Örüntü Tanıma

Yapay Zekâ ve Veri Dersleri Rehberi · Kısım II: Yapay Zekânın Dalları

Öznitelik çıkarma | Bayes kararı | Sınıflandırıcılar | Kümeleme | Başarım ölçme | Komşu dersler

Sercan KÜLCÜ · Giresun Üniversitesi · Bilgisayar Mühendisliği

© Sercan Külcü. Tüm hakları saklıdır. · Sunumun etkileşimli HTML sürümü: https://sercankulcu.github.io/files/ai_courses/sunumlar/Bolum_08_Oruntu_Tanima.html

GİRİŞ

Ders Hangi Soruyu Sorar?

> **Ölçümlerden bir nesnenin hangi sınıfa ait olduğuna nasıl karar verilir?**

Örüntü tanıma, ölçümlerden çıkarılan **öznitelik vektörüne** bakarak bir nesnenin hangi sınıfa ait olduğuna karar vermeyi inceler. Ders algılayıcıdan karara uzanan zinciri bütün olarak ele alır: ön işleme, öznitelik seçimi, sınıflandırıcı tasarımı ve hata çözümlemesi.

Bir bakışta

YAKLAŞIM istatistiksel karar kuramı	**GİRDİ** ölçüm ya da öznitelik vektörü	**ÇIKTI** sınıf etiketi
BAŞARI ÖLÇÜTÜ sınıflandırma hatası	**MATEMATİK YÜKÜ** orta-yüksek: olasılık, doğrusal cebir	**VERİ İHTİYACI** etiketli örnekler

Kısa Tarihçe

1936
Fisher doğrusal ayırtaç analizini yayımlar; makalede kullandığı Iris verisi bugün de ilk örnektir.

1962
Hu, öteleme, ölçek ve dönmeye karşı değişmeyen moment özniteliklerini tanıtır.

1967
Cover ve Hart, en yakın komşu kuralının hatasının Bayes hatasının iki katını aşmadığını gösterir.

1973
Duda ve Hart, alanın temel ders kitabı "Pattern Classification and Scene Analysis"i yayımlar.

1991
Turk ve Pentland özyüzlerle (eigenfaces) yüz tanımayı PCA üzerine kurar.

1998
LeCun ve arkadaşları LeNet-5'i ve MNIST el yazısı rakam verisini yayımlar.

DERS

Temel Kavramlar

Öznitelik vektörü
Nesneyi temsil eden ölçümler: uzunluk, renk, doku.

Öznitelik uzayı
Her örnek bir nokta; sınıflar bu uzayda bölgeler.

Bayes karar kuramı
Sonsal olasılığı en büyük sınıfı seçmek; en az hata.

Karar sınırı
Öznitelik uzayını sınıf bölgelerine ayıran yüzey.

Yoğunluk kestirimi
Parametrik (Gauss) ya da parametresiz (Parzen).

En yakın komşu
Yeni örneği en yakın etiketli örneklerin sınıfına ata.

Boyut indirgeme
PCA ve LDA ile az ama ayırt edici öznitelik.

Karışıklık matrisi
Hangi sınıfın hangisiyle karıştığını gösterir.

Tipik İş Akışı



Örnek: Bir posta ayırma makinesi zarftaki posta kodunu tarar, görüntüyü ikili hâle getirip rakamlara böler, her rakamdan biçim öznitelikleri çıkarır, en yakın komşu ya da SVM ile hangi rakam olduğuna karar verir ve yanlış okunan zarfları elle ayırmaya gönderir.

DERS

Derste Neler İşlenir?

1. Örüntü tanıma sistemi: algılama, öznitelik, karar
2. Bayes karar kuramı, risk ve kayıp
3. Normal dağılım ve ayırtaç fonksiyonları
4. Parametre kestirimi: en büyük olabilirlik, Bayes
5. Parametresiz yoğunluk kestirimi: Parzen
6. En yakın komşu kuralı ve uzaklık ölçüleri
7. Doğrusal ayırtaçlar ve algılayıcı
8. Destek vektör makineleri
9. Öznitelik çıkarma, seçme ve boyut laneti
10. PCA ve Fisher doğrusal ayırtaç analizi (LDA)
11. Sınıflandırıcıları değerlendirme ve karşılaştırma
12. Kümeleme ve Gauss karışım modelleri

Kazandırdığı beceriler

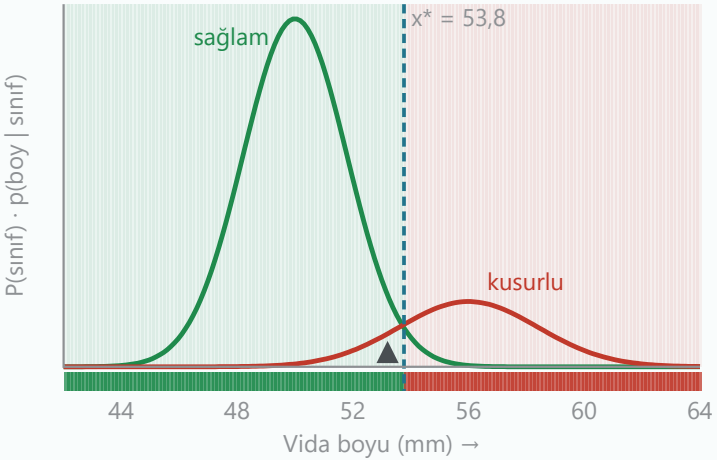
ölçümden öznitelik tasarlamak · Bayes sınıflandırıcı kurmak · scikit-learn ile k-NN ve SVM · OpenCV ile görüntüden öznitelik çıkarmak · karışıklık matrisiyle hata çözümlemek

Tipik bir dönemin akışıdır; sıra ve kapsam üniversiteden üniversiteye değişir.

DENE

Kusurlu Parçayı Ayıkla

Üretim hattında her vidanın boyu ölçülüyor; kusurlu vidalar ortalamada daha uzun ama dağılımlar örtüşüyor. Kusur oranını ve hata maliyetini değiştirin, Bayes karar sınırının nasıl kaydığını izleyin; grafiğe tıklayıp parça ölçün.



Kusurlu oranı (önsel) P(K) = **%20**

Kusurluyu kaçırmak, sağlamı atmaktan **1** kat pahalı

Başa dön

Karar sınırı: **53,8 mm**
Kaçan kusurlu %17,1 · boşa atılan sağlam %1,9
Toplam hata **%5,0** · 1000 parçada maliyet 50

Ölçülen parça (▲) 53,2 mm: P(K | boy) = %31,6 → **kabul**.
Maliyet artınca sınır sola kayar: toplam hata artar, pahalı hata azalır.

Örüntü tanıma kararı **olasılık ve maliyetle** verir: aynı ölçüm, kusur oranı ve hatanın bedeli değişince başka sınıfa atanır. Bayes kuralı beklenen kaybı en aza indiren sınırı verir.

KONUM

Ön Koşullar

Bu dersten önce bilinmesi yararlı olan dersler ve neden gerektikleri.

Ders	Neden gerekir
Olasılık ve İstatistik	Bayes karar kuramı, yoğunluk kestirimi
Programlama	Python ya da Java ile dizi, döngü ve fonksiyonlar
Doğrusal Cebir	vektör, uzaklık, özdeğer: PCA ve LDA'nın dili

KONUM

Bu Ders Sana Göre mi?

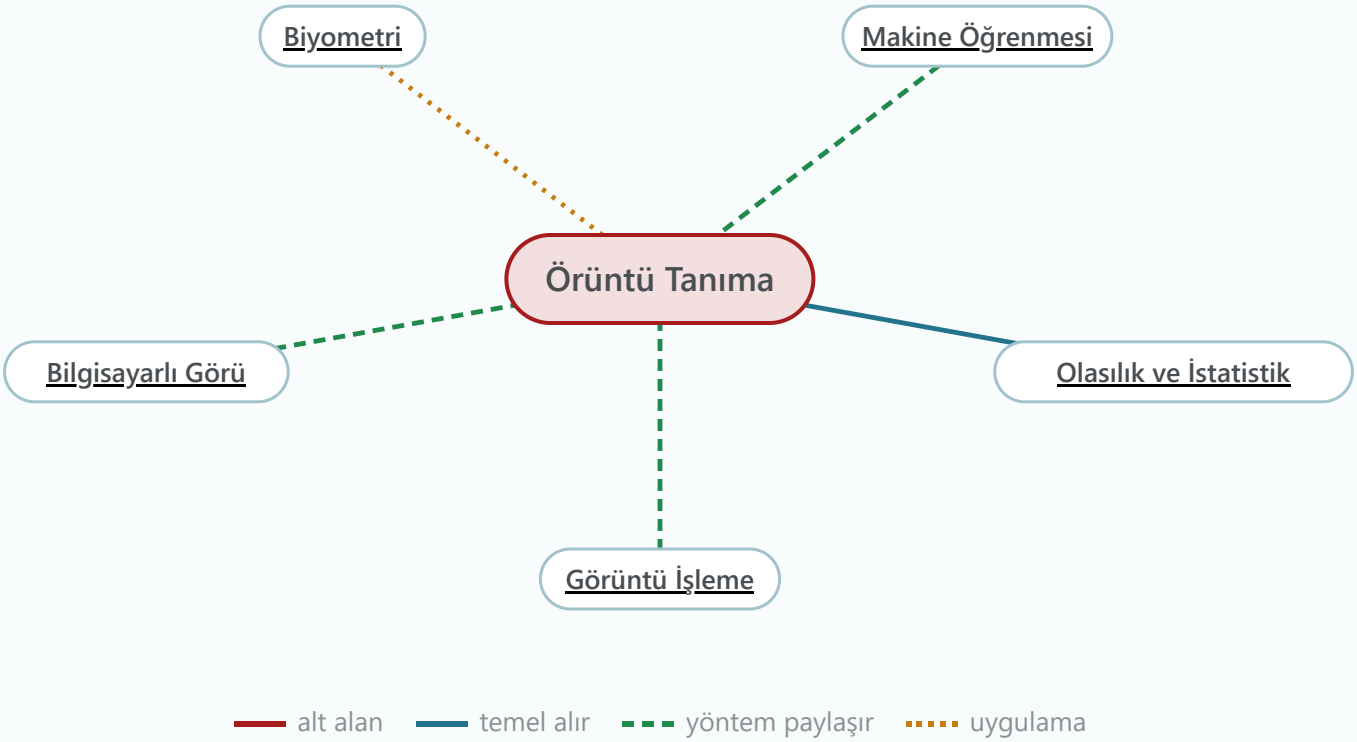
Kendine sor; "hayır" dediklerin için sağdaki sütun nereden başlayacağını gösterir.

Kendine sor	Hayırsa önce
Normal dağılımı, ortalamayı ve kovaryansı açıklayabiliyor musun?	Olasılık ve İstatistik
Öklid uzaklığını, özdeğeri ve özvektörü açıklayabiliyor musun?	Doğrusal Cebir
Bir fonksiyonun en küçük değerini türev alarak bulabiliyor musun?	Analiz dersindeki türev ve uç değer konuları
Python'da NumPy dizileriyle bir veri setini okuyup işleyebiliyor musun?	Programlamaya giriş, ardından NumPy'a kısa bir giriş

Çoğuna "evet" diyebiliyorsan derse hazırsın; bir ya da iki "hayır" dönem içinde kapanabilir.

KONUM

Dersin Komşuları



ÖRTÜŞME

Örtüştüğü Dersler

İlişki	Ders	Ortak konu
temel alır	Bu ders Olasılık ve İstatistik dersini temel alır	Bayes karar kuramı, yoğunluk kestirimi
uygulama	Biyometrik Sistemler bu dersin yöntemlerini uygular	eşleme, doğrulama ve tanımlama
yöntem paylaşır	Makine Öğrenmesi ile yöntem paylaşır	sınıflandırıcılar, karar sınırları
yöntem paylaşır	Görüntü İşleme ile yöntem paylaşır	öznitelik çıkarma, şablon eşleme
yöntem paylaşır	Bilgisayarlı Görü ile yöntem paylaşır	nesne sınıflandırma

Ayrıştığı Noktalar (1/2)

Ölçüt	Örüntü Tanıma	Makine Öğrenmesi	Bilgisayarlı Görü
Temel soru	Ölçümlerden bir nesnenin hangi sınıfa ait olduğuna nasıl karar verilir?	Bir program örneklerden nasıl genelleme yapar?	Bilgisayar bir görüntüde ya da videoda ne olduğunu nasıl anlar?
Yaklaşım	istatistiksel karar kuramı	istatistiksel öğrenme	görüntüden anlam çıkarma
Girdi	ölçüm ya da öznitelik vektörü	etiketli ya da etiketsiz örnekler	görüntü ya da video
Çıktı	sınıf etiketi	tahmin yapan model	etiket, kutu, derinlik, 3B yapı
Başarı ölçütü	sınıflandırma hatası	görülmemiş veride hata	doğruluk, mAP, IoU
Matematik yükü	orta-yüksek: olasılık, doğrusal cebir	yüksek: doğrusal cebir, olasılık, optimizasyon	yüksek: geometri, doğrusal cebir
Veri ihtiyacı	etiketli örnekler	orta ve çok	çok etiketli görüntü

- **Makine Öğrenmesi:** MÖ öğrenme algoritmalarını genel olarak inceler; bu ders ölçümden sınıf kararına giden tasarımı ve Bayes hatasını odağa alır.
- **Bilgisayarlı Görü:** Bilgisayarlı görü görüntüde sahneyi anlamaya çalışır; bu ders herhangi bir ölçüm türünden sınıf kararı vermeyi genel olarak ele alır.

Ayrıştığı Noktalar (2/2)

Ölçüt	Örüntü Tanıma	Görüntü İşleme	Biyometri
Temel soru	Ölçümlerden bir nesnenin hangi sınıfa ait olduğuna nasıl karar verilir?	Bir görüntü nasıl iyileştirilir, dönüştürülür ve parçalara ayrılır?	Bir kişi yüzünden, parmak izinden ya da sesinden nasıl tanınır?
Yaklaşım	istatistiksel karar kuramı	piksel düzeyinde dönüşüm	eşleme ve doğrulama
Girdi	ölçüm ya da öznitelik vektörü	görüntü	yüz, parmak izi, iris, ses
Çıktı	sınıf etiketi	işlenmiş görüntü ya da bölgeler	kimlik kararı
Başarı ölçütü	sınıflandırma hatası	görsel kalite, PSNR	FAR, FRR, EER
Matematik yükü	orta-yüksek: olasılık, doğrusal cebir	orta: evrişim, Fourier	orta
Veri ihtiyacı	etiketli örnekler	az; öğrenme şart değil	kişi başına az örnek

- **Görüntü İşleme:** Görüntü işleme görüntüyü iyileştirip dönüştürür, sonuç yine görüntüdür; bu dersin çıktısı bir sınıf etiketidir.
- **Biyometri:** Biyometri örüntü tanımanın bir uygulamasıdır; bu ders kişiye özgü değil, genel karar yöntemlerini öğretir.

UYGULAMA

Uygulama Alanları

El yazısı tanıma
Posta kodu, banka çeki, form okuma.

Optik karakter tanıma
Taranmış belgeyi düzenlenebilir metne çevirme.

Yüz ve parmak izi
Kimlik doğrulama ve tanımlama.

Endüstriyel denetim
Üretim hattında kusurlu ürünü ayırma.

Tıbbi tanı
EKG, kan tahlili ya da doku örneğini sınıflama.

Konuşmacı tanıma
Sesten kimin konuştuğunu bulma.

UYGULAMA

Riskler ve Etik

Rızasız yüz tanıma

Yüz tanıma kişileri habersiz izlemeye yarar; Clearview AI internetten topladığı yüzler için birçok Avrupa ülkesinde cezalandırıldı.

Gruplar arası hata farkı

Gender Shades çalışmasında ticari cinsiyet sınıflandırıcıları koyu tenli kadınlarda açık tenli erkeklerden çok daha sık yanıldı.

Eşit sayılan hatalar

Kanseri kaçırmak ile yanlış alarm aynı tartılırsa toplam hata düşük, sonuç tehlikeli olur; kayıp matrisi bilinçle seçilmeli.

Seçimde sızıntı

Öznitelikleri tüm veride seçip sonra çapraz doğrulama yapmak hatayı olduğundan çok düşük gösterir.

Ayrıntılı tartışma: Açıklanabilir Yapay Zekâ ve Yapay Zekâ Etiği.

UYGULAMA

Araçlar ve Örnek Veriler

Araçlar ve kütüphaneler

scikit-learn | NumPy | OpenCV | MATLAB Statistics and ML Toolbox | Weka | PRTools

Tanınmış veri setleri

Iris | MNIST | Optical Recognition of Handwritten Digits (UCI) | Wisconsin Breast Cancer | AT&T (ORL) yüz veri seti | EMNIST

UYGULAMA

Sık Karıştırılanlar

Yaygın düşünce	Doğrusu
Örüntü tanıma makine öğrenmesinin eski adıdır.	İki alan büyük ölçüde örtüşür; örüntü tanıma algılama ve öznitelik tasarımını da kapsayan mühendislik bakışıdır.
Sınıflandırıcı seçimi en önemli adımdır.	Çoğu problemde başarıyı belirleyen **öznitelik seçimidir**; kötü öznitelikle iyi sınıflandırıcı kurtarmaz.
Doğru öznitelikle hata sıfıra iner.	Sınıflar örtüşüyorsa en iyi karar bile Bayes hatası kadar yanılır.
Ne kadar çok öznitelik, o kadar iyi.	Boyut arttıkça örnekler seyrekleşir (boyut laneti); gereksiz öznitelik başarımı düşürür.

SONRASI

Bu Dersten Sonra

Bu dersi temel alan ya da genişleten dersler

Biyometrik Sistemler

İlgili meslekler

Görüntü ve sinyal işleme mühendisi

Makine öğrenmesi mühendisi

Biyometri uzmanı

Kalite denetim sistemleri mühendisi

Örüntü tanıma bilgisayarlı görü, konuşma tanıma ve biyometrinin ortak çekirdeğidir. Bu dersten sonra bir algılama alanına (görü, ses, biyometri) geçilebilir ya da öznitelikleri de veriden öğrenen derin öğrenmeye ilerlenebilir.

SONRASI

Kaynaklar

- Duda, Hart ve Stork, *Pattern Classification*, 2. baskı, Wiley, 2001.
- Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- Theodoridis ve Koutroumbas, *Pattern Recognition*, 4. baskı, Academic Press, 2009.
- Fukunaga, *Introduction to Statistical Pattern Recognition*, 2. baskı, Academic Press, 1990.

Son

Bölüm 8: Örüntü Tanıma

← Önceki bölüm · Alanın haritası · Sonraki bölüm →