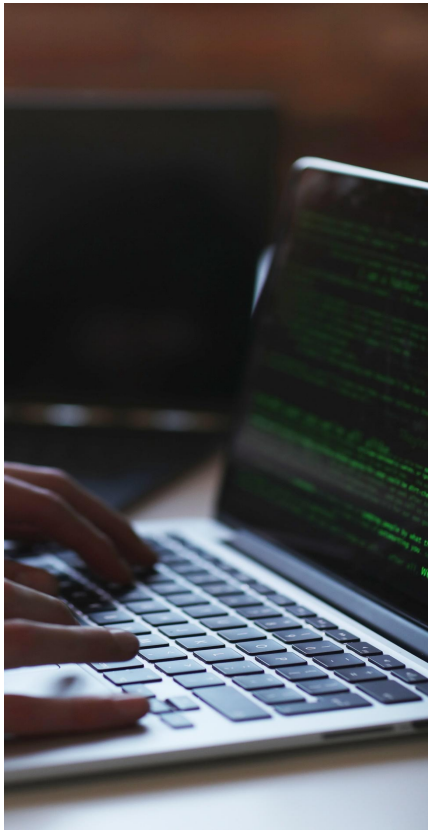
Course-Net
Empowering IT People

#EmpoweringITPeople

Introduction to Regression

Course-Net
Empowering IT People

#EmpoweringITPeople



APA YANG AKAN KITA PELAJARI

Regression and applications

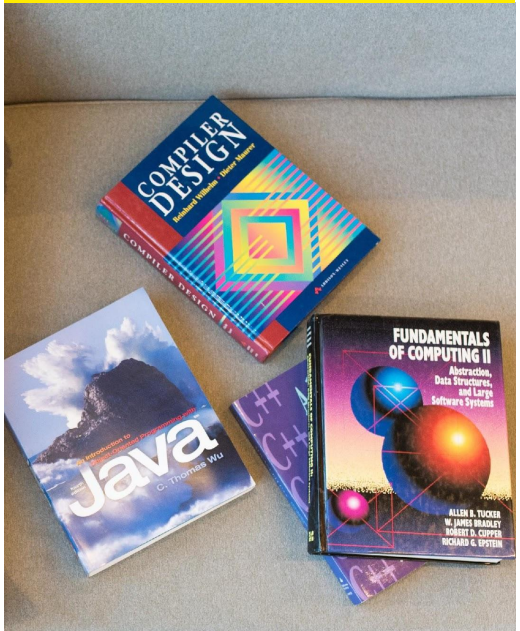
- What is Regression
- Regression Intuition
- Regression Algorithm
- Applications of Regression
- Regression Cases

Course-Net
Empowering IT People

#EmpoweringITPeople

Apa itu Regression Mengapa Kita Pelajari di Data Science ?

Categorical



Type of Data

Numeric

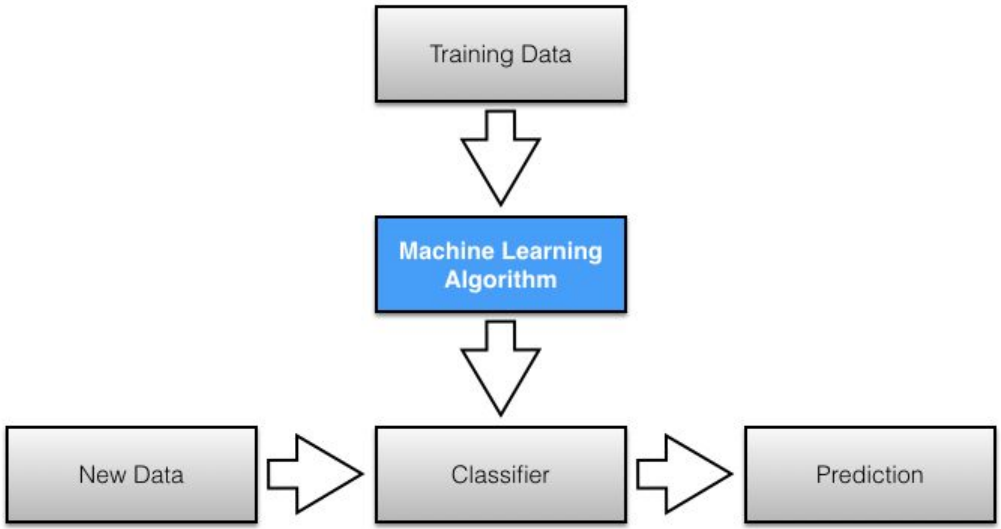
Contoh Data Numeric

House Price Prediction

Stock Price Prediction

Insurance Price Prediction

Supervised Learning Process



Regression Algorithm

1. Linear Regression
2. Multiple Linear Regression
3. Polynomial Regression
4. **Decision Tree Regression**
5. **Random Forest Regression**
6. **XGBoost Regression**

Course-Net
Empowering IT People

#EmpoweringITPeople

Regression Intuition

A Glance on Regression Data Example

Suburb	Address	Rooms	Type	Price
Abbotsford	49 Lithgow St	3	h	1490000
Abbotsford	59A Turner St	3	h	1220000
Abbotsford	119B Yarra St	3	h	1420000
Aberfeldie	68 Vida St	3	h	1515000
Airport West	92 Clydesdale Rd	2	h	670000

X(Independent) :

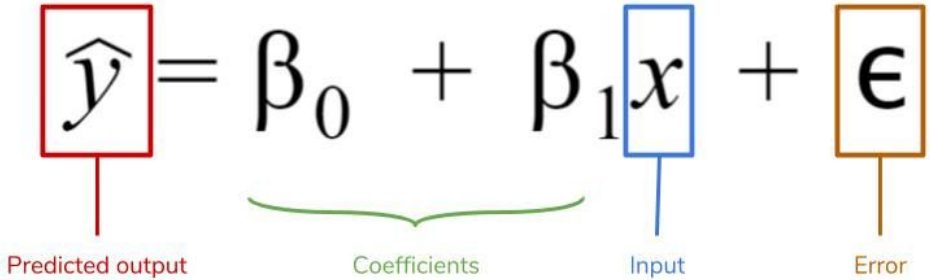
- Suburb
- Address
- Rooms
- Type
- ...

y(Dependent):

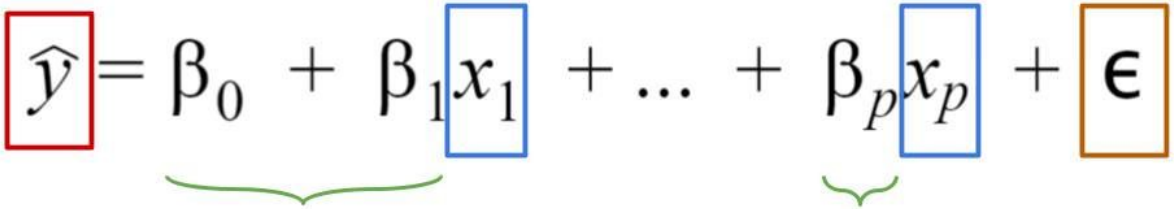
- Price

Regression Formula

Linear Regression: Single Variable

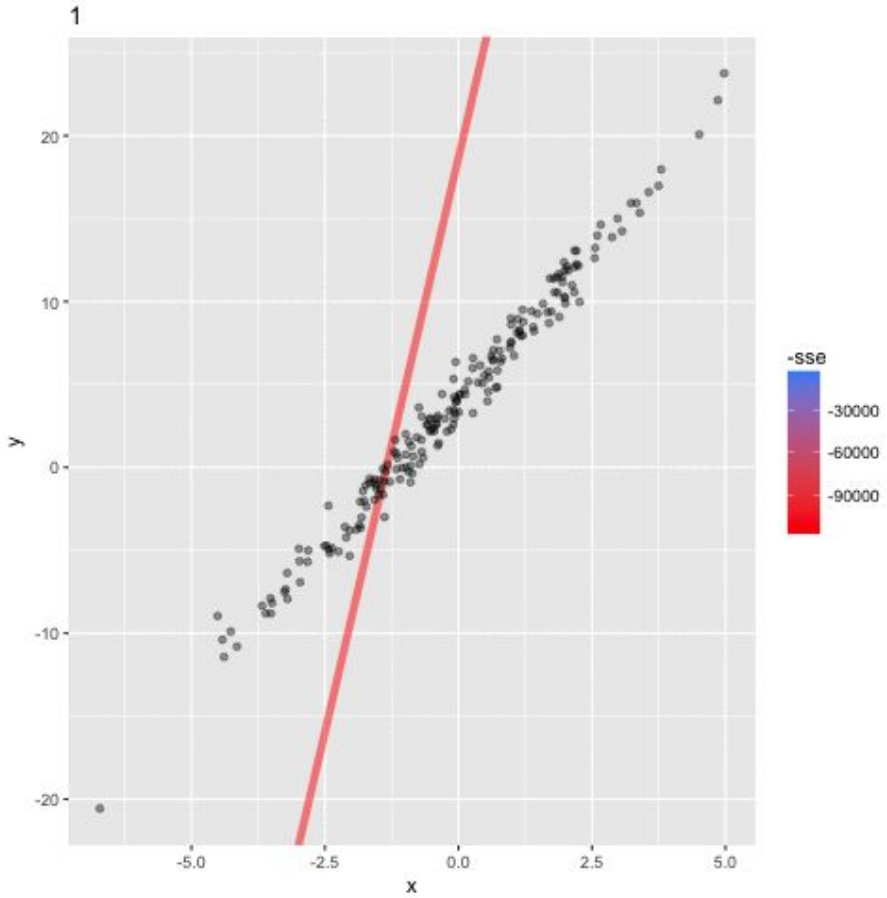


Linear Regression: Multiple Variables



Regression Formula

Regression Model	Equation
Simple linear	$Y = a + bX$
Quadratic	$Y = a + bX + bX^2$
Logarithmic	$Y = a + b \log X$
Exponential	$Y = ae^{bx}$ $e = 2.7183$


1
20
10
0
-10
-20
y
-5.0
-2.5
0.0
2.5
5.0
x
-sse
-30000
-60000
-90000

Cost Function

Cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between X and y

The objective of a ML model, therefore, is to find parameters, weights or a structure that minimises the cost function.

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

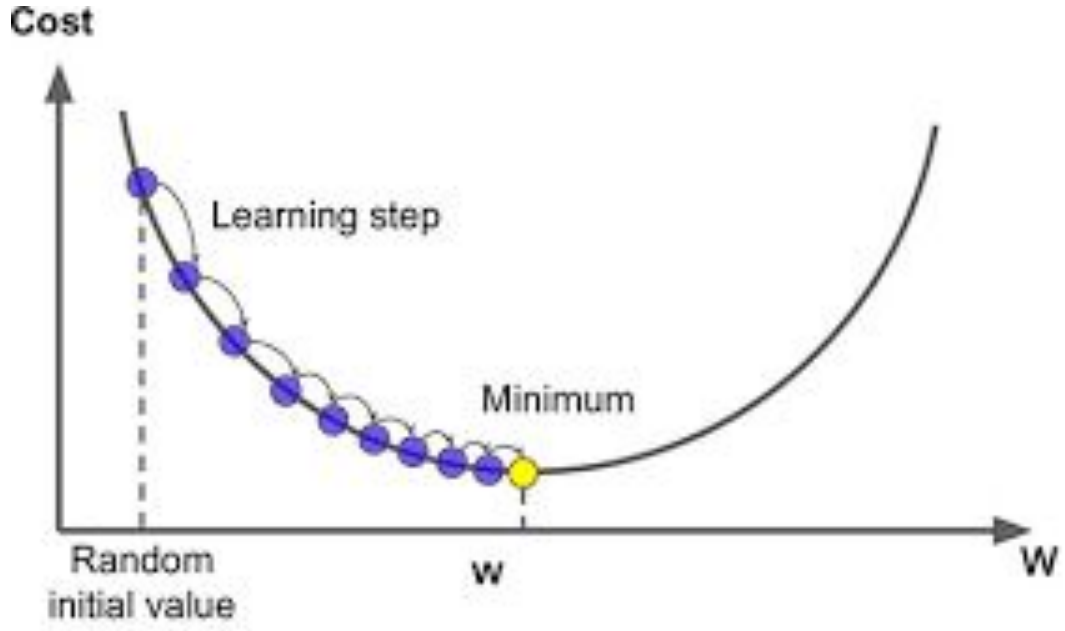
---

n = number of data
yi = actual data (test data)
yi_hat = prediction data

Error = Prediction - Actual

How to Minimize Cost Function?

Gradient Descent



Gradient Descent

repeat until convergence {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \cdot x^{(i)}$$

}

Course-Net
Empowering IT People

#EmpoweringITPeople

Regression Evaluation

Data yang akan di-training.

fitur_1	fitur_2	fitur_3	fitur_4	fitur_5	target
…	…	…	…	…	0.95
…	…	…	…	…	0.32
…	…	…	…	…	0.05
…	…	…	…	…	2.34
…	…	…	…	…	1.65
…	…	…	…	…	1.34

Setelah fase training, diperoleh model yang dapat memprediksi target.

fitur_1	fitur_2	fitur_3	fitur_4	fitur_5	target	prediksi_target
…	…	…	…	…	0.95	0.90
…	…	…	…	…	0.32	0.32
…	…	…	…	…	0.05	0.07
…	…	…	…	…	2.34	1.23
…	…	…	…	…	1.65	1.64
…	…	…	…	…	1.34	1.24

Target dan Hasil prediksi dievaluasi.
MAE,MSE,RMSE (Error Based)
R^2

fitur_1	fitur_2	fitur_3	fitur_4	fitur_5	target	prediksi_target	Evaluasi (selisih)
…	…	…	…	…	0.95	0.90	0.05
…	…	…	…	…	0.32	0.32	0.00
…	…	…	…	…	0.05	0.07	0.02
…	…	…	…	…	2.34	1.23	1.11
…	…	…	…	…	1.65	1.64	0.01
…	…	…	…	…	1.34	1.24	0.01

Regression Evaluation

Metrics for Regression :

1. R-Squared
2. Mean Squared Error (MSE)
3. Root Mean Squared Error (RMSE)
4. Mean Absolute Error (MAE)

R-Squared

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}},$$
$$= 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}.$$

Interval Nilai : -inf s.d. 1

buruk **baik**

Root Mean Squared Error (RMSE)

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

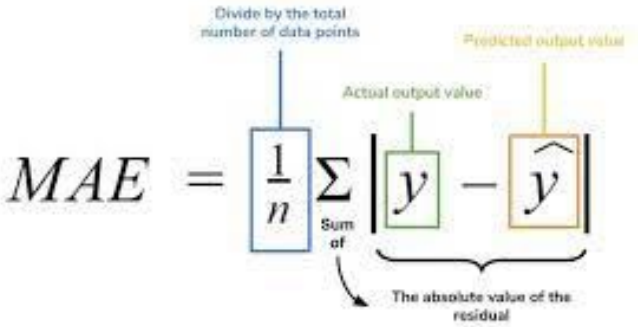
n = number of data
yi = actual data (test data)
yi_hat = prediction data

Error = Prediction - Actual

Interval Nilai : 0 s.d. +inf

baik **buruk**

Mean Absolute Error (MAE)



n = number of data
yi = actual data (test data)
yi_hat = prediction data

Error = Prediction - Actual

Interval Nilai : 0 s.d. +inf

baik **buruk**

Mean Squared Error (MSE)

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

n = number of data
yi = actual data (test data)
yi_hat = prediction data

Error = Prediction - Actual

Interval Nilai : 0 s.d. +inf

baik **buruk**