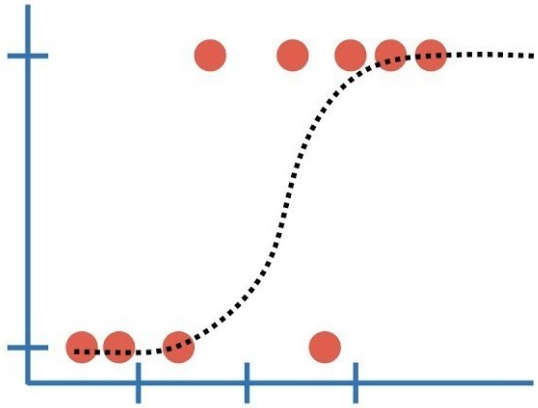
Logistic Regression using Python

Tushar B. Kute,
http://tusharkute.com



Logistic Regression

- Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary).
- Like all regression analyses, the logistic regression is a predictive analysis.
- Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.
- Remember: though the name of algorithm carries regression, it is used for classification.

Type of Logistic Regression

- Binary Logistic Regression
  - The categorical response has only two 2 possible outcomes. Example: Spam or Not.
- Multinomial Logistic Regression
  - Three or more categories without ordering. Example: Predicting which food is preferred more (Veg, Non-Veg, Vegan).
- Ordinal Logistic Regression
  - Three or more categories with ordering. Example: Movie rating from 1 to 5.

What we know?

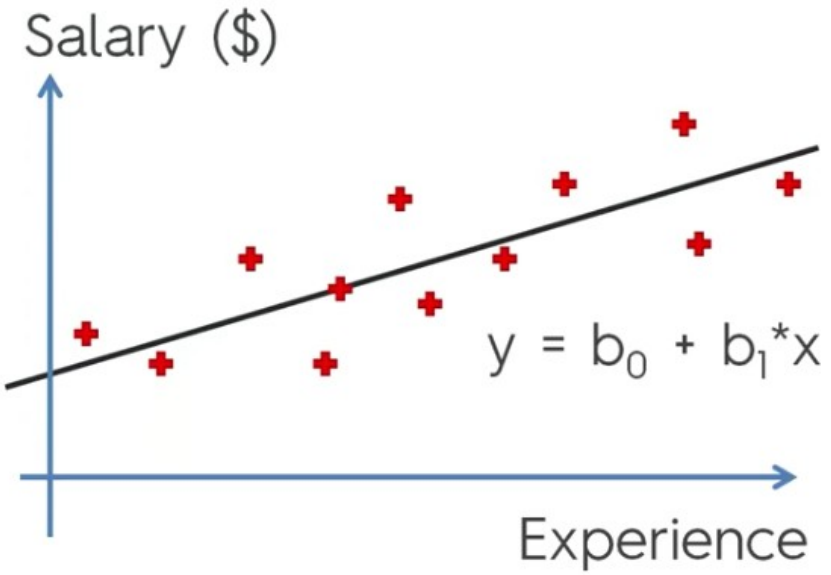
Linear Regression:

- **Simple:**

$$y = b_0 + b_1 * x$$

- **Multiple:**

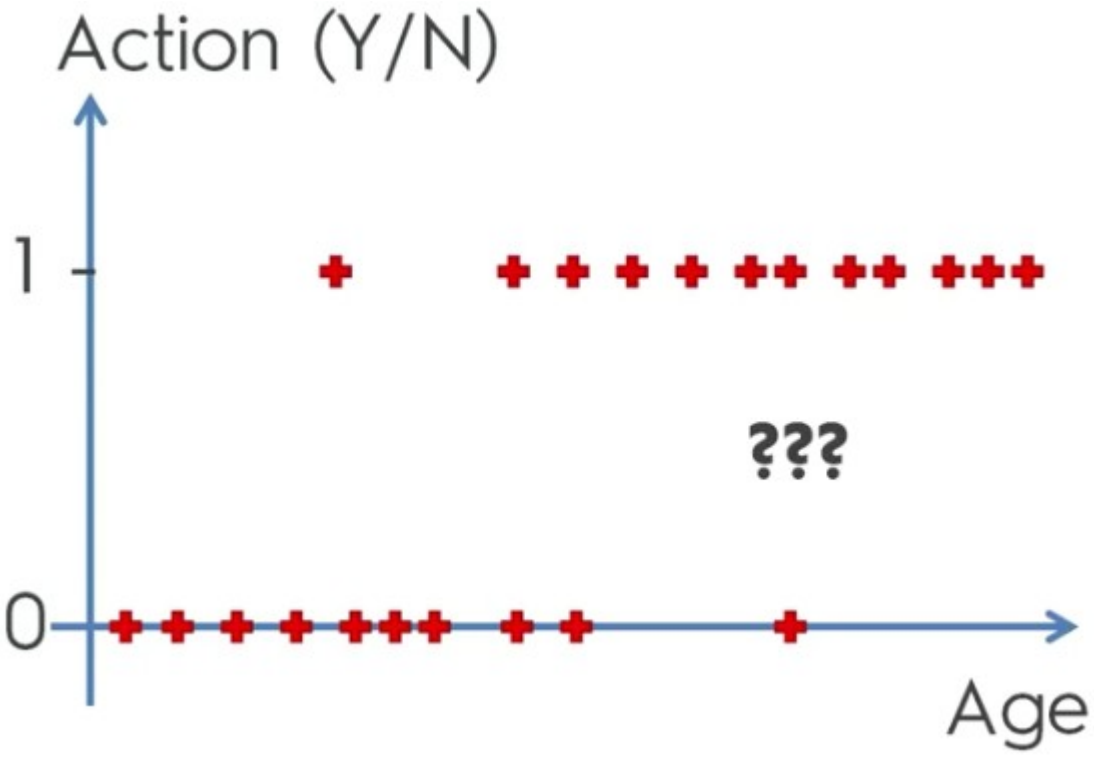
$$y = b_0 + b_1 * x_1 + \ldots + b_n * x_n$$



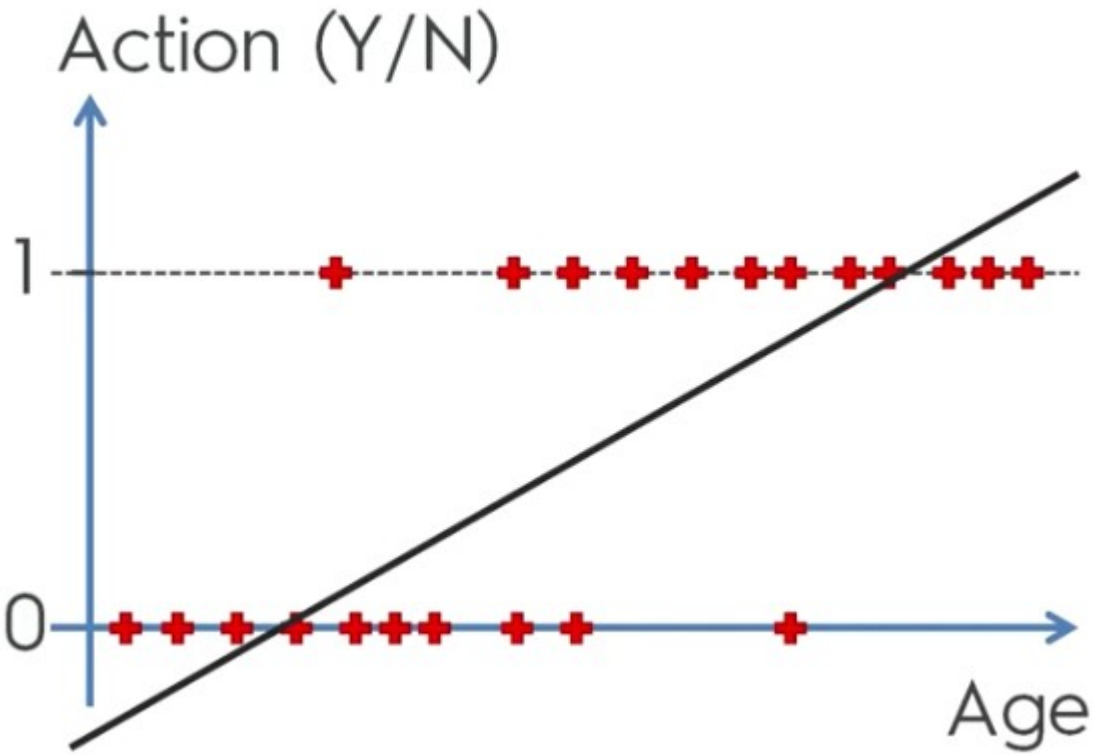
Example Reference: Super Data Science

A new problem

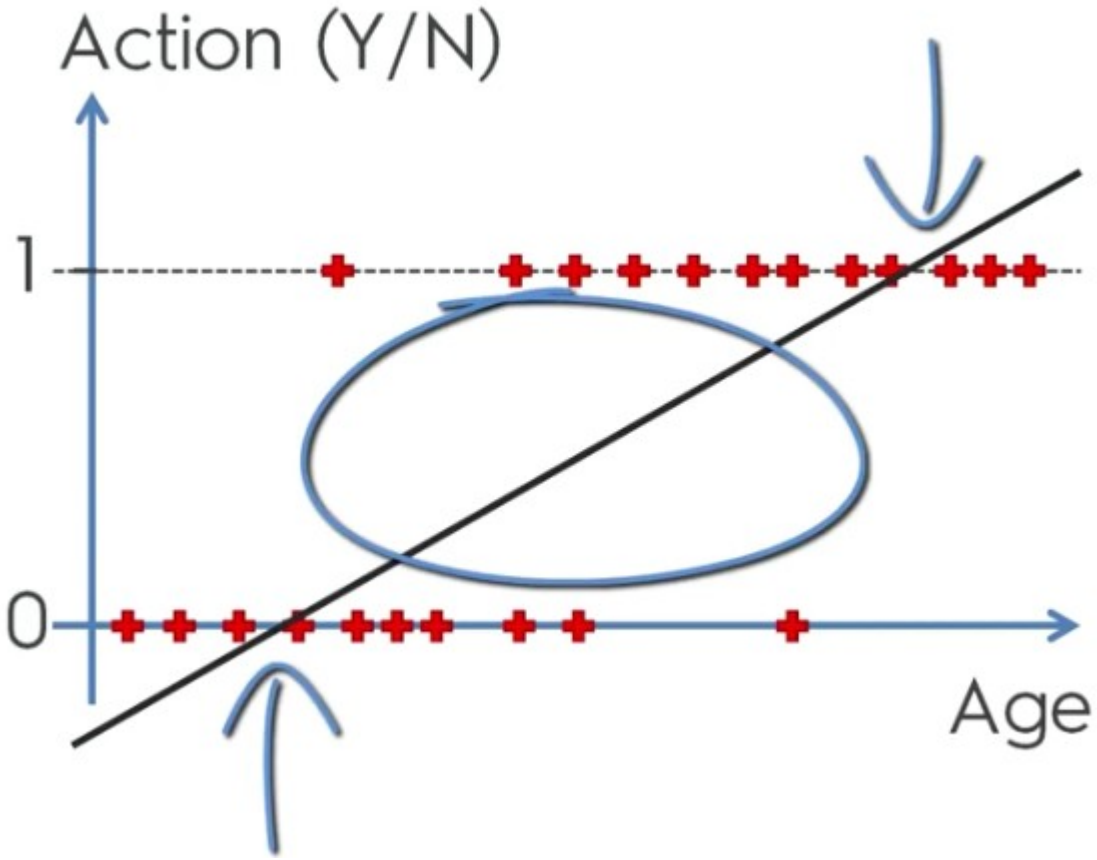
A company has provided an offer by email to their customers.



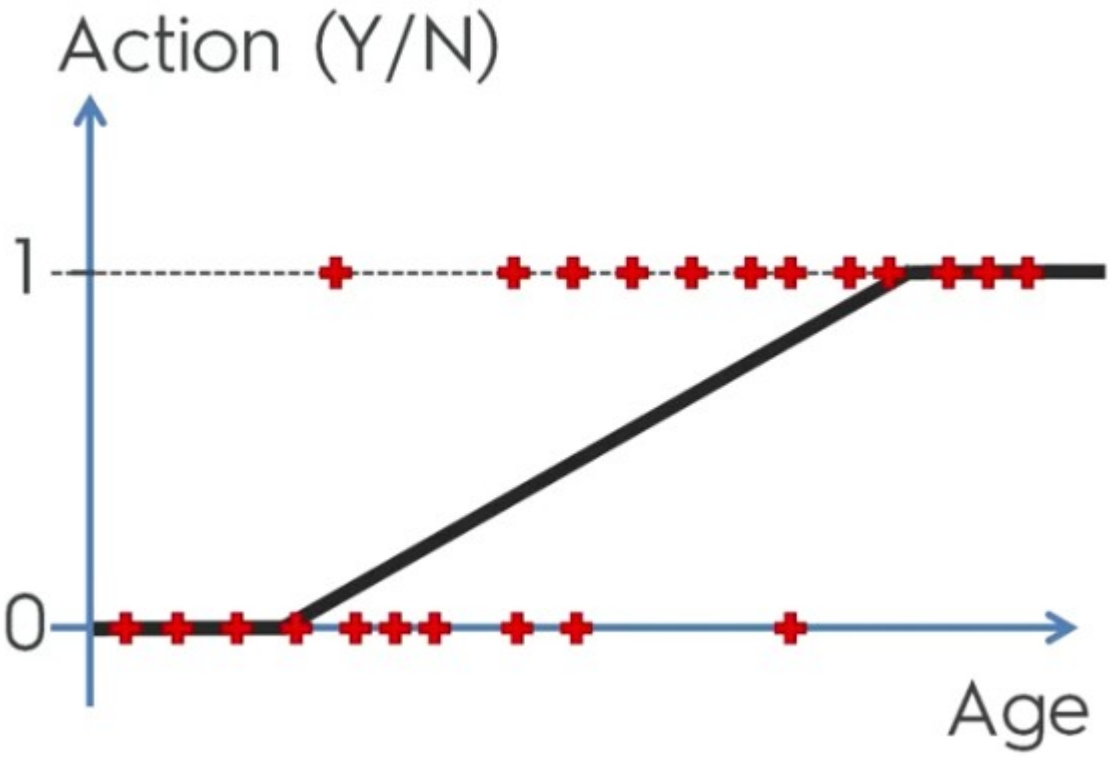
Apply Linear Regression



Apply Linear Regression



Apply Linear Regression



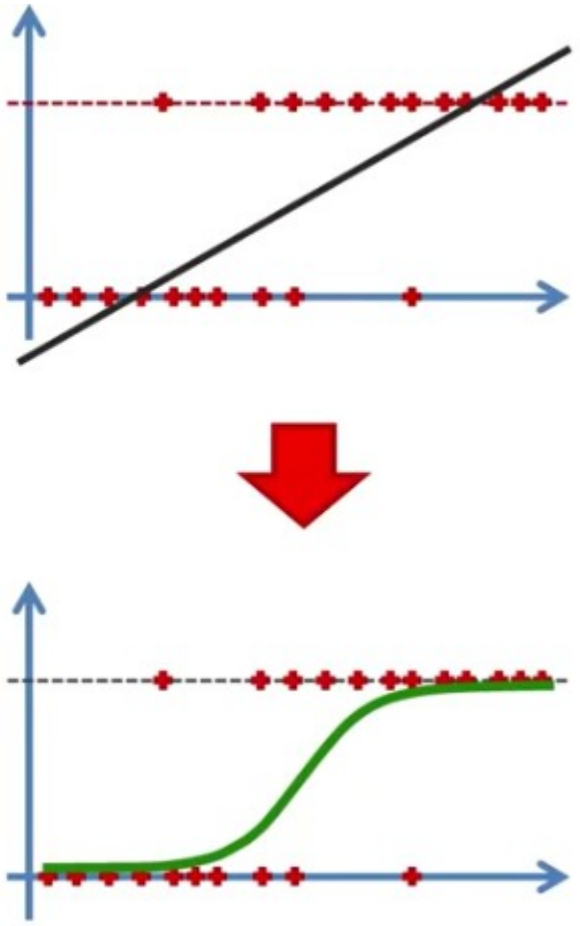
Logistic Regression

$$y = b_0 + b_1 * x$$

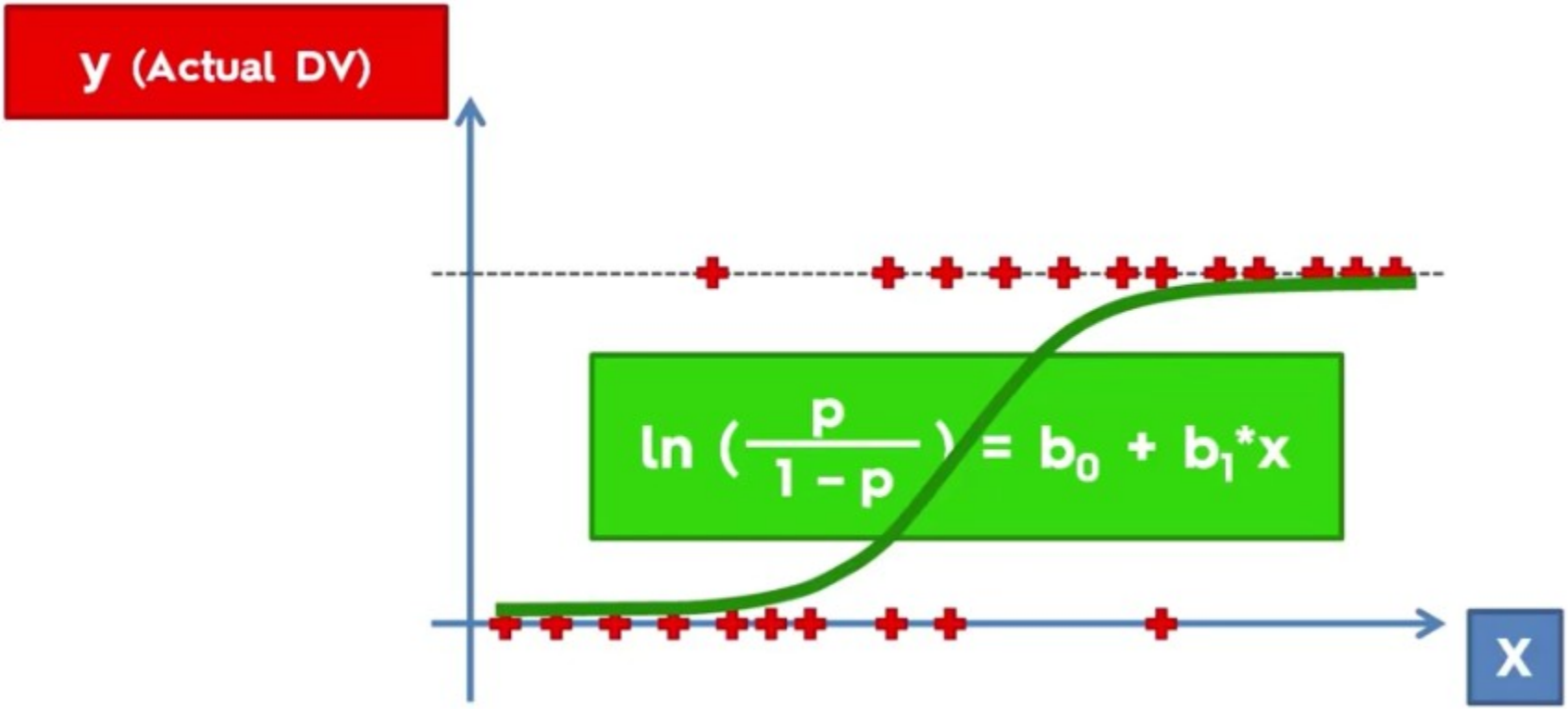
Sigmoid Function

$$p = \frac{1}{1 + e^{-y}}$$

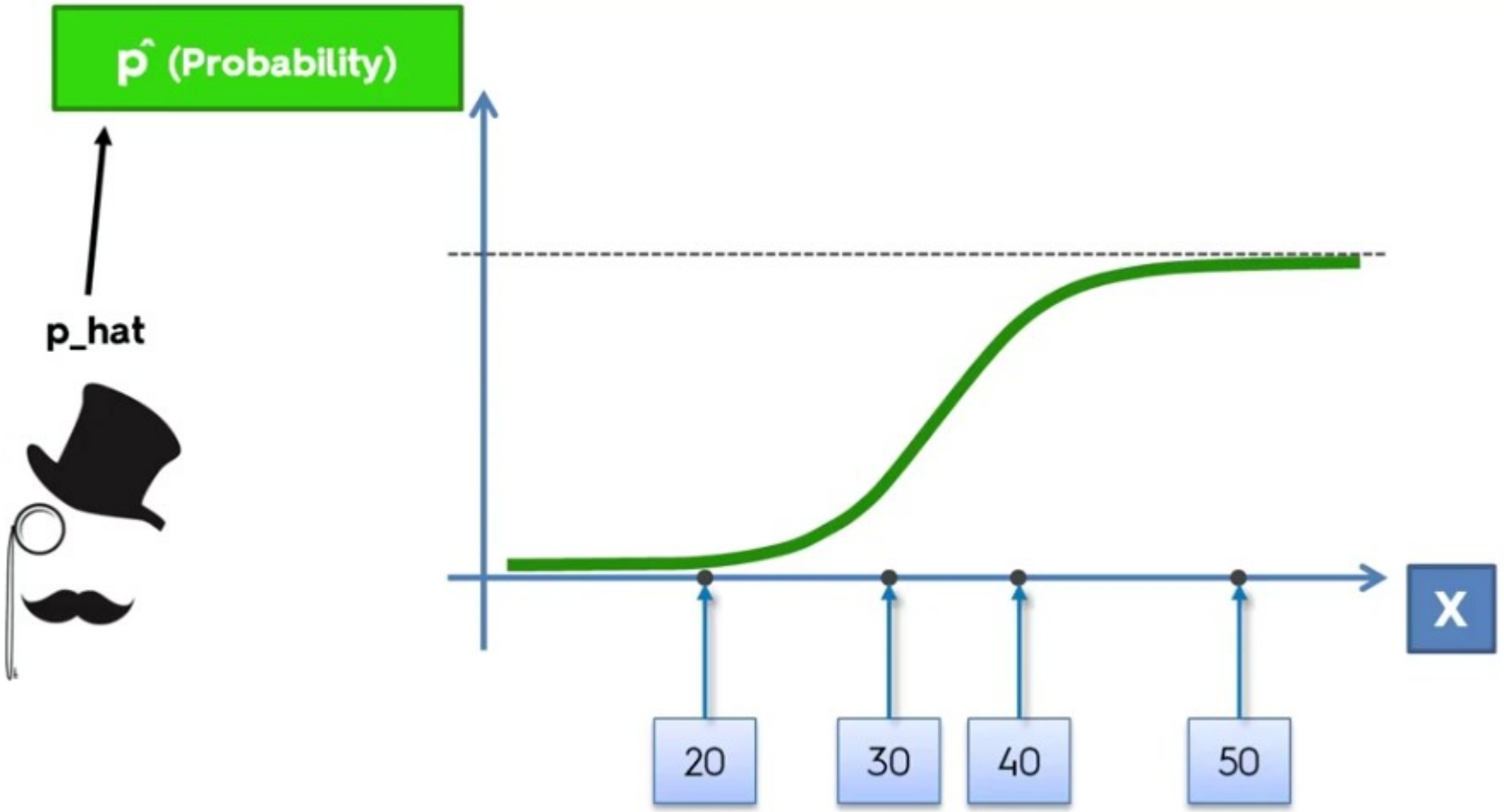
$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 * x$$



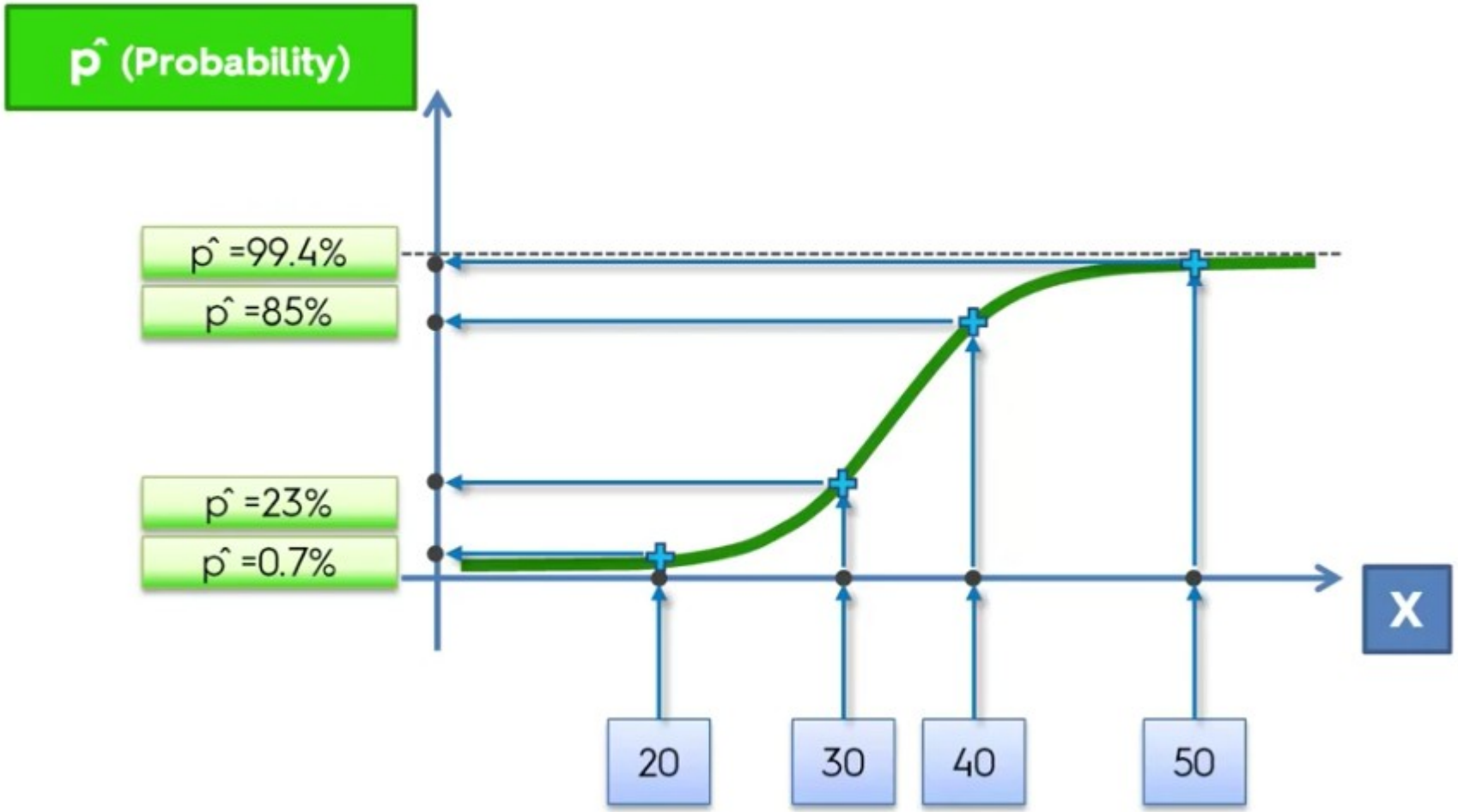
Logistic Regression – Logit Function



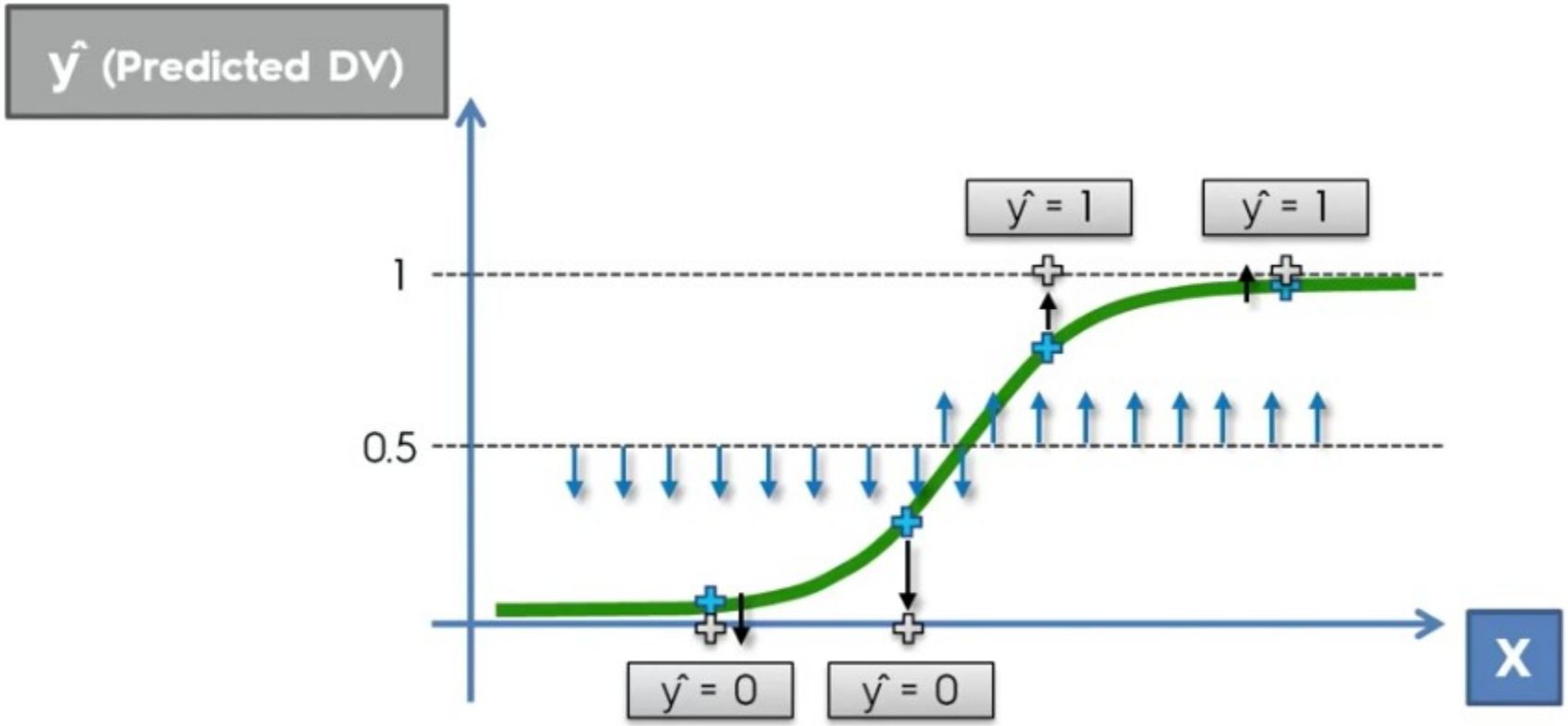
Logistic Regression



Logistic Regression – Probabilities



Logistic Regression – Prediction



Advantages

- Logistic Regression performs well when the dataset is linearly separable.
- Logistic regression is less prone to over-fitting but it can overfit in high dimensional datasets. You should consider Regularization (L1 and L2) techniques to avoid over-fitting in these scenarios.
- Logistic Regression not only gives a measure of how relevant a predictor (coefficient size) is, but also its direction of association (positive or negative).
- Logistic regression is easier to implement, interpret and very efficient to train.

Disadvantages

- Main limitation of Logistic Regression is the assumption of linearity between the dependent variable and the independent variables. In the real world, the data is rarely linearly separable. Most of the time data would be a jumbled mess.
- If the number of observations are lesser than the number of features, Logistic Regression should not be used, otherwise it may lead to overfit.
- Logistic Regression can only be used to predict discrete functions. Therefore, the dependent variable of Logistic Regression is restricted to the discrete number set.

Useful resources

- www.superdatascience.com
- www.mitu.co.in
- www.pythonprogramminglanguage.com
- www.scikit-learn.org
- www.towardsdatascience.com
- www.medium.com
- www.analyticsvidhya.com
- www.kaggle.com
- www.stephacking.com
- www.github.com

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com