

En Küçük Kareler Sınıflandırma

T.C. Trakya Üniversitesi
Mühendislik Fakültesi
Elektrik - Elektronik Mühendisliği Bölümü
Kontrol Anabilim Dalı

Dr. Öğr. Üyesi Işık İlber Sirmatel
sirmatel.github.io

Kaynak (source)

*Lecture Slides for Introduction to
Applied Linear Algebra: Vectors,
Matrices, and Least Squares.*

Stephen Boyd, Lieven Vandenberghe

Konu listesi

1. Sınıflandırma
2. En küçük kareler sınıflandırma
3. Çok sınıflı sınıflandırma

Bölüm 1

Sınıflandırma

Sınıflandırma

- ▶ sayısal olmayan değerler alan sonuçlar (*outcome*) ile veri uydurma yapmak istiyoruz
- ▶ bu tip sonuçlara bazı örnekler:
 - *DOĞRU* (*true*) veya *YANLIŞ* (*false*)
 - *SPAM* veya *SPAM DEĞİL*
 - *KEDİ*, *KÖPEK* veya *FARE*
- ▶ sonuç değerlerine etiket (*label*) veya kategori denir
- ▶ bu şekilde yapılan veri uydurmaya sınıflandırma (*classification*) denir

- ▶ sadece iki olası sonucun olduğu bir duruma bakalım
- ▶ bu sınıflandırmaya Boole veya iki-yönlü sınıflandırma denir
- ▶ sonuçları $+1$ (*DOĞRU*) ve -1 (*YANLIŞ*) olarak modelleyelim
- ▶ sınıflandırıcı formu: $\hat{y} = \hat{f}(x), f : \mathbb{R}^n \rightarrow \{-1, +1\}$

Sınıflandırma - Uygulama örnekleri

- ▶ spam e-posta saptaması: x bir e-postanın özniteliklerini (sözcük sayıları vs.) içerir
- ▶ finansal işlemde sahtecilik saptaması: x yapılması istenen bir işlemin özniteliklerini (işlem tutarı vs.) içerir
- ▶ belge sınıflandırma: x belgenin sözcük sayıları vektörü
- ▶ hastalık saptaması: x hastanın özniteliklerini ve tahlil sonuçlarını içerir
- ▶ dijital haberleşme alıcısı: y gönderilen bit; x alınan işaretten n adet ölçüm içerir

Öngörü hataları

- ▶ veri noktası (x, y) , öngörülen (tahmin edilen) sonuç

$$\hat{y} = \hat{f}(x)$$

- ▶ sadece dört olasılık vardır:

- doğru pozitif: $y = +1$ ve $\hat{y} = +1$

- doğru negatif: $y = -1$ ve $\hat{y} = -1$

(bu iki durumda öngörü doğrudur/hatasızdır (*correct*))

- yanlış pozitif: $y = -1$ ve $\hat{y} = +1$

- yanlış negatif: $y = +1$ ve $\hat{y} = -1$

(bu iki durumda öngörü yanlıştır/hatalıdır (*wrong*))

- ▶ hatalara birçok farklı isim verilir, örneğin: Tip 1 ve Tip 2

Karışıklık matrisi

- ▶ veri kümesi $x^{(1)}, \dots, x^{(N)}, y^{(1)}, \dots, y^{(N)}$ ve sınıflandırıcı \hat{f} verilsin
- ▶ dört sonuçtan her birini sayalım:

	$\hat{y} = +1$	$\hat{y} = -1$	toplam
$y = +1$	N_{tp}	N_{fn}	N_p
$y = -1$	N_{fp}	N_{tn}	N_n
hepsi	$N_{tp} + N_{fp}$	$N_{fn} + N_{tn}$	N

- ▶ köşegen-dışı terimlere öngörü hataları denir
- ▶ çeşitli hata oranları (*error rate*) ve ölçüm doğruluğu (*accuracy measure*) değerleri kullanılır
 - hata oranı: $\frac{N_{fp} + N_{fn}}{N}$
 - doğru pozitif oranı: $\frac{N_{tp}}{N_p}$
 - yanlış pozitif oranı: $\frac{N_{fp}}{N_n}$
- ▶ tasarlanan bir sınıflandırıcının başarımı bir test kümesi üzerindeki hata oranlarına göre değerlendirilir

Karışıklık matrisi - Örnek

- bir test kümesi üzerinde spam e-posta filtresi başarımı:

	$\hat{y} = +1$ (spam)	$\hat{y} = -1$ (spam değil)	toplam
$y = +1$ (spam)	95	32	127
$y = -1$ (spam değil)	19	1120	1139
hepsi	114	1152	1266

- hata oranı: $\frac{(19+32)}{1266} = \%4.03$
- yanlış pozitif oranı: $\frac{19}{1139} = \%1.67$

Bölüm 2

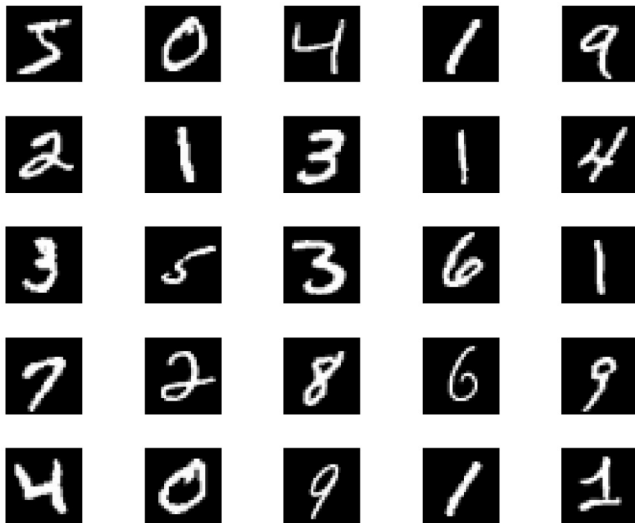
En küçük kareler sınıflandırma

En küçük kareler sınıflandırma

- ▶ standart en küçük kareler veri uydurma kullanarak model \tilde{f} 'i ikili (+1 ve -1) $y^{(i)}$ değerlerine uydurmak istiyoruz
- ▶ $y = +1$ olduğunda $\tilde{f}(x)$ +1'e yakın olmalı, $y = -1$ olduğunda ise +1'e yakın olmalı
- ▶ $\tilde{f}(x)$ bir sayıdır
- ▶ model olarak $\hat{f}(x) = \text{sign}(\tilde{f}(x))$ kullanalım
- ▶ $\tilde{f}(x)$ 'in büyüklüğü öngörüye olan güven (*confidence*) ile bağıntılıdır

Örnek: El yazısı rakam sınıflandırma

- ▶ **MNIST veritabanı:** 70000 adet 28×28 piksel görüntü (el yazısı rakam) içeriyor



Örnek: El yazısı rakam sınıflandırma

- ▶ veri kümesi eğitim kümesi (60000) ve test kümesi (10000) olarak ayrılmış
- ▶ sol alttaki piksel daima siyah (0)
- ▶ x 494-vektör: sabit 0 piksele ilave olarak sıfır-olmayan değerli 493 piksel değeri
- ▶ x 784-vektör değil çünkü bazı pikseller bütün görüntülerde 0 (siyah) olduğundan veriyi sadeleştirilmiş
- ▶ etiketler: görüntüdeki el yazısı rakam 0 ise $y = +1$, aksi halde $y = -1$

Örnek: El yazısı rakam sınıflandırma

en küçük kareler sınıflandırıcı sonuçları

- ▶ eğitim kümesi (hata oranı %1.6)

	$\hat{y} = +1$	$\hat{y} = -1$	toplam
$y = +1$	5158	765	5923
$y = -1$	167	53910	54077
hepsi	5325	54675	60000

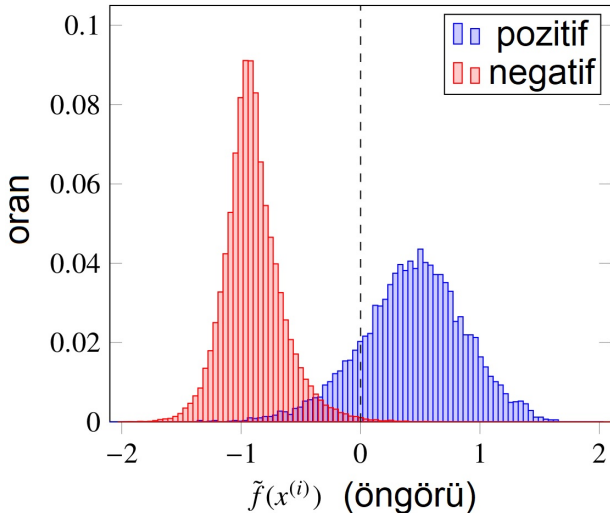
- ▶ test kümesi (hata oranı %1.6)

	$\hat{y} = +1$	$\hat{y} = -1$	toplam
$y = +1$	864	116	980
$y = -1$	42	8978	9020
hepsi	906	9094	10000

- ▶ önceden görülmemiş görüntüler üzerinde büyük ihtimalle %1.6 hata oranı elde ederiz

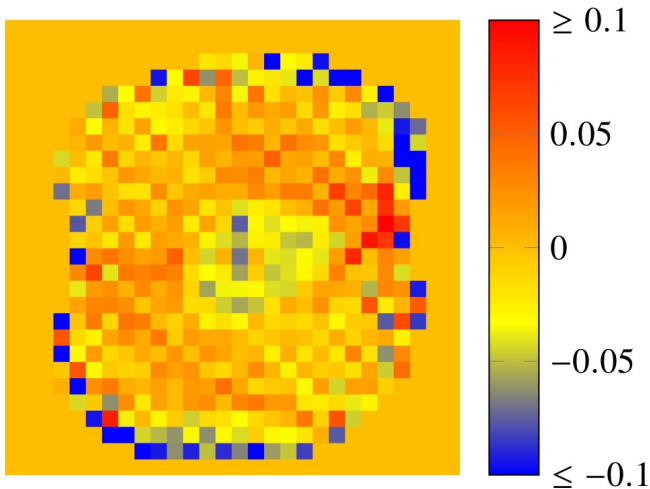
Örnek: El yazısı rakam sınıflandırma

en küçük kareler uydurma sonuçlarının dağılımı ($\tilde{f}(x^{(i)})$)
değerlerinin eğitim kümesi üzerindeki dağılımı)



Örnek: El yazısı rakam sınıflandırma

en küçük kareler sınıflandırıcısının katsayıları



Çarpık karar eşiği

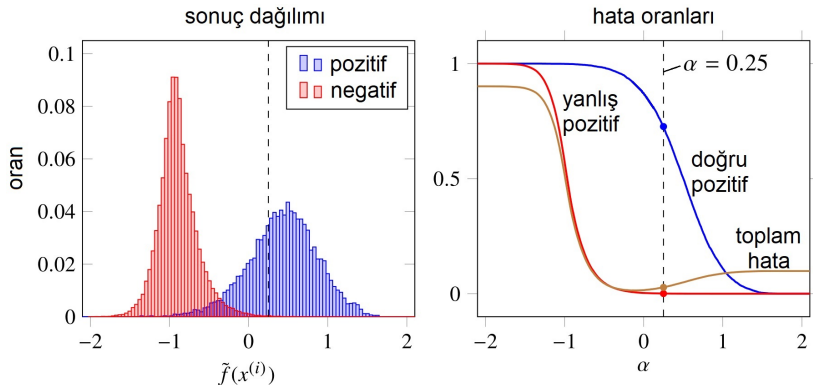
- ▶ $\hat{f}(x) = \text{sign}(\tilde{f}(x) - \alpha)$ formunda bir öngörücü (*predictor*) kullanalım:

$$\hat{f}(x) = \begin{cases} +1 & \tilde{f}(x) \geq \alpha \\ -1 & \tilde{f}(x) < \alpha \end{cases}$$

- ▶ α : karar eşiği (*decision threshold*)
- ▶ pozitif α için yanlış pozitif oranı daha düşük olur (bunu isteriz) ancak doğru pozitif oranı da daha düşük olur (bunu istemeyiz)
- ▶ negatif α için yanlış pozitif oranı daha yüksek olur (bunu istemeyiz) ancak doğru pozitif oranı da daha yüksek olur (bunu isteriz)
- ▶ doğru pozitive karşı yanlış pozitif oranlarını gösteren ödünleşim (*trade-off*) eğrisine “karar değerlendirme grafiği” (*receiver operating characteristic, ROC*) denir

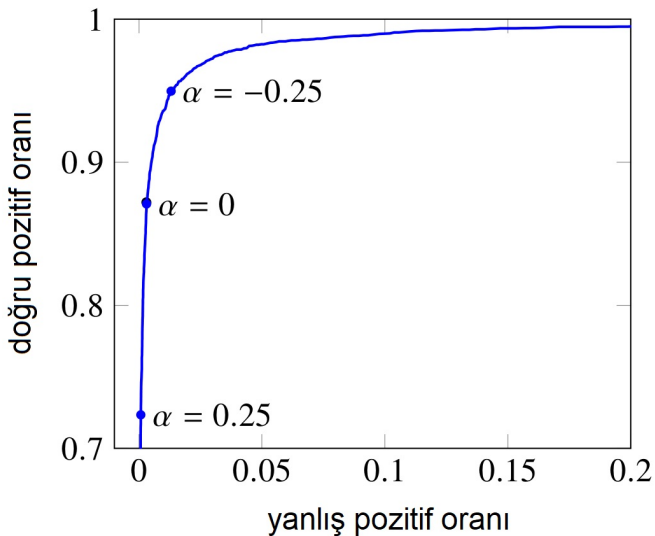
Örnek: El yazısı rakam sınıflandırma

karar eşiği α 'nın fonksiyonu olarak hata oranları



Örnek: El yazısı rakam sınıflandırma

karar değerlendirme grafiği



Bölüm 3

Çok sınıflı sınıflandırma

Çok sınıflı sınıflandırma

- ▶ K adet etiket ($K > 2$) var; etiket kümesi $\{1, 2, \dots, K\}$
- ▶ öngörücü: $\hat{f} : \mathbb{R}^n \rightarrow \{1, 2, \dots, K\}$
- ▶ verilen öngörücü ve veri kümesi için karışıklık matrisi $K \times K$ boyutludur
- ▶ bu matrisin bazı köşegen-dışı elemanları diğerlerinden çok daha kötü olabilir

Çok sınıflı sınıflandırma - Uyg. örnekleri

- ▶ el yazısı rakam sınıflandırma (piksel değerlerinden el yazısı rakamı tahmin etmek)
- ▶ pazarlama nüfus kesimi (*marketing demographic*) sınıflandırma (müşterilerin alışveriş geçmişlerinden nüfus kesimini tahmin etmek)
- ▶ hastalık tanılama (*diagnosis*) (tahlil sonuçları ve hasta özniteliklerinden, bir grup olası hastalık arasından doğru tanıyı tahmin etmek)
- ▶ çeviride sözcük seçimi (bağlam (*context*) özniteliklerinden, bir sözcüğün (birkaç seçenekli durumda) nasıl doğru çevrileceğini seçmek)
- ▶ belge konusu öngörüsü (sözcük sayısı histogramından belgenin konusunu tahmin etmek)

En küçük kareler çok sınıflı sınıflandırıcı

- ▶ her bir etiket için (diğer etiketlere karşı) bir en küçük kareler sınıflandırıcı kurulur
- ▶ çok sınıflı sınıflandırıcı

$$\hat{f}(x) = \operatorname{argmax}_{l \in \{1, \dots, K\}} (\tilde{f}_l(x))$$

formunda kurulur

- ▶ bu sınıflandırıcı, K adet $\tilde{f}_l(x)$ arasından en yüksek değeri olanın indisi l 'yi öngörü olarak seçer
- ▶ örneğin

$$\tilde{f}_1(x) = -0.7 \quad \tilde{f}_2(x) = +0.2 \quad \tilde{f}_3(x) = +0.8$$

için sınıflandırıcı $\hat{f}(x) = 3$ şeklinde öngörü yapar

Örnek: El yazısı rakam sınıflandırma

test kümesi için karışıklık matrisi

rakam	öngörü										toplam
	0	1	2	3	4	5	6	7	8	9	
0	944	0	1	2	2	8	13	2	7	1	980
1	0	1107	2	2	3	1	5	1	14	0	1135
2	18	54	815	26	16	0	38	22	39	4	1032
3	4	18	22	884	5	16	10	22	20	9	1010
4	0	22	6	0	883	3	9	1	12	46	982
5	24	19	3	74	24	656	24	13	38	17	892
6	17	9	10	0	22	17	876	0	7	0	958
7	5	43	14	6	25	1	1	883	1	49	1028
8	14	48	11	31	26	40	17	13	756	18	974
9	16	10	3	17	80	0	1	75	4	803	1009
hepsi	1042	1330	887	1042	1086	742	994	1032	898	947	10000

hata oranı yaklaşık %14 (eğitim kümesi için olanla aynı)

Örnek: El yazısı rakam sınıflandırma

yeni öznelikler ekleme

- ▶ 5000 adet rastgele (!) öznelik ekleyelim: $\max((Rx)_j, 0)$
(R : 5000×494 -matris, elemanları ± 1 , rastgele seçilmiş)
- ▶ 5494 boyutlu yeni öznelik vektörü için en küçük kareler sınıflandırma yapalım
- ▶ sonuçlar: eğitim kümesi hata oranı %1.5, test kümesi hata oranı %2.6
- ▶ yeni öznelikleri oluştururken biraz daha özenli olursak daha iyi sonuçlar almamız mümkün
- ▶ hatta, insanlar tarafından yapılan (!!) sınıflandırmadan bile daha iyi sonuçlar alınabilir

Örnek: El yazısı rakam sınıflandırma

test kümesi için karışıklık matrisi (yeni öznelikler ile)

rakam	öngörü										toplam
	0	1	2	3	4	5	6	7	8	9	
0	972	0	0	2	0	1	1	1	3	0	980
1	0	1126	3	1	1	0	3	0	1	0	1135
2	6	0	998	3	2	0	4	7	11	1	1032
3	0	0	3	977	0	13	0	5	8	4	1010
4	2	1	3	0	953	0	6	3	1	13	982
5	2	0	1	5	0	875	5	0	3	1	892
6	8	3	0	0	4	6	933	0	4	0	958
7	0	8	12	0	2	0	1	992	3	10	1028
8	3	1	3	6	4	3	2	2	946	4	974
9	4	3	1	12	11	7	1	3	3	964	1009
hepsi	997	1142	1024	1006	977	905	956	1013	983	997	10000