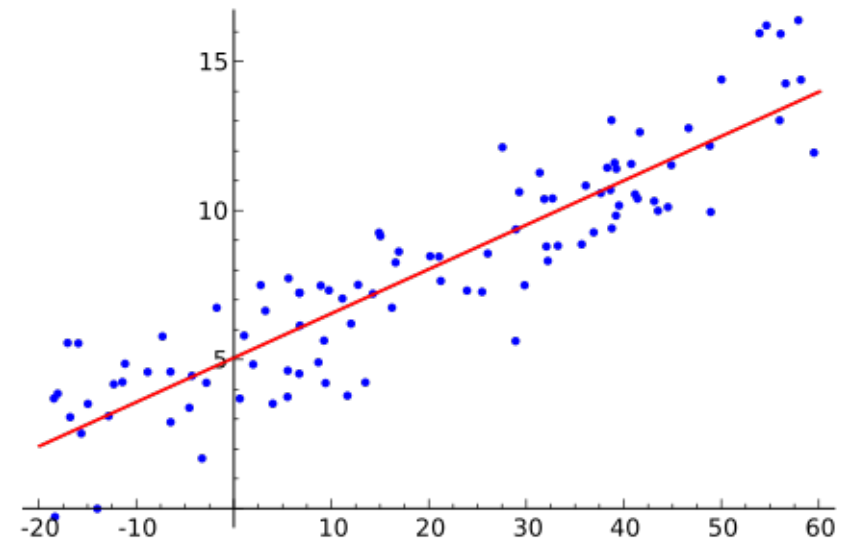# Basic Machine Learning: Linear Regression

Oleh: Muhammad Angga Muttaqien | Founder & Mentor

Goal
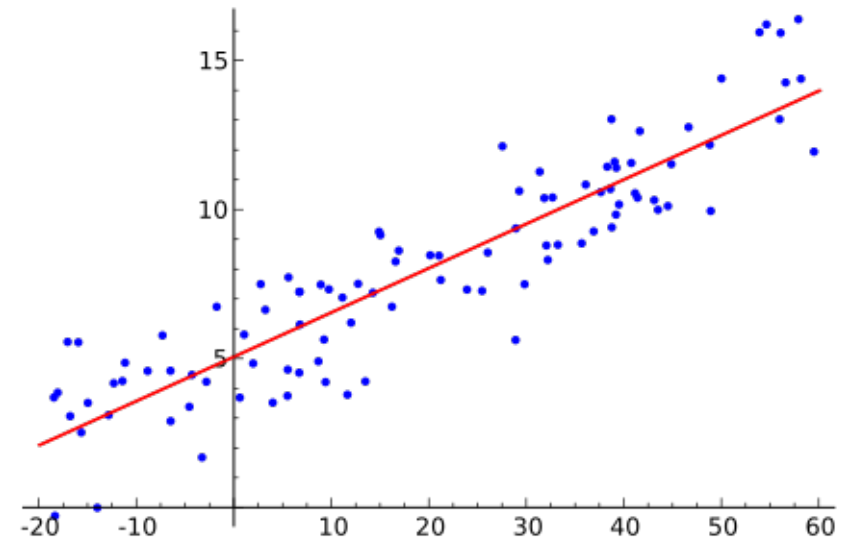
Goal

Understanding one of basic algorithm to easily create a smart system in AI which is linear regression algorithm.
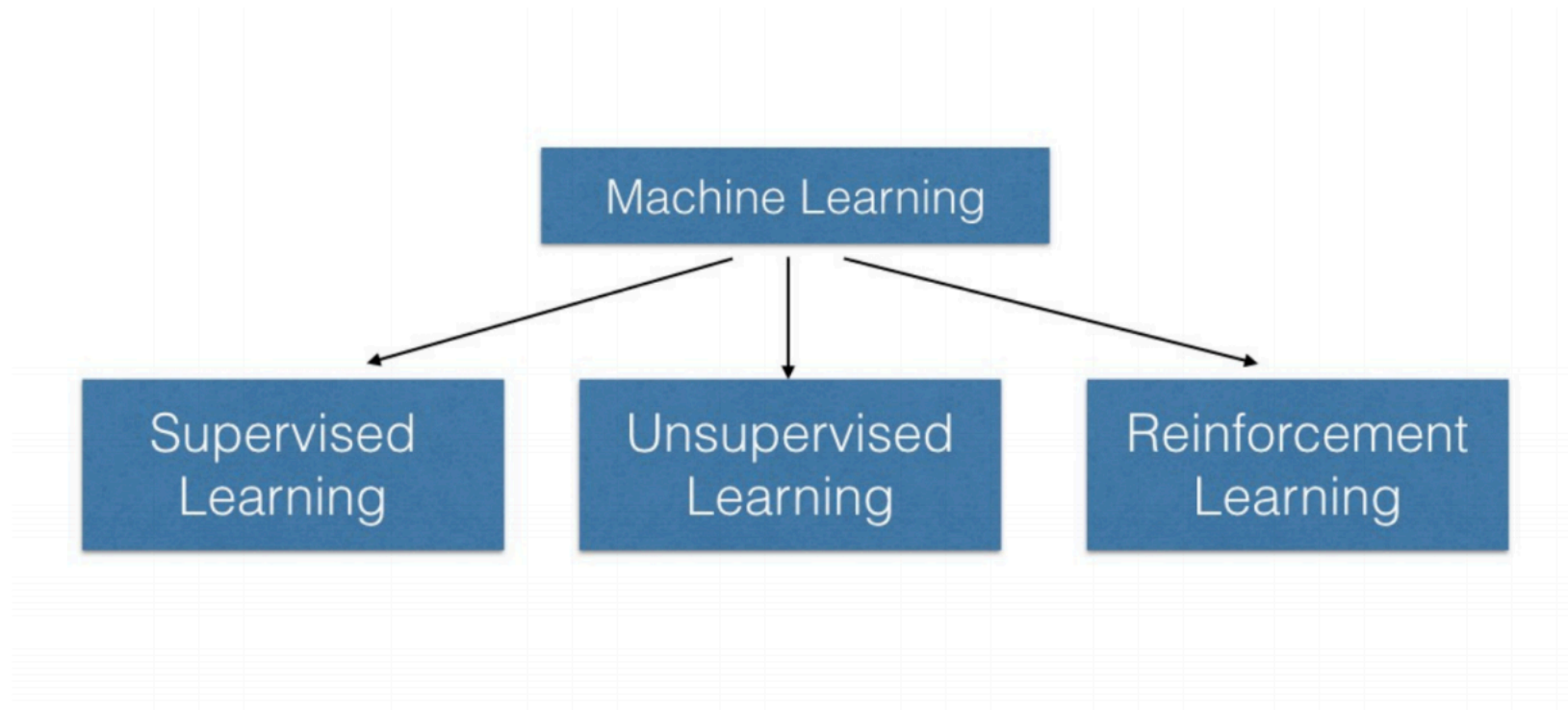
Outline

Outline

- Supervised vs Unsupervised Learning

- Linear Logistic Algorithm
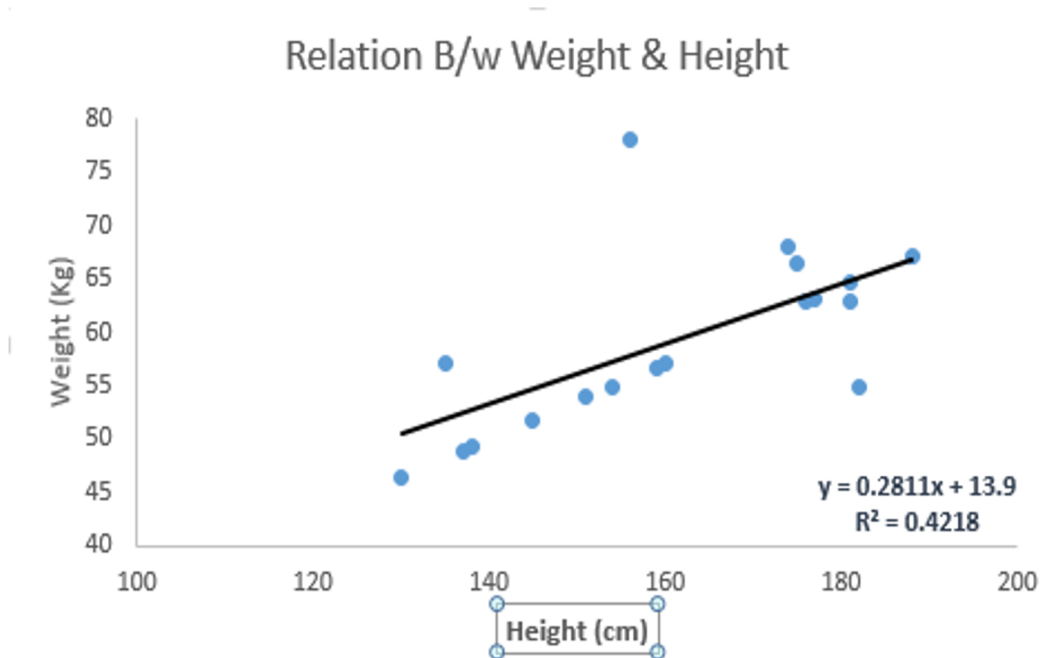 - Concept
  - Scikit Learn
  - Pros and Cons

Content

Supervised vs Unsupervised Learning
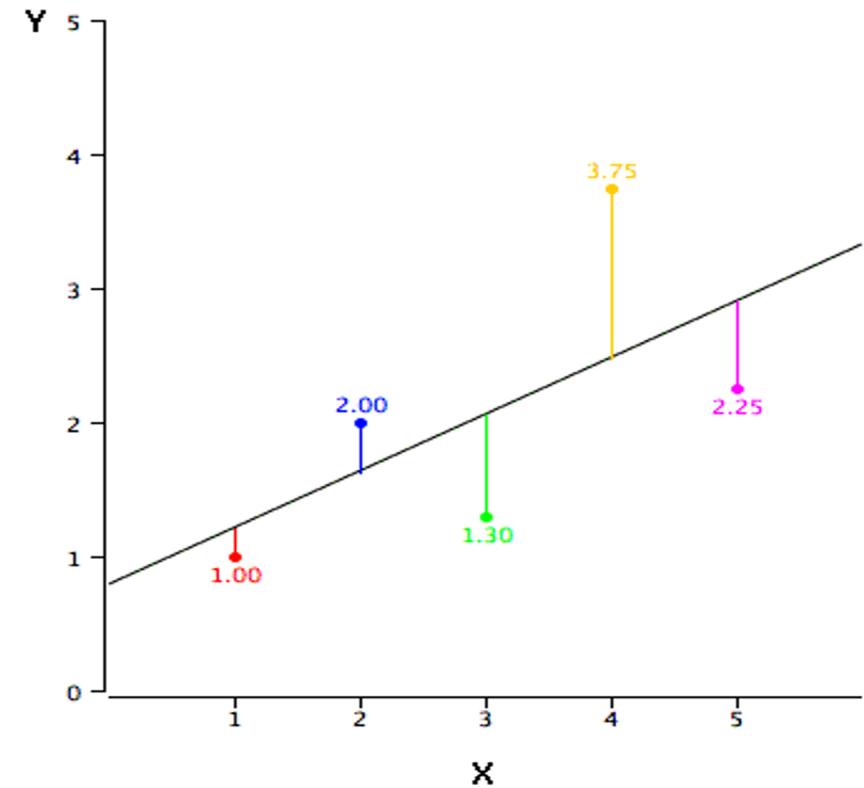
Linear Logistic Algorithm (Concept)

- Linear regression is usually among the first few topics which people pick while learning predictive modelling

- In this technique:
  - The dependent variable is continuous
  - Nature of regression line is linear

- Now, the question is *"How do we obtain best fit line?"*

Relation B/w Weight & Height

$y = 0.2811x + 13.9$

$R^2 = 0.4218$

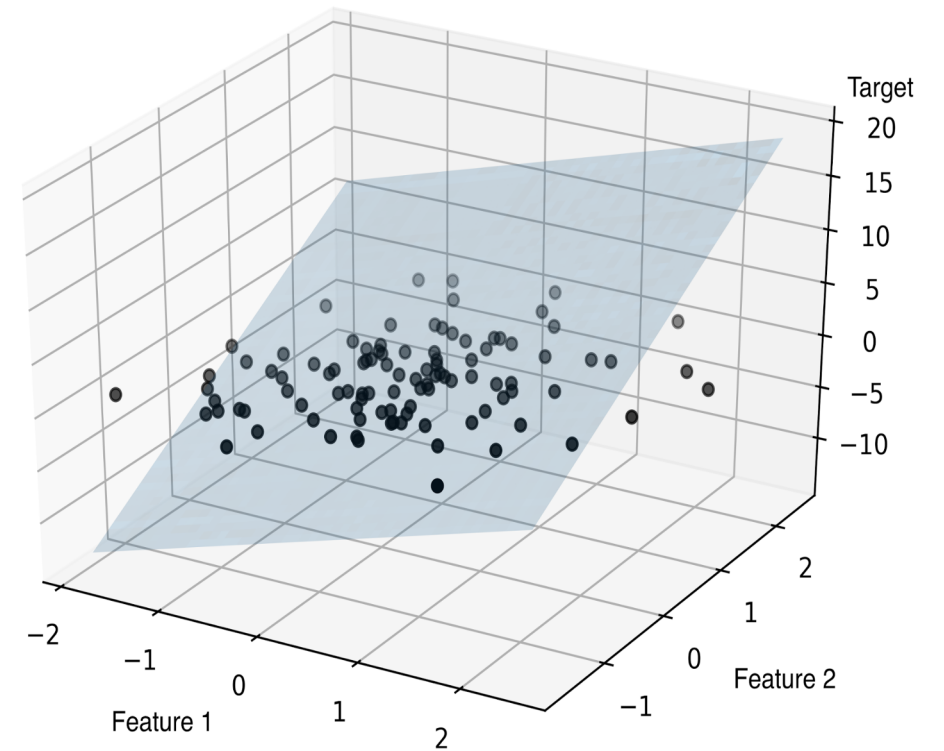# Linear Logistic Algorithm (Concept)

- How to obtain best fit line (value of a and b)?

- This task can be easily accomplished by Least Square Method

- We can evaluate the model performance using the metric R-square

Linear Logistic Algorithm (Concept)

- **Simple linear regression**: one-to-one relationship between the input variable and the output variable

- **Multiple linear regression**: many-to-one relationship, instead of just using one input variable, you use several

$$y = w_0 x_0 + w_1 x_1 + \ldots + w_m x_m = \sum_{i=0}^{m} w_i x_i = w^T x$$

Linear Logistic Algorithm (Scikit Learn)

- **Fit_intercept** : boolean, optional, default True

  - Whether to calculate the intercept for this model. If set to False, no intercept will be used in calculations (e.g. data is expected to be already centred).

- **Normalize** : boolean, optional, default False

  - This parameter is ignored when fit_intercept is set to False. If True, the regressors X will be normalized before regression by subtracting the mean and dividing by the l2-norm. If you wish to standardize, please use sklearn.preprocessing.StandardScaler before calling fit on an estimator with normalize=False

`class sklearn.linear_model. LinearRegression (fit_intercept=True, normalize=False, copy_X=True, n_jobs=1)`

Linear Logistic Algorithm (Scikit Learn)

- **Fit** : Estimates the best representative function for the the data points. With that representation, you can calculate new data points

- **Predict** : Utilizing incoming data points to find the new output based on model representation from the fit method

- **Score:** Returns the coefficient of determination R^2 of the prediction.

Linear Logistic Algorithm (Pros and Cons)

Pros:
1. Easy to understand
2. Easy to implement and achieve good scores
3. The ability to identify outliers or anomalies

Cons:
1. Linear regression is limited to linear relationships
2. Linear Regression Is Sensitive to Outliers

Assignment 2

o Lakukan Data Visualization, Data Preprocessing dan Data Modelling dengan menggunakan datasets *boston_housing.csv* untuk membangun AI yang bisa memprediksi harga rumah

o Setelah melakukan proses *training*, lakukan evaluasi dan kesimpulan dari *accuracy* yang berhasil dicapai

Thanks!