öğrenme**
  - Eğitim örnekleri, yeni örneğe en çok benzeyen örnek için aranır
  - Örneklerin kendileri bilgiyi temsil eder
  - **Örnek-tabanlı** öğrenme olarak da adlandırılır

## Kurallardan farkı (ağaçlar vb.):

- Sadece örnekleri saklayın; işi ertelemek ("tembel" öğrenme)
- Kural oluşturmanıza ve saklamanıza gerek yoktur.
- yenisine en yakın olanı bulmak için mevcut örneklerden çalışın.



# Örnek Tabanlı Gösterim

- Benzerlik işlevi neyin "öğrenildiğini" tanımlar
- **Yöntemler:**
  - en yakın komşu (sınıflandırılması gereken örneği bulmak için bir uzaklık fonksiyonunu kullanır)
  - k-en yakın komşu (birden fazla komşuyu kullanır; k-komşulardan çoğunluğunu kullanarak sınıflandırma yapar)

*Yöntemin eleştirisi: Hiçbir yapı "öğrenilmez" - örnekler kalıpları tanımlamaz.*

*Savunucuları diyor ki: uzaklık fonksiyonuyla birleştirilmiş örnekler "yapıdır".*



# Mesafe Fonksiyonu

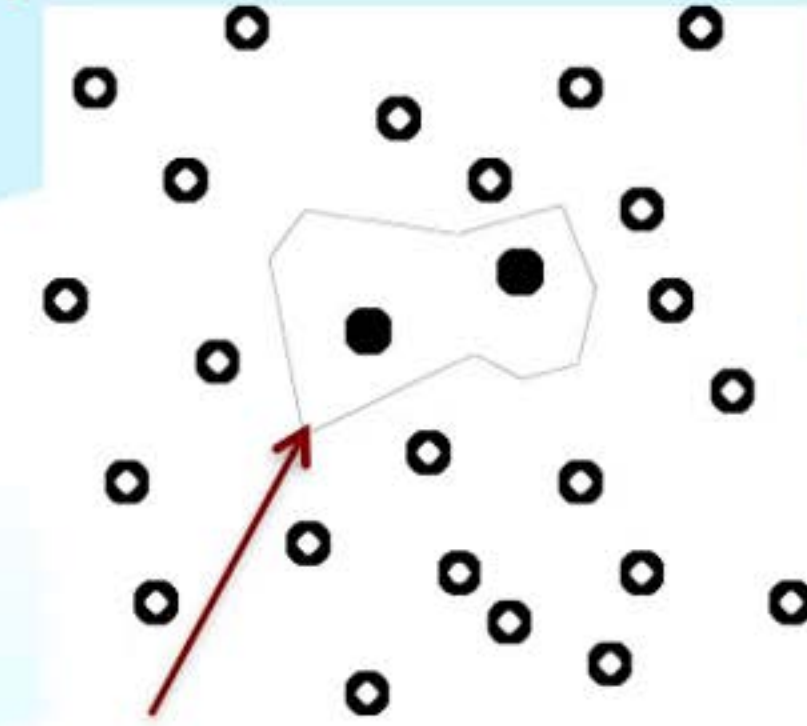
- **En basit durum:** yalnızca **bir** sayısal öznitelik
  - Uzaklık, ilgili iki öznitelik değeri (i.e. burada fonksiyon) arasındaki farktır
- **Birkaç sayısal öznitelik:** normalde, **Öklid mesafesi** kullanılır ve öznitelikler normalleştirilir
- **Nominal öznitelikler:** değerler farklıysa uzaklık 1'e, eşitse 0'a ayarlanır

**Tüm öznitelikler eşit derecede önemli mi?**

- Öznitelikleri ağırlıklandırmak gerekli olabilir

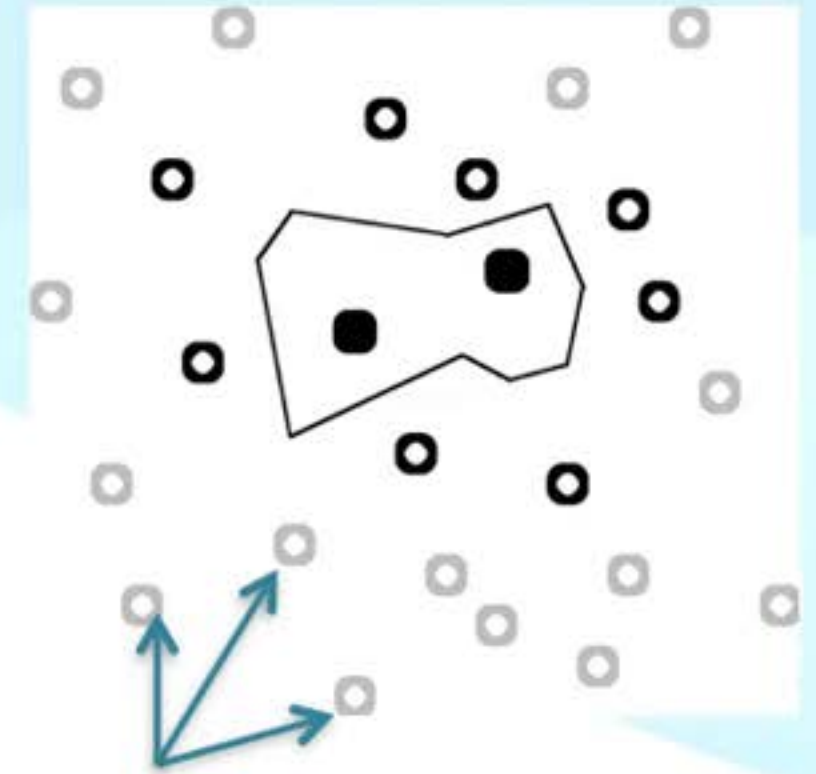
İki sınıf:  
dolu daireler ve  
açık daireler

# Öğrenme Prototipleri



En yakın  
komşuyu  
bölün

Örnek alanını  
bölümlemenin  
farklı yolları  
(yani, her  
sınıfın  
yalnızca  
kritik  
örneklerini  
kaydedebilirsiniz)



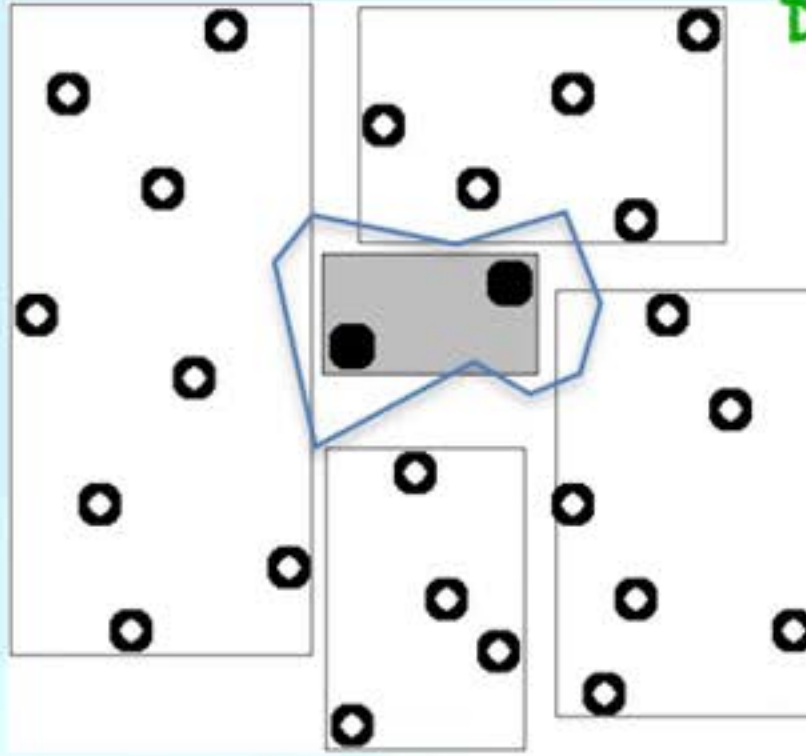
Gri olanları atın

- Yalnızca bir kararda yer alan örneklerin saklanması gerekir (tüm örnekleri depolamak istemiyoruz)
- Gürültülü örnekler filtrelenmelidir
- Fikir: yalnızca prototip örnekleri kullanın



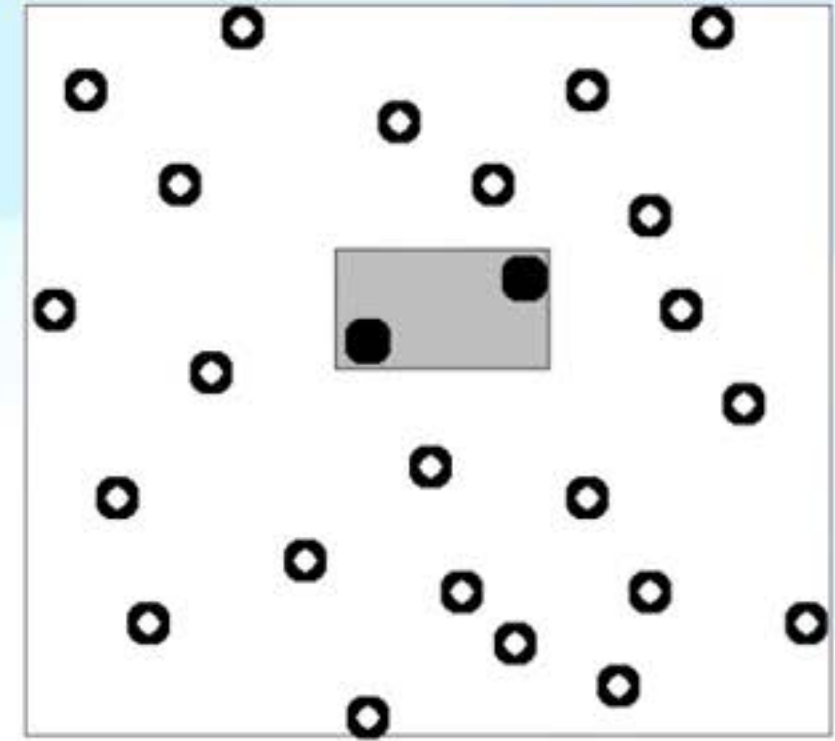
# Daha İleri Gidersek: Dikdörtgenlerle

## Genelleştirin



Dikdörtgenler  
aynı sınıfı  
kapsar

İç içe  
yerleştirme, iç  
bölgenin  
farklı bir  
sınıfa sahip  
olmasını  
sağlar

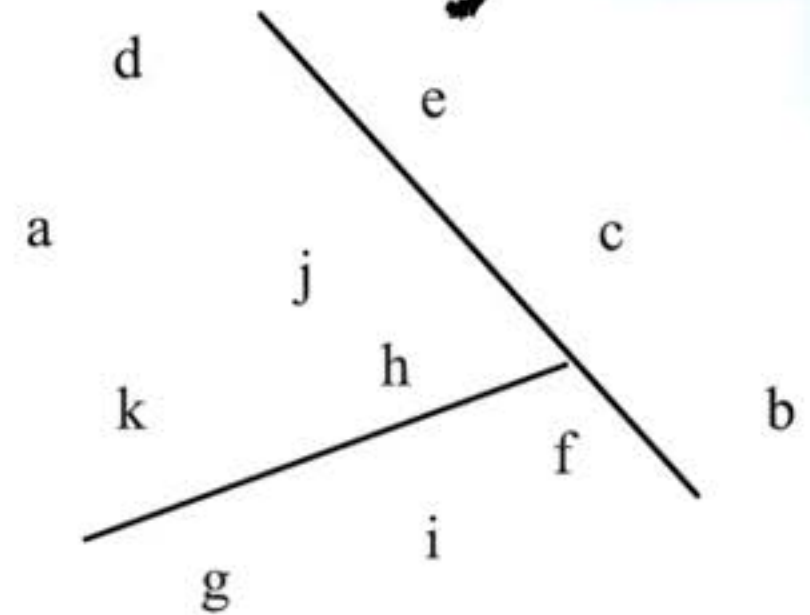


- Aynı dikdörtgene düşerse, aynı sınıf
  - ancak önceki slayttaki en yakın komşudan farklı karar sınırı.
- Dikdörtgenlerin dışında en yakın komşu kuralı kullanılır
- Dikdörtgenler kuraldır! (Ancak "normal" kurallardan daha muhafazakar olabilirler.)
- İç içe dikdörtgenler istisnalar dışında kurallardır
- *Not: Nominal bölgeleri görselleştirmek zordur, çok boyutludur*

# Küme Gösterimi I

- Bir küme öğrenildiğinde, çıktı diyagram şeklini alır
- **En basit durum:**
  - her örneğe bir küme numarası atayın, örnekleri ve bölümleri düzenleyin.

**Basit 2-D  
Gösterimi**

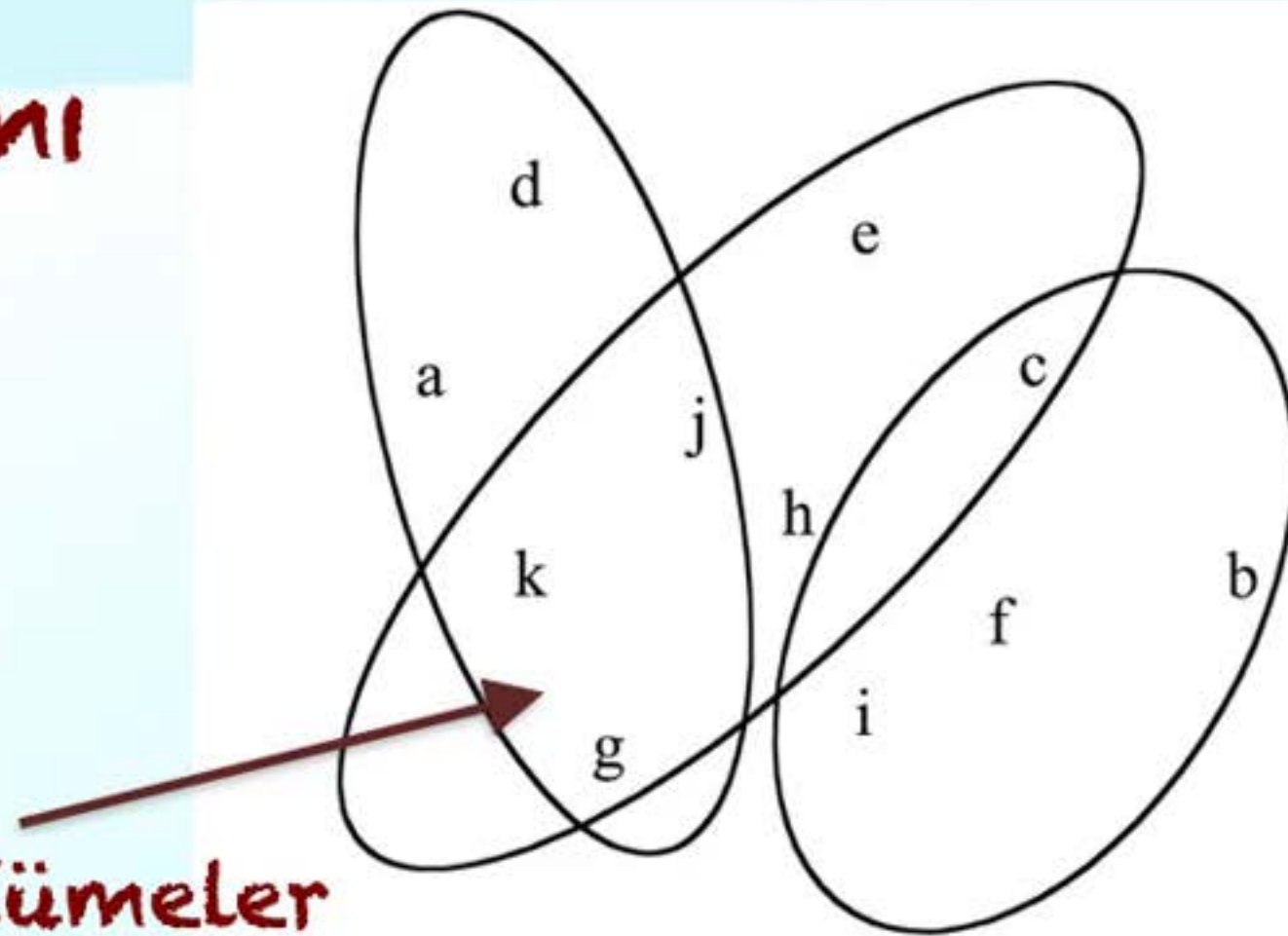




# Kümeleri Temsil Eden II

- Bazı kümeleme algoritmaları kümelerin çakışmasına izin verir.

Venn  
Diyagramı



Çakışan Kümeler



# Kümeleri Temsil Etme III

## Olasılıksal Atama

Bazı algoritmalar olasılıkları kullanır. Her örnek için 1, 2 veya 3 kümelere bir üyelik derecesi vardır.

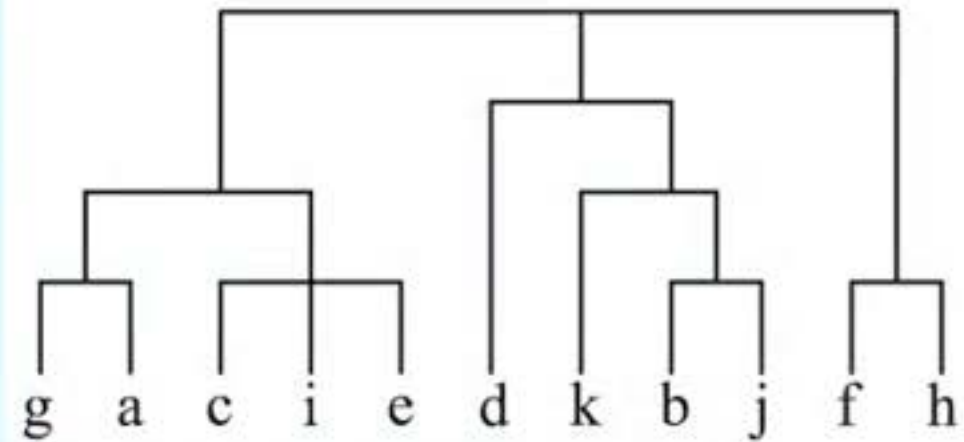
	1	2	3
a	0.4	0.1	0.5
b	0.1	0.8	0.1
c	0.3	0.3	0.4
d	0.1	0.1	0.8
e	0.4	0.2	0.4
f	0.1	0.4	0.5
g	0.7	0.2	0.1
h	0.5	0.4	0.1
...			



# Kümeleri Temsil Eden IV

Burada, diyagramın  
"yaprakları" ndaki  
kümeler, daha  
yüksek  
düzeylerden daha  
sıkı kümelenir.

## Dendrogram



NOT: Dendron  
Yunanca ağaç  
kelimesidir