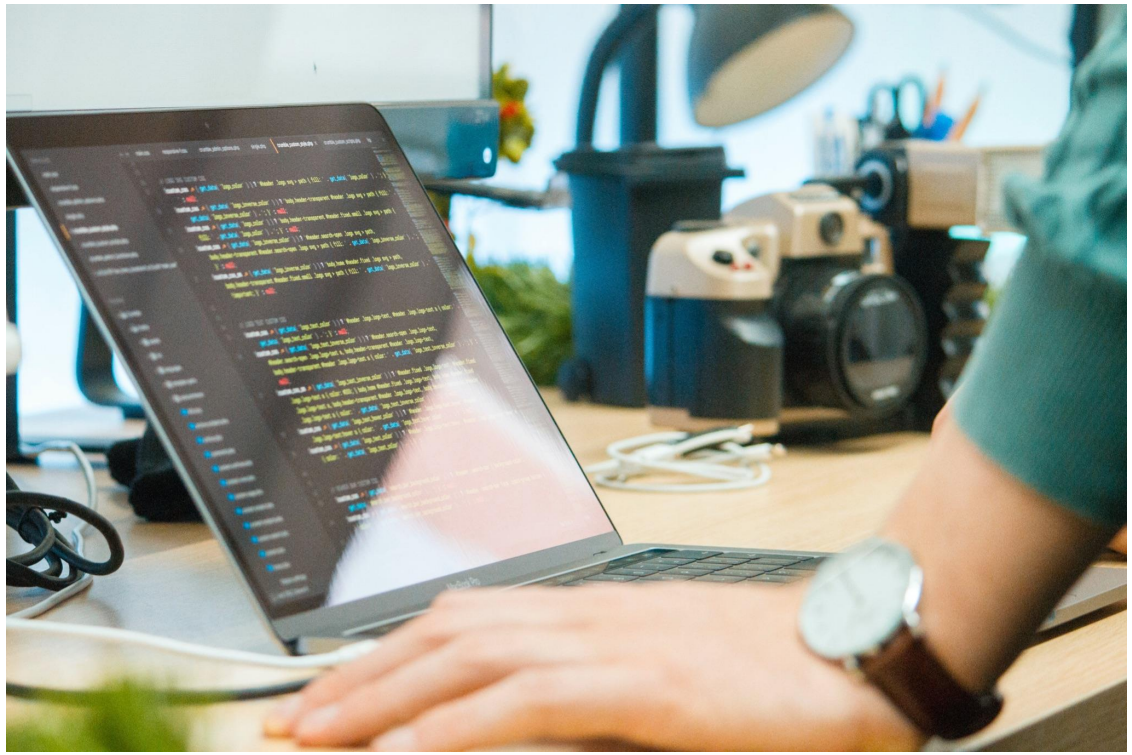


Linear & Logistic Regression

Indonesia AI
AI for Everyone, AI for Indonesia

Objective & Outline

Objective & Outline

Indonesia AI
AI for Everyone, AI for Indonesia

Objective

Understanding two of basic algorithms to easily create a smart system in AI which are linear & logistic regression algorithms.

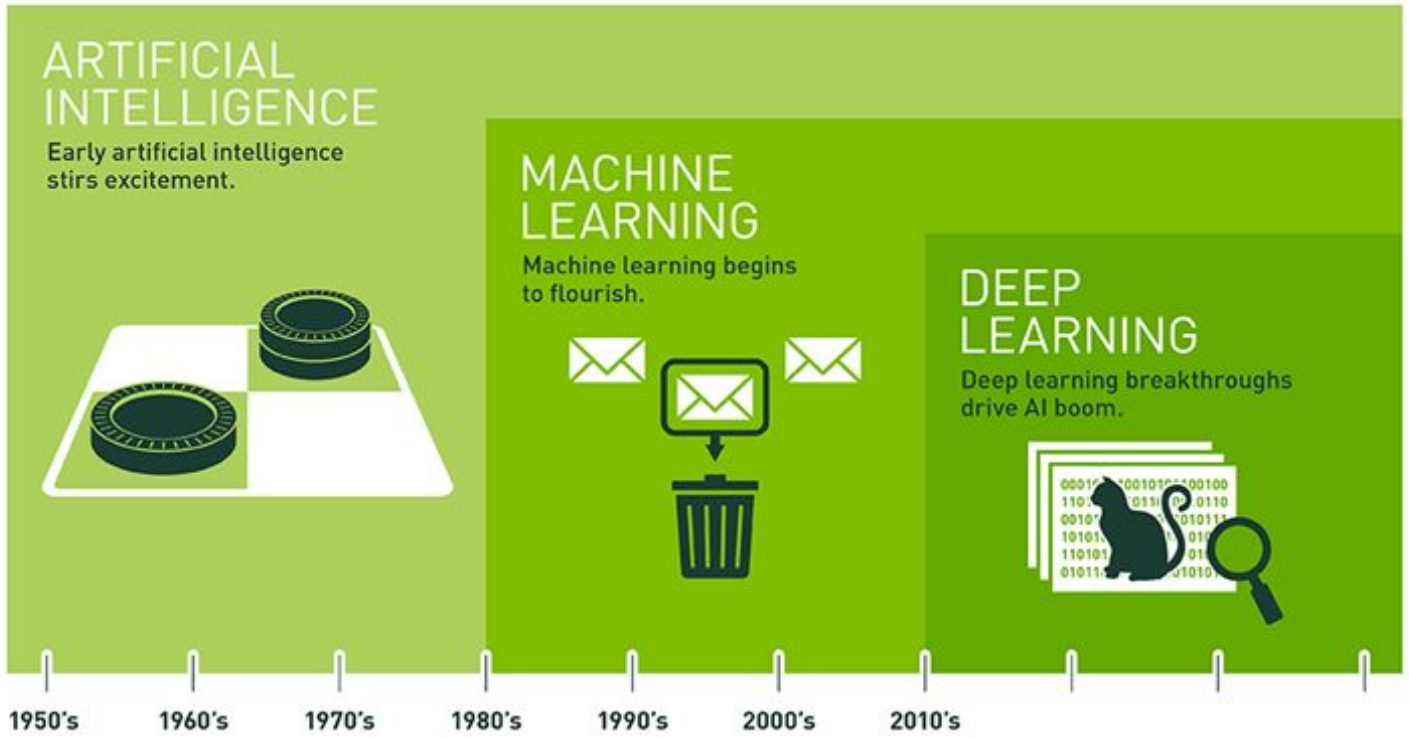
Outline

- AI, Machine Learning & Deep Learning
- Supervised and Unsupervised Learning
- Linear Regression Algorithm
- Logistic Regression Algorithm

AI, Machine Learning and Deep Learning

Indonesia AI
AI for Everyone, AI for Indonesia

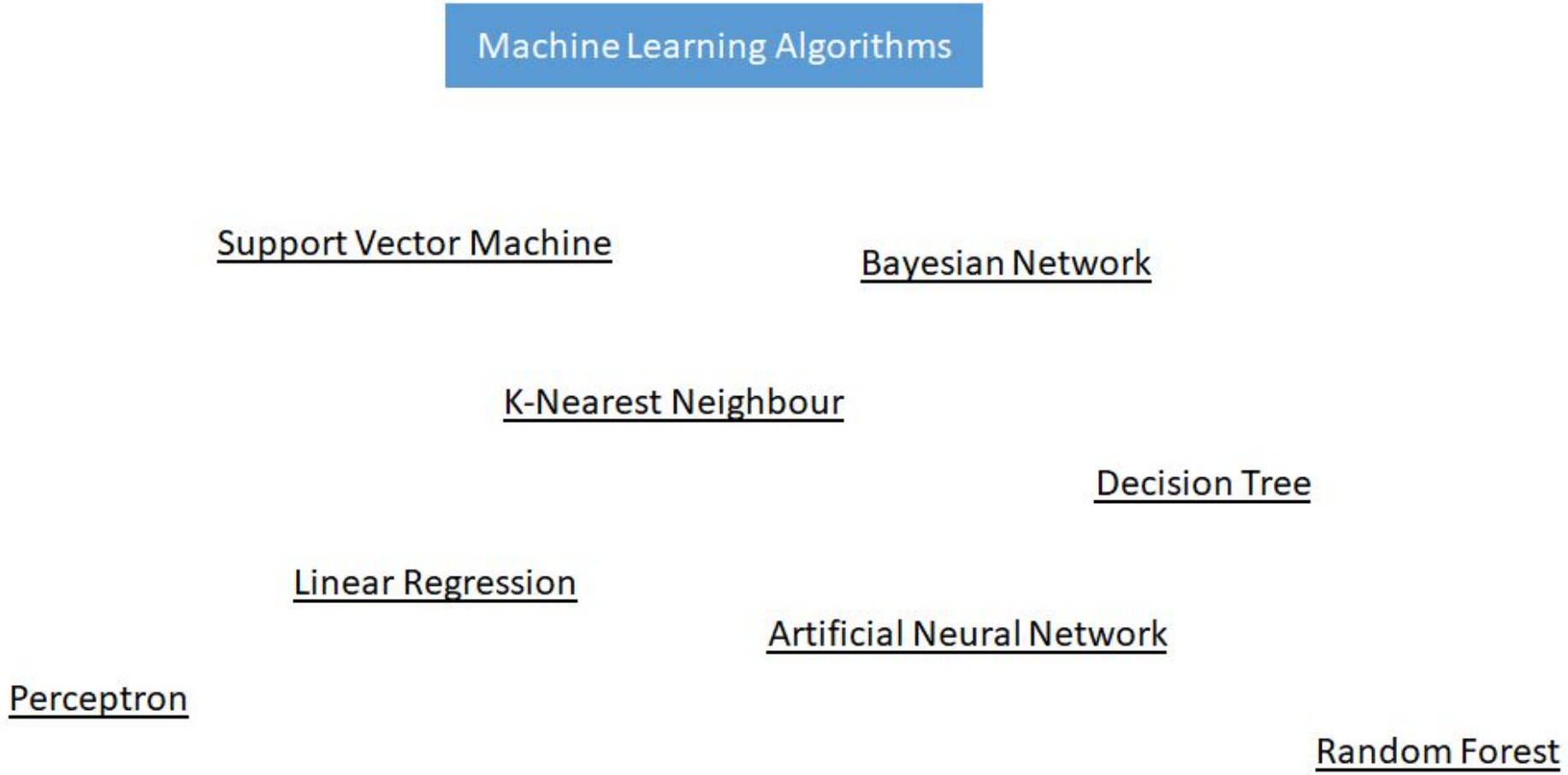
AI, Machine Learning and Deep Learning



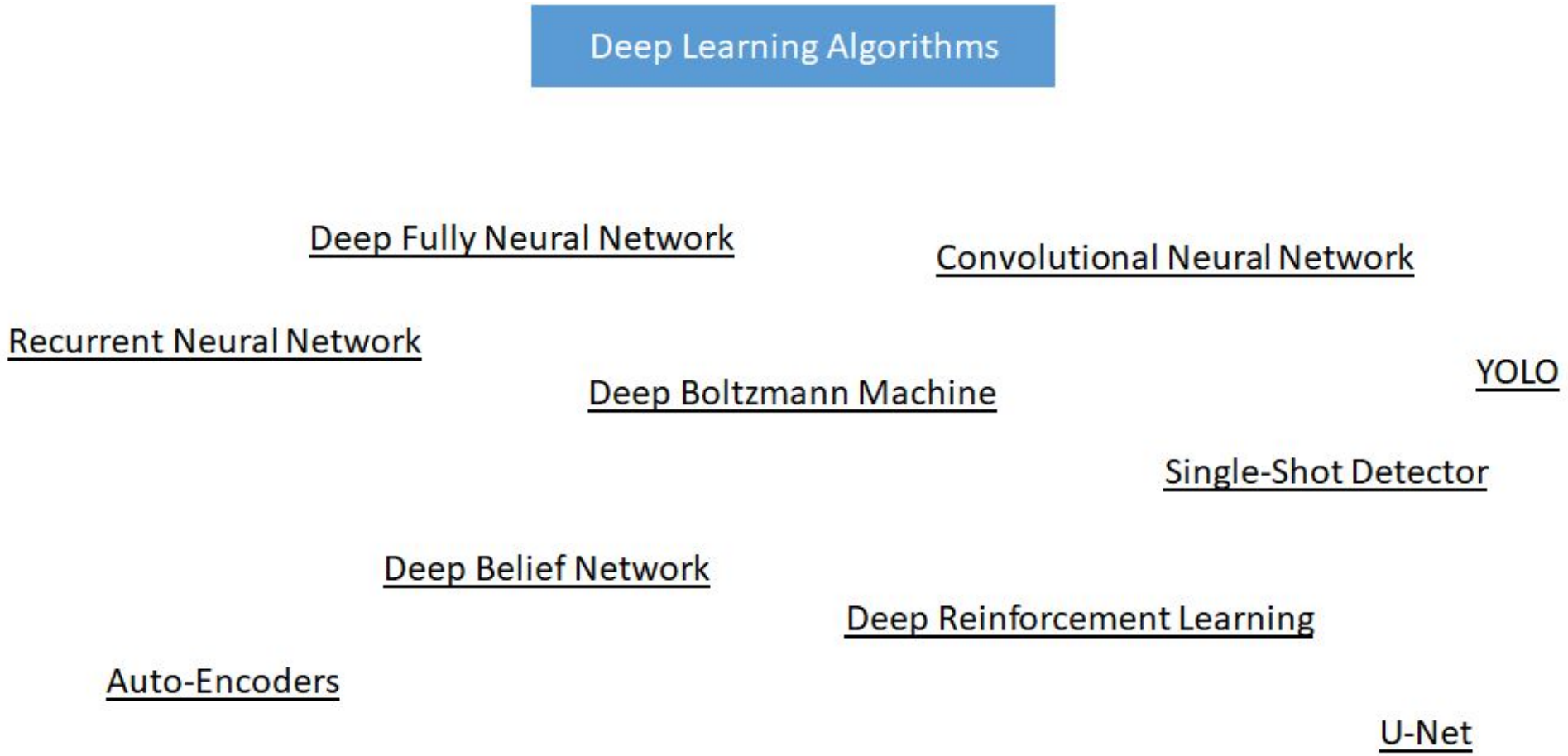
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

AI, Machine Learning and Deep Learning

Indonesia AI
AI for Everyone, AI for Indonesia



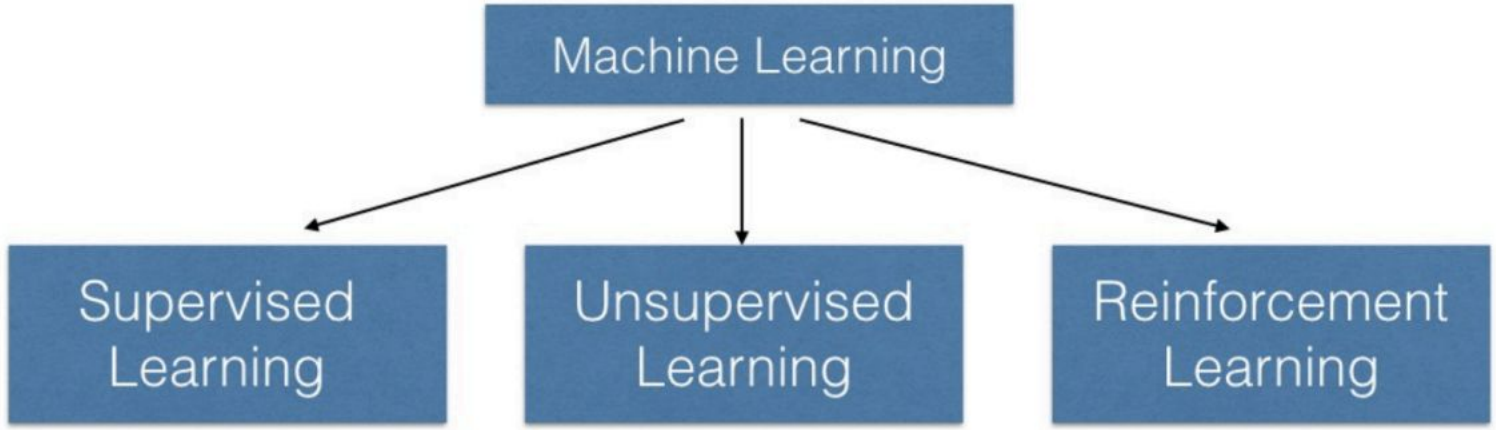
AI, Machine Learning and Deep Learning



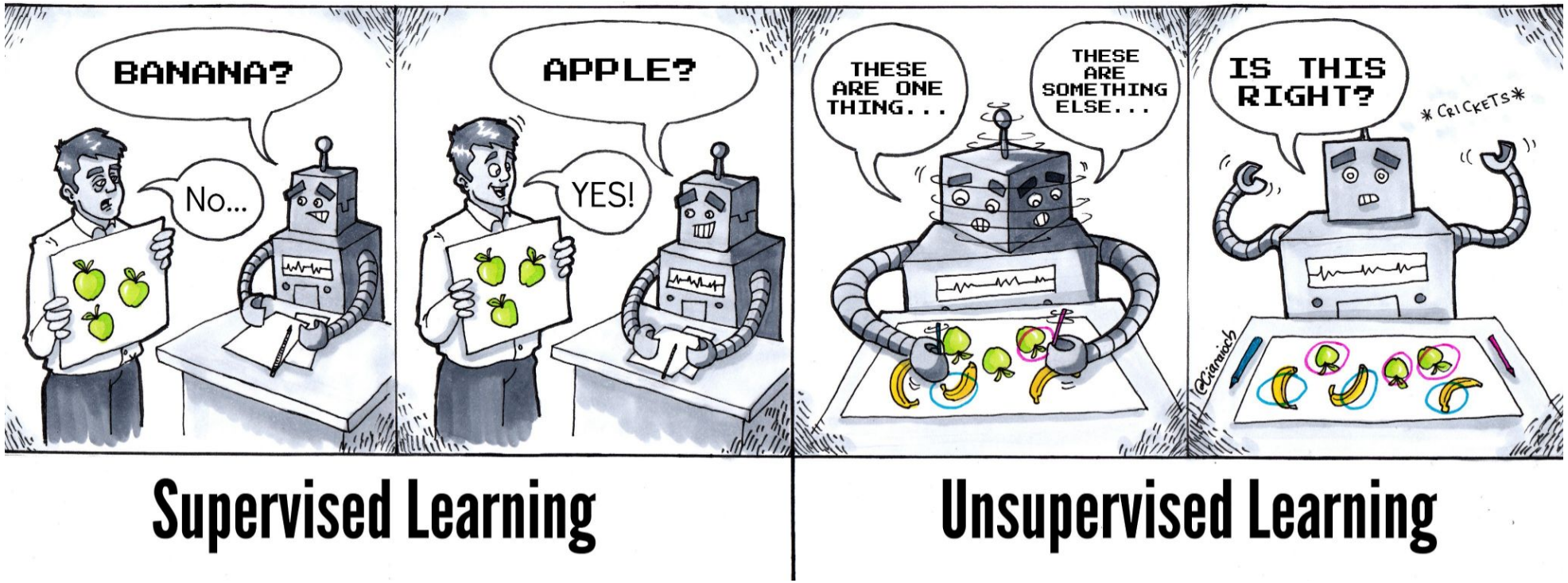
Supervised and Unsupervised Learning

Indonesia AI
AI for Everyone, AI for Indonesia

Supervised and unsupervised learning



Supervised and unsupervised learning



Quiz Session

Quiz

Indonesia AI
AI for Everyone, AI for Indonesia

What do you think about the current development of AI and how it affects your life?

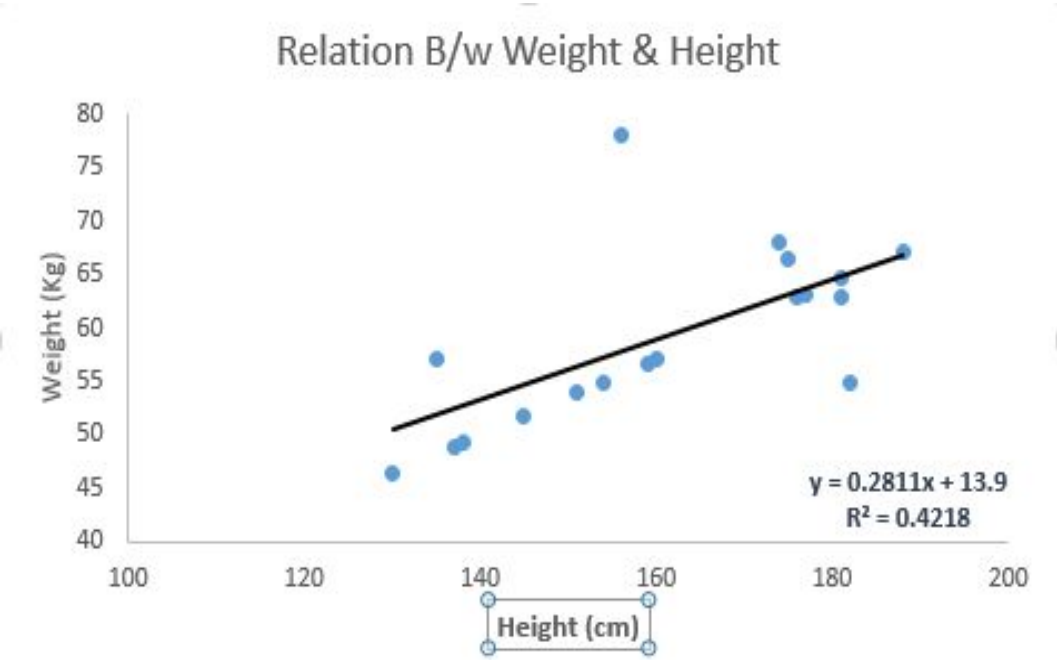
***

Linear Regression Algorithm

Linear Regression Algorithm (Concept)

Indonesia AI
AI for Everyone, AI for Indonesia

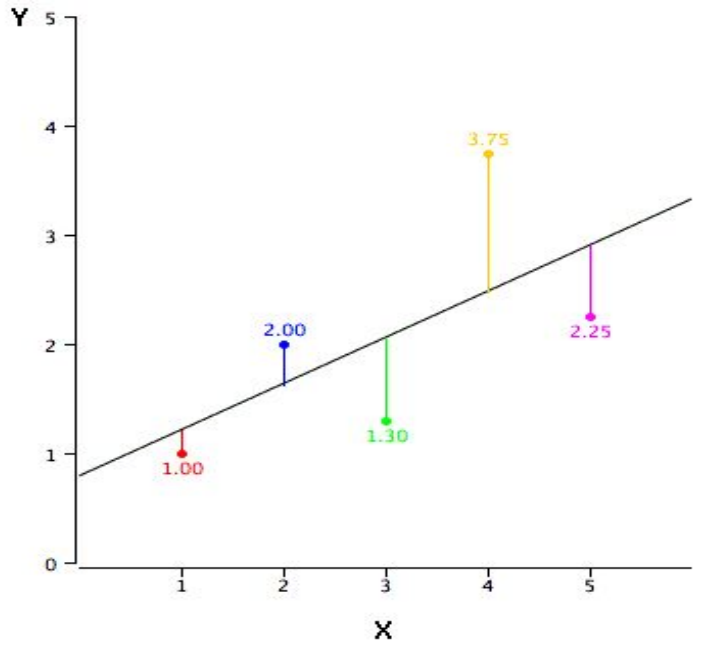
- Linear regression is usually among the first few topics which people pick while learning predictive modelling
- Also called **Hello World** of Machine Learning
- It works for **regression** problem so in this technique the dependent or target variable (weight) is continuous
- Let say we want to make AI to predict what is the weight of patient based on the data of his/her height? So the AI will be supposed to create the **straight line**
- Now, the question is *"How does it obtain the best straight line?"*



Linear Regression Algorithm (Concept)

Indonesia AI
AI for Everyone, AI for Indonesia

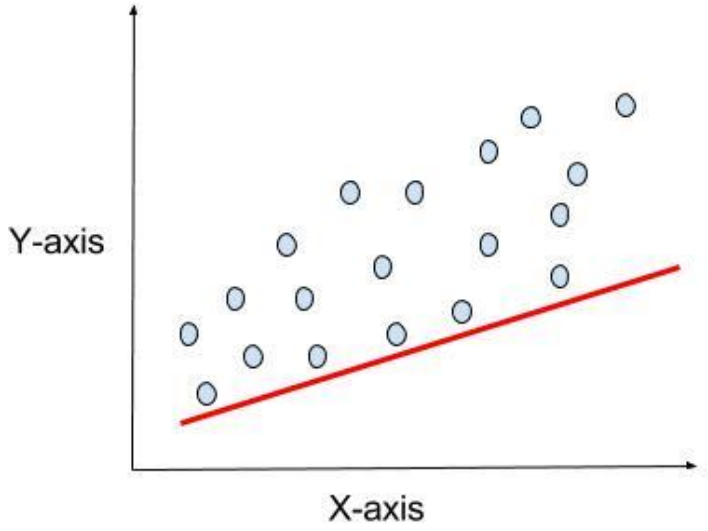
- How to obtain best straight line (value of a and b)?
- This task can be easily accomplished by calculating the **distance error** from all the data points to the straight line (Least Square Method)
- Then slowly but surely it will find the **correct position** of the straight line
- We can evaluate the model performance using the metric in statistic such as R-square or RMSE



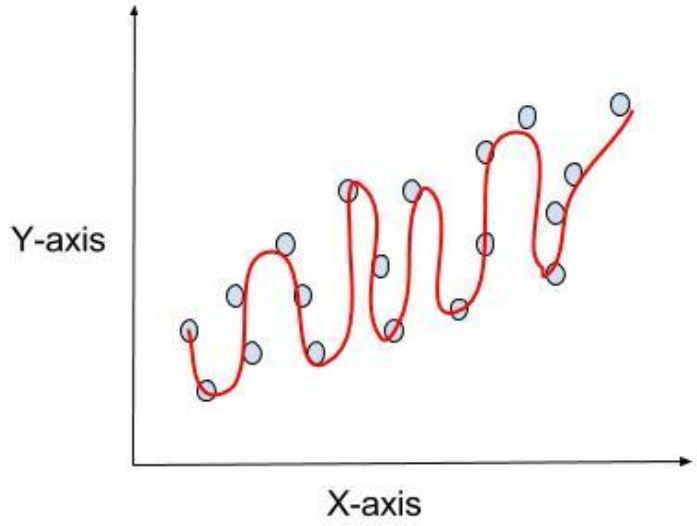
Linear Regression Algorithm (Concept)

Indonesia AI
AI for Everyone, AI for Indonesia

- The model should fit to the data points with the most less error
- We expect the line should follow the trend of the data points called **Good fitting**, No underfitting, no overfitting!



Underfitting



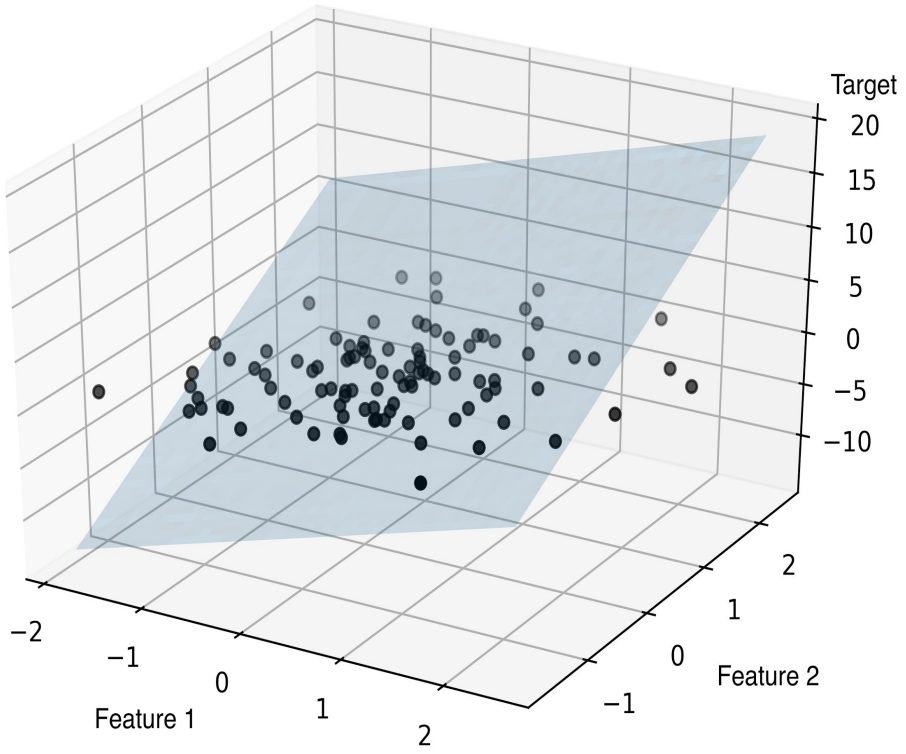
Overfitting

Linear Regression 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 Regression Algorithm (Sklearn)

Indonesia AI
AI for Everyone, AI for Indonesia

- **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 Regression Algorithm (Sklearn)

- **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 Regression Algorithm (Pros & Cons)

Indonesia AI
AI for Everyone, AI for Indonesia

Pros:

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

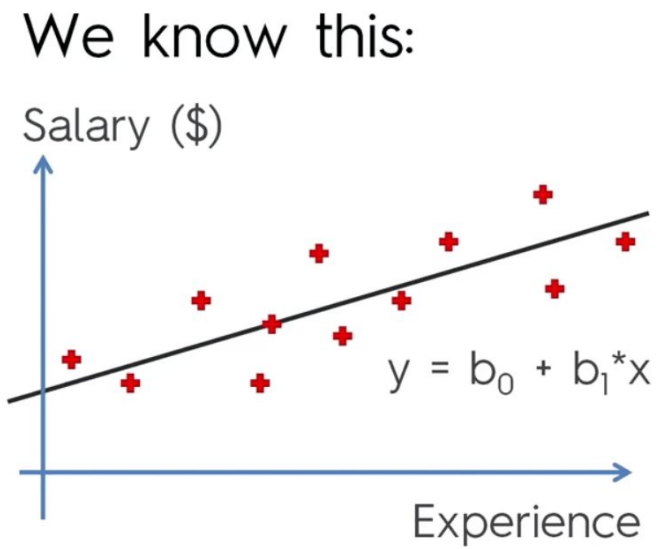
Cons:

1. Linear Regression is limited to linear relationships
2. Linear Regression is sensitive to outliers

Logistic Regression Algorithm

Logistic Regression Algorithm (Concept)

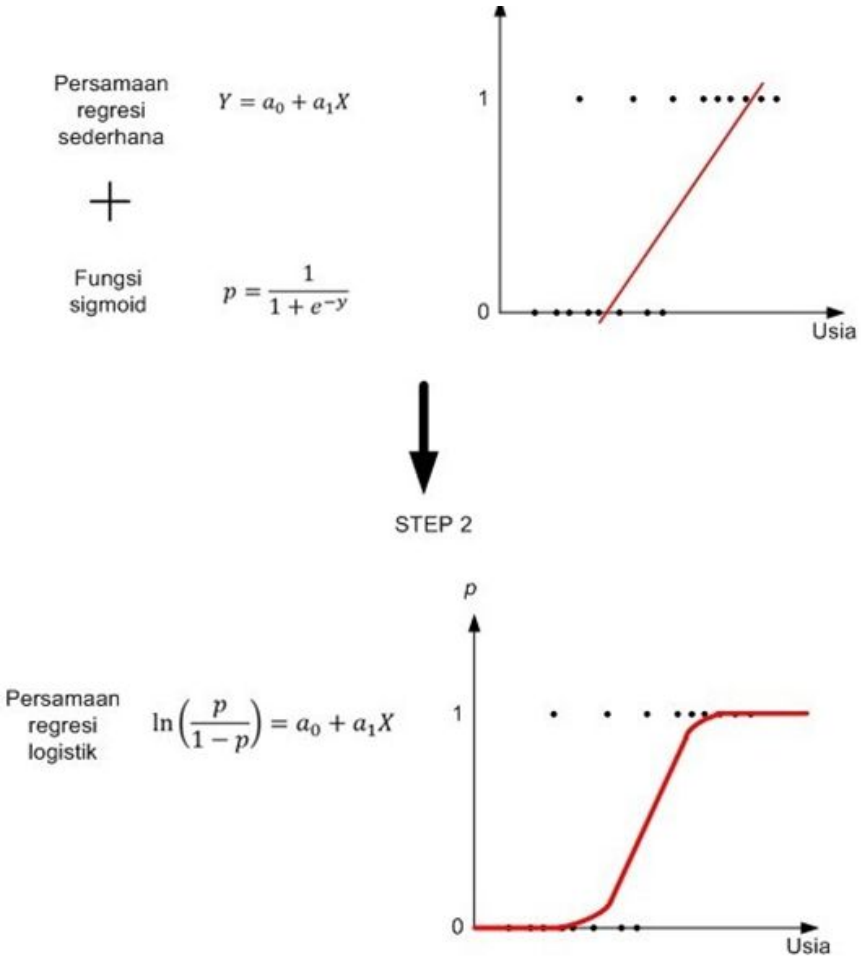
- Why not use Linear Regression algorithm?
- It is not a regression model, it works for **classification** problem instead
- Regression: how much employees should be paid?
-
- Classification: Will employees be paid or not?



Logistic Regression Algorithm (Concept)

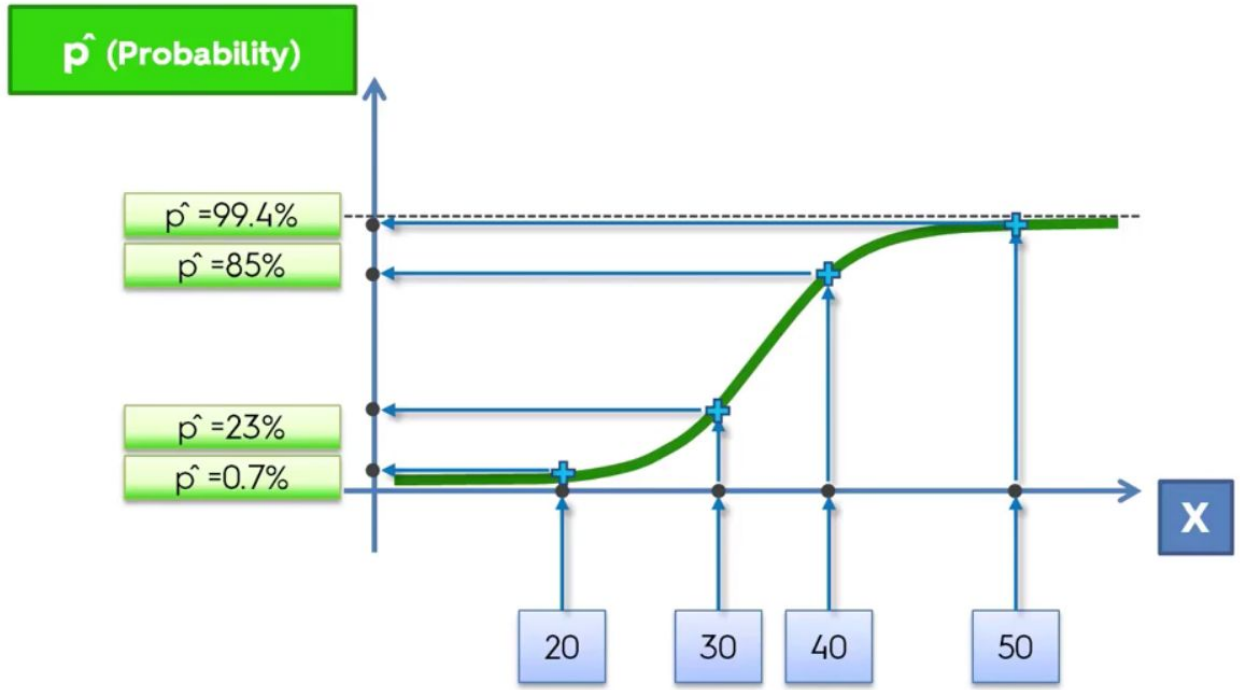
- Logistic regression predicts the probability of occurrences of an event by fitting data to a logit function by using sigmoid function]

- Apply sigmoid function to logistic regression since we have to think of it as probabilities!



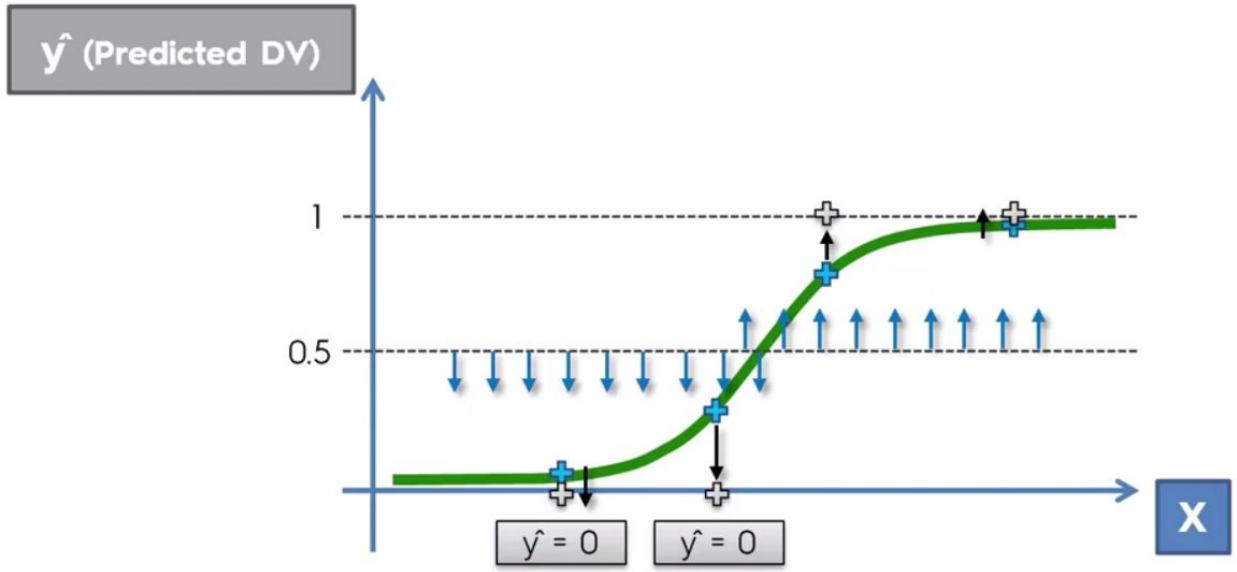
Logistic Regression Algorithm (Concept)

- Logistic regression seen from probabilities point of view
- So it will output the probability of Yes (1) is (let's say) 82% and the probability of No (2) is 18%



Logistic Regression Algorithm (Concept)

- Threshold determine whether one data is considered as one class (Yes) or the other (No)
- By default, almost every machine learning algorithm set the threshold as 0.5
- This has become the fundamental concept of Deep Learning



Quiz Session

Quiz

Indonesia AI
AI for Everyone, AI for Indonesia

You are predicting whether an email is spam or not. Based on the features, you obtained an estimated probability to be **0.75**. What's the meaning of this estimated probability? The **threshold** to differ the classes is **0.5**.

***

Assignment 2

- Lakukan Data Visualization, Data Preprocessing dan Data Modelling dengan menggunakan datasets The Boston Housing Datasets untuk membangun AI yang bisa memprediksi harga rumah

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

Indonesia AI
AI for Everyone, AI for Indonesia

Terima Kasih!

Indonesia AI | AI for Everyone, AI for Indonesia

contact@aiforindonesia.org