

Veri Madenciliği

Güz 2023

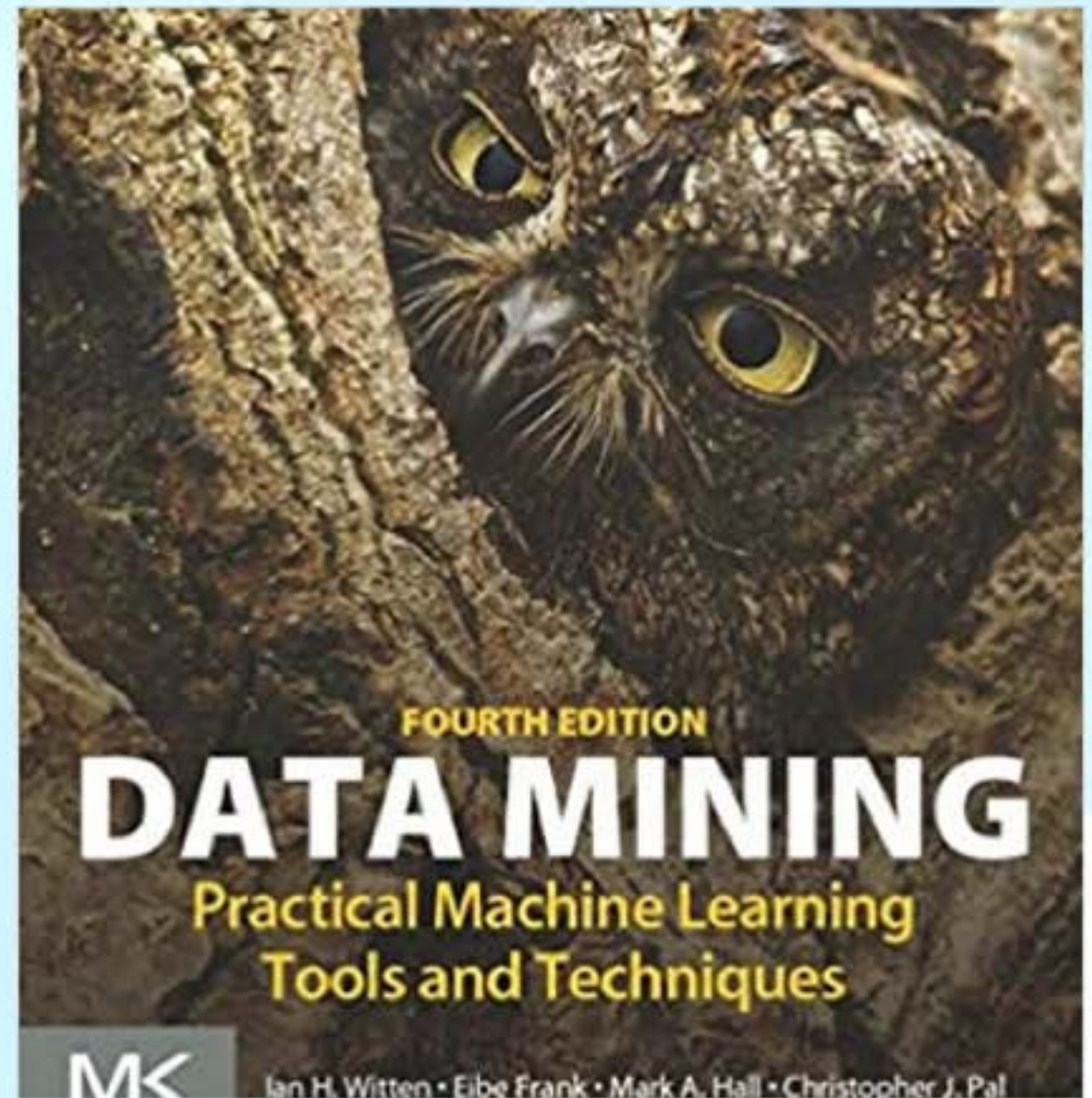
Ders 12

Temel Algoritmalar

1R

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)



Denetimli ve Denetimsiz Öğrenme

- **Denetimli öğrenme (sınıflandırma)**
 - **Denetim:** Eğitim verilerine (gözlemler, ölçümler vb.) gözlemlerin sınıfını gösteren **etiketler** eşlik eder.
 - Yeni veriler eğitim kümesine göre sınıflandırılır
- **Denetimsiz öğrenme (kümeleme)**
 - Eğitim verilerinin sınıf etiketleri bilinmiyor
 - Verilerde sınıfların veya kümelerin varlığını belirlemek amacıyla bir dizi ölçüm, gözlem vb.

Tahmin Sorunları:

Sayısal Tahmine Karşı Sınıflandırma

■ Sınıflandırma

- kategorik sınıf etiketlerini tahmin eder (ayrık veya nominal)
- Eğitim kümesine ve sınıf etiketlerine göre bir model oluşturur.
- Yeni verileri sınıflandırmada modeli kullanır

■ Sayısal Tahmin

- sürekli-değere sahip fonksiyonları modeller.
- Bilinmeyen veya eksik değerleri tahmin etmek için fonksiyon kullanır

Sınıflandırma:

İki Aşamalı İşlem

Adım 1: Model yapımı:

Önceden belirlenmiş sınıflar kümesini tanımlama

- Her örneğin önceden tanımlanmış bir sınıfa ait olduğu varsayılır.
- **Eğitim seti:** model yapımı için kullanılan örnekler
- **Model:** sınıflandırma kuralları, karar ağaçları veya matematiksel formüller olarak temsil edilir

Sınıflandırma: İki Aşamalı İşlem

Adım 2: Model kullanımı:

Gelecekteki veya bilinmeyen nesneleri sınıflandırmak için

- Modelin doğruluğunu tahmin edin
 - "Test örneğinin bilinen etiketi" modelden elde edilen sınıflandırılmış sonuçla karşılaştırılır
 - **Doğruluk oranı**: test seti örneklerinin doğru sınıflandırılan yüzdesi
 - **Test seti**: eğitim kümesinden bağımsız (aksi takdirde aşırı uyum gösterir (overfitting))
- **Doğruluk kabul edilebilir düzeydeyse**, sınıf etiketleri bilinmeyen veri örneklerini sınıflandırmak için modeli kullanın.

Süreç: Model Yapımı



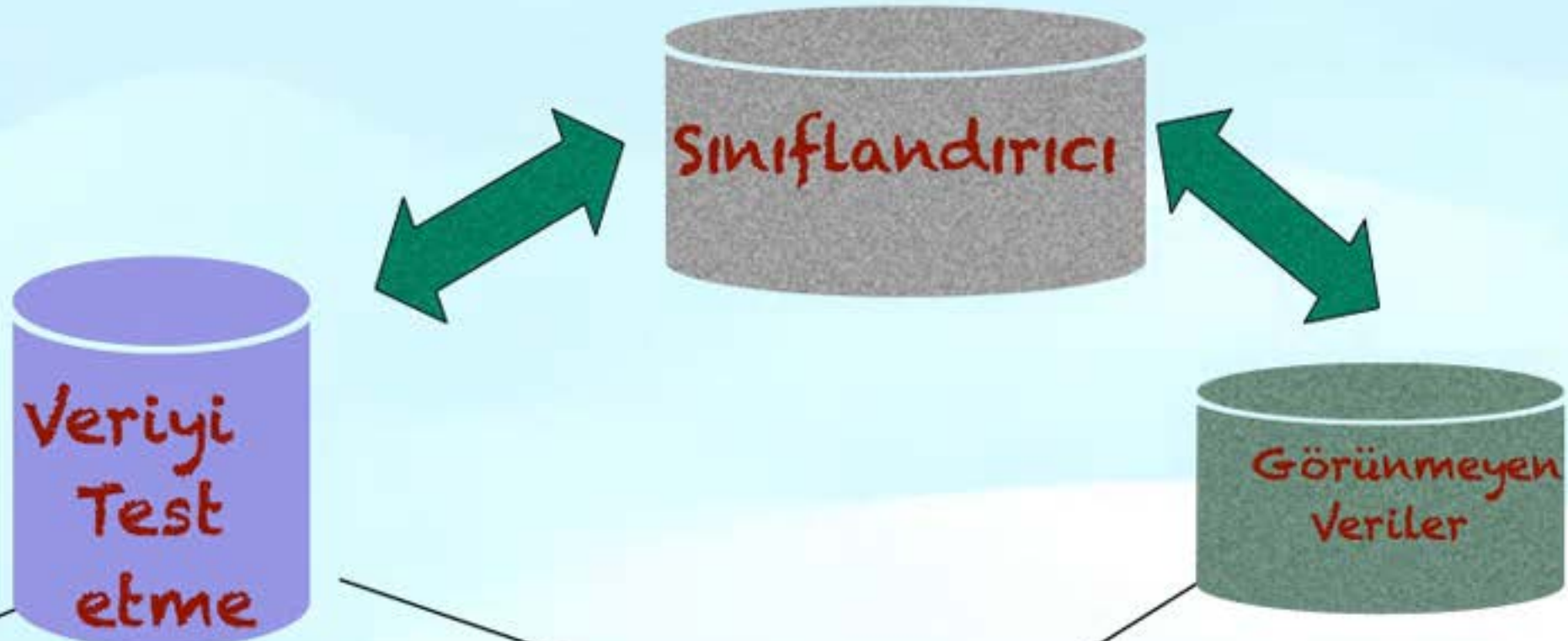
Sınıflandırma
Algoritma

İsim	Ünvan	Yıl	Kadrolu mu?
Mehmet	Dr. Öğr. Üyesi	3	Hayır
Merve	Dr. Öğr. Üyesi	7	Evet
Bilal	Profesör	2	Evet
Jale	Doç. Dr.	7	Evet
Deniz	Dr. Öğr. Üyesi	6	Hayır
Ali	Doç. Dr.	3	Hayır

Sınıflan
dırıcısı
(Model)

IF ünvan = profesör
OR yıl > 6
THEN kadrolu = evet

İşlem: Tahminde Modeli Kullanma



İsim	Ünvan	Yıl	Kadrolu mu?
Tamer	Dr. Öğr. Üyesi	2	Hayır
Melisa	Doç. Dr.	7	Hayır
Gökhan	Profesör	5	Evet
Cihan	Dr. Öğr. Üyesi	7	Evet

(Haluk, Profesör, 4)

Kadrolu?

EVET

Önce Basitlik

- **İyi Haber:** Basit algoritmalar genellikle çok iyi çalışır!
- Birçok basit yapı türü vardır, örneğin.:
 - Tüm işi bir öznelite yapar
 - Tüm özneliteler eşit ve bağımsız olarak katkıda bulunur
 - Ağırlıklı bir doğrusal kombinasyon yapabilir
 - **Örnek-tabanlı:** birkaç prototip kullanılır
 - Basit mantıksal kurallar kullanılır
 - Denetimsiz bile olabilir
- Yöntemin başarısı ilgi alanına bağlıdır

İlkel Kuralların Çıkarılması

1R: 1 seviyeli bir karar "**ağacı**" öğrenir

- yani, hepsinin belirli bir özelliği test ettiği kurallar

Temel fikir

- Her değer için bir dal oluşturun
- Her dal en sık görülen sınıfı atar
- **Hata oranı:** karşılık gelen dallarının çoğunluk sınıfına ait olmayan örneklerinin oranı
- En düşük hata oranına sahip özneliği seçin

...nominal nitelikler olduğunu varsayalım

1R için Söзде-kod

Her öznitelik için,

Özniteliğin her değeri için aşağıdaki gibi bir kural yapın:

- ▶ her sınıfın ne sıklıkta görüldüğünü sayın
- ▶ en sık görülen sınıfı bul
- ▶ kuralın o sınıfı bu öznitelik-değerine atamasını sağlayın

Kuralların hata oranını hesaplayın

En küçük hata oranına sahip kuralı seçin

Not: "eksik" ayrı bir öznitelik değeri olarak kabul edilir
Yani, görünüm değeri bilinmeyen bir örnek olsaydı,
güneşli, kapalı, yağmurlu ve eksik olmak üzere 4
değer olurdu.

Hava Durumu Niteliklerini Hatırlama

Görünüm	Sıcaklık	Nemi	Rüzgarlı	Oyun
Güneşli	Sıcak	Yüksek	Yanlış	Hayır
Güneşli	Sıcak	Yüksek	Doğru	Hayır
Bulutlu	Sıcak	Yüksek	Yanlış	Evet
Yağmurlu	Hafif	Yüksek	Yanlış	Evet
Yağmurlu	Serin	Normal	Yanlış	Evet
Yağmurlu	Serin	Normal	Doğru	Hayır
Bulutlu	Serin	Normal	Doğru	Evet
Güneşli	Hafif	Yüksek	Yanlış	Hayır
Güneşli	Serin	Normal	Yanlış	Evet
Yağmurlu	Hafif	Normal	Yanlış	Evet
Güneşli	Hafif	Normal	Doğru	Evet
Bulutlu	Hafif	Yüksek	Doğru	Evet
Bulutlu	Sıcak	Normal	Yanlış	Evet
Yağmurlu	Hafif	Yüksek	Doğru	Hayır

Hava Durumu Özelliklerini Değerlendirme

<u>Öznitelik</u>	<u>Kurallar</u>	<u>Hatalar</u>	<u>Toplam-Hatalar</u>
------------------	-----------------	----------------	-----------------------

Sıcaklık

Nem

Rüzgârlı