

Örüntü Tanıma

Emir Öztürk

Oğuz Kırat

Giriş

- ML yöntemlerine giriş
- Denetimli ve denetimsiz öğrenme
- Yanlılık (bias) ve variance
- Overfitting/underfitting

Makine öğrenmesi

- Bir işte performansın artması için deneyim kazanmak «öğrenme»
- Klasik bir programlama işleminde oluşturulan kod girdiye bir çıktı üretir
- Makine öğrenmesinde ise oluşturulan kod girdiye bir kod (model) üretir.

Makine öğrenmesi

- Makine öğrenmesinin kullanılmasının farklı ortamlarda farklı amaçları bulunmaktadır.
 - İnsanın olmadığı / olamadığı yerlerde
 - İnsanın iş yükünü azaltmak için
 - İnsanların tespit edemediği örüntülerin tespitinde
 - İnsanların işleyemeyeceği kadar büyük bir verinin işlenmesinde

Makine öğrenmesi

- Örüntülerin tanınması
 - Yüz, ses, el yazısı, medikal görüntüler
- Örüntülerin oluşturulması
 - Resim oluşturma, ses sentezleme
- Örüntü üzerindeki anomali tespiti
 - Görüntü olmak zorunda değil
 - Beklenmeyen bir işlem
 - Paket analizi
 - Banka hesap giriş çıkışları
- Gelecek tahmini
 - Hava, iş stoğu, finans

Öğrenme Türleri

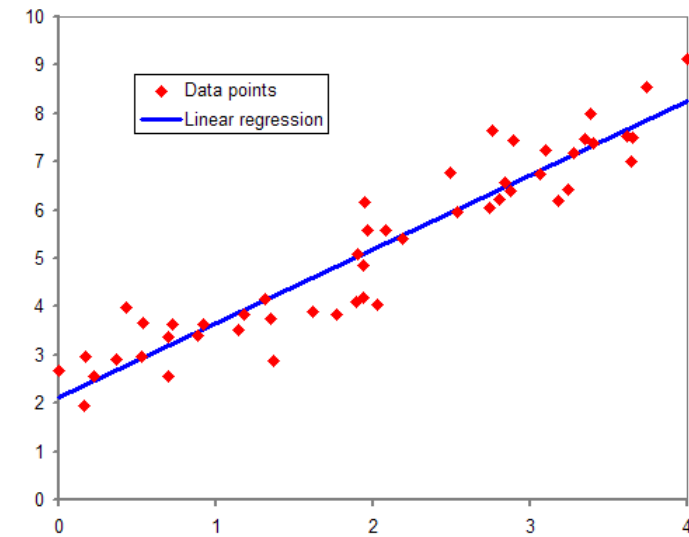
- Denetimli öğrenme
 - Veri ve istenilen çıktı
- Denetimsiz öğrenme
 - Veri
- Yarı denetimli öğrenme
 - Veri ve belirli çıktılar
- Takviyeli öğrenme
 - Aksiyonlara verilen ödüller

Denetimli öğrenme

- Bir girdi verisine karşılık istenilen çıktının bilindiği durum
- Kişiler ve biyometrik verileri
- Yazılar ve kime ait oldukları / türleri
- Sesler ve bu seslerin kime / neye ait olduklarının tanımlı olması
- Sensör verileri ve bu verilerin üretmesi gereken çıktı

Denetimli öğrenme - Regresyon

- Sürekli verilerin eldesinde kullanılır
- Girdi verisine karşılık bir çıktı değeri üretilir
- Verilerin bir sınırı yoktur
- Eğri uydurma örneği verilebilir
- Lineer
- Polinomial



Denetimli öğrenme – Lineer Regresyon

X	Y
0	0
1	5
2	12
3	15
4	24
5	25
6	36

Denetimli öğrenme – Lineer Regresyon

X	Y	X*5	Kayıp (12)
0	0	0	0
1	5	5	0
2	12	10	2
3	15	15	0
4	24	20	4
5	25	25	0
6	36	30	6

Denetimli öğrenme – Lineer Regresyon

X	Y	X*6	Kayıp (9)
0	0	0	0
1	5	6	1
2	12	12	0
3	15	18	3
4	24	24	0
5	25	30	5
6	36	36	0

Denetimli öğrenme – Linear Regresyon

```
import matplotlib
matplotlib.use('TkAgg')
import matplotlib.pyplot as plt
from sklearn import linear_model
import pandas as pd

df = pd.read_csv("/Users/emirozturk/Desktop/deneme.csv")

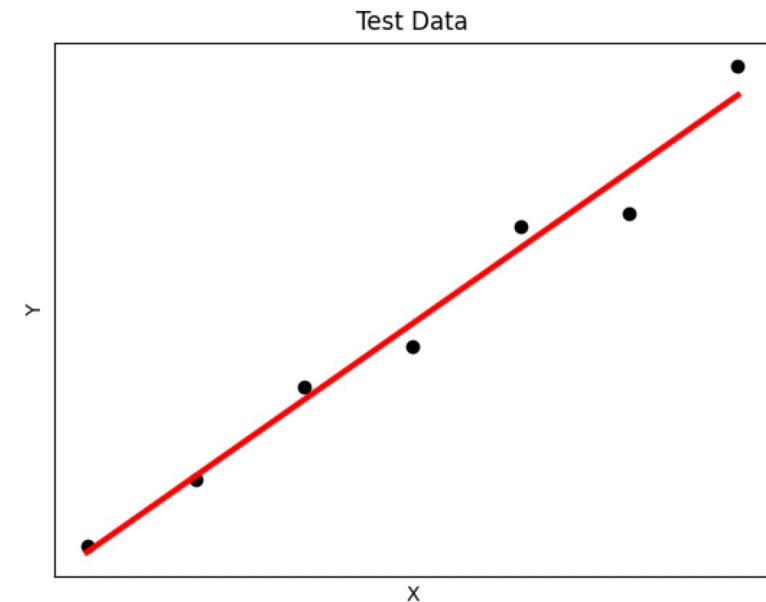
Y = df['y'].to_numpy().reshape(7, 1)
X = df['x'].to_numpy().reshape(7, 1)

plt.scatter(X, Y, color='black')
plt.title('Test Data')
plt.xlabel('X')
plt.ylabel('Y')
plt.xticks(())
plt.yticks(())

regr = linear_model.LinearRegression()
regr.fit(X, Y)

print(f'Linear equation: Y = {regr.coef_[0][0]} * X + {regr.intercept_[0]}')

plt.plot(X, regr.predict(X), color='red', linewidth=3)
plt.show()
```



$$Y = 5.714285714285713 * X + -0.4285714285714235$$

Denetimli öğrenme – Lineer Regresyon

X	Y	Regr.	Kayıp (10,3)
0	0	-0,4285714	0,42857143
1	5	5,28571429	0,28571429
2	12	11	1
3	15	16,7142857	1,71428571
4	24	22,4285714	1,57142857
5	25	28,1428571	3,14285714
6	36	33,8571429	2,14285714

Denetimli Öğrenme - Sınıflandırma

- Kategorik verilerin eldesinde kullanılır
- Girdi verisine karşılık bir kategori / sınıf değeri üretilir
- Veriler eğitim esnasında etiketli olarak verilir
- Etiketler başından belli olduğu için çıktı verisinin kümesi sınırlıdır
- Veri girdisinden sonra etiket olasılıkları çıktı olarak elde edilir

Denetimli Öğrenme - Sınıflandırma

X	Y	Y>15?
0	0	0
1	5	0
2	12	0
3	15	0
4	24	1
5	25	1
6	36	1

Denetimli Öğrenme - Sınıflandırma

X	Y	Y>15?
0	0	0
6	36	1
1	5	0
2	12	0
4	24	1
3	15	0
5	25	1

```
Accuracy: 0.6666666666666666
Test set predictions: [0 0 1]
Test set actual values: [1 0 1]
Logistic regression coefficients: [0.86590451]
Intercept: -3.7574970236705445
```


Denetimsiz Öğrenme

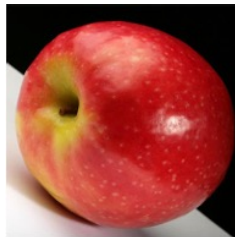
- Girdi verisi bulunur fakat bir çıktı verisi bulunmaz
- Verilen verinin ayrılması, kümelenmesi veya gruplanması amaçlanır

Denetimsiz Öğrenme - Kümeleme

- Verilen verinin gruplanması için belirli yöntemler kullanılır
- Küme sayısı verilebilir veya benzerliğe göre otomatik belirlenebilir
- Sınıf sayısı bilinmediğinden küme sayısına göre sonuçlar değişebilir

Denetimsiz Öğrenme - Kümeleme

- 2 sınıf?



Denetimsiz Öğrenme - Kümeleme

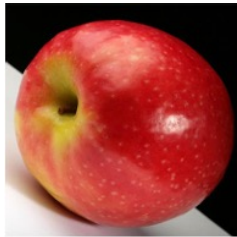


Denetimsiz Öğrenme - Kümeleme

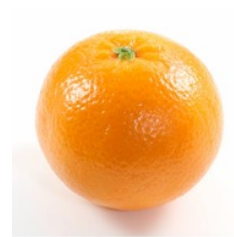
- 3 sınıf?



Denetimsiz Öğrenme - Kümeleme



Denetimsiz Öğrenme - Kümeleme



Denetimsiz Öğrenme - Kümeleme

- 3 sınıf?



https://sites.astro.caltech.edu/~george/aybi199/Donalek_Classif.pdf

Takviyeli Öğrenme

- Herhangi bir girdi veya çıktı verisi eğitim için toplu bir şekilde verilmez
- Agent
- Ortam
- Ödül

Takviyeli Öğrenme



Yöntem Seçimi

- Girdi ve çıktı verisi mevcutsa
 - Eğer veriler sürekli ise
 - Regresyon
 - Eğer etiketli veriler mevcutsa
 - Sınıflandırma
- Girdi verisi mevcut fakat çıktı verisi bulunmuyorsa
 - Eğer verilerin gruplanması isteniyorsa
 - Kümeleme
- Eğer veri üretimi / boyut azaltımı gibi çıktı verisi bulunmadan karar alınması gerekiyorsa?

Verisetinin Hazırlanması - Variance

- Verinin istenilen tüm örneklerle sahip olması gerekir
- Verinin her örnekten dengeli veriye sahip olması gerekir
- Veri sınıflandırma aşamasında bias olmamalı
- Eğitim, validasyon ve test olarak bölünmeli
 - Bölünen veride de aynı denge korunmalı
- Verilerin bir ön işleme tabi tutulması gerekir
- Her veri türü için farklı ön işlemler mevcuttur

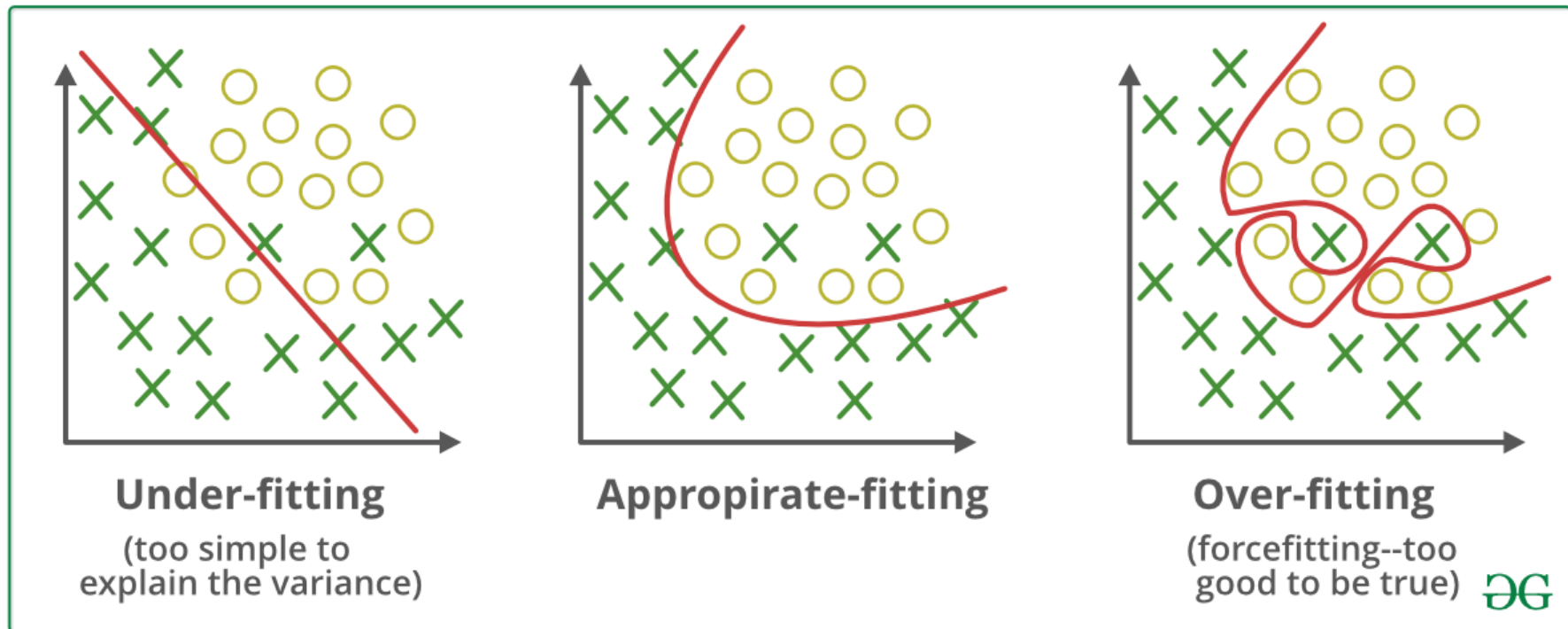
Bias

- Bir çıktıya iyi veya kötü yönde yönelim
- Objektifliğin bozulması anlamına gelir
- İsteyerek veya istemeden olabilir
- Farkındalık ile
 - Gender bias
 - Racial bias
 - Age bias
- Farkında olmadan
 - Affinity bias
 - Beauty bias
 - Halo effect (Bir iyi özelliğe göre tüm özelliklere karar vermek)
 - Horns effect (Negative halo)

Başarının Ölçümü

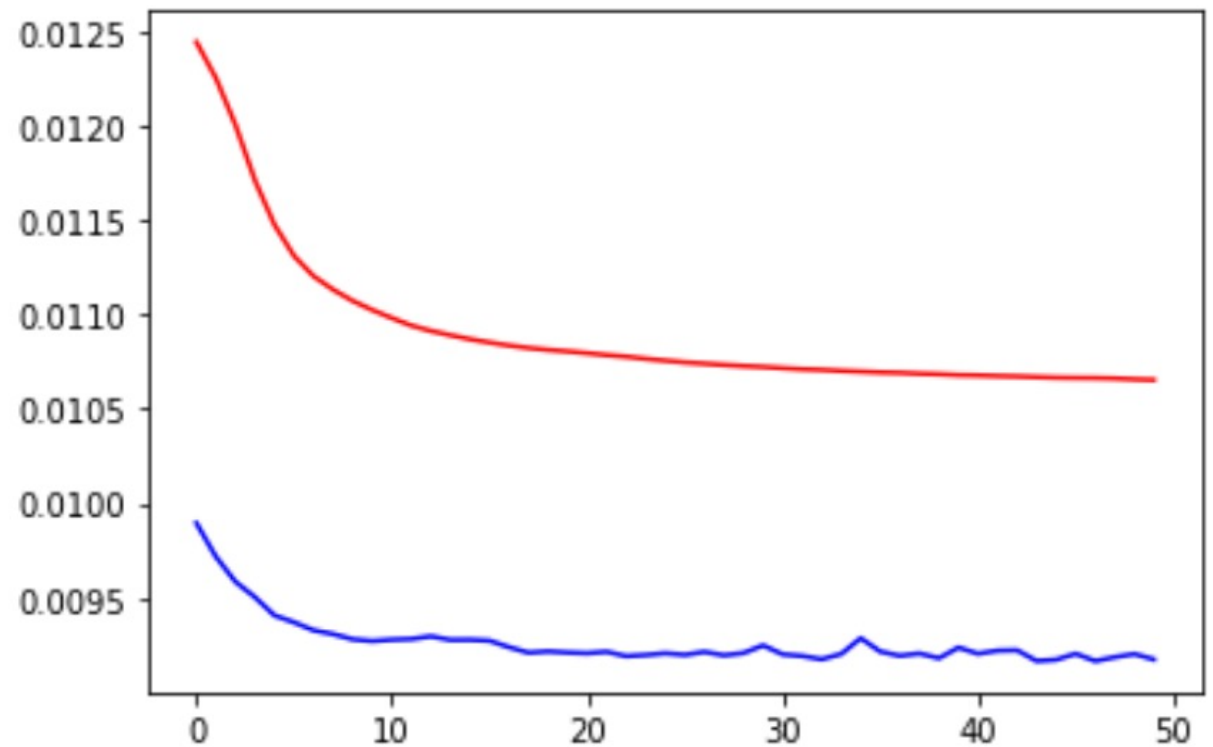
- Accuracy
 - F1 Score
 - RoC
- Loss
 - MSE, RMSE, MAE, PAE
- Kümeleme için skor ve indeksler
- Takviyeli öğrenme için «ödül»

Overfitting ve Underfitting



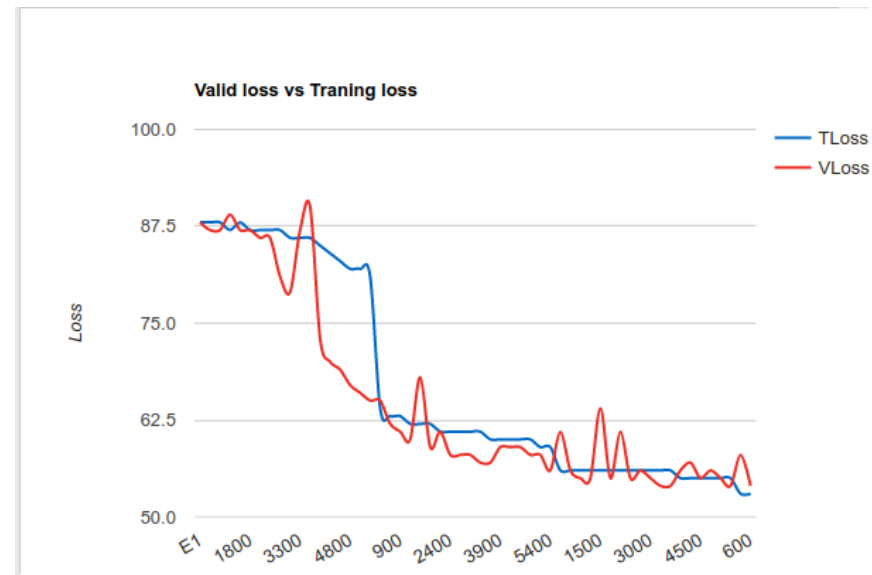
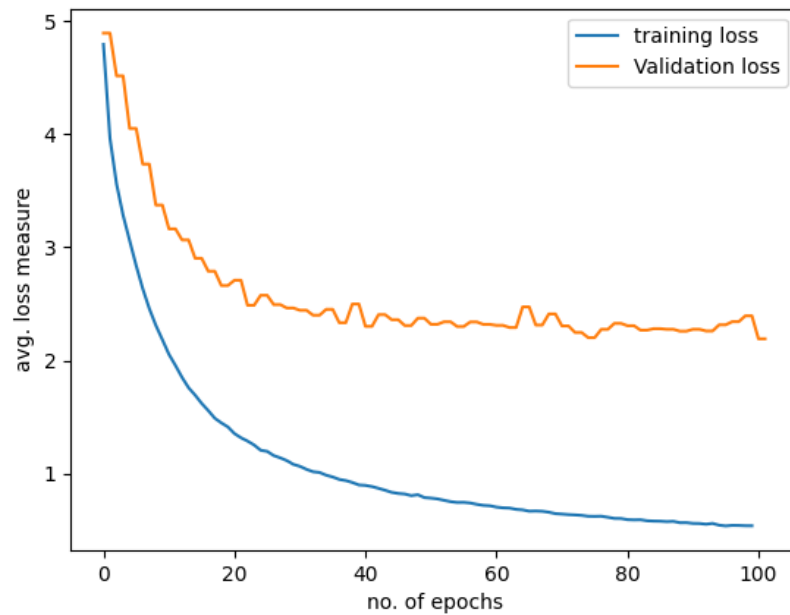
Overfitting ve Underfitting

- Kırmızı Train ise
- Kırmızı Test ise



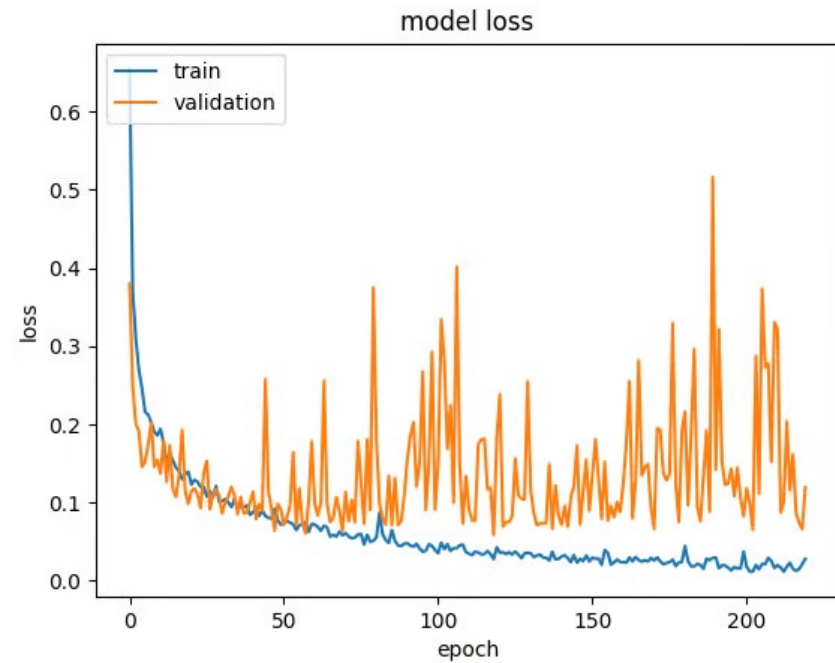
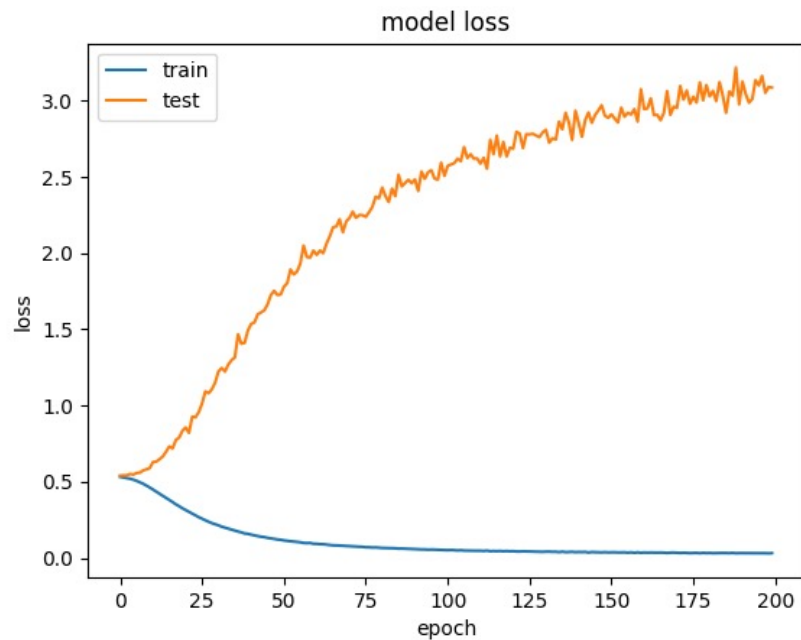
Overfitting ve Underfitting

- Sorun?



Overfitting ve Underfitting

- Sorun?



Overfitting ve Underfitting

- Sorun?

