

DR. ÖĞR. ÜYESİ SİNAN UĞUZ



ÖĞRENME

MAKİNE.

TEORİK YÖNLERİ VE

python

UYGULAMALARI

ile Bir **YAPAY ZEKA** Ekolü

BÖLÜM 7

DOĞRUSAL REGRESYON

"İşlerinizi başkalarına yaptırmayın. Çünkü kendi istedikleri gibi yaparlar."

Cahit Arf

❖ 7.1 En Küçük Kareler Metodu

❖ 7.2 Eğim İniş Algoritması

❖ 7.3 Yığın Eğim İniş Algoritması

❖ 7.4 Rastgele Eğim İniş Algoritması

❖ 7.5 Mini-Yığın Eğim İniş Algoritması

❖ 7.6 Doğrusal Regresyon ile Örnek Python Uygulaması

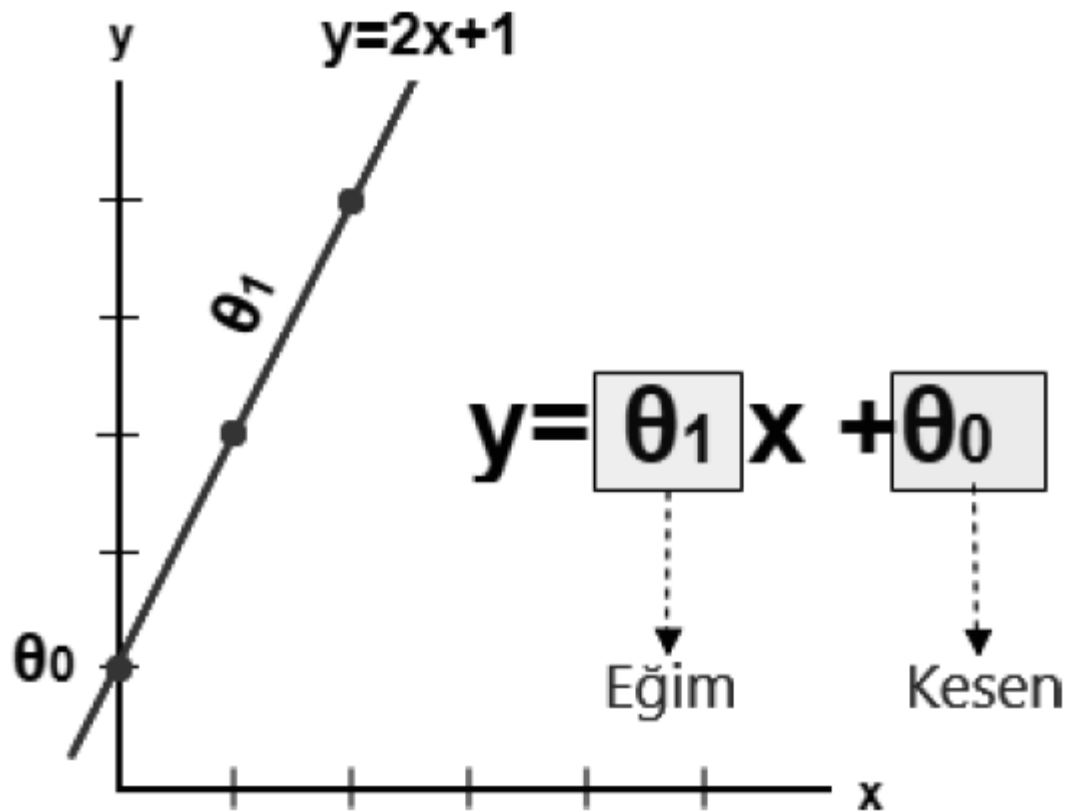
❖ 7.7 Bölüm Özeti

DOĞRUSAL REGRESYON

Doğrusal (linear) modeller, girdi özniteliklerinin doğrusal bir fonksiyon ile temsil edilerek, çıktı özniteliğinin tahmin edilmesine olanak tanıyan modellerdir.

Doğrusal modellerin geçmişi yüzyıldan fazla bir süreye dayanır ve bu modeller geniş bir uygulama alanına sahiptirler.

İstatistikte sayısal bir değerin tahmin edilmesi süreci için regresyon ifadesi kullanılır (aslında regresyon kelimesi, İngilizcede "önceki duruma döndürme" manasına gelmektedir ve tahmin etmek ile bir ilgisi yoktur).



Şekil 7.1: Bir doğrunun denklemi ve grafiği

7.1 En küçük Kareler Metodu

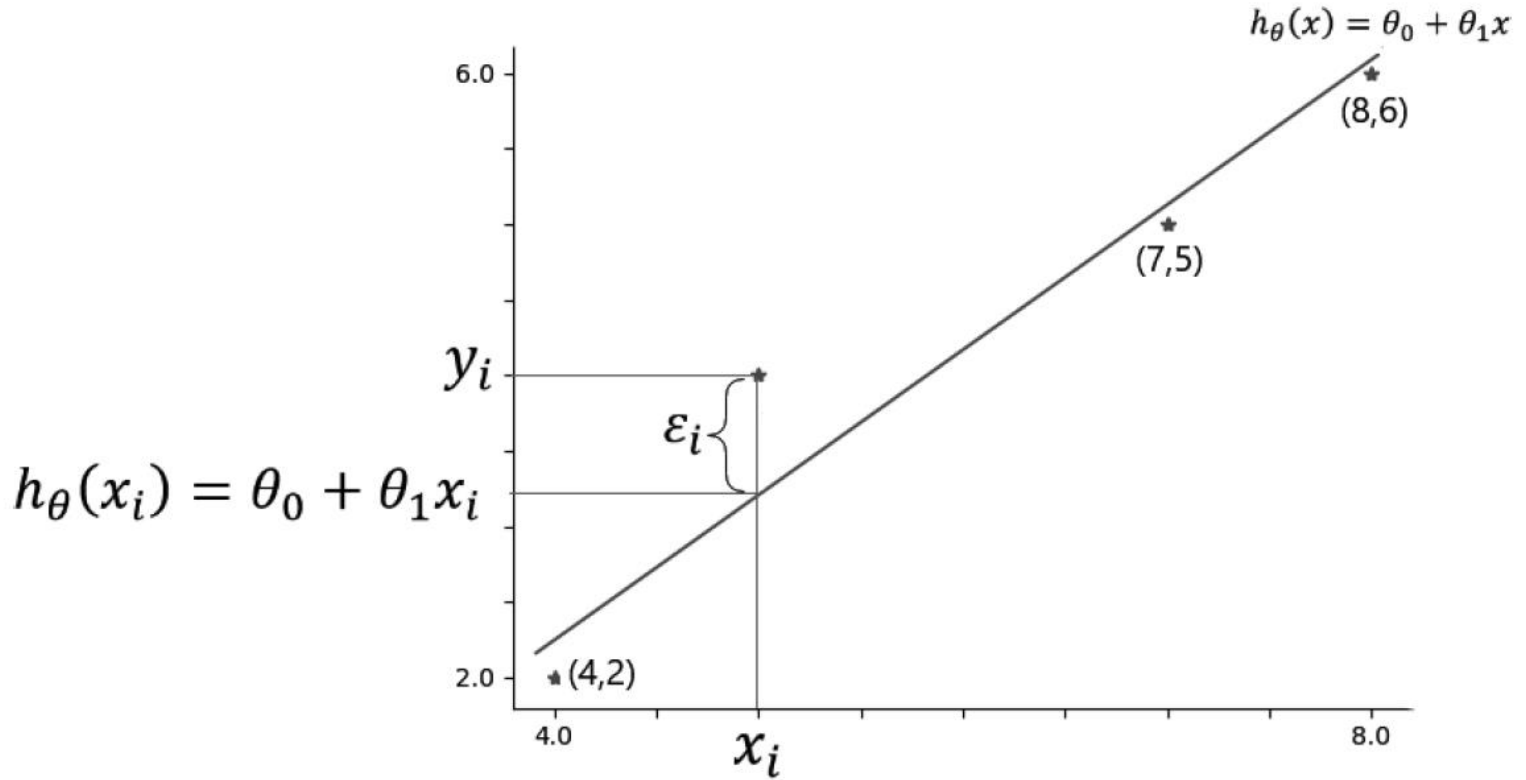
Regresyon analizi aslında, bağımlı ve bağımsız değişkenlerden oluşan veri setini, en iyi temsil eden fonksiyonu bulma sürecidir.

En iyi fonksiyon ile kastedilen nedir? Bu sorunun cevabı, 1795 yılında Gauss tarafından geliştirilen en küçük kareler metodu ile açıklanmaya çalışılacaktır.

Tablo 7.1'de bir ürünün renklerine göre aldığı tercih puanlarını içeren veriler görülmektedir.

Tablo 7.1

| Renk | Renk kodu | Tercih puanı |
|---------|-----------|--------------|
| Kırmızı | 4 | 2 |
| Yeşil | 5 | 4 |
| Mavi | 7 | 5 |
| Sarı | 8 | 6 |

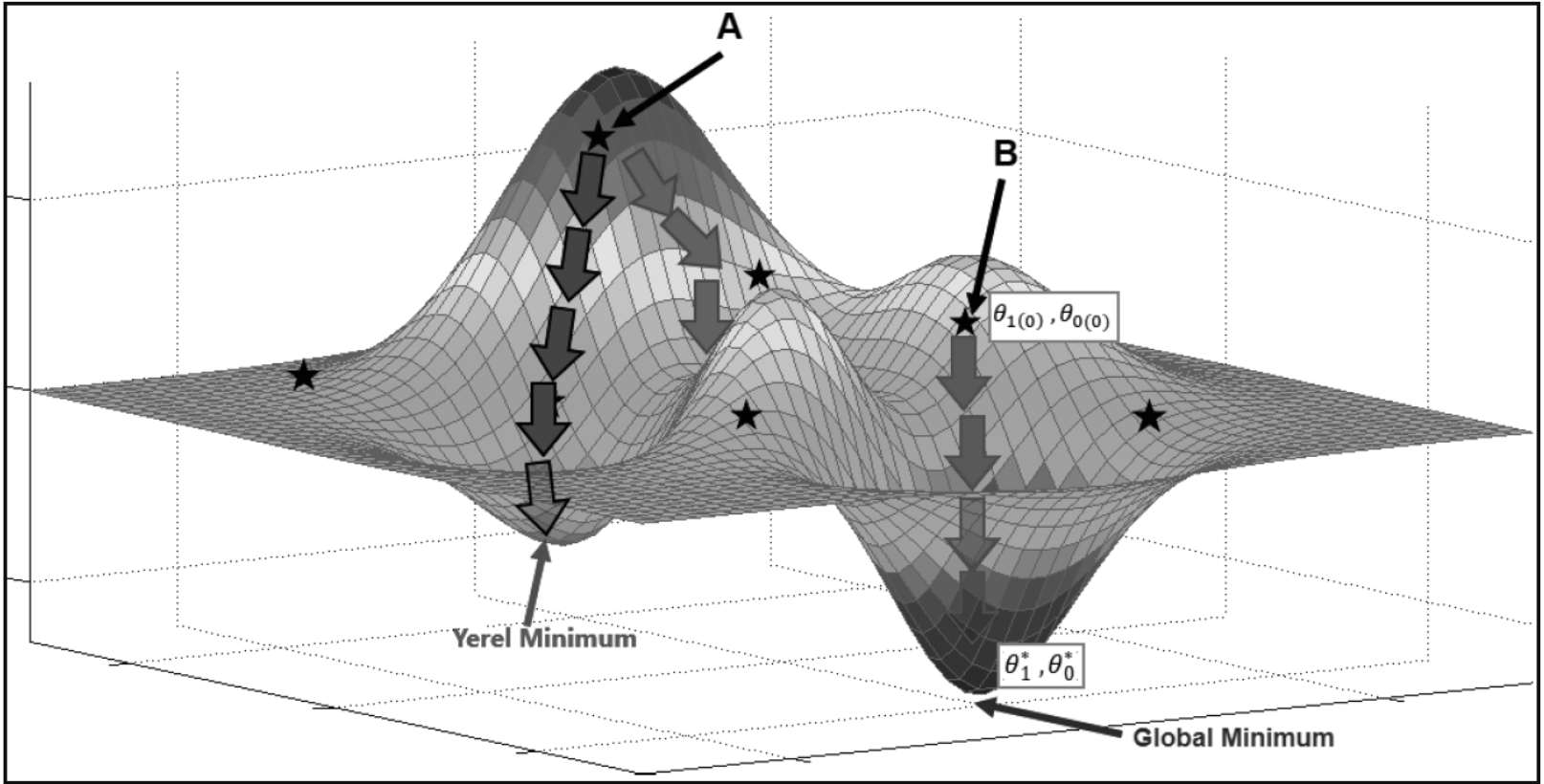


Şekil 7.2: En küçük kareler metodu ve hata kavramı

7.2 Eğim İniş Algoritması

En küçük kareler metodunun anlatıldığı bir önceki bölümde, hataların karelerinin toplamı ile bir fonksiyon oluşturulmuş ve bu fonksiyonun minimum noktasının bulunması halinde en küçük hata değerinin elde edildiğinden bahsedilmişti.

Bir fonksiyonun minimum noktasının bulunması aslında bir en iyileme (optimizasyon) problemidir ve bunu gerçekleştirmek için çeşitli en iyileme algoritmaları mevcuttur.



Şekil 7.3: Eğim iniş algoritması

7.3 Yığın Eğitim İniş Algoritması

Yığın eğitim iniş algoritmasında (Batch gradient descent), eğitim setindeki her bir örnek için hata hesaplanır.

Bu algoritmayı farklı kılan ise tüm örneklerin değerlendirildikten sonra modelin güncellenmesinin yapılmasıdır.

Daha az güncellemenin yapıldığı bir eğitim iniş algoritması çeşididir.

7.4 Rastgele Eğim İniş Algoritması

Yığın eğim iniş algoritmasının temel problemi olarak, algoritmanın her adımında eğimini hesaplamak için tüm eğitim veri setinin kullanılıyor olması gösterilebilir.

Bu durumda büyük veri setleri için algoritmanın yavaş çalışması söz konusu olmaktadır.

Rastgele eğim iniş algoritması (Stochastic gradient descent algorithm) ise bu durumun aksine, her adımda eğitim setinden rastgele bir örneği seçer ve sadece bu örneğe göre eğimi hesaplar.

7.5 Mini-Yığın Eğim İniş Algoritması

Bir diğer eğim iniş algoritması çeşidi de mini-yığın eğim iniş algoritmasıdır (mini-batch gradient descent algorithm).

Bu algoritmayı anlamak için önceki algoritmaların iyi anlaşılmış olması önemlidir.

Bu algoritmada, yığın eğim iniş algoritmasındaki gibi tüm eğitim setine dayanan veya rastgele eğim iniş algoritmasındaki gibi sadece bir örneğe dayanan eğim hesaplaması yapmak yerine, rastgele seçilen ve mini-yığın olarak adlandırılan örnekler kümesine dayanarak eğim hesaplaması gerçekleştirilir.

7.6 Doğrusal Regresyon ile Örnek Python Uygulaması

Bu bölümde, bir veri setinin doğrusal regresyon ile modellenerek, modele ait denklemin grafiğinin çizilmesi ve performansının Python kütüphaneleri yardımıyla elde edilmesi üzerinde durulmuştur.

Liste 7.1'de öncelikle, veriseti adlı DataFrame nesnesinin pandas kütüphanesi yardımıyla oluşturulması görülmektedir.

⇒ Veri setine aşağıdaki bağlantıdan ulaşılabilir.

<https://github.com/mubaris/potential-enigma/blob/master/headbrain.csv>

Liste 7.1

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

```
# Veri setinin alınması
veriseti = pd.read_csv('DogrusalRegresyon.csv')
veriseti.head()
```

```
>>>
```

| | Cinsiyet | Yas_Araligi | Bas_cevresi(cm^3) | Beyin_agirligi(gr) |
|---|----------|-------------|-------------------|--------------------|
| 0 | 1 | 1 | 4512 | 1530 |
| 1 | 1 | 1 | 3738 | 1297 |
| 2 | 1 | 1 | 4261 | 1335 |
| 3 | 1 | 1 | 3777 | 1282 |
| 4 | 1 | 1 | 4177 | 1590 |

7.7 Bölüm Özeti

Bir doğrusal regresyon modeli, girdi özniteliği ve çıktı özniteliği arasında doğrusal bir ilişki olduğu varsayımına dayanarak, bu öznitelikleri en iyi temsil eden fonksiyonun en küçük kareler metoduna dayanan bir teknik ile bulunması şeklinde ifade edilebilir.

En küçük kareler metodunda, iki boyutlu bir düzlemde veri setine ait tüm noktaları temsil eden en uygun doğruyu çizmek hedeflenir.