

Makine Öğrenmesi - 1

by Sefa Isci



Makine Öğrenmesi Giriş

İnsanoğluna insan üstü fayda!

Makine Öğrenmesi Nedir?

Bilgisayarların insanlara benzer şekilde öğrenmesini sağlamak amacıyla çeşitli algoritma ve tekniklerin geliştirilmesi için çalışılan bilimsel çalışma alanıdır.

Makine Öğrenmesi Nedir?

Ülke	Yaş	Maaş	Satın Alma Durumu
Türkiye	55	72000	Hayır
İspanya	32	60000	Evet
Almanya	42	50000	Evet
Fransa	45	30000	Hayır
Türkiye	35	32000	Evet
Almanya	32	35000	Hayır
İspanya	50	70000	Hayır
Almanya	29	32000	Evet
Türkiye	30	33000	Evet
Fransa	40	40000	Evet

Makine Öğrenmesi Nedir?

Ülke	Yaş	Maaş	Satın Alma Durumu
Türkiye	55	72000	Hayır
İspanya	32	60000	Evet
Almanya	42	50000	Evet
Fransa	45	30000	Hayır
Türkiye	35	32000	Evet
Almanya	32	35000	Hayır
İspanya	50	70000	Hayır
Almanya	29	32000	Evet
Türkiye	30	33000	Evet
Fransa	40	40000	Evet

- 32 yaşında 30 Bin TL geliri olan bir Türkiye vatandaşı ürün alır mı?
- 46 yaşında 32 Bin TL geliri olan bir Fransız ürünü alır mı?
- Türkiye'ye en benzer davranışı gösteren ülke hangisidir?

Makine Öğrenmesi Nedir?

Arabam Ne Kadar Eder?



$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_j X_{ij} + \dots + \beta_p X_{ip} + \varepsilon_i$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

$$\hat{y} = \boxed{b_0} + \boxed{b_1} \boxed{x_1} + \boxed{b_2} \boxed{x_2} + \dots + \boxed{b_p} \boxed{x_p}$$

↙

$$\text{Araç fiyatı} = \boxed{80000} - \boxed{0,2} * \boxed{60000} - \boxed{2500} * \boxed{\text{Hasar Durumu}} + \dots + \boxed{6000} * \boxed{\text{Vites Türü}}$$

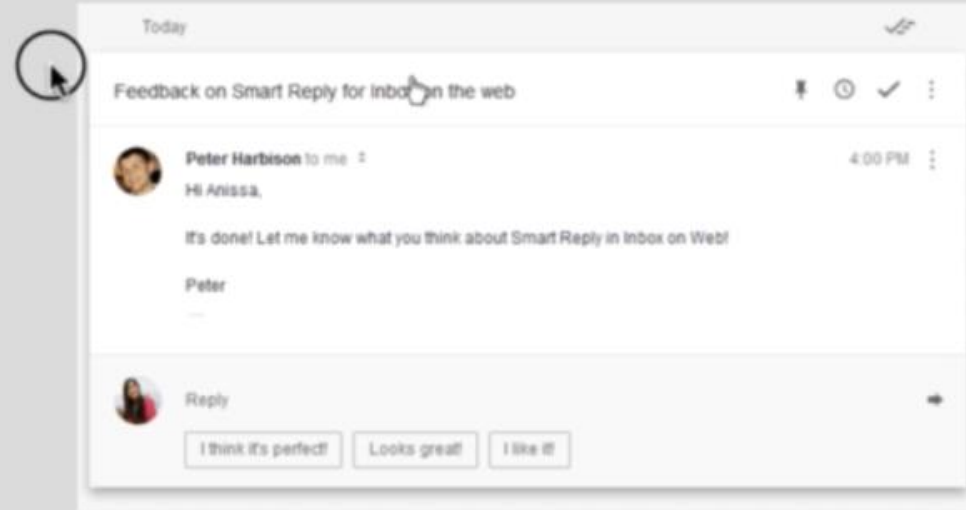
Gerçek Hayat Örnekleri

- fotoğraftan araç tanıma uygulaması



Gerçek Hayat Örnekleri

- LinkedIn, letgo ve gmail hazır cevap uygulamaları



Gerçek Hayat Örnekleri

- Chatbotlar ve kişisel asistanlar (Siri ve diğerleri)



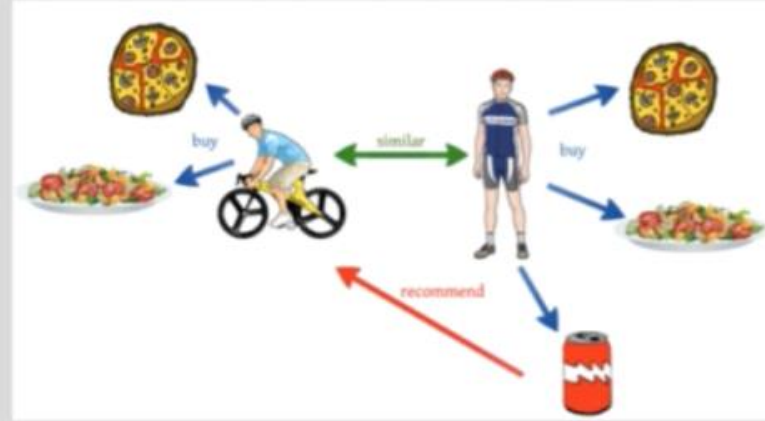
Gerçek Hayat Örnekleri

■ Facebook



Gerçek Hayat Örnekleri

- Netflix, Amazon ve E Ticarete tavsiye sistemleri



Gerçek Hayat Örnekleri

- Spam Engelleme



Gerçek Hayat Örnekleri

- Cümle tamamlama ve smile tamamlama



Gerçek Hayat Örnekleri

- Uber



Gerçek Hayat Örnekleri

- Sahtekarlık Önleme Çalışmaları (Fraud Prevention)



Gerçek Hayat Örnekleri

- Kredi Başvurusu Değerlendirme



Daha Gerçek Makine Öğrenmesi



Terminoloji

Makine öğrenmesinde sık kullanılan kavramlara kısa bir bakış

Kavramlar

Bağımlı Değişken &
Bağımsız Değişken



	y1	x1	x2	x3	x4
1	8.04	10	10	10	8
2	6.95	8	8	8	8
3	7.58	13	13	13	8
4	8.81	9	9	9	8
5	8.33	11	11	11	8
6	9.96	14	14	14	8
7	7.24	6	6	6	8
8	4.26	4	4	4	19
9	10.84	12	12	12	8
10	4.82	7	7	7	8
11	5.68	5	5	5	8

Kavramlar

Öğrenme Türleri

Gözetimli Öğrenme

Gözetimsiz Öğrenme

Yarı Gözetimli Öğrenme

	y1	x1	x2	x3	x4
1	8.04	10	10	10	8
2	6.95	8	8	8	8
3	7.58	13	13	13	8
4	8.81	9	9	9	8
5	8.33	11	11	11	8
6	9.96	14	14	14	8
7	7.24	6	6	6	8
8	4.26	4	4	4	19
9	10.84	12	12	12	8
10	4.82	7	7	7	8
11	5.68	5	5	5	8

	x1	x2	x3	x4
1	10	10	10	8
2	8	8	8	8
3	13	13	13	8
4	9	9	9	8
5	11	11	11	8
6	14	14	14	8
7	6	6	6	8
8	4	4	4	19
9	12	12	12	8
10	7	7	7	8
11	5	5	5	8

Kavramlar

Problem Türü:
Regresyon vs
Sınıflandırma

	y1	x1	x2	x3	x4
1	8.04	10	10	10	8
2	6.95	8	8	8	8
3	7.58	13	13	13	8
4	8.81	9	9	9	8
5	8.33	11	11	11	8
6	9.96	14	14	14	8
7	7.24	6	6	6	8
8	4.26	4	4	4	19
9	10.84	12	12	12	8
10	4.82	7	7	7	8
11	5.68	5	5	5	8

	y1	x1	x2	x3	x4
1	1	10	10	10	8
2	0	8	8	8	8
3	1	13	13	13	8
4	1	9	9	9	8
5	1	11	11	11	8
6	1	14	14	14	8
7	1	6	6	6	8
8	0	4	4	4	19
9	1	12	12	12	8
10	0	7	7	7	8
11	0	5	5	5	8

Kavramlar

Değişken Türleri



- Değişken Türleri
 - Sayısal Değişkenler (nicel, kantitatif)
 - Kategorik Değişkenler (nitel, kalitatif)
- Ölçek Türleri
 - Sayısal değişkenler için: Aralık ve Oran
 - Kategorik değişkenler için: Nominal ve Ordinal

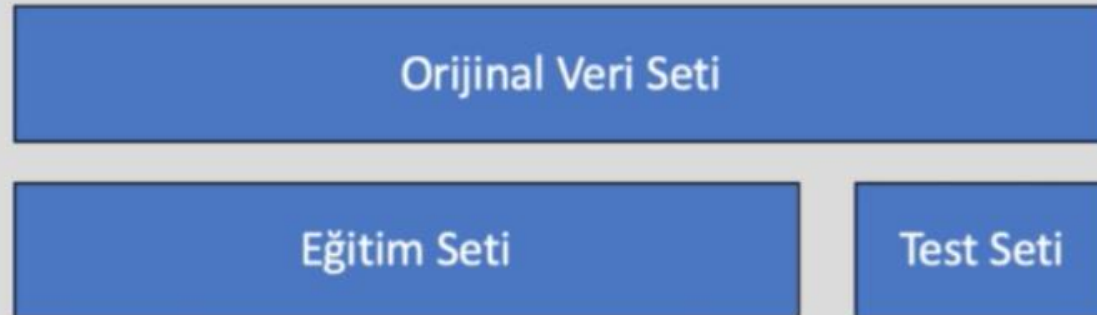


Screenshot saved

The screenshot was added to your
OneDrive.
OneDrive

Kavramlar

Test-Train Ayrımı



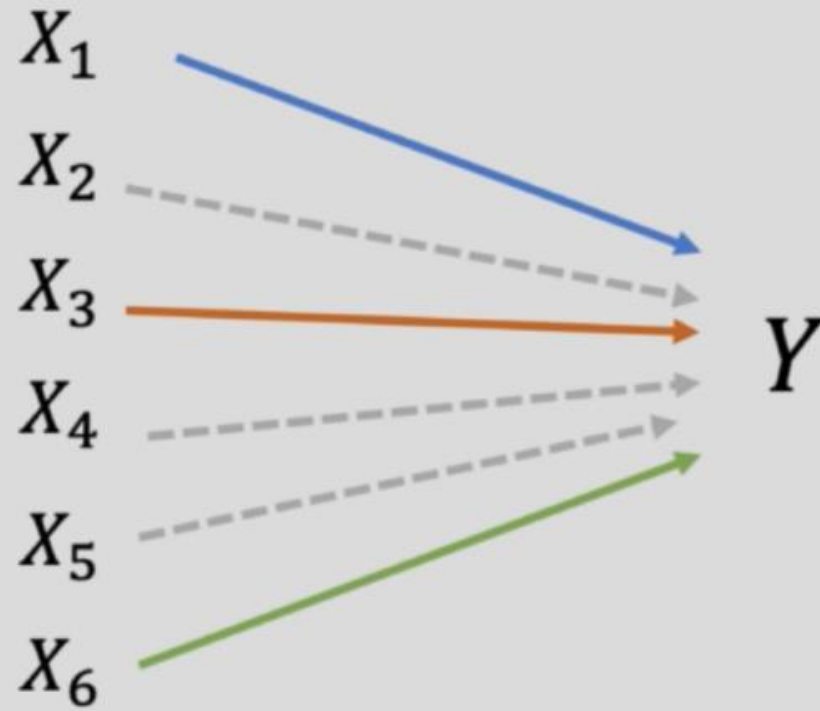
Kavramlar

Değişken Mühendisliği
(Feature Engineering)



Kavramlar

Değişken Seçimi
(Variable Selection)



Kavramlar

Model Seçimi



Model Seçimi için İki Durum Söz Konusu

- **Birincisi:** Oluşabilecek değişken kombinasyonları ile oluşturulan modeller arasından en iyi modelin seçilmesi
- **İkincisi:** Kurulan birbirinden farklı modeller arasından model seçimi

Model Neye Göre Seçilir?

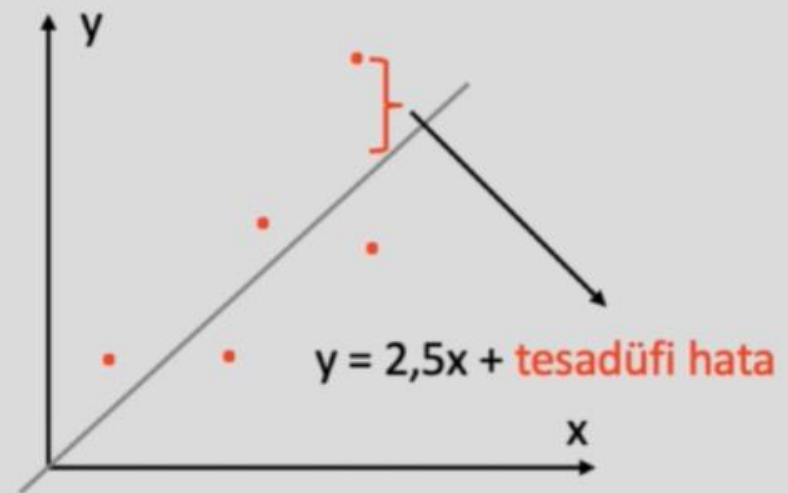
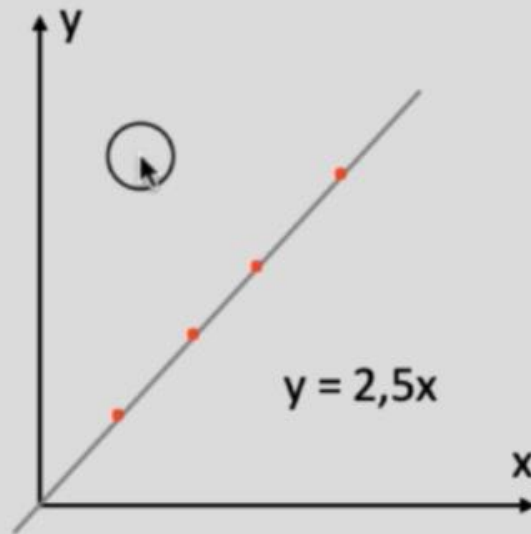
- Regresyon için açıklanabilirlik oranı ve RMSE benzeri bir değer.
- Sınıflandırma için doğru sınıflandırma oranı benzeri bir değer.

Kavramlar

Aşırı Öğrenme
(Overfitting)

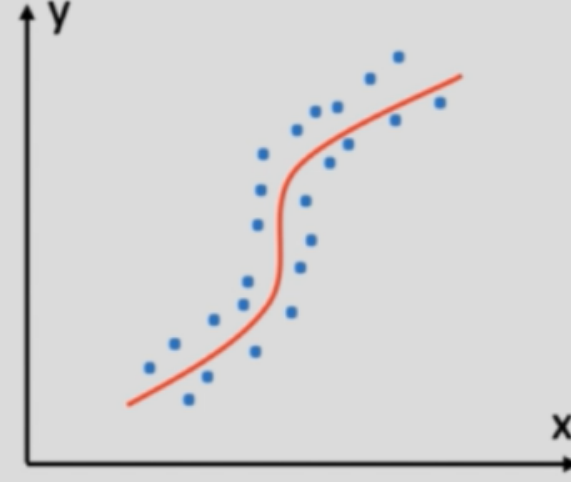
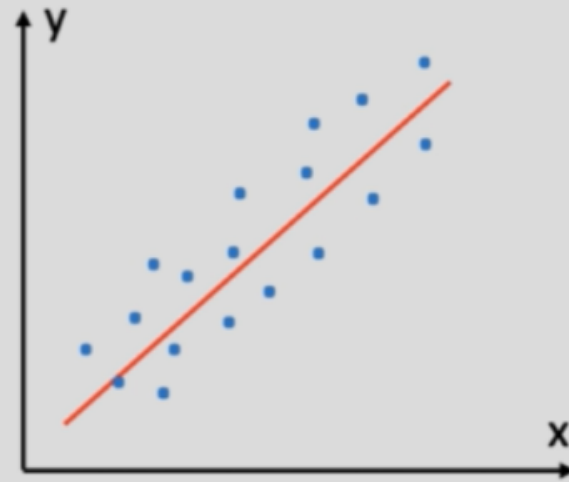


Deterministik Modeller vs Stokastik Modeller



Olasılıksal Model: $y = 2,5x + \text{tesadüfi hata}$

Doğrusal Modeller vs Doğrusal Olmayan Modeller

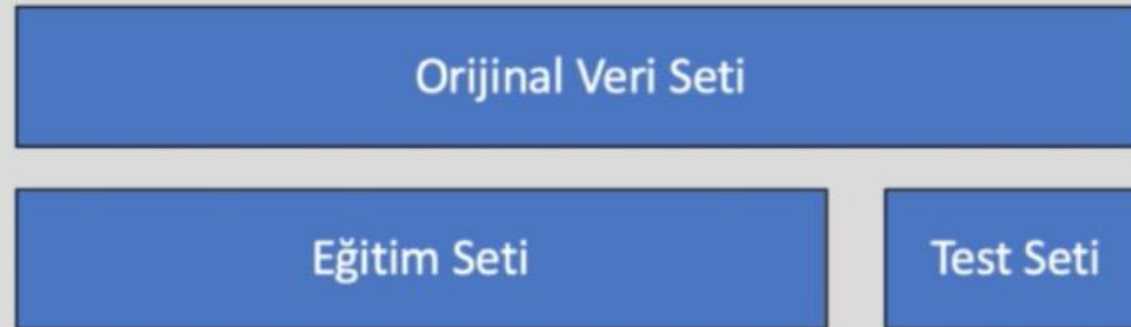


Model Doğrulama Yöntemleri

Modellerin ürettiği sonuçların doğru değerlendirilmesi çalışmaları

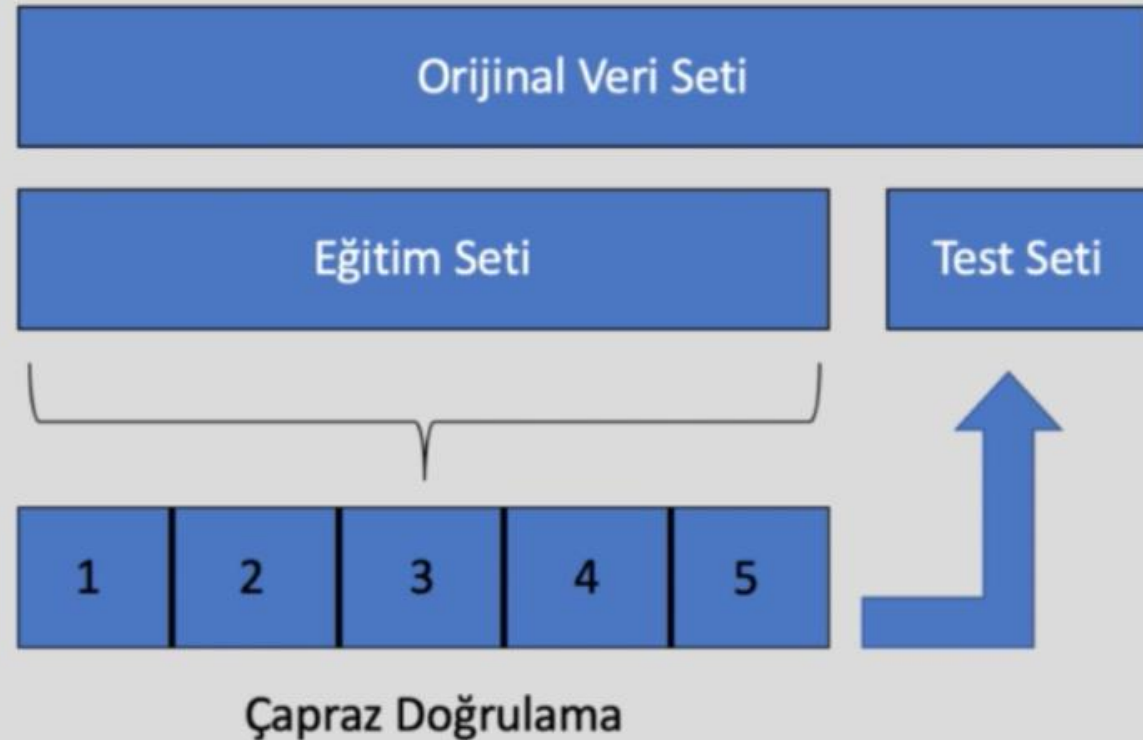
Model Doğrulama Yöntemleri

Holdout Yöntemi



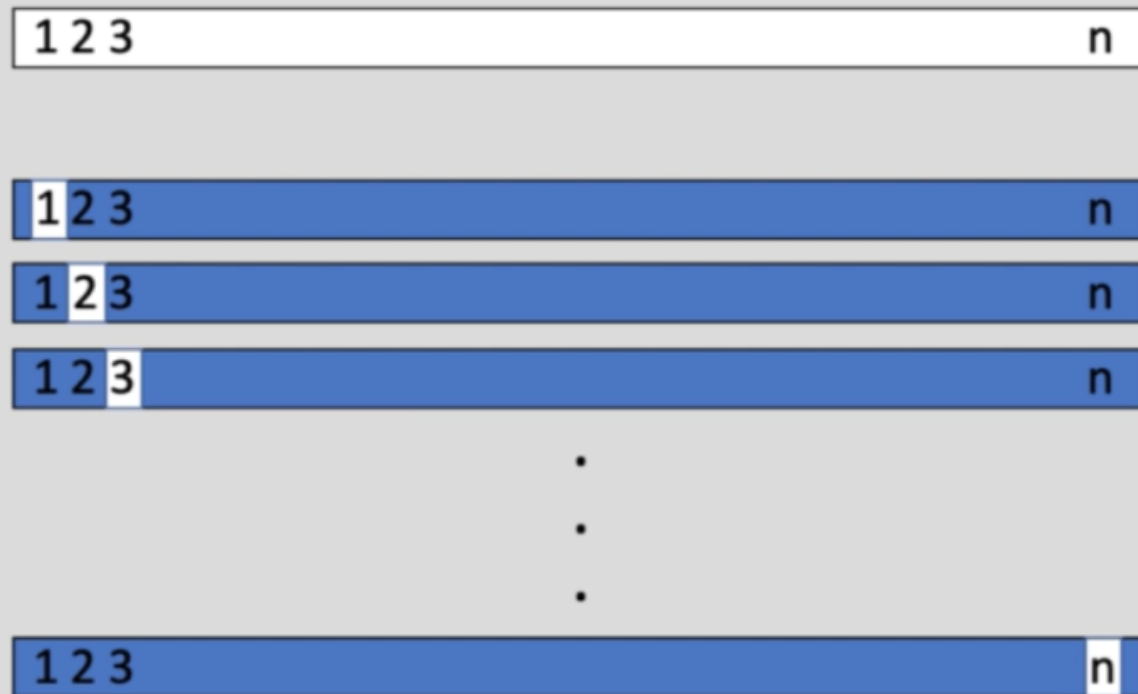
Model Doğrulama Yöntemleri

K-Katlı Çapraz Doğrulama (k fold cross validation)



Model Doğrulama Yöntemleri

Leave One Out



Model Doğrulama Yöntemleri

Bootstrap

Orijinal Veri Seti

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Bootstrap 1	x_3	x_2	x_{10}	x_4	x_1	x_8	x_7	x_6	x_9	x_5
Bootstrap 2	x_8	x_{10}	x_1	x_6	x_3	x_2	x_7	x_9	x_5	x_4
Bootstrap 3	x_1	x_4	x_2	x_3	x_{10}	x_5	x_7	x_9	x_6	x_8

Model Başarı Değerlendirme Yöntemleri

Modellerin Tahmin Başarılarının Değerlendirilmesi

Model Başarı Değerlendirme - Regresyon

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Model Başarı Değerlendirme - Regresyon

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Gerçek Değerler Tahmin Edilen Değerler

Gözlem Sayısı

The diagram illustrates the Mean Squared Error (MSE) formula. The formula is $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$. The fraction $\frac{1}{n}$ is enclosed in a blue box, with a blue arrow pointing from it to the text 'Gözlem Sayısı' (Number of Observations) below. The term y_i is enclosed in a green box, with a green arrow pointing from it to the text 'Gerçek Değerler' (Actual Values) above. The term \hat{y}_i is enclosed in an orange box, with an orange arrow pointing from it to the text 'Tahmin Edilen Değerler' (Predicted Values) above.

Model Başarı Değerlendirme - Regresyon

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Model Başarı Değerlendirme - Regresyon

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$



Screenshot saved

The screenshot was added to your
OneDrive.
OneDrive

Model Başarı Değerlendirme - Regresyon

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

	y	x1	x2	x3	y_sapka	farklar	kareler
1	8.04	10	10	10	7.00	1.04	1.0816
2	6.95	8	8	8	6.00	0.95	0.9025
3	7.58	13	13	13	8.00	-0.42	0.1764
4	8.81	9	9	9	9.00	-0.19	0.0361
5	8.33	11	11	11	8.00	0.33	0.1089
6	9.96	14	14	14	9.00	0.96	0.9216
7	7.24	6	6	6	7.21	0.03	0.0009
8	4.26	4	4	4	4.10	0.16	0.0256
9	10.84	12	12	12	10.00	0.84	0.7056
10	4.82	7	7	7	5.00	-0.18	0.0324
11	5.68	5	5	5	6.00	-0.32	0.1024

+

$$4.094 / 11 = 0.3721818$$

Model Başarı Değerlendirme - Sınıflandırma

Karışıklık Matrisi

		Tahmin Edilen Sınıf	
		Sınıf = 1	Sınıf = 0
Gerçek Sınıf	Sınıf = 1	a	b
	Sınıf = 0	c	d

a: True Pozitif (TP)

d: True Negatif (TN)

c: False Pozitif (FP)

b: False Negatif (FN)

Doğruluk: $(TP+TN) / \text{Hepsi}$

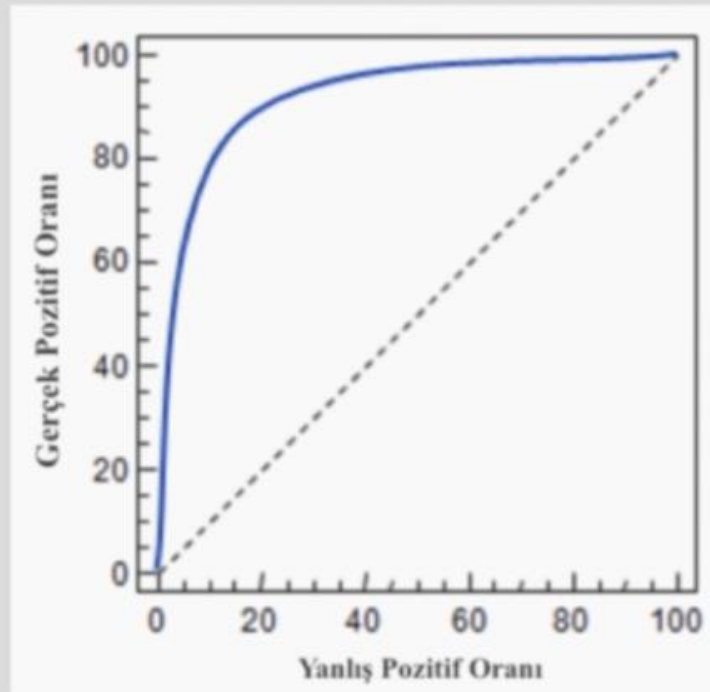
Hata Oranı: $(FN+FP) / \text{Hepsi}$

Kesinlik: $TP / (TP+FP)$

Anma: $TP / (TP + FN)$

Model Başarı Değerlendirme - Sınıflandırma

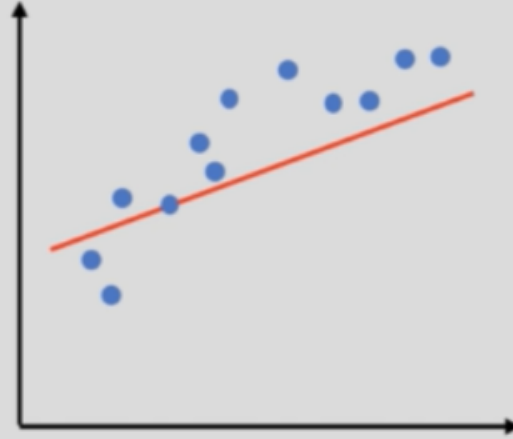
ROC Eğrisi



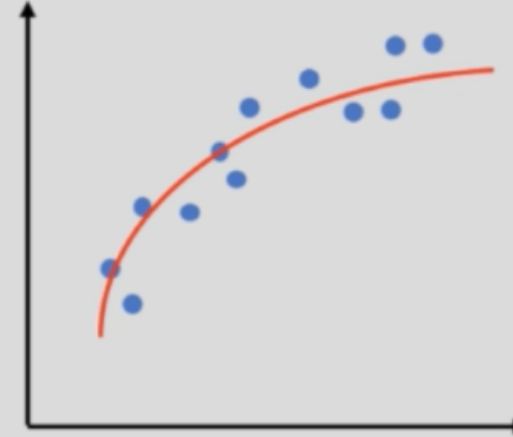
Yanlılık - Varyans Değiş Tokuşu (Bias-Variance Tradeoff)

Modellerin Tahmin Başarılarının Değerlendirilmesi

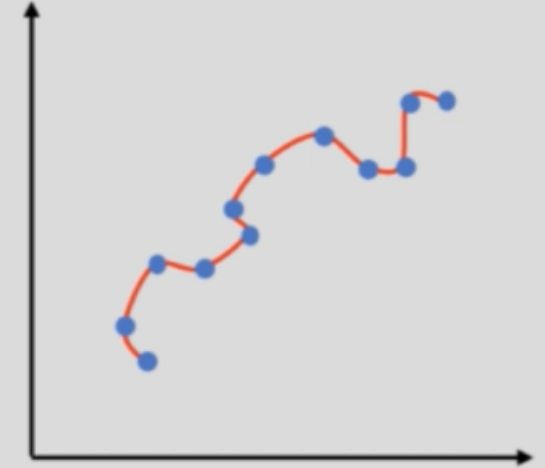
Bias-Variance



Underfitting
Yüksek Yanlılık

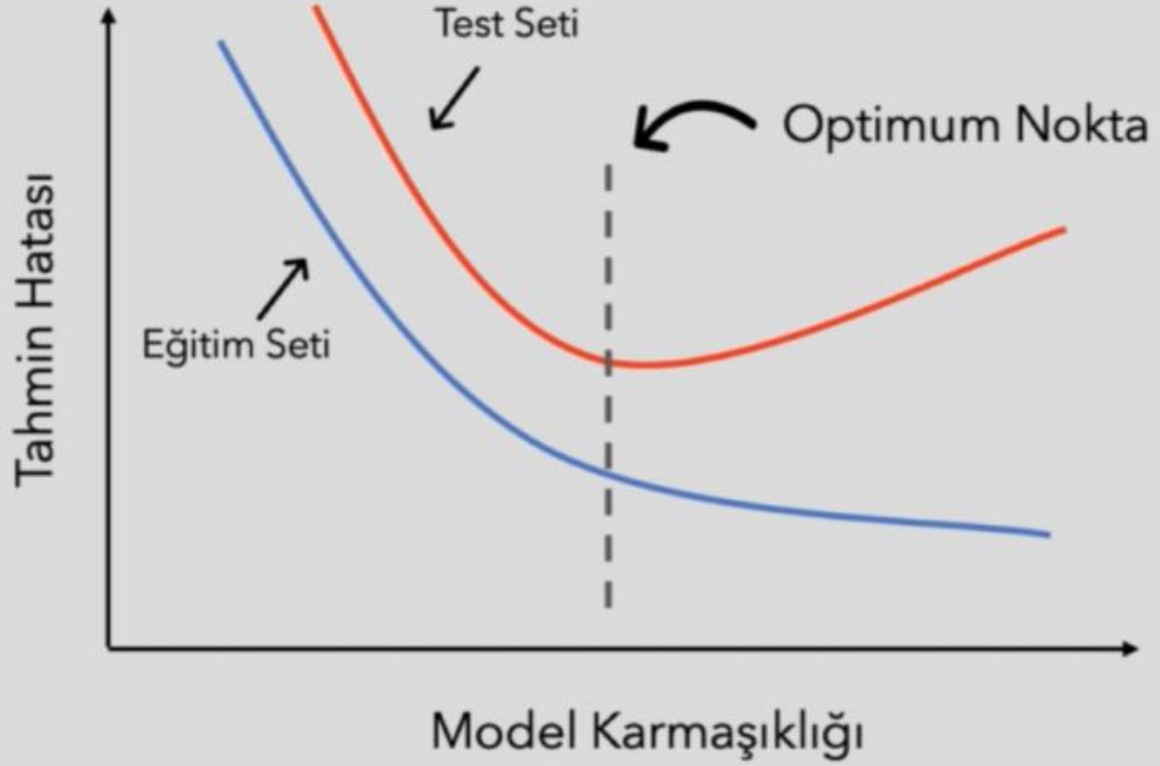


Doğru Model
Düşük Yanlılık, Düşük Varyans



Overfitting
Yüksek Varyans

Eğitim Hatası vs Test Hatası



Parametre, Hiperparametre, Parametre Tuning, Model Tuning

Model Tahmin Performansını Arttırmak