



BÜYÜK VERİ UYGULAMALARI – DERS 4

Doç. Dr. Yuriy Mishchenko

PLAN

- Lineer modeller
- Öznitelik seçilmesi
- Model oluşturulması

BÜYÜK VERİ VE MAKİNE ÖĞRENME



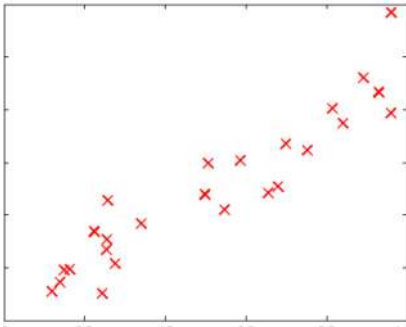
BÜYÜK VERİ VE MAKİNE ÖĞRENME

- Büyük veri uygulamaları, tahmin veya tespit etme olarak formüle edilmiş iş karar verme problemleri kapsamaktadır
- Bu tip iş karar verme problemleri ile ilgili karışık durumların tahmin veya tespit etmek için, Makine Öğrenme (Machine Learning, ML) yöntemleri kullanılmaktadır



BÜYÜK VERİ VE MAKİNE ÖĞRENME

- Makine öğrenme, karmaşık *ilişkisel – yani neden-sonuç – durumları* modelleyen parametrik modellerdir
(*Parametrik model, biçimi genel ama çok sayıda ayarlanabilen “parametre” içeren herhangi matematiksel modeldir*)
- Makine öğrenme, bu tip modellerin parametreleri tarihsel veriden belirtme yöntemleri ortaya çıkartarak, iş kararları ile ilgili karmaşık saklı desenler için matematiksel model bulur ve ilgili gelecek durumları tahmin etme imkanını sağlar



$$\{y = h(x; \theta)\}$$

Tarihsel veri

Model ailesi

θ^*

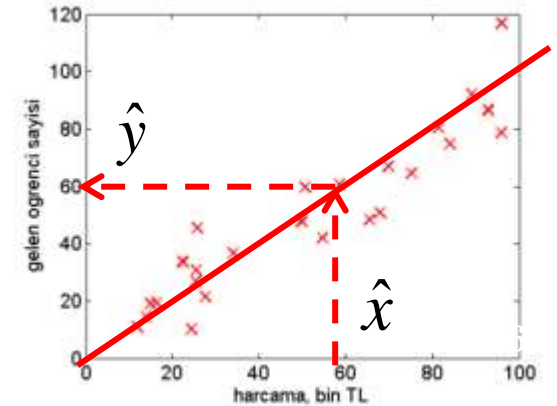
Model öğrenimi

$$y = h(x; \theta^*)$$

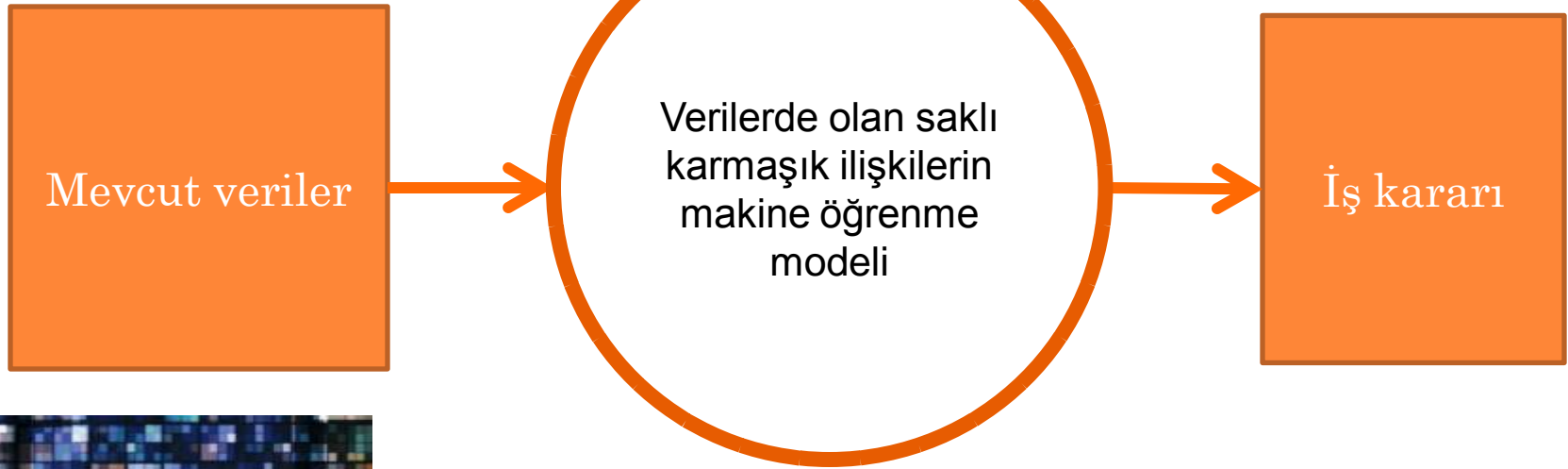
Spesifik model

İş kararları

$$\hat{y} = h(\hat{x}; \theta^*)$$



BÜYÜK VERİ VE MAKİNE ÖĞRENME



BÜYÜK VERİ VE MAKİNE ÖĞRENME

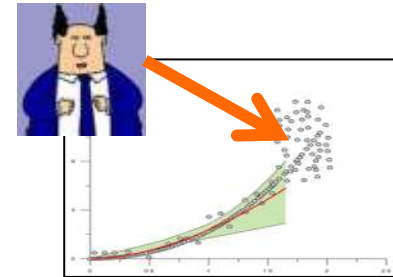
Gerçek problem ve ilgili faktörler veya değişkenler

$a, b, X, \mu, Z \dots$

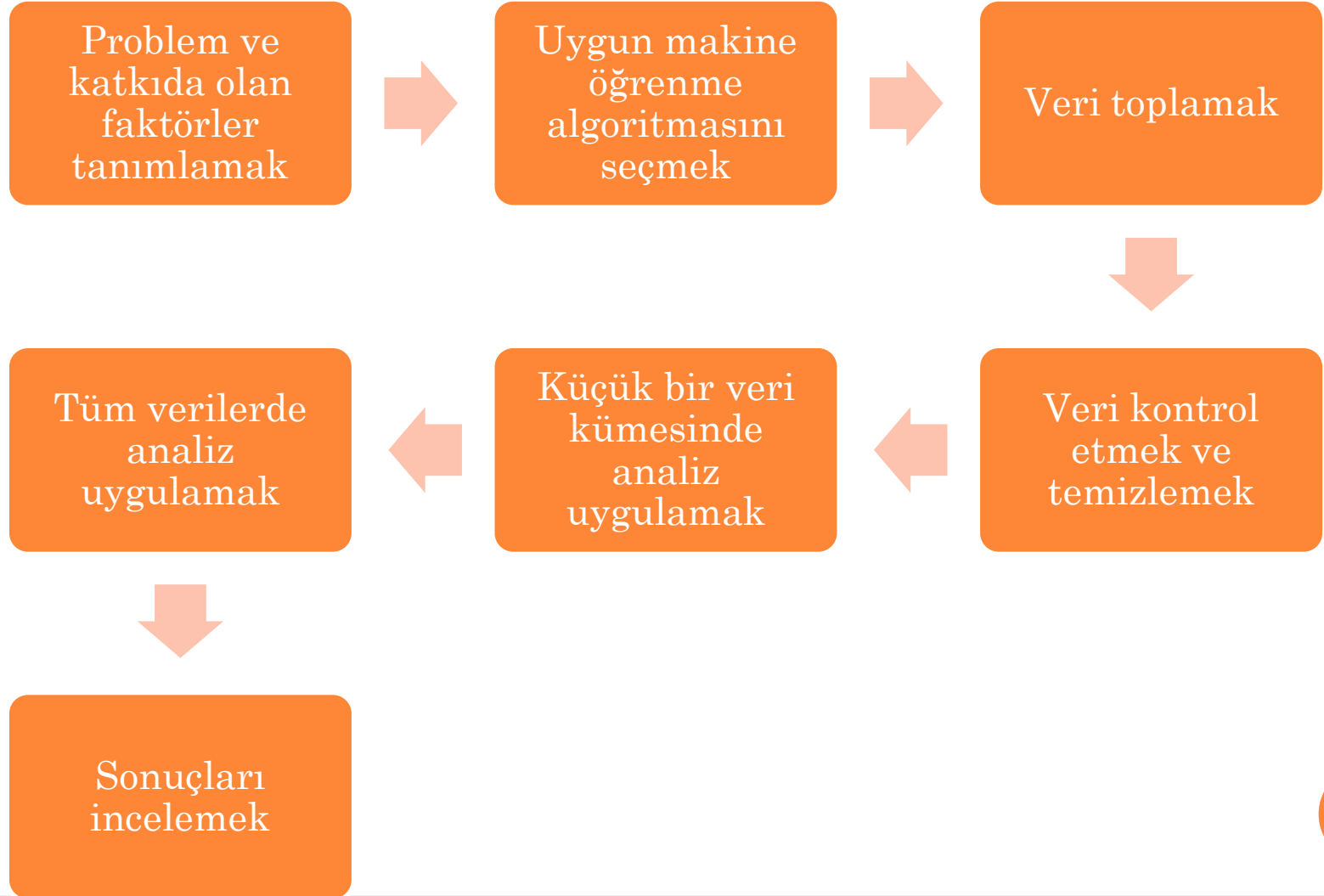
İlişkiyi ifade eden parametrik model

$$Z = a_1 X + a_2 Y + a_3 XY + \dots$$

İş karar etme için tahminler ve tespitler



TIPIK UYGULAMA SÜRECİ



MAKİNE ÖĞRENME

- Makine öğrenmeye ait bileşenleri
 - **Temel makine öğrenme problem türleri**
 - Veri/öznitelik mühendisliği
 - Modelin oluşturulması
 - **Model eğitimi**
 - Performans değerlendirilmesi
 - Tahminlerin üretilmesi

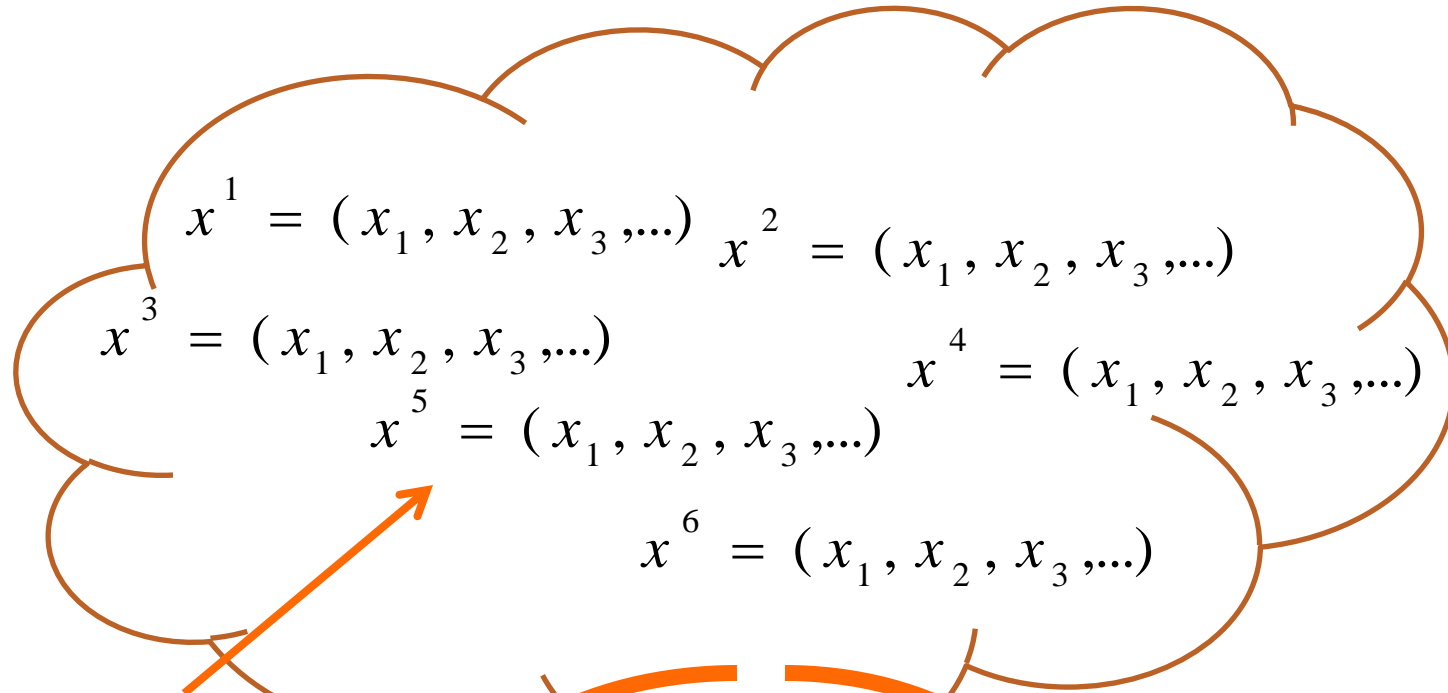
MAKİNE ÖĞRENME PROBLEMLERİNİN ANA TIPLERİ

- Denetimli
 - Sınıflandırma
 - İki sınıflı (binary) sınıflandırma
 - Çok sınıflı sınıflandırma
 - Regresyon
- Denetimsiz
 - Kümeleme
 - İlişkisel öğrenim
 - Boyut azaltma

MAKİNE ÖĞRENME PROBLEMLERİNİN ANA TIPLERİ

- Sınıflandırma problemi, sonuç ayrık olduğu zaman elimizde vardır
 - Müsteri ürün “satın alacak”/”satın almayacak” mı?
 - Kitap “bilim kurgu”/”roma”/”belgesel”/ “şiir” midir?
 - Sistem “normal durum”/”arzalı durum” da mı?
 - Email “spam”/”spam değil” mi?
- Regresyon problemi, sonuç sürekli olduğu zaman elimizde vardır
 - Reklam harcamasına göre satılacak ürün sayısı
 - Evin özelliklerine göre satış fiyatı
 - Yarınki hava sıcaklığı

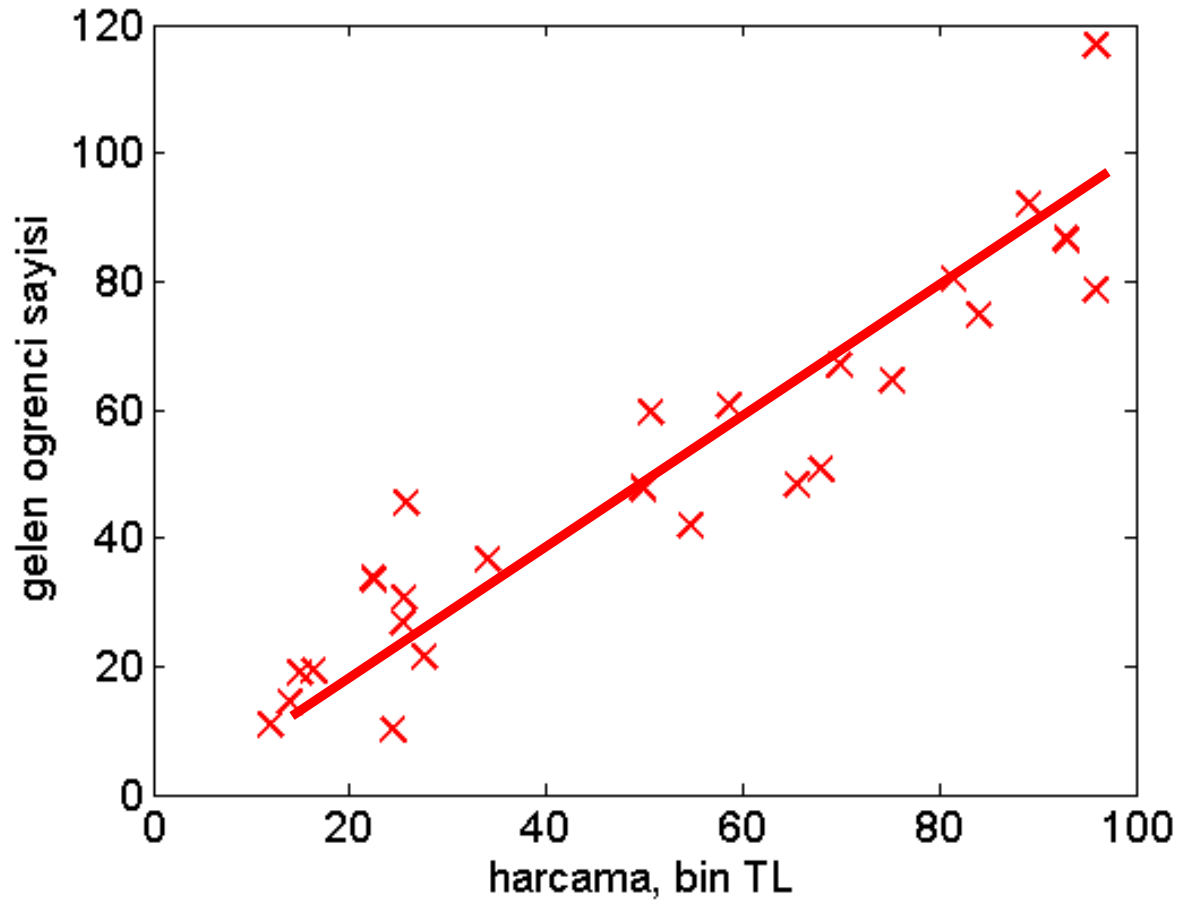
SİNİFLANDIRMA



**GEÇMİŞTEKİ
ÖNEKLER/VERİLER**

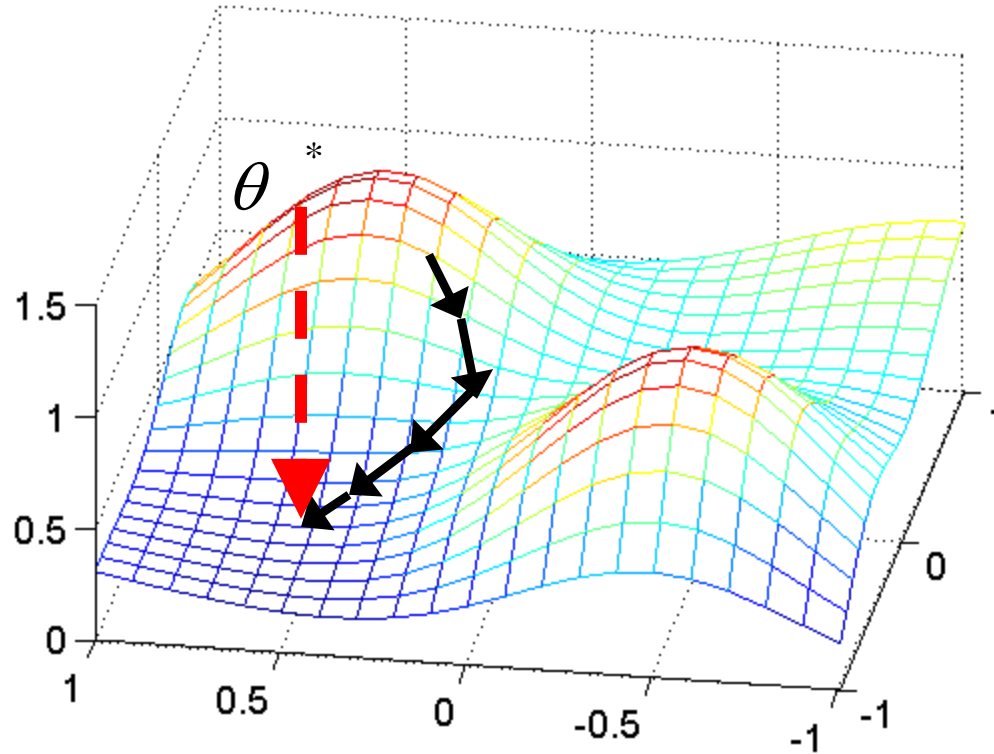


REGRESYON



MODEL PARAMETRELERİNİN BELİRTİLMESİ

Modelin kalitesi fonksiyonu –
maliyet fonksiyonu J



○ Bu derste

- ~~Temel makine öğrenme problem türleri~~
- **Veri/öznitellik mühendisliği**
- **Model oluşturulması**
- ~~Model eğitimi~~
- Performans değerlendirilmesi
- Tahminlerin üretilmesi

LINEER MODELLER



LINEER MODELLER

- Lineer veya doğrusal modeller bir sonuç Y ve neden X_i değişken sırasını bağlayan, aşağıdaki şekilde olan modeller

$$\begin{aligned} Y &= \theta_0 + \theta_1 \cdot X_1 + \theta_2 \cdot X_2 + \dots \\ &+ \theta_{n+1} \cdot f(X_1) + \theta_{n+2} \cdot f(X_2) + \dots \\ &+ \theta_{m+1} \cdot f(X_1, X_2) + \theta_{m+2} \cdot f(X_2, X_3) + \dots \end{aligned}$$



LINEER MODELLER

- Lineer modeller ilişkisel - yani sonuç-neden modeller dir, yani belirli neden faktörlerinin listesi (x_1, x_2, \dots) için bir sonuç y tahmin ediyor
- Lineer modelden belirli (y, x_1, x_2, \dots) bir örnek

$$\begin{aligned} y &= \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \dots \\ &+ \theta_{n+1} \cdot f_1(x_1) + \theta_{n+2} \cdot f_2(x_2) + \dots \\ &+ \theta_{m+1} \cdot g_1(x_1, x_2) + \theta_{m+2} \cdot g_2(x_2, x_3) + \dots \end{aligned}$$



LINEER MODELLER

- Fark edelim ki bu model, X 'lere göre lineer veya doğrusal değil – “lineer model” deki “lineer” kelime modelin parametrelerine doğrusal olarak bağlı olduğunu ifade eder (analizinde önemli kolaylık sağlar)

$$\begin{aligned} Y &= \theta_0 + \theta_1 \cdot X_1 + \theta_2 \cdot X_2 + \dots \\ &+ \theta_{n+1} \cdot f(X_1) + \theta_{n+2} \cdot f(X_2) + \dots \\ &+ \theta_{m+1} \cdot f(X_1, X_2) + \theta_{m+2} \cdot f(X_2, X_3) + \dots \end{aligned}$$



LINEER MODELLER

- Lineer modellerin ana avantajları:
 - Teorisi çok iyi geliştirilmiş, özellikleri ve davranışı iyi bilinir
 - Parametrelerin tahmin etme problemi iyi geliştirilmiş ve kolaydır
 - İfade edilebilen ilişkiler çok geniş ve çeşitli, özellikle fonksiyonel terimler dikkate alırsak
 - Spesifik veriler için tahmin problemi çok büyük ölçekte çözülebilir – üz binlerce faktör ve parametre içeren lineer modelleri günümüzdeki bilgisayarlarla başarmak mümkündür



LINEER MODELLER

- Lineer olmayan terimler içeren lineer modeller sahte-değişkenleri tanımlayarak standart bir şekile götürülebilir

$$f(X_1) = X_{n+1}$$

$$f(X_2) = X_{n+2}$$

...

$$Y = \theta_0 + \theta_1 \cdot X_1 + \theta_2 \cdot X_2 + \dots$$



LINEER MODELLER

- Böylece, orijinal neden değişkenlerin değerleri varsa - (x_1, x_2, \dots) - böyle ek-değişkenler de hesaplanabilir
- Bu şekilde orijinal X-değişkenler büyütülmüş X-değişkenlere genişletilir

$$\left\{ \begin{array}{l} (x_1, x_2, \dots, x_n) \rightarrow (x_1, x_2, \dots, x_n, \bar{x}_{n+1} = f_1(x_1), \dots) \\ (x_1, x_2, \dots, x_n) \rightarrow (x_1, x_2, \dots, x_n, \bar{x}_{n+1} = f_1(x_1), \dots) \\ (x_1, x_2, \dots, x_n) \rightarrow (x_1, x_2, \dots, x_n, \bar{x}_{n+1} = f_1(x_1), \dots) \\ (x_1, x_2, \dots, x_n) \rightarrow (x_1, x_2, \dots, x_n, \bar{x}_{n+1} = f_1(x_1), \dots) \\ \dots \end{array} \right\}$$



LINEER MODELLER

- Genişletilmiş veri kümesinde ilişki standart lineer model olarak ifade edilebilir

$$y = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \dots + \theta_n x_n + \theta_{n+1} \bar{x}_{n+1} + \dots$$



LINEER MODELLER

- Bu nedenle,
 1. Lineer modellerde tahmin etme problemi sadece standart lineer model biçimi için çözülür
 2. Modelinizde herhangi lineer olmayan terimler varsa, bu terimler temel değişkenlerden hesaplayıp veri kümesine eklemeniz lazım



LINEER MODELLER

Y	X ₁	X ₂	X ₃	f ₁ (X ₁)	f ₂ (X ₂)
2	1	2	1.5		
-1	-1	2	1		
.8	.5	.5	.3		
-.1	-.1	.2	1.2		
8	3	-1	0		
5	2	2	4		
6	-3	4	3		
.9	.7	2	3		
7	2	1	2		
0	.1	.1	0		
-.1	-.1	.5	0		



Y	X ₁	X ₂	X ₃	X ₄	X ₅
2	1	2	1.5	3	4
-1	-1	2	1	3	4
.8	.5	.5	.3	2	.25
-.1	-.1	.2	1.2	3	.04
8	3	-1	0	2	1
5	2	2	4	1	4
6	-3	4	3	2	16
.9	.7	2	3	.7	4
7	2	1	2	1	1
0	.1	.1	0	3	.01
-.1	-.1	.5	0	3	.25

LINEER MODELLER

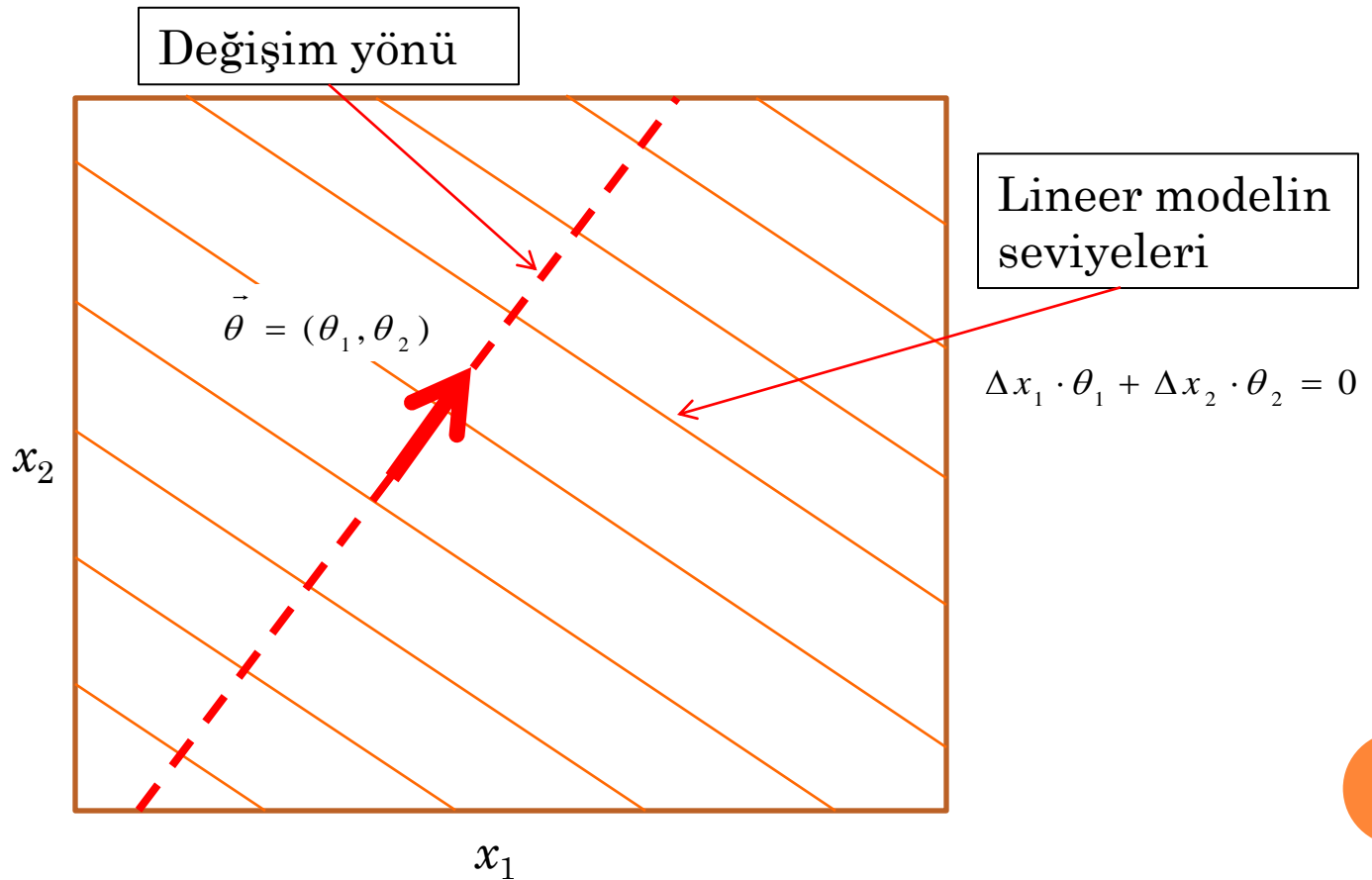
- Linear modelin standart şekli

$$Y = \theta_0 + \theta_1 \cdot X_1 + \theta_2 \cdot X_2 + \dots + \theta_m \cdot X_m$$



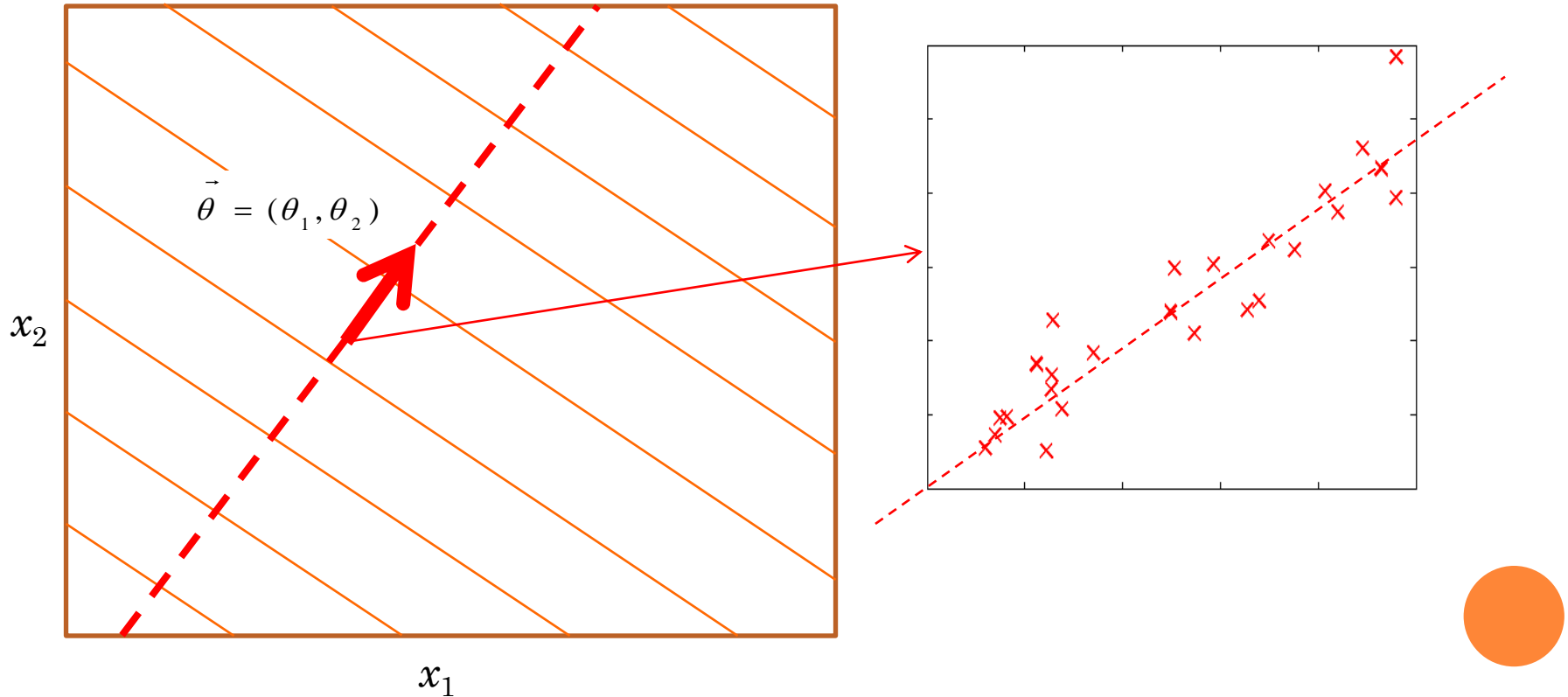
LINEER MODELLER

- Lineer model bir doğrusal yapı oluşturur ve bağlıdır



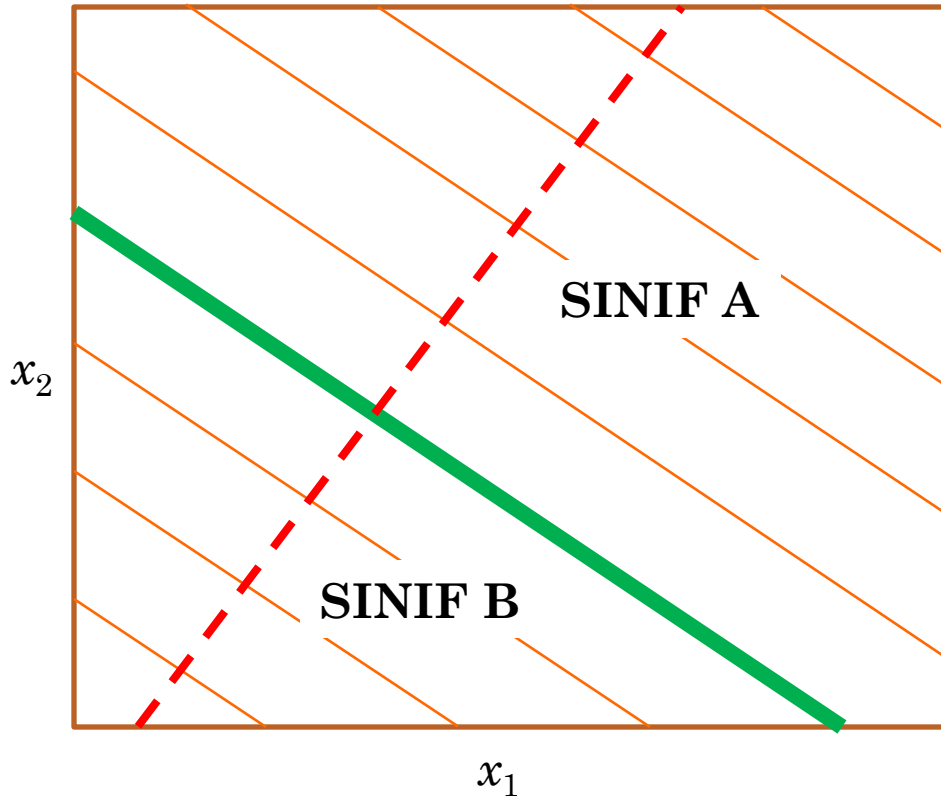
LINEER MODELLER

- Regresyon problemlerinde lineer model değişim yönünde lineer olarak değişiyor



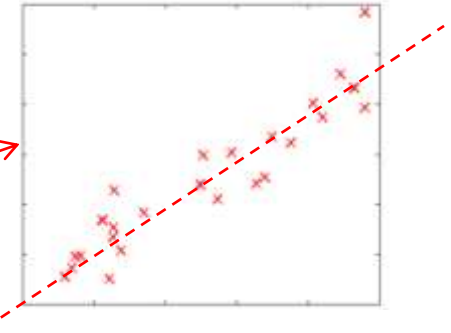
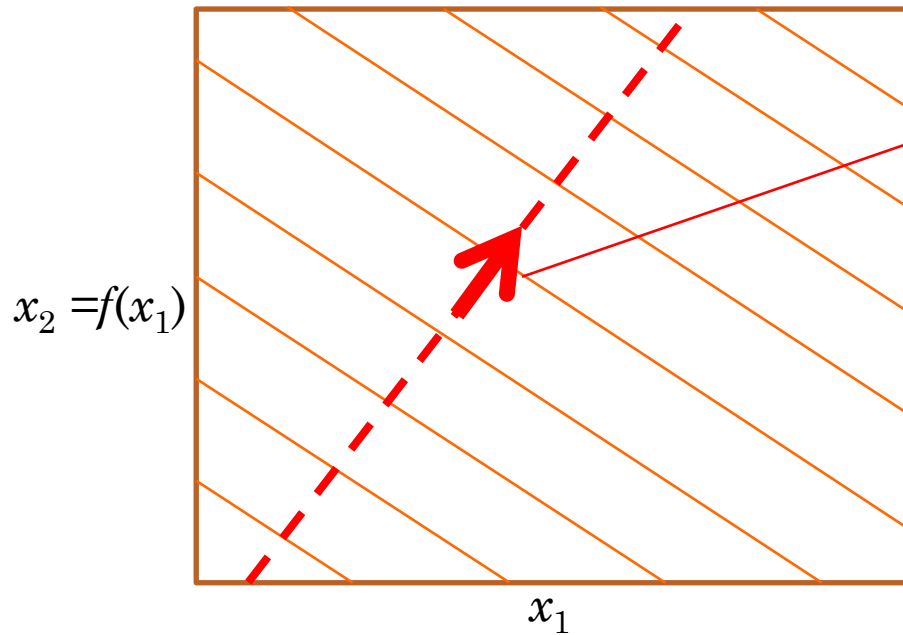
LINEER MODELLER

- Sınıflandırma problemlerinde lineer model bir doğrusal sınıf sınırı oluşturur



LINEER MODELLER

- **Önemli nokta:** bu yapı özelliklerin sahte-uzayında geçerlidir, yani bu uzaya tüm eklediğimiz, fonksiyonel olarak bağlı olan sahte-özellikler de dahildir !!!



$(x_1, f(x_1))$ yönündedir,
yani

$$y = x_1 \theta_1 + f(x_1) \theta_2$$



ÖZNİTELİK SEÇİLMESİ



ÖZİNİTELİKLER

- Lineer modeldeki y deęişkenine sonuç ve x_i deęişkenlerine **özınitelik (feature)** denir
- Yani (x_1, x_2, \dots, x_n) **özınitelik sırası** veya **özınitelik vektörü (feature vector)** bizim için ilginç olan belirli durumu karakterize eder, y o durumun tahmin etmek istediđimiz deęeri dir

$$y = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \dots + \theta_n x_n$$



ÖZNİTELİKLER

○ Örnek:

- y yarınki hava durumu olsa (x_1, x_2, \dots) bugünkü sıcaklık (x_1) , nem (x_2) , basınç (x_3) , rüzgar hızı (x_4) , rüzgar yönü (x_5) vb parametreleri listeleyen bir “bugünün özellikleri” sırası olabilir
- Lineer modelde biz tahmin edilen değişkeni y ve öznitelik vektörü (x_1, x_2, \dots) aşağıdaki şekilde bağlamak istiyoruz

$$y = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \dots + \theta_n x_n$$



ÖZNETELIKLER

○ Başka bir örnek:

- y üniversiteye gelecek sayısı olsa, (x_1, x_2, \dots) reklam harçamalarımız (x_1) , liselerden mezun sayısı (x_2) , lise mezunların ortalama notları (x_3) , kayıt günün sıcaklığı (x_4) , eğitim ücretimiz (x_5) vb parametreleri listeleyen “özelliklerimiz” sırası olabilir
- Yinde bu durumda biz tahmin edilen öğrenci sayısı öznetelik vektörü (x_1, x_2, \dots) kullanılarak aşağıdaki şekilde belirtiyoruz

$$y = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \dots + \theta_n x_n$$



ÖZİNİTELİKLER

- Modelimiz baya basit olduğuna göre kolayca anlaşılır ki, öznelilik seçimi modelimizin başarısı için son derecede önemli
- Öznelilikler şu genel gruba bölünebilir
 - Temel öznelilikler
 - Bileşik öznelilikler
 - Ortak etki öznelilikler

TEMEL ÖZNİTELİKLER

- Durumumuzu karakterize eden orijinal özellikler **temel öznitelik kümesini** oluşturur
- Bunlar, hava sıcaklığı örneğindeki bugünkü hava sıcaklığı, rüzgar hızı, nem ve sayre ve öğrenci sayısı örneğindeki reklam harcamaları, lise mezunu sayısı ve sayre gibi dir
- Bunları, problemi tanımlayan uzman seçmek zorundadır; bu tip özelliklerin problemin doğasıyla alakalı olması düşünülmesi genelde gerekir

TEMEL ÖZNİTELİKLER

- **Temel öznitelikler üç tipten olabilir:**
 - **Kategorik**
 - **İkili**
 - **Sayısal**

TEMEL ÖZNİTELİKLER

- **Temel öznitelikler:**
 - Sonuçla alakalı olmalı
 - Tekrarlanan olmalı

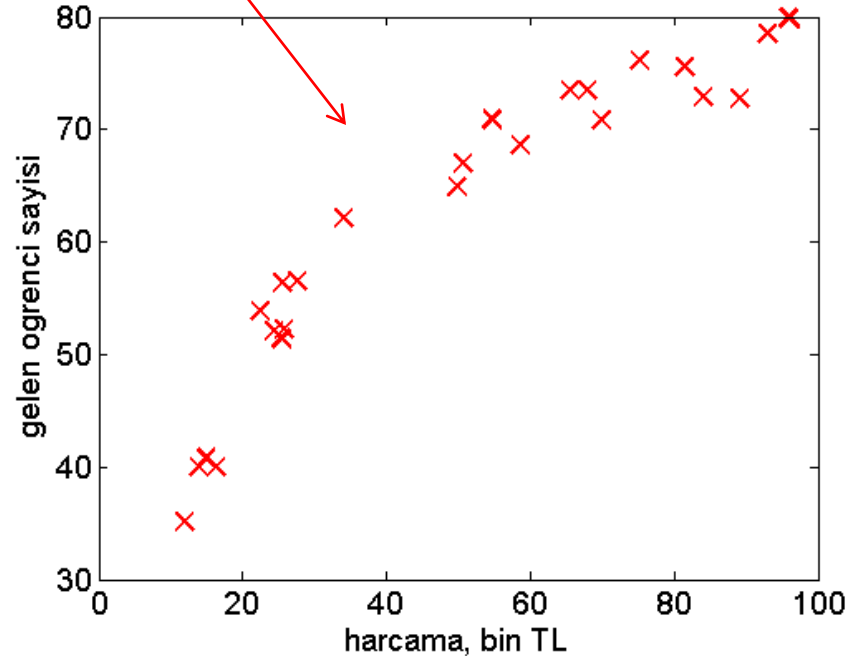
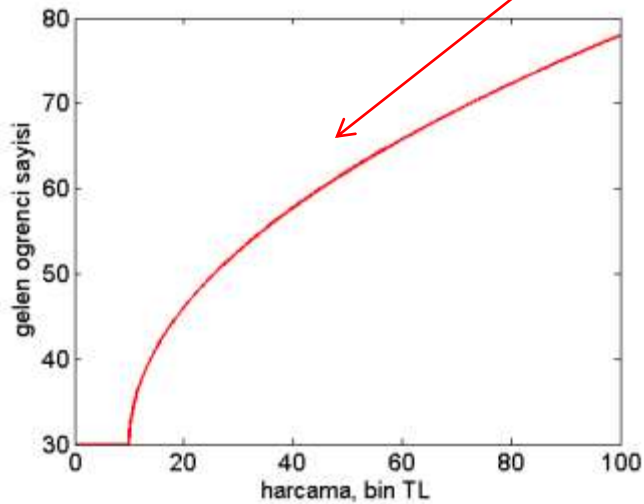
BİLEŞİK ÖZNETELİKLER

- Temel öznetelikler birçok durumda fazla dar bir model sağlar, mesela hava sıcaklığı örneğinde bunlar yarınki hava sıcaklığı doğrusal şekilde sadece bugünkü, dünkü, vb sıcaklığı, neme ve sayre bağıllığı ifade edebilir
- Birçok durumda bu yeterli değil
- Öyle ise temel öznetelik vektörüne, temel özneteliklere lineer olmayan fonksiyonel şekilde bağı olan **bileşik öznetelikler** eklenebilir

BİLEŞİK ÖZNİTELİKLER

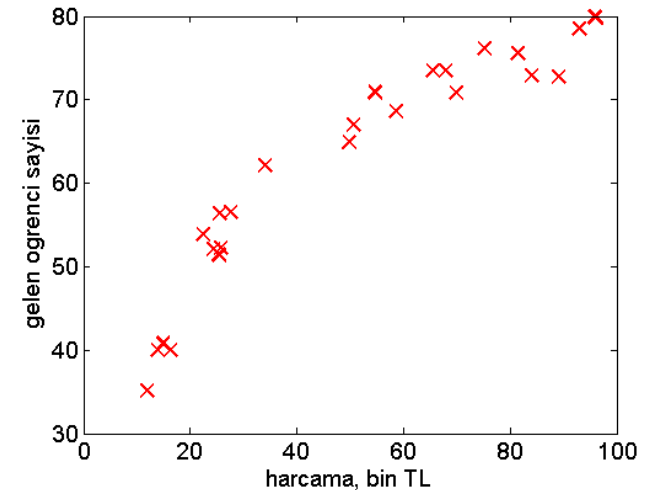
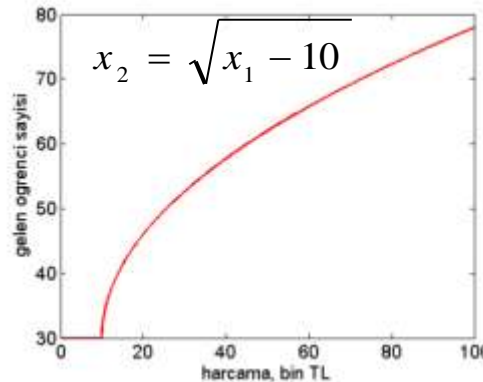
- Verilerde lineer olmayan ilişkinin gözlenmesi

$$y \sim 30 + 0.16 \sqrt{x - 10}$$



BİLEŞİK ÖZİNİTELİKLER

- Orijinal/temel x_1 öz niteliğe ek olarak, yeni bileşik x_2 özelliği şöyle tanımlayalım: $x_2 = \sqrt{x_1 - 10}$
- Öz nitelik vektörlerimizin hepsinde x_1 kullanılarak x_2 değerleri de hesaplayıp, öz nitelik vektörlerimize ekleyelim – $(x_1) \rightarrow (x_1, x_2)$
- y -modelimiz için şimdi, (x_1, x_2) üzerinde iki boyutlu lineer model ayarlayalım



BİLEŞİK ÖZİNİTELİKLER

- Genişletilmiş öz nitelik uzayında model lineer, fakat orijinal öz niteliklerin uzayında lineer değildir ve lineer olmayan ilişki modellemeye yol açıyor
- Model parametrelerin bulunması için $(\theta_0, \theta_1, \theta_2)$, lineer modellerin güçlü yöntemleri ve yazılımları direk kullanılabilir

$$h(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$


$$\rightarrow \theta_0 + \theta_1 x_1 + \theta_2 \sqrt{x_1 - 10}$$



BİLEŞİK ÖZİNİTELİKLER

- Bileşik öz nitelikler elle seçilmeli ve problemin doğasının uzman tarafından anlaşılması bunun için çok önemlidir
- Bununla beraber birçok sayıda standart seçinek de vardır, örneğin çok sık kullanılan bileşik öz nitelik **polinom öz nitelikler dir**

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$


$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1^3 + \theta_5 x_1 x_2^2 + \dots$$



ORTAK ETKİ ÖZNETELİKLERİ

- “Ortak etkiler” bileşik özneteliklerin özel bir durum
- Ortak etki demek ki, iki yada daha çok faktör beraber çalışır, beraber ise sonucu farklı şekilde etkiliyor



ORTAK ETKİ ÖZNETELİKLERİ

○ Örnek:

- Müşterin kredi riski hesaplamak için “içel” ve “gelir” iki temel öznetelik kullanabiliriz
- Temel lineer modelde, kredi riskine gelirin etkisi içelden bağımsız olması varsayılıyor
- Fakat, iki ayrı içelde aynı gelir kredi riskini çok farklı şekilde etkileyebilir
- Bu durumda “içel” ve “gelir” özneteliğin ortak etkisi vardır, yani bu faktörler beraber dikkate alınması gerekiyor



ORTAK ETKİ ÖZNETELİKLERİ

- Ortak etkileri modellerde belirli bileşik veya polinom öznitelikleri kullanılarak temsil edilebilir
- Örneğin, gelir grubu ve ilçe ortak etkisi ile incelemek için, $x_{ilçe}x_{gelir}$ gibi yeni özellikler tanımlanıp modele eklenebilir

$X_{ig} = X_i X_g$ yeni (ortak) etki özniteligidir

$$y = \theta_0 + \theta_i \cdot x_i + \theta_g \cdot x_g + \theta_{ig} \cdot x_i x_g + \dots$$



ORTAK ETKİ ÖZNİTELİKLERİ

○ Örnek:

- x_{ig} sadece “akdeniz” ve “düşük gelir” durumda 1 ve *tek bu durumda* modelimize ek olarak (ortak) katkı θ_{ig} ekliyor

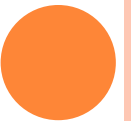
x_i	x_g	$x_{ig}=x_i x_g$	Açıklama
0	0	0	Akdeniz değil, yüksek gelir
1	0	0	Akdeniz, yüksek gelir
0	1	0	Akdeniz değil, düşük gelir
1	1	1	Akdeniz, düşük gelir

$$y = \theta_0 + \theta_i \cdot x_i + \theta_g \cdot x_g + \theta_{ig} \cdot x_{ig} + \dots$$



TEKRARLANAN ETKİLERİ

- Baze durumunda bir modelde neredeyse aynı etki yaratan iki veya daha çok öznelik vardır
- Örneğin, önceli örnekte “gelir”, “ilçe” ve “gelir-ilçe” öznelik var olduğu takdirde aynı etki içeren iki terim oldu, yani x_i ve x_{ig}



TEKRARLANAN ETKİLERİ

- Bu durumdan genellikle sorun çıkmaz
- Bir modelde iki benzer etki varsa, lineer model parametrelerinin çözümünde sonuçları daha iyi anlatabilecek faktörü seçilecek ve modele eklenecek; diğer yakın ama biraz daha kötü faktörünün θ -parametresi sıfıra yakın veya tam eşit olacaktır



TEKRARLANAN ETKİLERİ

- x_i ve $x_i x_g$ özellikleri için, eğer $x_i x_g$ sonuçları daha iyi anlatıyorsa, lineer modelde $x_i x_g$ faktörü seçilmiş olup ona yüksek θ değeri, x_i daha düşük θ değeri atanacak

$$y = \theta_0 + \theta_i \cdot x_i + \boxed{0 \cdot x_i x_g} + \dots$$

yada

$$y = \theta_0 \boxed{+ 0 \cdot x_i} + \theta_{ig} \cdot x_i x_g + \dots$$



ÖZNİTELİKLERİN NORMALLEŞTİRİLMESİ

- Aşağıdaki gibi elde edilmiş özniteliklerin genellikle normalleştirilmesi gerekiyor
- Bu adım mevcut makine öğrenme çözümleri otomatik olarak zaten yapıyor, ama genel olarak bu adımın ana fikri – tüm özniteliklerin benzer olması
- Mesela hava durumu karakterize eden sıcaklık (~ 20) ve basınç (~ 800) tipik değerleri çok farklı olabilir – bu gereksiz farkı çözüm algoritmaları olumsuz şekilde etkileyebilir



ÖZNİTELİKLERİN NORMALLEŞTİRİLMESİ

- Bu sorunu gidermek için, modelin parametreleri aranması başlamadan önce her bir özneliğin değerleri benzerliği sağlamak için uygun sayısal faktör ile bölünüyor
- En tipik normalleştirme adımı, öznelik değerlerden ortalama değeri çıkartmak ve sonra standart sapması ile bölmek
- Böylece tüm öznelikler **sıfır-merkezli** ve **tipik değişim 1** olarak kaydediliyor – bu dönüşüm lineer modeli tabi ki herhangi şekilde etkilemiyor



ÖZNİTELİKLERİN NORMALLEŐTİRİLMESİ

- İkinci çok tipik öznitelik normalleştirilmesi yöntemi – “whitening” – biz bu konuya bakmayacağız



MODEL

OLUŐTURULMASI



MODEL OLUŐTURMA

- AŐağıdaki y6nelliklerine g6re belirli model i6in 6z niteliklerin oluŐturulması 6z nitelik m6hendisliđi diya adlandırılan faaliyetin bir alt faaliyetidir
- Herhangi problemi makine 6đrenme yaklaŐımı 66zlemek i6in 6zelliklerin dođru se6ilmesi ve dođru Őekilde temsil edilmesi 6ok 6nemli
- AŐağıda bu problemi biz 6rnek 6zerinde inceleyeceđiz



MODEL OLUŐTURMA

- Problem: Bir banka kredi vermeden önce müşterisinin kredi riski belirtmek istemektedir



MODEL OLUŐTURMA

- İlk önce, temel açıklayıcı faktörler veya deęişkenler veya öznitelikler belirtmemiz gerekiyor
- Makine öğrenme çok büyük modelleri kullanabilmesi için, birçok böyle faktör seçebiliriz ve aslında aklımıza gelen herhangi öyle alakalı faktör temel öznitelik kümesine eklemek mümkündür



MODEL OLUŐTURMA

- Örnek olarak burada, kredi riski etkileyen faktörler (X'ler)
 - Müőterinin oturan ilçe
 - Müőterinin geliri
 - Müőterinin önceki kredileri



MODEL OLUŐTURMA

- Modelin sonucu olarak (Y) kredi riski srekli bir sayı olarak deęerlendirmek istiyoruz – pozitif deęerler yksek risk, negatif deęerler dŐk risk demek olacaktır



MODEL OLUŐTURMA

1. öznitelik – müşteri oturan ilçe, olabilir deęerler:

- Akdeniz
- Toroslar
- Yenişehir
- Çiftlikköy
- Mezitli
- ...



MODEL OLUŐTURMA

MüŐteri oturan ilçe :

- Akdeniz
- Toroslar
- YeniŐehir
- Çiftlikköy
- Mezitli
- ...



MODEL OLUŐTURMA

- Bunun gibi deęişkenlere “kategorik” denir; demek bu deęişken aslında durumun birkaç kategoriden olduğunu belirtiyor (akdeniz, yenişehir, vb)
- Bunun gibi deęişken modelde temsil etmek için, kategorik deęişkenin tüm mümkün deęerleri için ayrı bir ikili (0-1 deęerinde olan) öznitelik tanımlanıyor



MODEL OLUŐTURMA

Müşteri oturan ilçe:

- Burada, bütün ilçeler için ayrı ikili x -öznitelik tanımlanması gerekir,
- Bu özellik 0 veya 1 değerini alıp müşterinin ilgili ilçede oturup oturmadığını ifade edecek
- $x=1$ – müşteri ilgili ilçede oturur, $x=0$ oturmaz anlamına gelecek



MODEL OLUŐTURMA

MüŐteri oturan ilçe:

- Akdeniz (x_1 0 veya 1 olabilir)
- Toroslar (x_2 0 veya 1 olabilir)
- YeniŐehir (x_3 0 veya 1 olabilir)
- Çiftlikköy (x_4 0 veya 1 olabilir)
- Mezitli (x_5 0 veya 1 olabilir)
- Tece (x_6 0 veya 1 olabilir)
- Gözne (x_7 0 veya 1 olabilir)
- Davultepe (x_8 0 veya 1 olabilir)
- Bahçeli (x_9 0 veya 1 olabilir)
- Adanalıođlu (x_{10} 0 veya 1 olabilir)



MODEL OLUŐTURMA

MüŐteri oturan ilçe:

- Bu özniteliklerden sadece bir tane 1 olabilir
- 1 olan öznitelik, müŐterinin ilçesini kodlanır



MODEL OLUŐTURMA

MüŐteri oturan ilçe:

- Akdeniz (x_1): öđrenci Akdeniz'den geliyorsa, $x_1 = 1$
- Toroslar (x_2)
- YeniŐehir (x_3)
- Çiftlikköy (x_4)
- Mezitli (x_5)
- Tece (x_6)
- Gözne (x_7)
- Davultepe (x_8)
- Bahçeli (x_9)
- Adanalıođlu (x_{10})



MODEL OLUŐTURMA

MüŐteri oturan ilçe:

- Akdeniz (x_1)
- **Toroslar (x_2): öđrenci Toroslar'dan geliyorsa, $x_2 = 1$**
- YeniŐehir(x_3)
- Çiftlikköy (x_4)
- Mezitli (x_5)
- Tece (x_6)
- Gözne (x_7)
- Davultepe (x_8)
- Bahçeli (x_9)
- Adanaliođlu (x_{10})



MODEL OLUŐTURMA

2. öznitelik - müşteriinin geliri:

- Müşterinin geliri normalde sürekli bir deęişken olmalıdır (örneğin – ayda 500TL, 1000 TL, 1500 TL, 2000 TL, 2500 TL olarak karakterize edilebilir)



MODEL OLUŞTURMA

Lineer model kontekstinde bu bir sorun yaratabilir;

$$y = \theta_0 + \theta_1 \cdot x + \dots$$

- Eğer müşterinin geliri direk kullanırsak, gelirin kredi riskine etkisi doğrusal şekilde varsaymış olacağız, ama bu gerçekten iyi mi fikir?
- Eğer adayın geliri iki kat yüksek, bunun kredi riskine etkisi iki kat düşük olur mu?



MODEL OLUŐTURMA

- Bu durumda daha mantıklı yaklaŐım, müŐterinin geliri “düşük”, “orta” vb şekilde **kategorik olarak** temsil etmektir, örneğın
 - Düşük gelir
 - Orta-düşük gelir
 - Orta gelir
 - Orta-yüksek gelir
 - Yüksek gelir
 - Çok yüksek gelir



MODEL OLUŐTURMA

- Önceki ilçe özniteliđi gibi, böyle temsil edilen gelir birkaç ikili öznitelik kullanılarak modelimizde temsil edilecektir
- 1'de olan gelir-kategori özniteliđi müşterinin gelir grubunu belirtecektir



MODEL OLUŐTURMA

MüŐterinin geliri:

- Düşük gelir ($x_{11}=0$ veya 1)
- Orta-düşük gelir ($x_{12}=0$ veya 1)
- Orta gelir ($x_{13}=0$ veya 1)
- Orta-yüksek gelir ($x_{14}=0$ veya 1)
- Yüksek gelir ($x_{15}=0$ veya 1)
- Çok yüksek gelir ($x_{16}=0$ veya 1)



MODEL OLUŐTURMA

- MüŐterinin gelirinin kredi riskine etkisi bu Őekilde olacak “ $+θ_{11}x_{11} + θ_{12}x_{12} + θ_{13}x_{13} + θ_{14}x_{14}$ ”
- Böylece, farklı gelir grupların kredi riskine etkileri tamamen bağımsız olup, modelimiz daha esnek Őekilde kredi riskine gelirin katkısı deęerlenderebilecektir
 - Düşük gelir grubün ($x_{11}=1$) riske etkisi $θ_{11}$
 - Düşük-orta gelir grubün ($x_{12}=1$) riske etkisi $θ_{12}$
- Modelimiz iki kat daha düşük gelirin kredi riskine etkisi 2 kat düşük varsaymayacaktır



MODEL OLUŐTURMA

MüŐterinin önceki kredileri:

- MüŐterinin kredi tarihi, burada müŐteri tarafından alınmış ve ödenip ödenmemiş kredi sayısı olarak ifade edeceđiz
- Bu bilginin kredi riskine etkisi doğrusal olarak düşüneceđiz
- Böylece, bu faktör modelimizde **sürekli öznitelik** olarak girecektir



MODEL OLUŐTURMA

Sonuçta olan kredi risk modelimiz:

$$Risk = \theta_0 + \sum_{k=1}^{18} \theta_k \cdot X_k$$

x_1 - x_{10} müşteri oturan ilçe (ikili, 0 veya 1)

x_{11} - x_{16} müşteri gelir grubu (ikili, 0 veya 1)

x_{17} - x_{18} müşterinin önceki ödenmiş ve ödenmemiş kredi sayısı (iki sürekl  öznitelik)



- **Öbür derste**
 - ~~Temel makine öğrenme problem türleri~~
 - ~~Veri/öznitelik mühendisliği~~
 - ~~Model oluşturulması~~
 - **Model eğitimi**
 - **Modellerin performans değerlendirilmesi**
 - Tahminlerin üretilmesi
- Ayrıca, Amazon ML kullanılarak bir modelin oluşturulup denetlenmesi örneği