

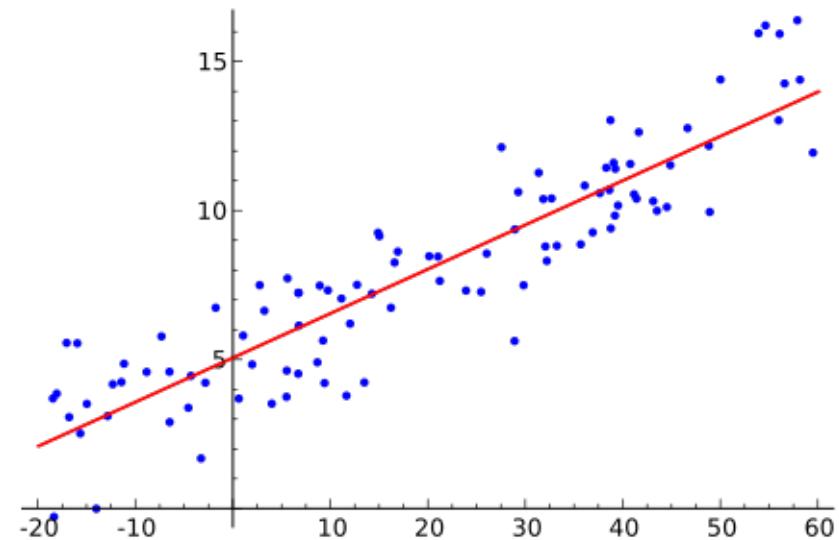
# Basic Machine Learning: Linear & Logistic Regression

Oleh: Muhammad Angga Muttaqien | Founder & Mentor

# Goal

# Goal

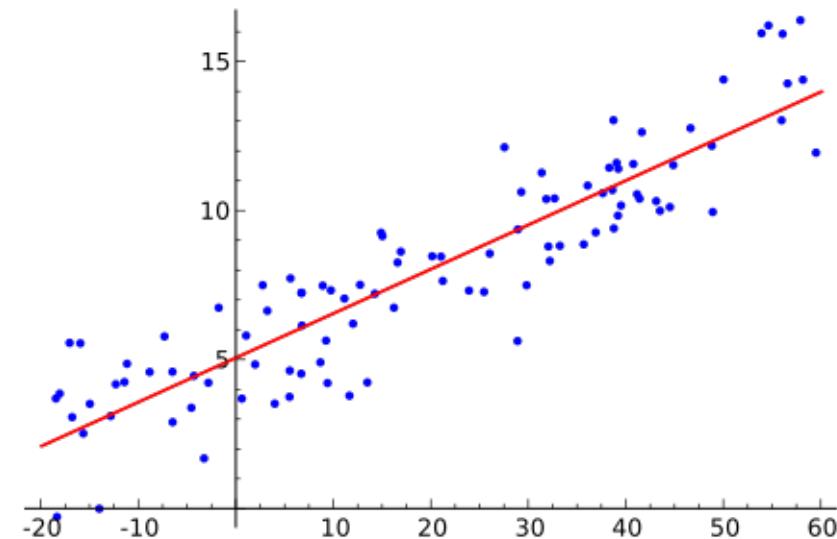
Understanding one of basic algorithm to easily create a smart system in AI which is linear & logistic regression algorithm.



# Outline

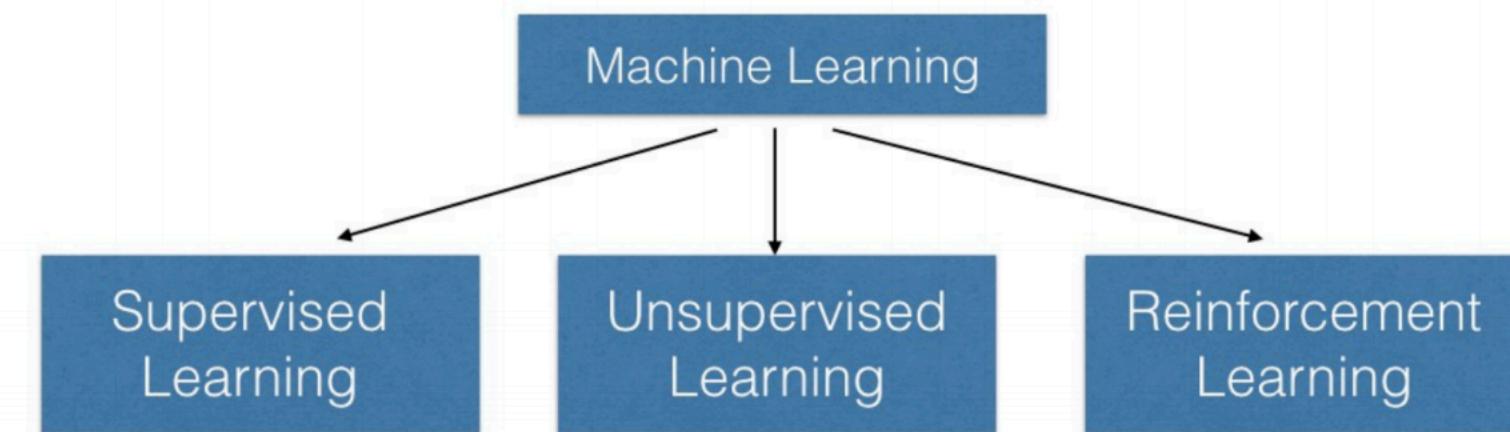
# Outline

- Supervised vs Unsupervised Learning
- Linear & Logistic Regression Algorithm
  - Concept
  - Scikit Learn
  - Pros and Cons



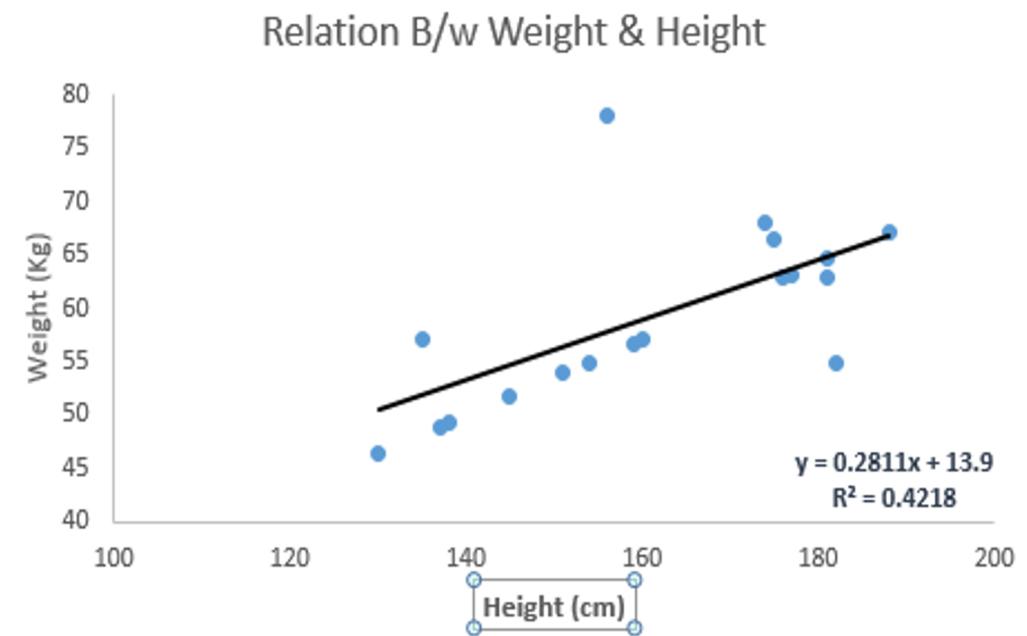
# Content

# Supervised vs Unsupervised Learning



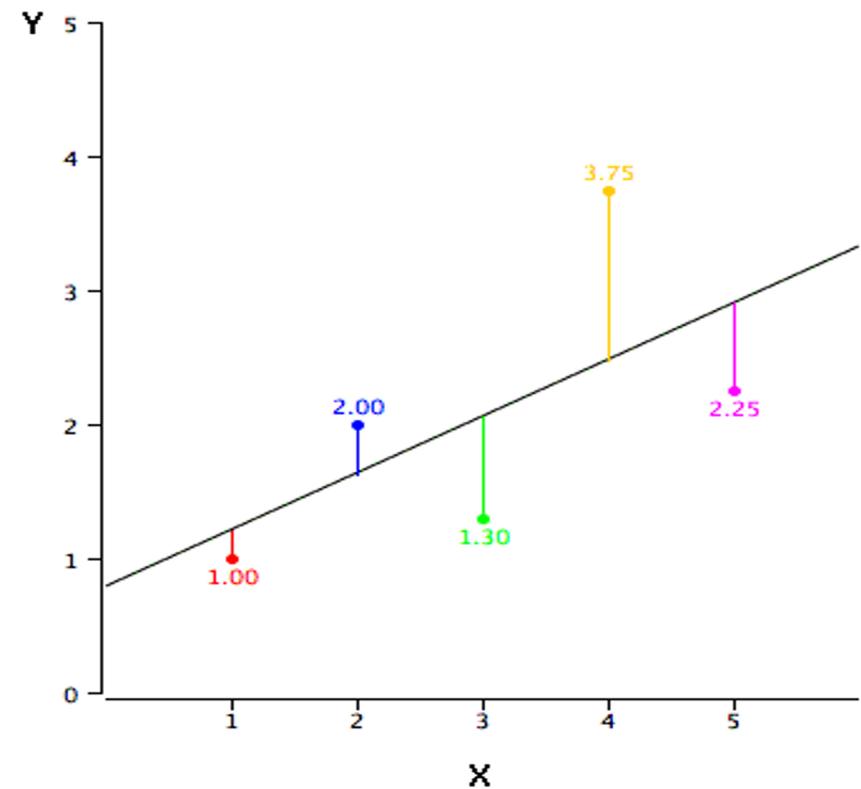
# Linear Regression Algorithm (Concept)

- 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?*”



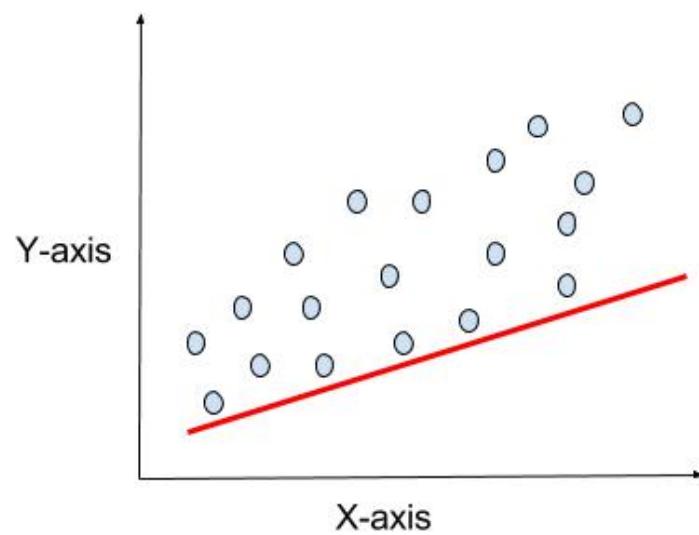
# Linear Regression Algorithm (Concept)

- 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

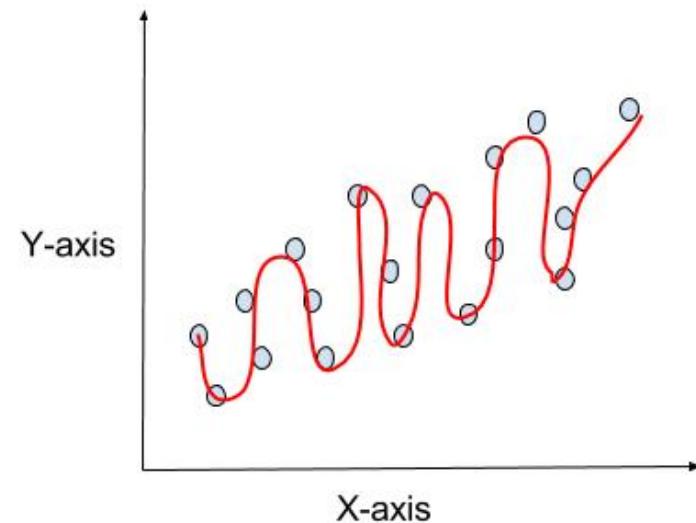


# Linear Regression Algorithm (Concept)

- The model should fit to the datapoint with the most less error
- No underfitting, no overfitting



Underfitting

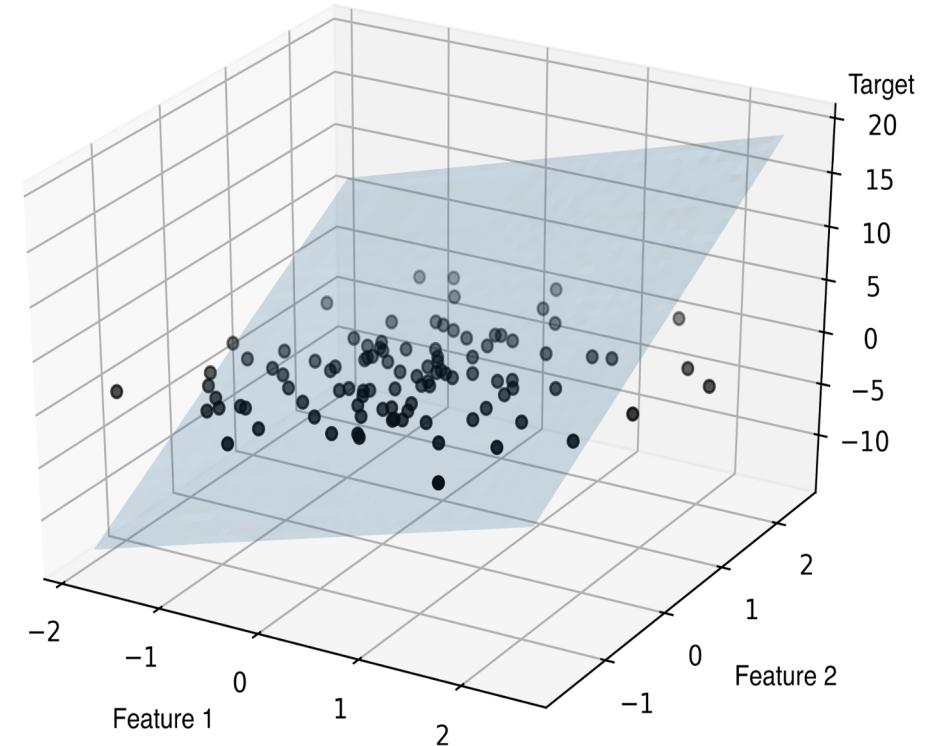


Overfitting

# 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

$$y = w_0x_0 + w_1x_1 + \dots + w_mx_m = \sum_{i=0}^m w_i x_i = w^T x$$



# Linear Regression 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 Regression 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 Regression 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

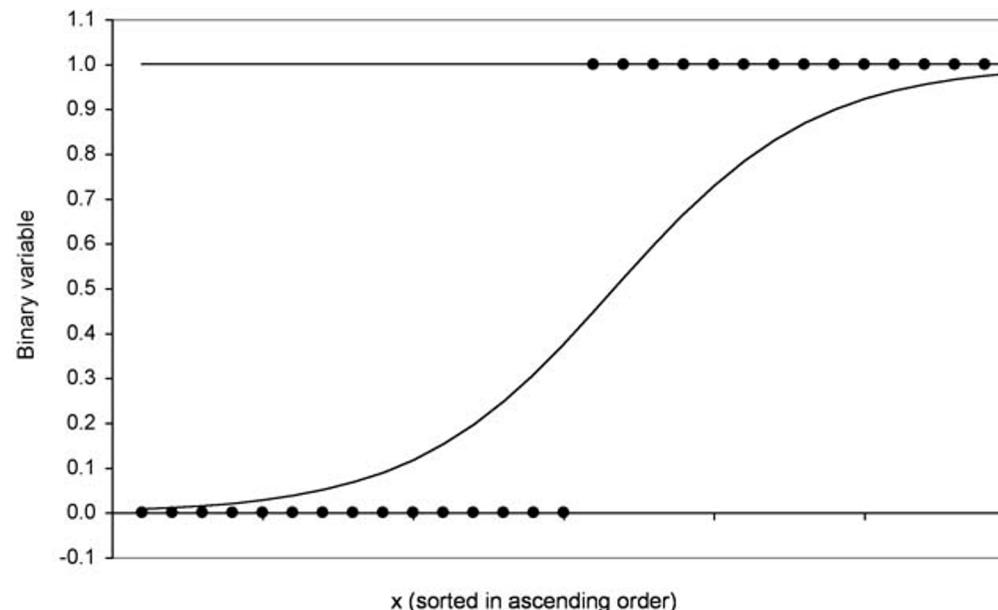
# Assignment 2

- Lakukan Data Visualization, Data Preprocessing dan Data Modelling dengan menggunakan datasets *boston\_housing.csv* untuk membangun AI yang bisa memprediksi harga rumah
- Setelah melakukan proses *training*, lakukan evaluasi dan kesimpulan dari *accuracy* yang berhasil dicapai



# Logistic Regression Algorithm (Concept)

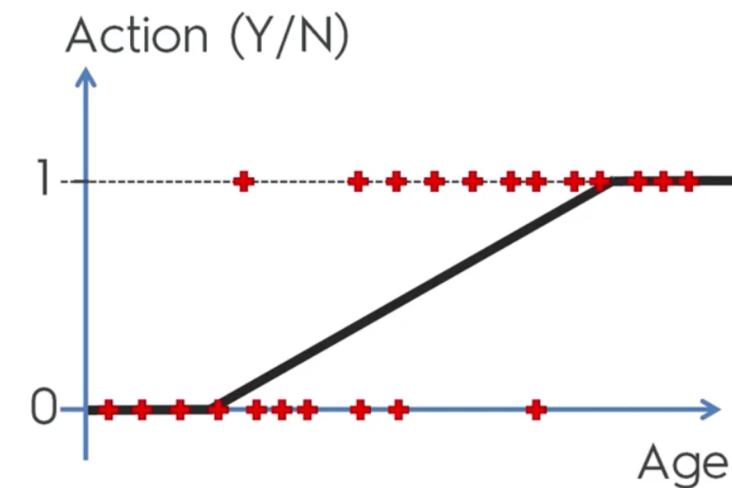
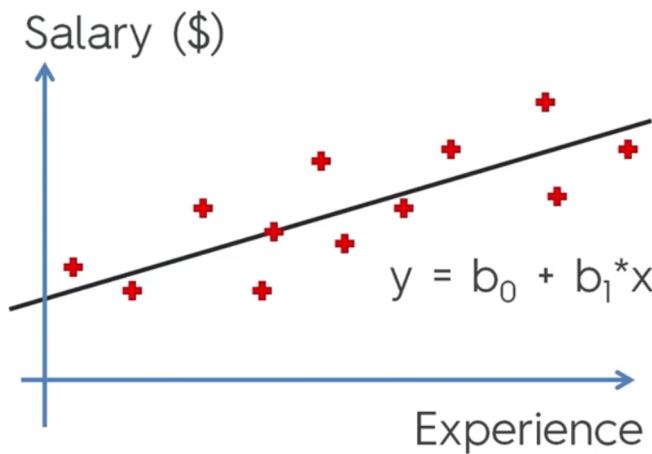
- ❑ It is not a regression model
- ❑ Logistic regression predicts the probability of occurrences of an event by fitting data to a logit function by using sigmoid function



# Logistic Regression Algorithm (Concept)

Why not use Linear Regression algorithm?

We know this:

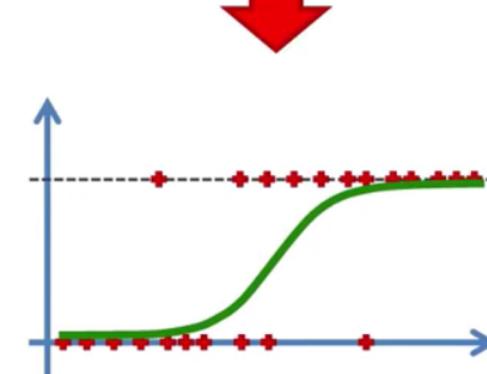
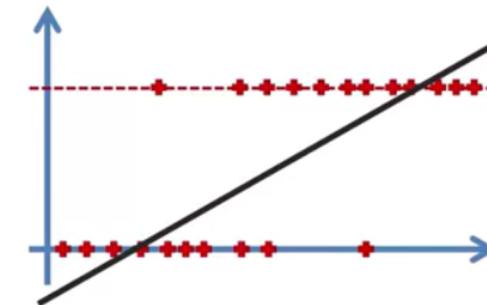


# Logistic Regression Algorithm (Concept)

Apply sigmoid function to logistic regression.

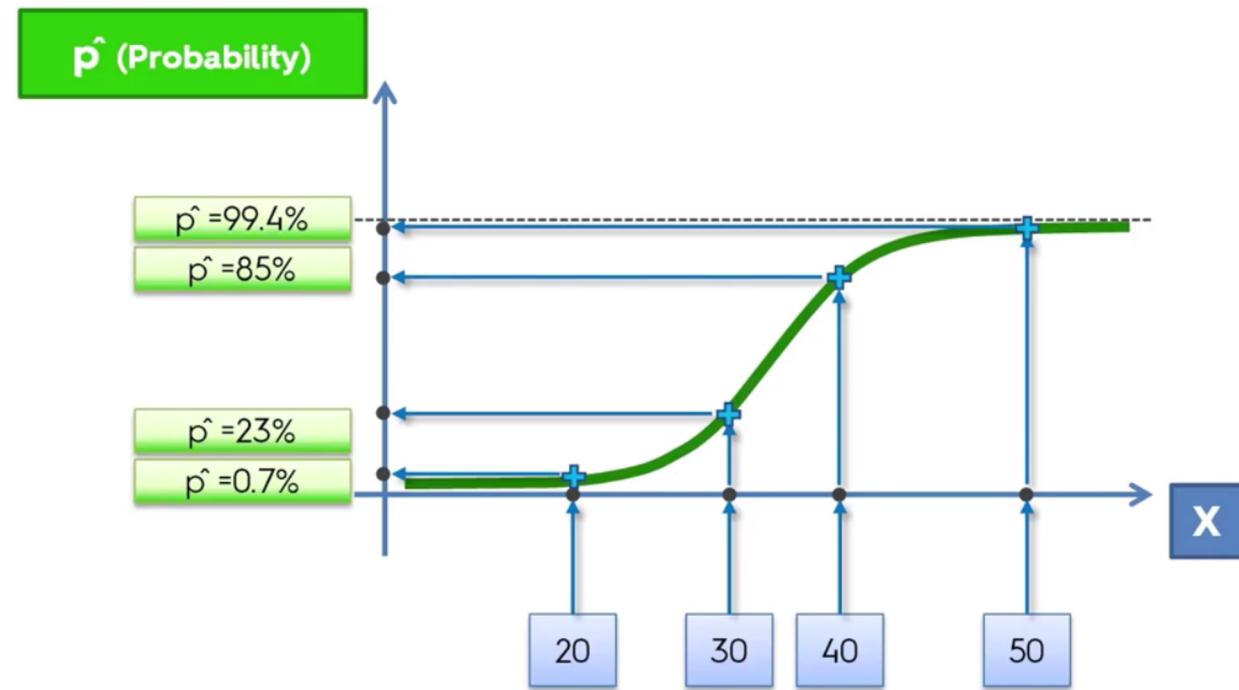
Think of it as probabilities!

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$



# Logistic Regression Algorithm (Concept)

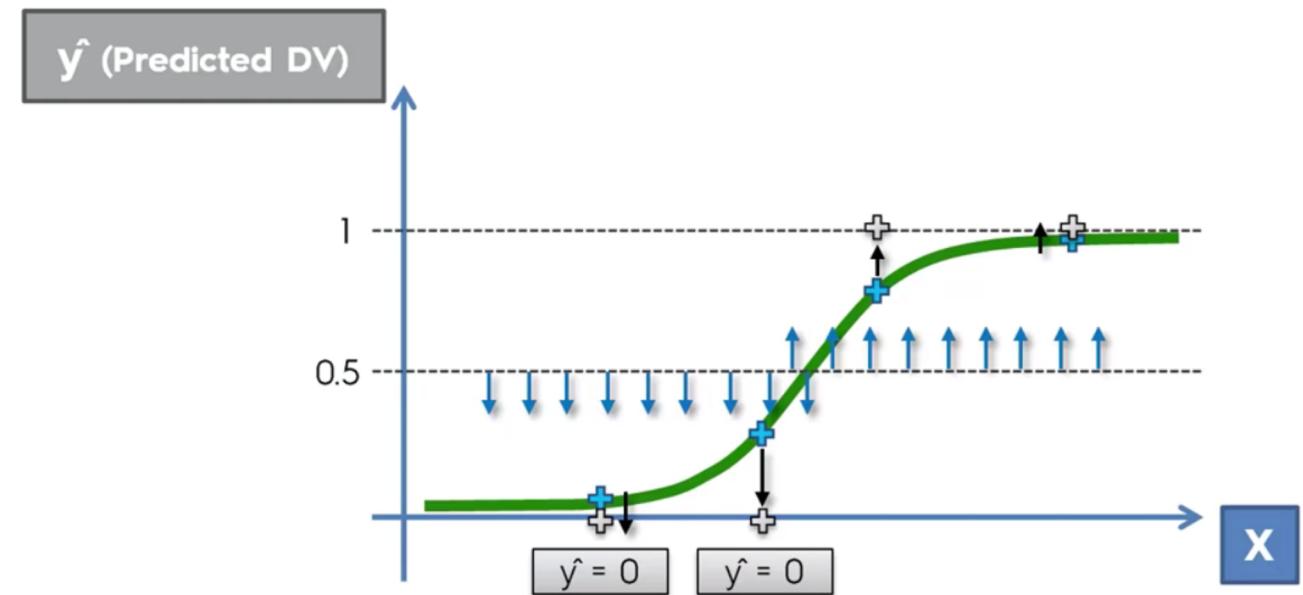
Logistic regression can be seen from probabilities point of view



# Logistic Regression Algorithm (Concept)

Threshold determine whether one data is considered as one class or the other.

This is the fundamental concept of Deep Learning.



Thanks!