









BATCH : B150 Data Science
LESSON : Machine Learning
DATE : 05.07.2023
SUBJECT : Supervised Learning

 techproeducation
 techproeducation
 techproeducation
 techproeducation
 techproedu

 techproeducation.com  info@techproeducation.com  +1 (917) 768-7466



MACHINE LEARNING - 2



Makine Öğrenmesi – 2



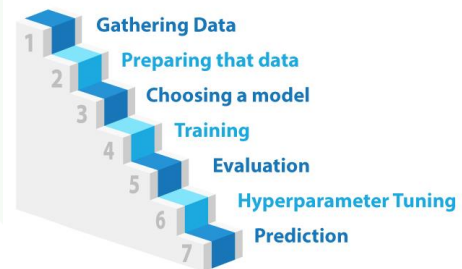
Overall Table of Contents



General Content

- ✓ Supervised Learning Algorithm
- ✓ Supervised Algorithm practices Python application
- ✓ Projects Solutions

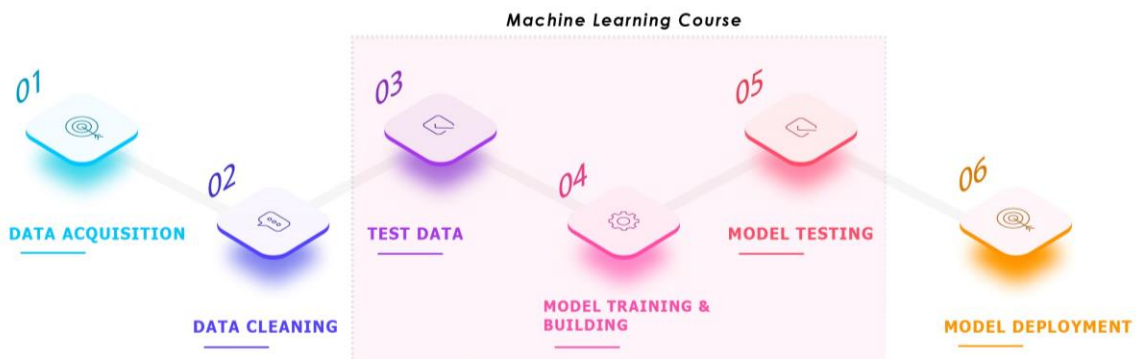
7 steps of Machine Learning

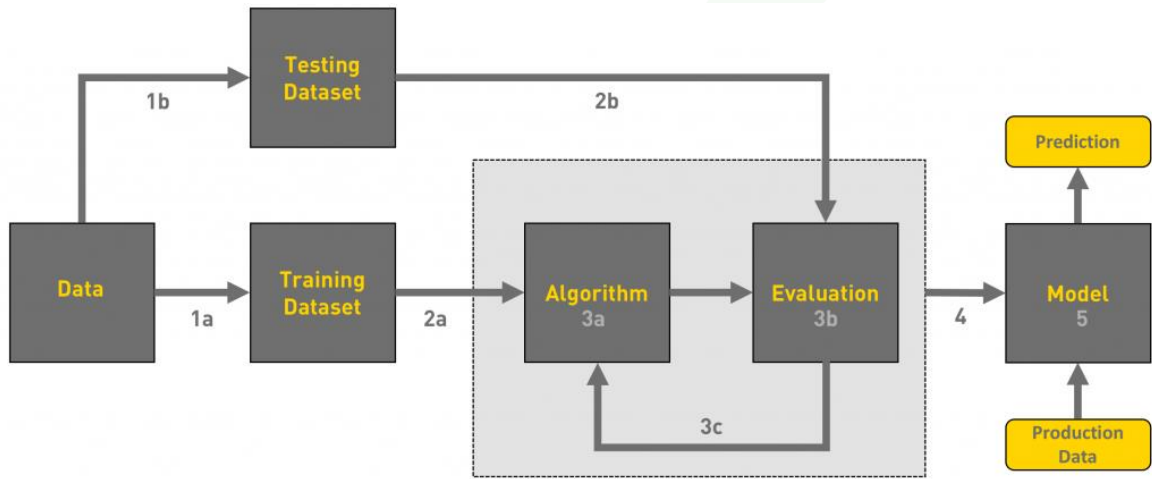


Where are We? Run Navigation !!

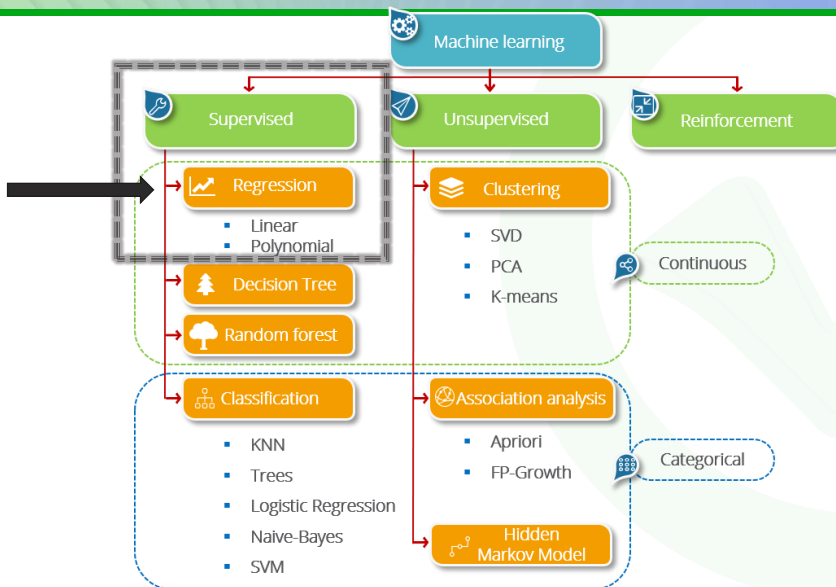
Where are we?

DATA SCIENCE





More Focus





SUPERVISED LEARNING



Supervised Learning

Supervised Learning Processes



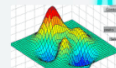
Prepare Data

Data preparation is crucial for any data analysis. If your data is messy, there's no way you can make sense of it.



Choose an algorithm

Choosing the most appropriate algorithm for specific problem is the most crucial task here



Fit a Model

Model fitting is a procedure that takes three steps- Function, Error Function and Parameter to minimise the difference.



Choose a validation method

Depending on the nature of data, choosing a validation set can be the most important step.



TUUM – Test Update Use Model

Examine fit and update until satisfied. Use fitted model for prediction



Output

Every output that the model provides, along with the new data that facilitated the output, becomes the new input-output combination that is fed as training data into the model for learning.

www.AILabPage.com

AILabPage



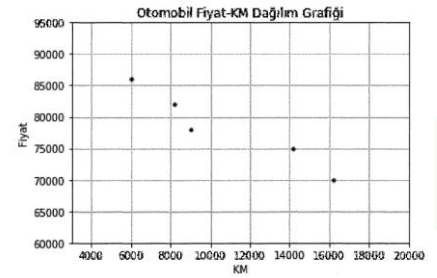
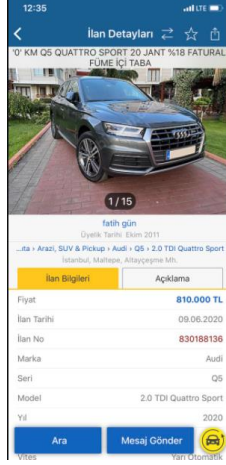
Supervised Learning



Regresyon Kavramını

- Scott'ı tanıyalım, problemini öğrenelim
- 2.el araba alım-satım için değerlendirmeler

Sıra	Fiyat (₺)	KM
1	86000	6000
2	82000	8200
3	78000	9000
4	75000	14200
5	70000	16200

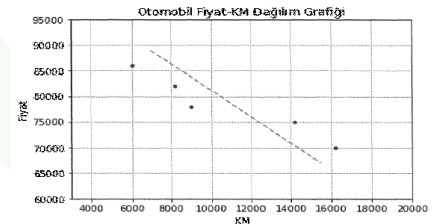
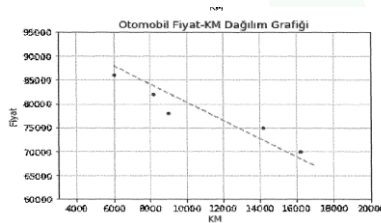
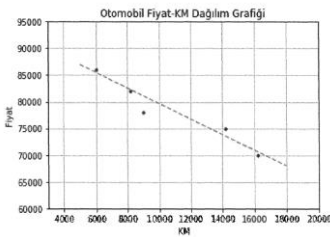


Supervised Learning



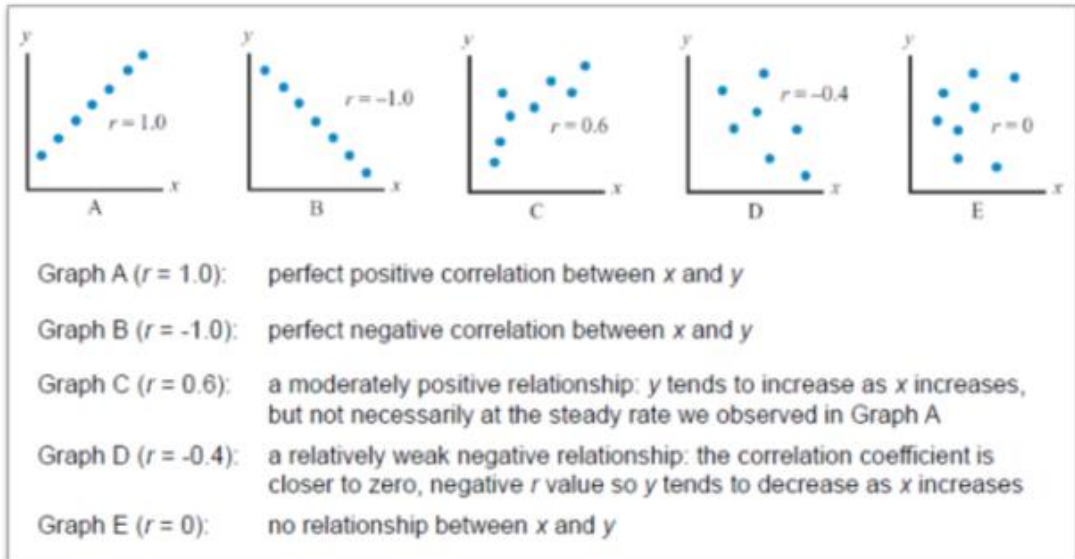
Regresyon Kavramı

- Grafikleri değerlendirelim





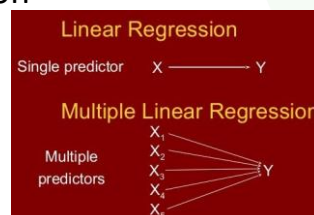
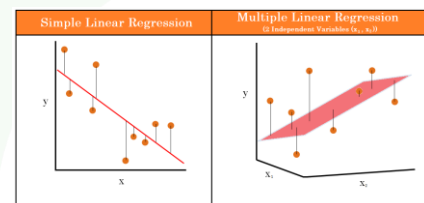
Supervised Learning



Supervised Learning – Linear Regression

Lineer Regresyon

- Dependent variable – Target (Bağımlı değişken – Hedef değişken)
- Independent variable – Features (Bağımsız değişken)
- Simple linear regression
- Multiple Linear regression



Independent variable

Intercept (bias)

$$y = xw + b$$

Depended variable

Slope

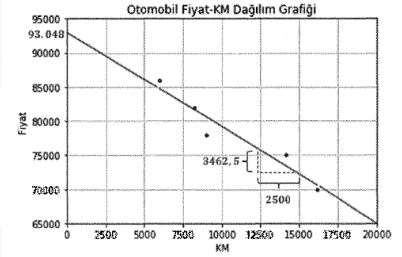
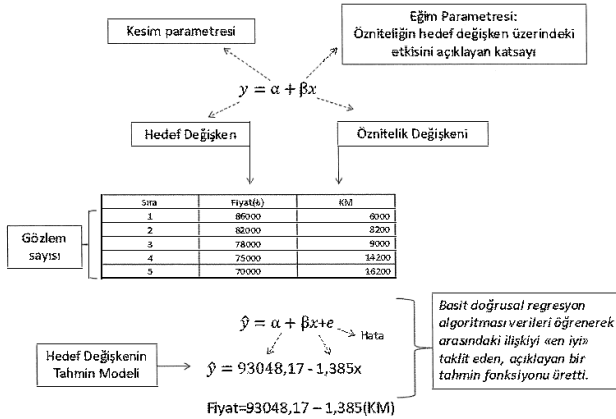


Supervised Learning



Simple Linear Regression

Formül --> $Y = a + bx$
Katsayıların tahmini



$$\beta = \frac{-3462,5}{2500} = -1,385$$

Yorumu !?



Simple Linear Regression



Scott'ın almak istediği 9000 km'lik aracın fiyat tahminini de yine bu modelden yapabilir:

$$\text{Fiyat} = 93.048 - 1,385 (9000) = 80.583$$

Yine bu modele göre yeni ilanda yer alan 12.000 km'lik aracın fiyat tahminini yapabiliriz:

$$\text{Fiyat} = 93.048 - 1,385 (12000) = 76.478$$

Aynı şekilde Scott'ın 10.000 km'deki aracı için bir satış fiyatı belirleyebiliriz:

$$\text{Fiyat} = 93.048 - 1,385 (10000) = 79.198$$

- Peki, bu algoritma makine Öğrenmesi ile alternatifler arasından en doğru olanı nasıl belirledi?

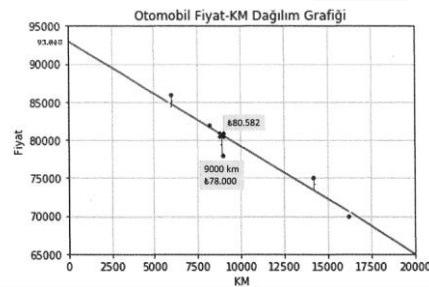
Sıra	Fiyat (₺)	KM	Tahmin Edilen Fiyat (₺)
1	86000	6000	84737
2	82000	8200	81690
3	78000	9000	80582
4	75000	14200	73379
5	70000	16200	70609



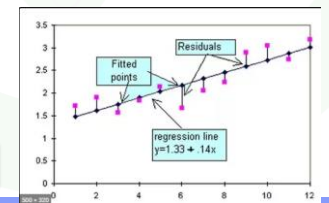
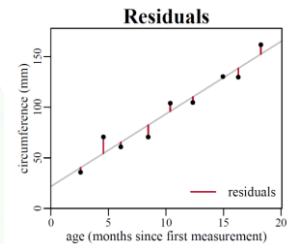
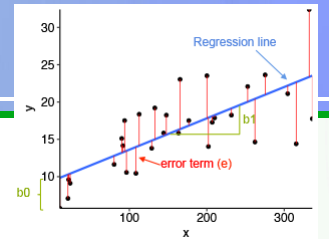
Supervised Learning

Simple Linear Regression

- ✓ Mantığı anlayalım
- ✓ Residuals
- ✓ Cost function



Neden tam olarak değeri tahmin edemedik ??

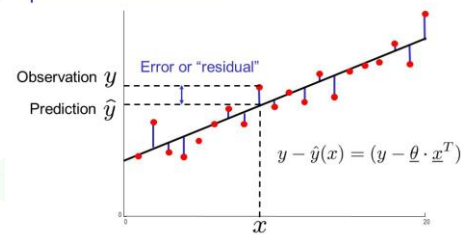


Supervised Learning - RESIDUALS

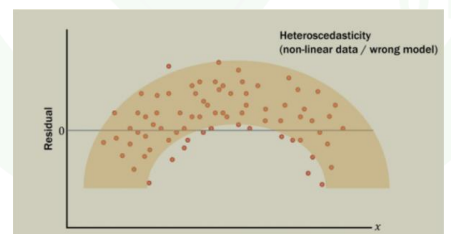
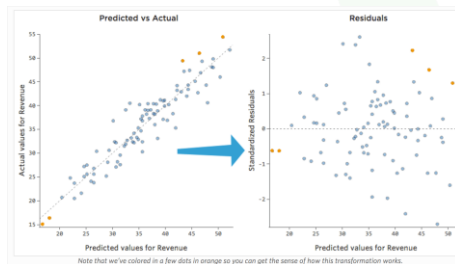
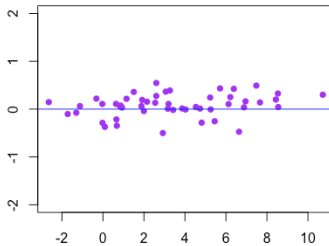
- Residuals are sometimes called “errors.”
- It means that there is some unexplained difference – an error that isn't explained by the regression line.

$$\text{Residual} = y_{\text{test}} - y_{\text{pred}}$$

Measuring error



Good Residual Plot

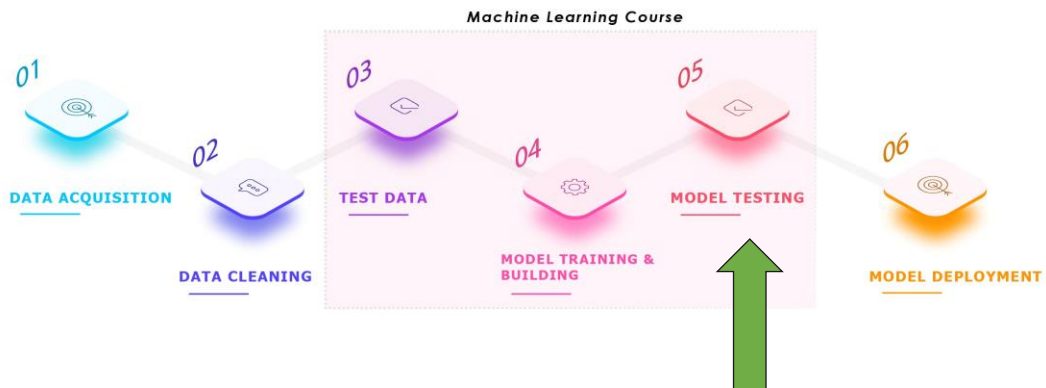




Where are We ? Run Navigation !!

Where are we?

DATA SCIENCE



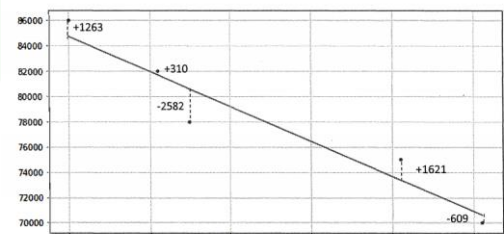
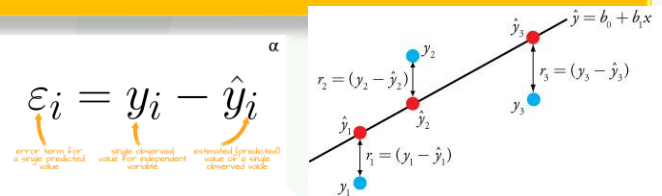
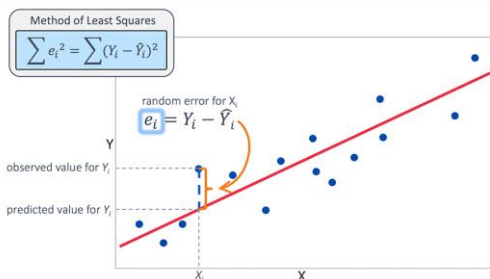
Supervised Learning

Error Metrics

Regresyon Modelleri İçin
Performans Değerlendirme
Ölçütleri (Error Metrics)

- Residual - Error

Madem bilgi kaybından kaynaklanan bir maliyet söz konusu, bunu nasıl hesaplarız? Başka bir ifadeyle modelin iyi olup olmadığını nasıl ölçeriz?





Supervised Learning

Regresyon Modelleri İçin Performans Değerlendirme Ölçütleri

✓ Mean Absolute Error (MAE) Ortalama Mutlak Hata

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

n : Gözlem Sayısı
 y_j : Her bir gerçek değer
 \hat{y}_j : Her bir tahmin edilmiş değer

$$MAE = \frac{1}{n} \sum |y - \hat{y}|$$

Divide by the total number of data points
 Predicted output value
 Actual output value
 Sum of
 The absolute value of the residual

✓ Mean Squared Error (MSE) Ortalama Kare Hata

$$MSE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$$

n : Gözlem Sayısı
 y_j : Her bir gerçek değer
 \hat{y}_j : Her bir tahmin edilmiş değer

Mean Squared Error

- Average (squared) deviation from truth $\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$
- Very sensitive to outliers
 - 99 exact, 1 off by \$10
 - all 100 wrong by \$1
- Sensitive to mean / scale
 - $\mu_y = 1/2, \Sigma y_i \dots$ good baseline
- Relative squared error (Weka)
 - $\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n y_j^2}$
 - MSE of predictor
 - MSE when using the mean as a predictor



Supervised Learning

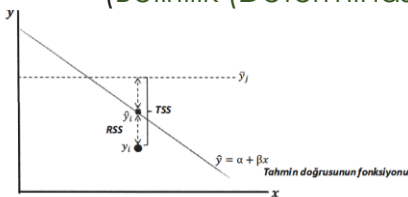
Regresyon Modelleri İçin Performans Değerlendirme Ölçütleri

✓ Root Mean Square Error (RMSE) Kök Ortalama Kare Hata

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

✓ Coefficient of Determination (R Squared-R2) – Adjusted R2 (Belirlilik (Determinasyon) Katsayısı – Düzeltilmiş Belirlilik Katsayısı)



$$R^2 = 1 - \frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y})^2}$$

Kalıntı Kareleri Toplamı (Residual Sum of Squares-RSS)
 Toplam Kareler Toplamı (Total Sum of Squares- TSS)
 Gerçek değerlerinin ortalaması
 Düzeltilmiş $R^2 = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1} \right]$
 k: özdeşlik sayısı



**It is time to coding for
Linear Regression**



Supervised Learning

- ✔ **Linear Regresyon modelleme sonrası Genel Değerlendirme**
 - ✔ **Scott ne alemde şimdi ? NE düşünüyor ?**
 - ✔ **Örneğin sonuç $R^2 = 0,92$ ise ne anlama gelir ?**
 - ✔ **Test seti için sonuçlar nasıl ?**
 - ✔ **Sihirli küreye benzeyen model !**
 - ✔ **Scott yeni alım-satımlar peşinde**
 - ✔ **Model güvenilir mi acaba?**

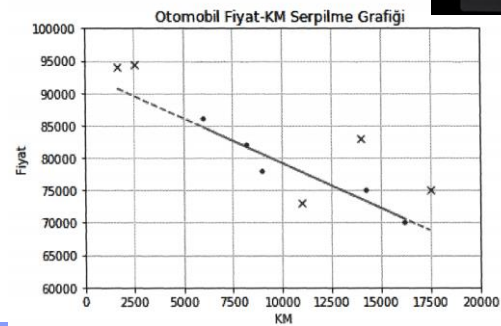
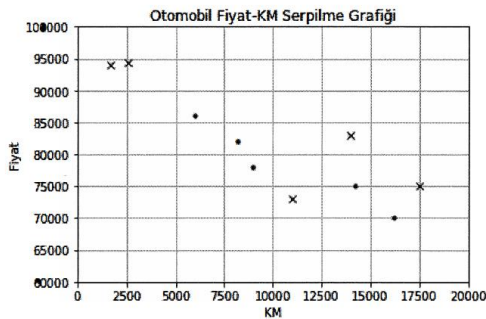




Supervised Learning

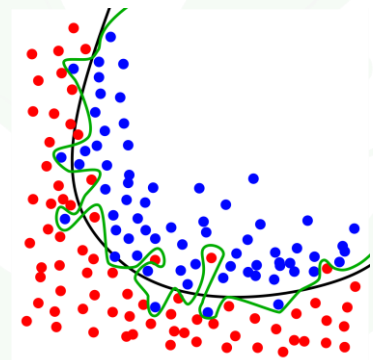
Linear Regresyon Test set modelleme sonrası Genel Değerlendirme

- ✓ Test setinin sonuçlarını inceleyelim
- ✓ Scott'ın sihirli küresi halen çalışıyor mu ?
- ✓ Test seti örneğin $R^2=0,55$ ise ne demek bu ?



Supervised Learning

- ✓ Overfitting ve Underfitting Sorunları
 - ✓ Ezberleme – Eksik Öğrenme (Aşırı uyum – Yetersiz uyum)
- ✓ Overfitting nedenleri
 - ✓ Aşırı veri ile öğrenme (Bu matematik ders notlarını iyi ezberleyip hocasının sınavda aynı soruları sorması üzerine iyi not alan ancak yeni problemleri çözemeyen öğrencinin durumu gibidir)
 - ✓ Aşırı bilgi yüklemesi (Bu durum, un ve şeker gibi malzemelerle ilgili her şeyi bilen ama helva yapamayan kişinin durumuna benzetilebilir)



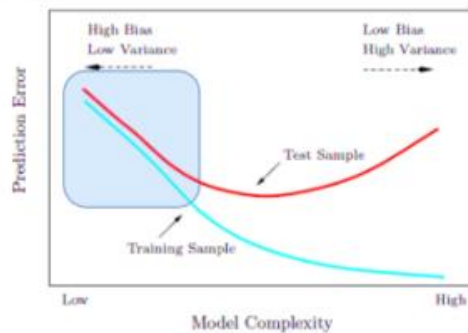


Supervised Learning

- Overfitting ve Underfitting Sorunları
- Ezberleme – Tembellik

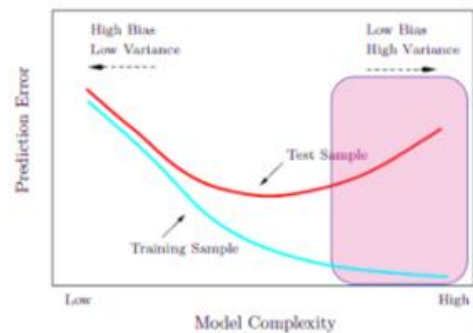
UNDERFITTING

High training error and **high** test error



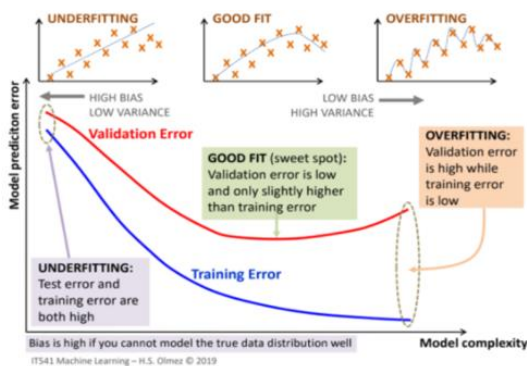
OVERFITTING

Low training error and **high** test error



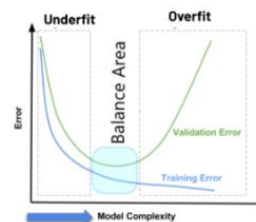
Supervised Learning

- Overfitting ve Underfitting Sorunları
- Ezberleme – Tembellik



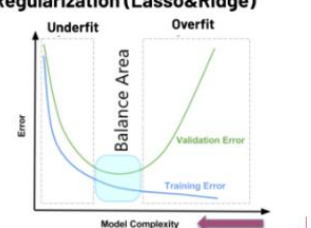
UNDERFITTING ?

- Find a **more complex** model



OVERFITTING ?

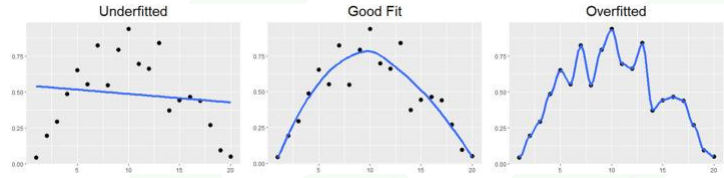
- Decrease the number of parameters
- More training data / **Cross Validation**
- Regularization (Lasso&Ridge)**





Supervised Learning

- ✓ Underfitting nedenleri
 - ✓ Veri yetersizliği
 - ✓ Eksik Bilgi



«Eksik bilgi» kavramını anlamak için bir kitap alıntısı:

«.....Bunun için tek bir siyah kuğu yeterlidir»

Avustralya kıtası keşfedilmesinde önce kuğularla ilgili yazılan tüm araştırma kitapları ile eğitim gören bir zoolog için bir kuğuyu tanımlayan öznelilikleri "kanat ölçüleri", "uzunluğu", "kilosu" ve "beyaz" olmasıdır. Zoologun bu modeli ile Avustralya'nın siyah kuğularını öngörememesi gayet normaldir

Nassim Nicholas Taleb

SİYAH KUĞU



Olasızsız görünenin etkisi



7. KASIM



Supervised Learning

Gerçek dünya olaylarını hep lineer ilişki ile modelleyebilir miyiz ? Bazı olaylar eğri ile daha iyi ifade edilebilir miydi?

Simple Polynomial Regression

- Quadratic – 2nd order

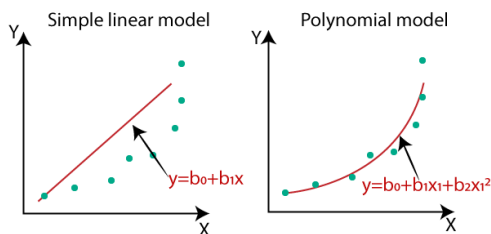
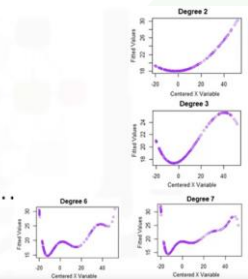
$$\hat{Y} = b_0 + b_1 x_1 + b_2 (x_1)^2$$

- Cubic – 3rd order

$$\hat{Y} = b_0 + b_1 x_1 + b_2 (x_1)^2 + b_3 (x_1)^3$$

- Higher order

$$\hat{Y} = b_0 + b_1 x_1 + b_2 (x_1)^2 + b_3 (x_1)^3 + \dots$$



Simple Linear Regression

$$y = b_0 + b_1 x_1$$

Multiple Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Polynomial Linear Regression

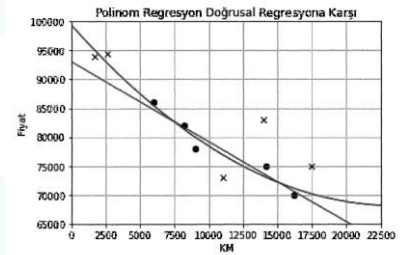
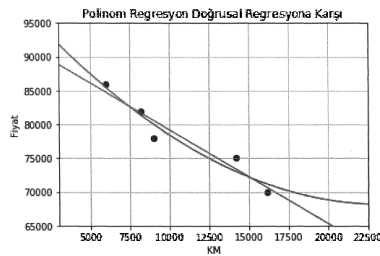
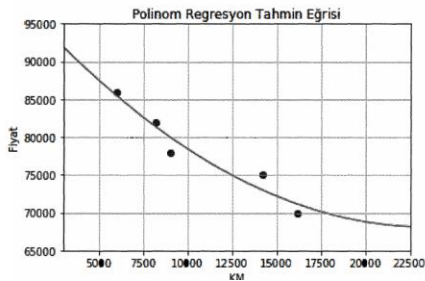
$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \dots + b_n x_1^n$$



Supervised Learning

Simple Polynomial Regression

- 2.El Alım-satım örneğine uygulaması
- Sonuçlar nasıl? daha iyi mi ?

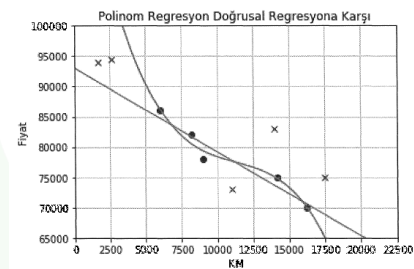


Supervised Learning

Simple Polynomial Regression

- Polinom derecesini artırırsak ne olur ?
- Çoklu regresyona geçiş

```
88 # üzerindeki başarısına bakalım:
89 polinom_derecesi = PolynomialFeatures(degree = 3)
90 X_train_polinom = polinom_derecesi.fit_transform(X_train)
91 X_test_polinom = polinom_derecesi.transform(X_test)
```

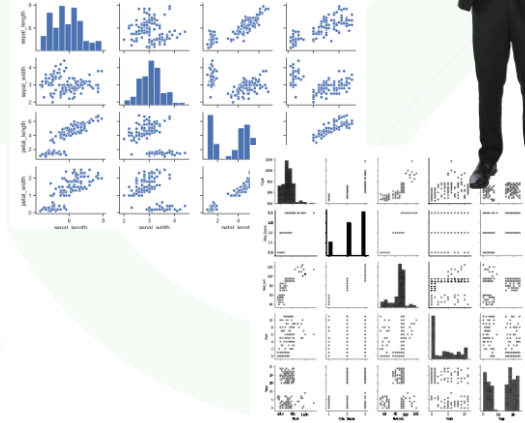
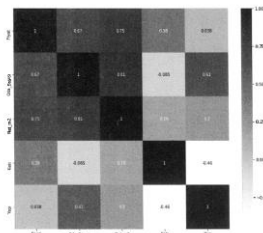
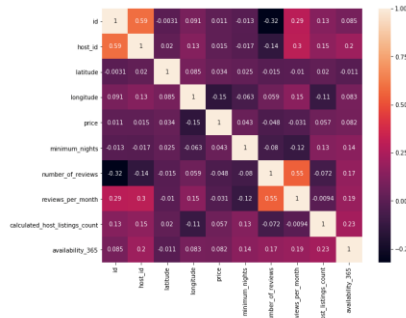




Supervised Learning

Multiple Linear Regression

- ✓ Emlak Piyasasına bir dalsak nasıl olur ?
- ✓ Heatmap ..
- ✓ Pairplot..



Supervised Learning

Multiple Linear Regression

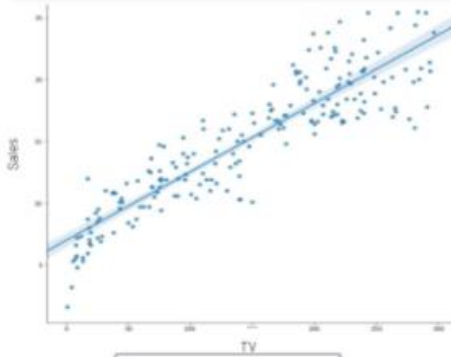
- ✓ Model Sonuçları ve yorumlanması
 - ✓ Parametre tahmininden yola çıkarak eğitim veri seti üzerinden üretilen tahmin modeli aşağıdaki şekildedir:

$$y = 305,35 + 118,35 \text{ (Oda Sayısı)} + 1,133 \text{ (Net m2)} + 7,096 \text{ (Katı)} - 4,665 \text{ (Yaşı)}$$
 - ✓ R2 için hangi değerler iyi sayılabilir ?
- ✓ Modele tahmin yaptırılım
 - ✓ Tahmin sonucu yorum



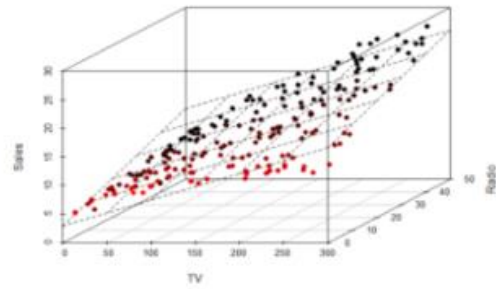
Supervised Learning

Simple Linear Regression



$$y = b_0 + b_1 \cdot x_1$$

Multiple Linear Regression



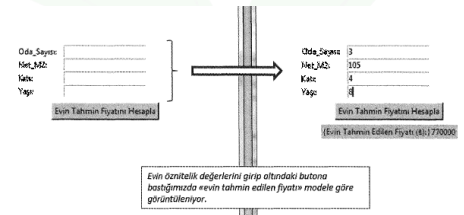
Dependent variable (DV) Independent variables (IVs)

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n$$



ARAŞTIRMA KONUSU

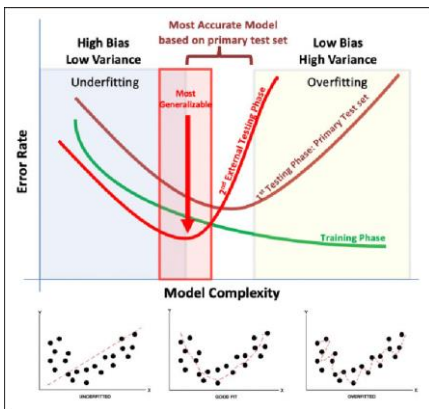
Python Tkinter kütüphanesi ile Interface (arayüz) oluşturma





Supervised Learning

BIAS – VARIANCE TRADE OFF (Yanlılık – Varyans ilişkisi) ve Model Complexity



En iyi model ??

Eğitim verisini %90 öngörüp test verisini %60 öngören bir model mi daha iyidir, yoksa eğitim verisini %60 öngörüp test verisini %59 öngören bir model mi daha iyidir?

*Bias : Yanlılık (Verileri dengelemek)
Variance: Esneklik*

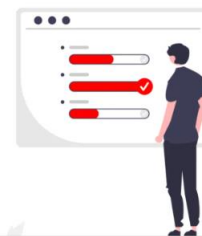
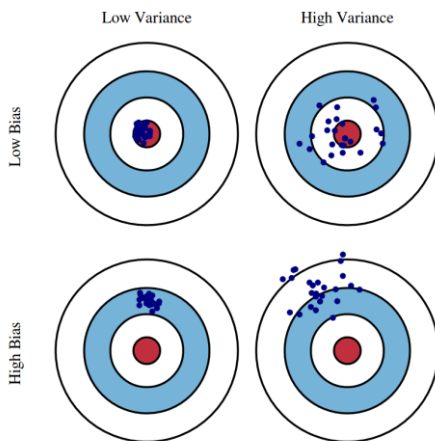
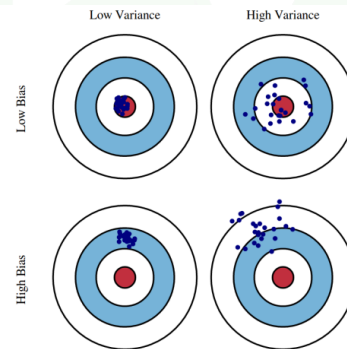
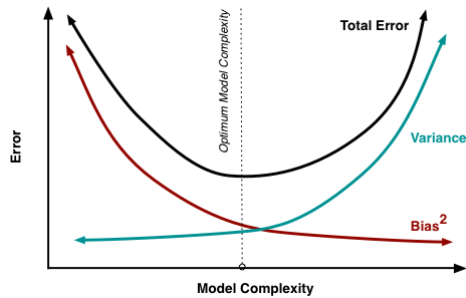


Supervised Learning

BIAS – VARIANCE TRADE OFF (Yanlılık – Varyans ilişkisi) ve Model Complexity

✓ Borsadan bir örnek

✓ Hangisine yatırım (90-60 ? | 60-59 ?)



BIAS VS VARIANCE TRADEOFF

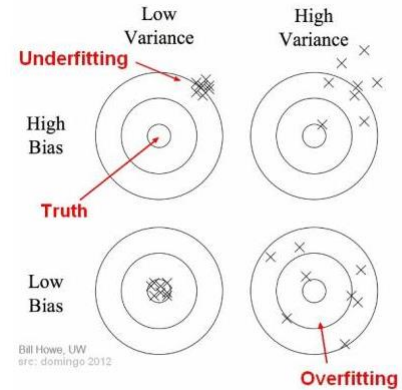
What is Bias vs Variance Tradeoff and Why is it important?



Supervised Learning

BIAS – VARIANCE TRADE OFF (Yanlılık – Varyans ilişkisi) ve Model Complexity

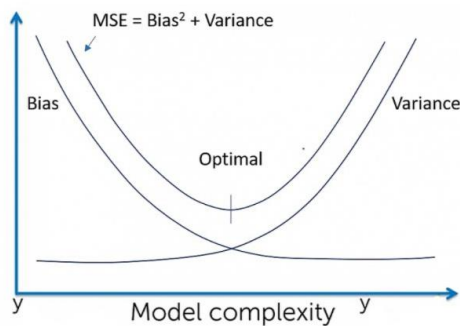
✓ Emlak örneği üzerinden anlayalım



Supervised Learning

BIAS – VARIANCE TRADE OFF (Yanlılık – Varyans ilişkisi) ve Model Complexity

✓ $MSE = \text{Varyans} + \text{bias}(\text{yanlılık})^2$



- Bias (Yanlılık) Tanımı
- Variance tanımı

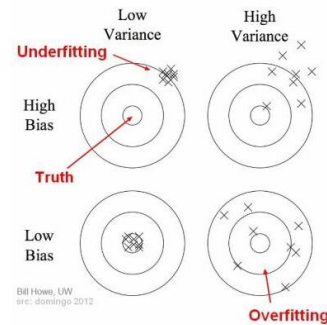
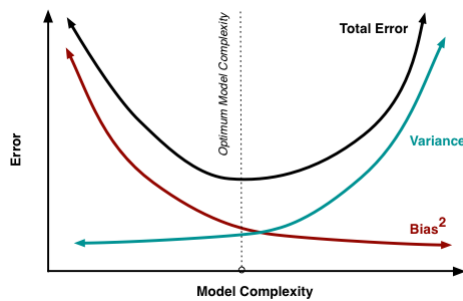
Bias, train set üzerinden üretilen modelin eğitim verisindeki hedef değişkenini ne kadar hatayla öngörebildiği ile ilişkilidir. Varyans ise eğitim verisi üzerinden üretilen modelin test verisindeki hedef değişkenini ne kadar hatayla öngörebildiği ile ilişkilidir. Bu ikisi arasında modelin karmaşıklığı üzerinden bir çeşit alışveriş ya da değiş-tokuş (**trade off**) ilişkisi vardır



Supervised Learning

BIAS – VARIANCE TRADE OFF (Yanlılık – Varyans ilişkisi) ve Model Complexity

- ✓ Somut olarak Bias ve Variance nedir ?
- ✓ Optimumu nedir o zaman?



Supervised Learning

Overfitting var.. Azaltmak istiyoruz ve En baştan neler yapabiliriz ?

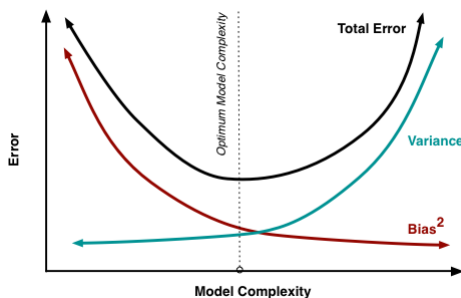
1. Collecting more data
2. ilk madde mümkün değilse data augmentation (veri artırımı)
3. Regularization
4. Decrease the parameters
5. Cross validation



Supervised Learning

REGULARIZATION

- ✓ Multicollinearity problemi
- ✓ High variance ve Overfitting problemi



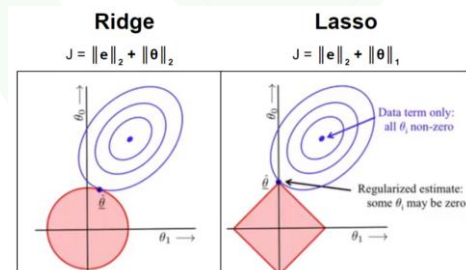
Grafikte, modelin complexity karmaşıklığı arttıkça bias (yanlılık) ve varyans birbirine zıt yönde hareket etmektedir. Optimum nokta ise gerektiği kadar feature (öznitelik) ile target (hedef) değişkenini öngören, veri setinin tamamında (train ve test verisi) hata oranının en aza düştüğü model için varyansın ve biasın (yanlılığın) el sıkışıp anlaştığı yerdir



Supervised Learning

REGULARIZATION

- ✓ Optimum nokta nasıl belirlenecek o zaman ?
 - ✓ Multicollinearity ye bak
 - ✓ Feature ları ağırlıklandır
- ✓ Regularization tanımı (regülatör hatırlayalım)
- ✓ Regularization tipleri
 - ✓ L1 (Lasso)
 - ✓ L2 (Ridge)





Supervised Learning

REGULARIZATION

✓ Neden regularization

LASSO Regression

- ✓ Penalty (Cezalandırma)
- ✓ Shrinkage (Sıkıştırma)

$$RSS = \sum_{j=1}^n (y_j - \hat{y}_j)^2 \rightarrow RSS_{CV} = \sum_{j=1}^n (y_j - \hat{y}_j)^2 + \text{Ceza}$$

Doğrusal Regresyon Modelinin Maliyet Fonksiyonu

Doğrusal Regresyon Modelinin Çekidüzen Verilmiş (ÇV) Maliyet Fonksiyonu

$$RSS_{LASSO} = \sum_{j=1}^n (y_j - \hat{y}_j)^2 + L1 \sum_{j=1}^p |\beta_j|$$

LASSO L1 uygulaması ile çekidüzen verilmiş maliyet fonksiyonu

p : öznitelik sayısı

Maliyet Fonksiyonu

Ceza Terimİ



Types of Regularization in ML

01
Ridge
Regression

02
Lasso
Regression



Supervised Learning

$$RSS_{LASSO} = \sum_{j=1}^n (y_j - \hat{y}_j)^2 + L1 \sum_{j=1}^p |\beta_j|$$

✓ RIDGE Regression

$$RSS_{RIDGE} = \sum_{j=1}^n (y_j - \hat{y}_j)^2 + L2 \sum_{j=1}^p (\beta_j)^2$$

Scikit-learn kütüphanesindeki L1 hiperparametresi «alpha» olarak isimlendirilmiştir. Sembolü (alfa) şeklindedir. Kesim parametresi intercept ile karıştırmayın

$$C = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^F |\beta_j|$$

L1 normalleştirme

$$C = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^F \beta_j^2$$

L2 normalleştirme

✓ Elastic-Net Regression



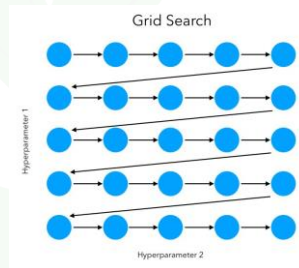
Supervised Learning

Hyperparameter leri tek tek mi deneyeceğiz ?

✓ GRID SEARCH İLE Hyperparameter Optimization

Bir algortimanın en iyi sonucu verecek hiperparametrelerini deneme yanılma yolu yerine, tek seferde bize verecek sihirli değneğin adı **Grid Search**'dür.

- ✓ Grid Search uygulaması (notebook ile)
 - ✓ Evaluation metrics seçilmelidir (R2 gibi)
 - ✓ Lasso ve Ridge için en iyi modeli oluşturacak Hyperparameters nedir?



Supervised Learning

Hayat zikzaklarla dolu bir yol ve doğrusal regresyon o zikzakları öngörmekten aciz.. bu çerçevede uyguladığımız Lasso, Ridge ve ElastikNet algoritmaları doğrusal regresyonun "ayakları daha yere basan" versiyonlarıydı.

Multi Polynomial Regression

Eğer birden fazla feature değişkeni varsa feature'lerin uygun polinom derecesini nasıl tespit ederiz?

İki öznelilik değişkeni olan çoklu doğrusal regresyon modeli

$$y = \alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

İki öznelilik değişkeni olan 2. dereceden çoklu polinom regresyon modeli

$$y = \alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 \times X_2 + \varepsilon$$

Özellikler Özelliklerin polinom dereceleri Etkileşim Terimi

POLİNOM DÖNÜŞÜMÜ:
Doğrusal regresyon modeline özelliklerin polinom derecelerini ve etkileşim terimlerini dahil ederek polinom dönüştürme yapıyoruz.

3. dereceden polinom regresyon oluştursaydık feature'lerin küplerini de modele dahil edecektik.

Üç feature olsaydı bunun da karesini ve diğer iki feature ile etkileşim terimlerini modele ekleyecektik.

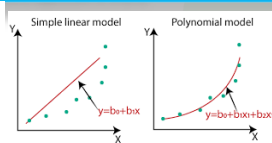
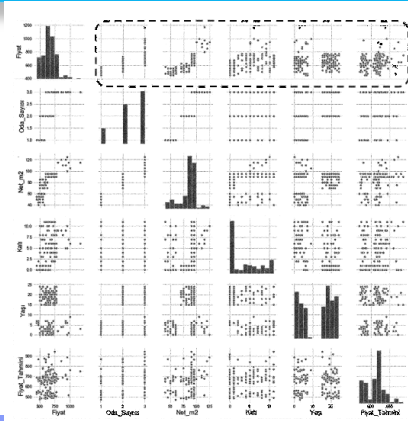
uygun polinom derecesini belirlemek için grid search yapıldığında hata skoru ölçütü için RMSE yaygın kullanıma sahiptir.



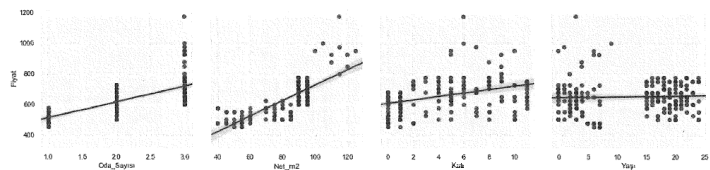
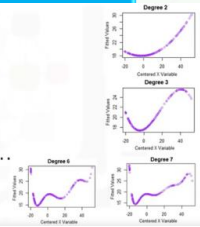
Supervised Learning

Uygun Polinom Derecesinin Tespiti

- ✓ Her feature için polinom derecesi aynı mı?
- ✓ Pairplot a bir bakalım..



- Quadratic – 2nd order
 $\hat{Y} = b_0 + b_1 x_1 + b_2 (x_1)^2$
- Cubic – 3rd order
 $\hat{Y} = b_0 + b_1 x_1 + b_2 (x_1)^2 + b_3 (x_1)^3$
- Higher order
 $\hat{Y} = b_0 + b_1 x_1 + b_2 (x_1)^2 + b_3 (x_1)^3 + \dots$

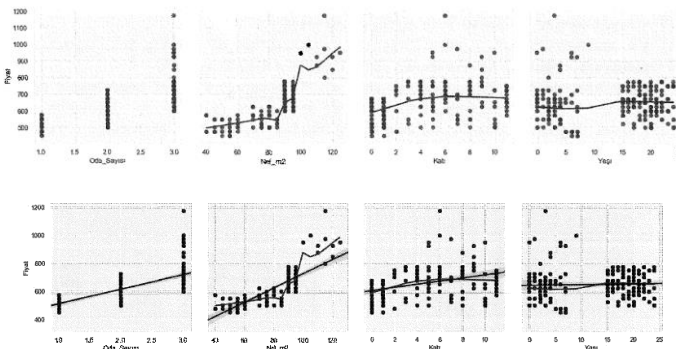


Supervised Learning

Uygun Polinom Derecesinin Tespiti

Lowess Eğrisi

```
5 g = sns.PairGrid(df, y_vars=["Fiyat"], x_vars=["Oda_Sayisi", "Net_m2", "Katı", "Yaşı"],
7 height=4)
6 g.map(sns.regplot, lowess=True, color=".3")
8
```



ÖNEMLİ NOKTALAR :

- 1) Tüm feature'lerin hedef değişkenle ilişkisi doğrusal olmayabilir .
- 2) feature'lerin hedef değişkenle ilişkisi doğrusal olmasa bile hepsinin doğrusal olmama (polinom) derecesi farklı olabilir.
- 3) polinom regresyon yaparken her bir feature in hedef değişkenle ilişkisinin, doğrusal olup olmadığını, değilse derecesini dikkate almak gerekir.

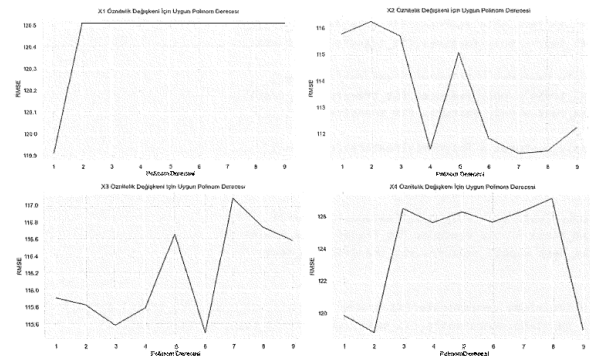


Supervised Learning

Uygun Polinom Derecesinin Tespiti

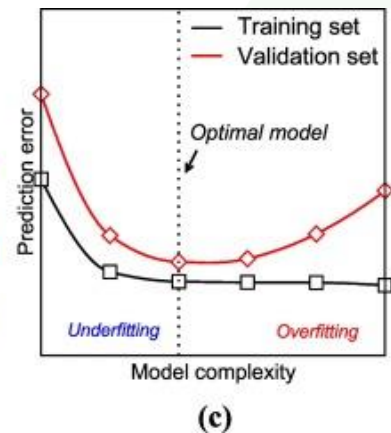
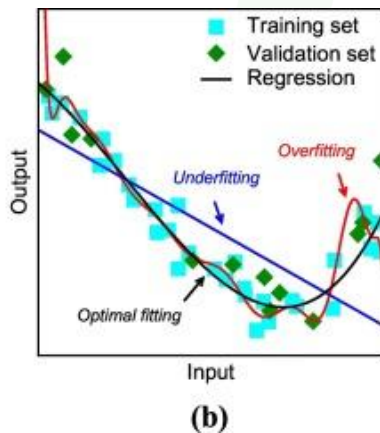
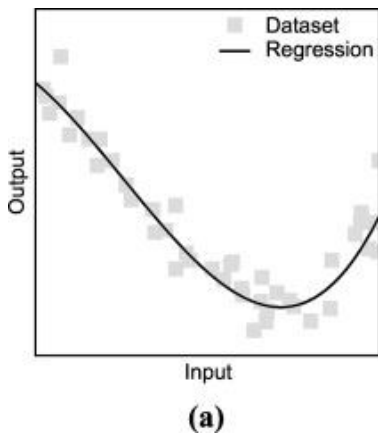
- ✓ RMSE ile polinom derecesi tespiti
- ✓ Polinom regresyonların avantaj-dezavantajları

```
##### Kod Bloğu #####
rmse = []
degrees = np.arange(1, 10)
min_rmse, min_deg = 1e30, 0
for deg in degrees:
    poly_features = PolynomialFeatures(degree=deg, include_bias=False)
    X_poly_train = poly_features.fit_transform(X1_train, X2_train, ... için değiştir!
    poly_reg = LinearRegression()
    poly_reg.fit(X_poly_train, y_train)
    X_poly_test = poly_features.fit_transform(X1_test, X2_test, ... için değiştir!
    poly_predict = poly_reg.predict(X_poly_test)
    poly_rmse = mean_squared_error(y_test, poly_predict)
    poly_rmse = np.sqrt(poly_rmse)
    rmse.append(poly_rmse)
    if min_rmse > poly_rmse:
        min_rmse = poly_rmse
        min_deg = deg
# Grafik: Tüm polinom dereceleri için RMSE skorları:
plt.figure(figsize=(10,8))
plt.title('Denetlik Değişkeni İçin Uygun Polinom Derecesi')
sns.lineplot(x=degrees, y=rmse)
plt.xlabel('Polinom Derecesi')
plt.ylabel('RMSE')
#####
```



Supervised Learning

Uygun Polinom Derecesinin Tespiti



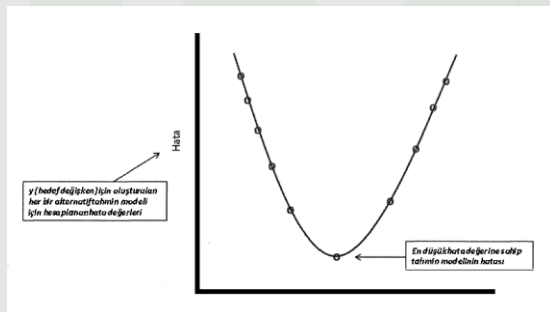


Supervised Learning

GRADIENT DESCENT Algoritmaları İle Model Hatası Minimizasyonu (Dereceli Azalma – Meyilli iniş)

✓ $y = F(x) = a + Bx + \text{error}$ (en iyi a ve B bulmak)

$$\text{Maliyet} = \text{MSE} = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$$



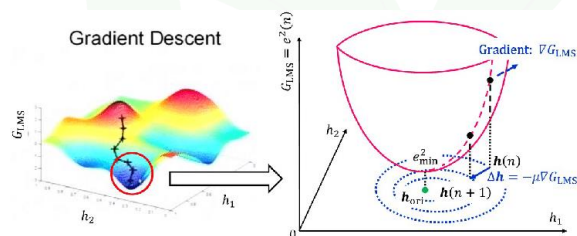
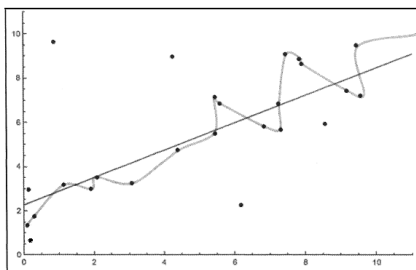
✓ Bu model, gerçekten en düşük hataya sahip olan model midir?



Supervised Learning

GRADIENT DESCENT (Model Hatası Minimizasyonu)

✓ Durum tespiti

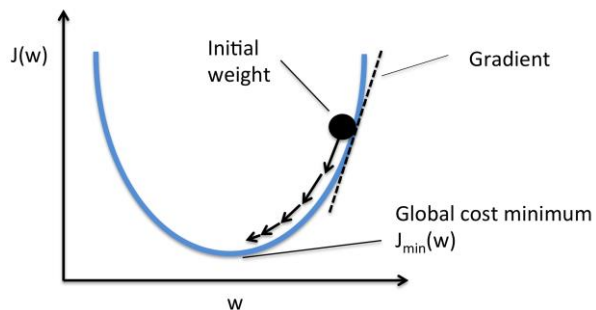
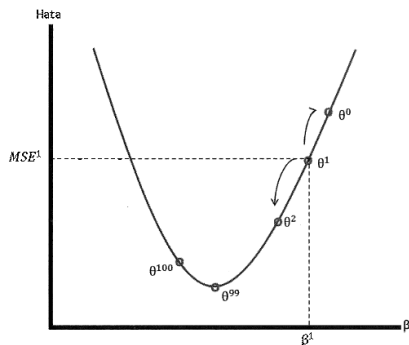




Supervised Learning

GRADIENT DESCENT (Model Hatası Minimizasyonu)

- ✓ B ve hata parabolü
- ✓ Vadi örneği

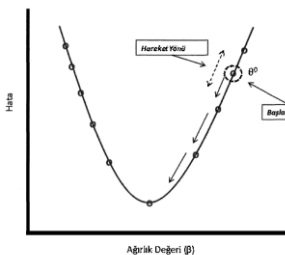
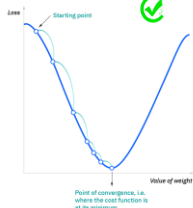


Supervised Learning

GRADIENT DESCENT (Model Hatası Minimizasyonu)

✓ Temel adımlar

- ✓ başlangıç noktasının seçimidir
- ✓ başlangıç noktasındaki eğimini (gradyanını) hesaplama
- ✓ Learning rate seçimi

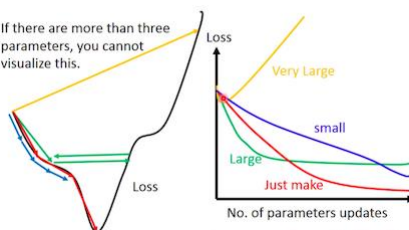


Learning Rate

$$\theta^i = \theta^{i-1} - \eta \nabla C(\theta^{i-1})$$

Set the learning rate η carefully

If there are more than three parameters, you cannot visualize this.





Supervised Learning

GRADIENT DESCENT (Model Hatası Minimizasyonu)

✓ Çeşitleri

- ✓ **Batch Gradient Descent-BGD** (tüm eğitim seti için teker teker hataları hesaplayarak optimuma ulaşmaya çalışır),
- ✓ **Stochastic Gradient Descent-SGD** (Adım boyları belirlenerek tüm eğitim veri seti için değil sadece belirlenen adımlara karşılık gelen veriler için hataları hesaplar.) ,
- ✓ **Minibatch Gradient Descent-MGD** (Batch ve Stochastic Gradient Descent optimizasyonunun bileşimidir. Eğitim veri setini küçük parçalara ayırır ve her biri için ayrı ayrı hataları hesaplar) .

PARAMETERS	BATCH GD ALGORITHM	MINI BATCH ALGORITHM	STOCHASTIC GD ALGORITHM
ACCURACY	HIGH	MODERATE	LOW
TIME CONSUMING	MORE	MODERATE	LESS



Supervised Learning

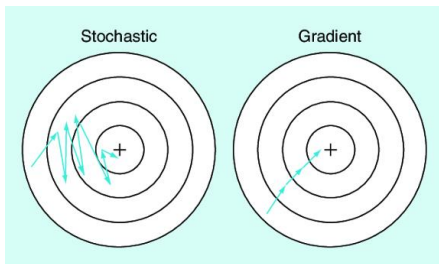
GRADIENT DESCENT (Model Hatası Minimizasyonu)

✓ Stochastic Gradient Descent-SGD

✓ Starting Point

✓ Learning rate

Dikkat !! 



- Gradient Descent algoritması doğrusal regresyon gibi model oluşturma algoritmaları için optimizasyon yapmaya yarıyor. Yoksa başlı başına bir model oluşturma algoritması değildir.
- Doğrusal regresyon algoritması ile üretilen tahmin modeli gerçekten de en optimal model olabilir. Ancak veri setinin durumuna bağlı olarak bundan bazen emin olamayabiliriz. Bu gibi durumlarda Grad.Desc ile optimizasyon elde ettiğimiz modeli teyit etmemize yardımcı olur

<https://medium.com/deep-learning-turkiye/gradient-descent-nedir-3ec6afcb9900>



Supervised Learning

İlave Bilgi olarak Standardizasyon Konusu

✓ Bazı algoritmalar feature'lerin ölçeğine duyarlıdır

Standartlaştırma:

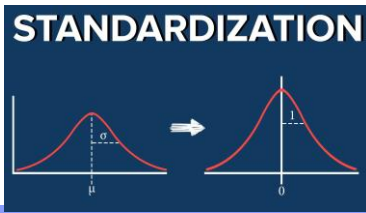
Bir değişkenin içindeki her bir gözlemin değerini kendi aritmetik ortalamasından farkını alıp standart sapmasına bölersek o gözlemi standartlaştırmış oluruz. Mesela y'nin (hedef değişkeni) gözlemlerini aşağıdaki şekilde standartlaştırırız:

y_j (y değişkenindeki j gözlemi) için standartlaştırılmış değer.

$$s_j = \frac{y_j - \bar{y}}{\hat{\sigma}_y}$$

y değişkeninin tüm değerlerinin aritmetik ortalaması.

y değişkeninin standart sapması



Standardization

- To achieve a data which mean = 0 and deviation = 1
- Z-score normalization

$$X_{\text{changed}} = \frac{X - \mu}{\sigma}$$

Normalization

- Shrink all data values between 0-1

$$X_{\text{changed}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Minimum-Maksimum Yöntemiyle Standartlaştırma:

Bir değişkenin içindeki her bir gözlemin değerini o değişken içindeki en düşük değere sahip gözlemden çıkarıp bulunan değeri de değişkenin en yüksek ve en düşük değerleri arasındaki farka bölerek o gözlemi standartlaştırmış oluruz.

y_j (y değişkenindeki j gözlemi) için minimum-maksimum yöntemiyle standartlaştırmış değer.

$$mms_j = \frac{y_j - y_{\min}}{y_{\max} - y_{\min}}$$

y değişkeni içindeki minimum değer

y değişkeni içindeki maksimum değer



Supervised Learning

FEATURE SELECTION

- ✓ Çok sayıda faktörü/değişkeni feature olarak kullanmak gerçekten gerekli mi?
- ✓ Ortaya çıkan sorunlar
- ✓ Feature seçiminin sonuçları

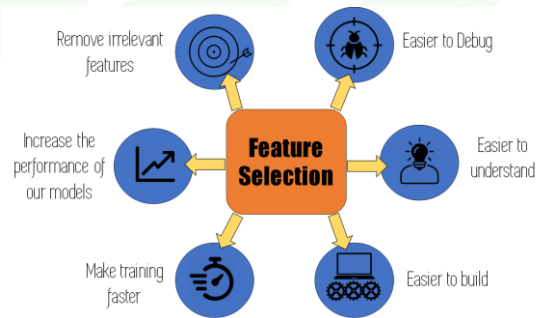
All Features



Feature Selection



Final Features

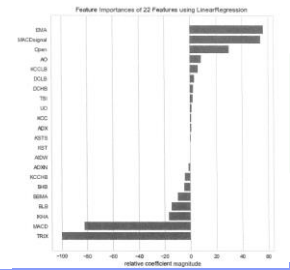
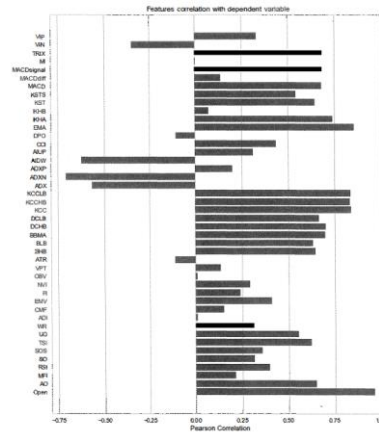
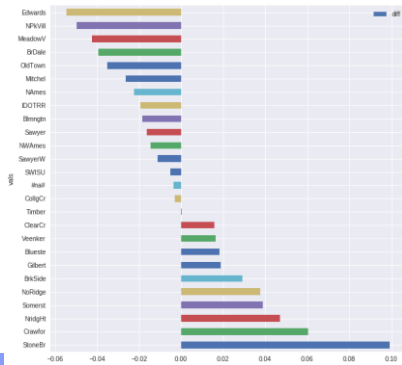




Supervised Learning

FEATURE SELECTION

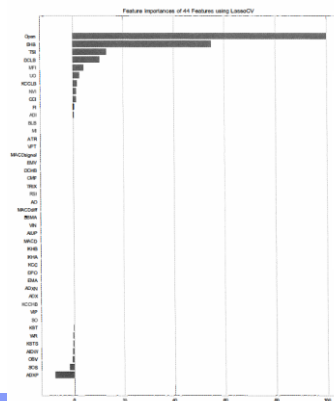
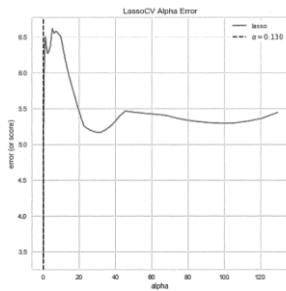
- ✓ Çok sayıda faktörü/değişkeni feature olarak kullanmak gerçekten gerekli mi?
- ✓ Ortaya çıkan sorunlar
- ✓ Feature seçiminin sonuçları



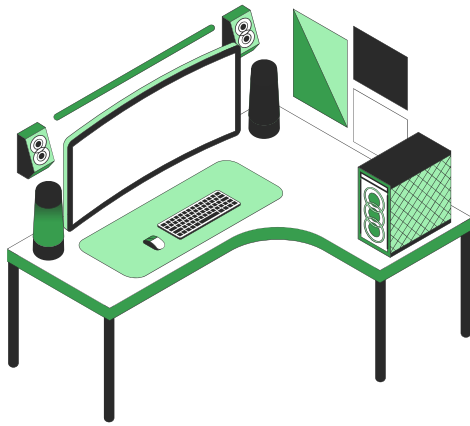
Supervised Learning

FEATURE SELECTION

- ✓ Lasso Cross Validation Algoritmasıyla ile Feature Seçimi
- ✓ LassoCV kullanımı



Öznitelik Seçim Tekniği	Seçilen Öznitelik Sayısı	Eğitim Seti Performansı	Test Seti Performansı
Korelasyon (Pearson)	22	0.93	0.84
Hipotez Testi (p-değeri)	36	0.94	0.71
Lasso	19	0.93	0.86
RFE	8	0.90	0.87



Do you
have any
questions?

Send it to us! We hope you learned
something new.