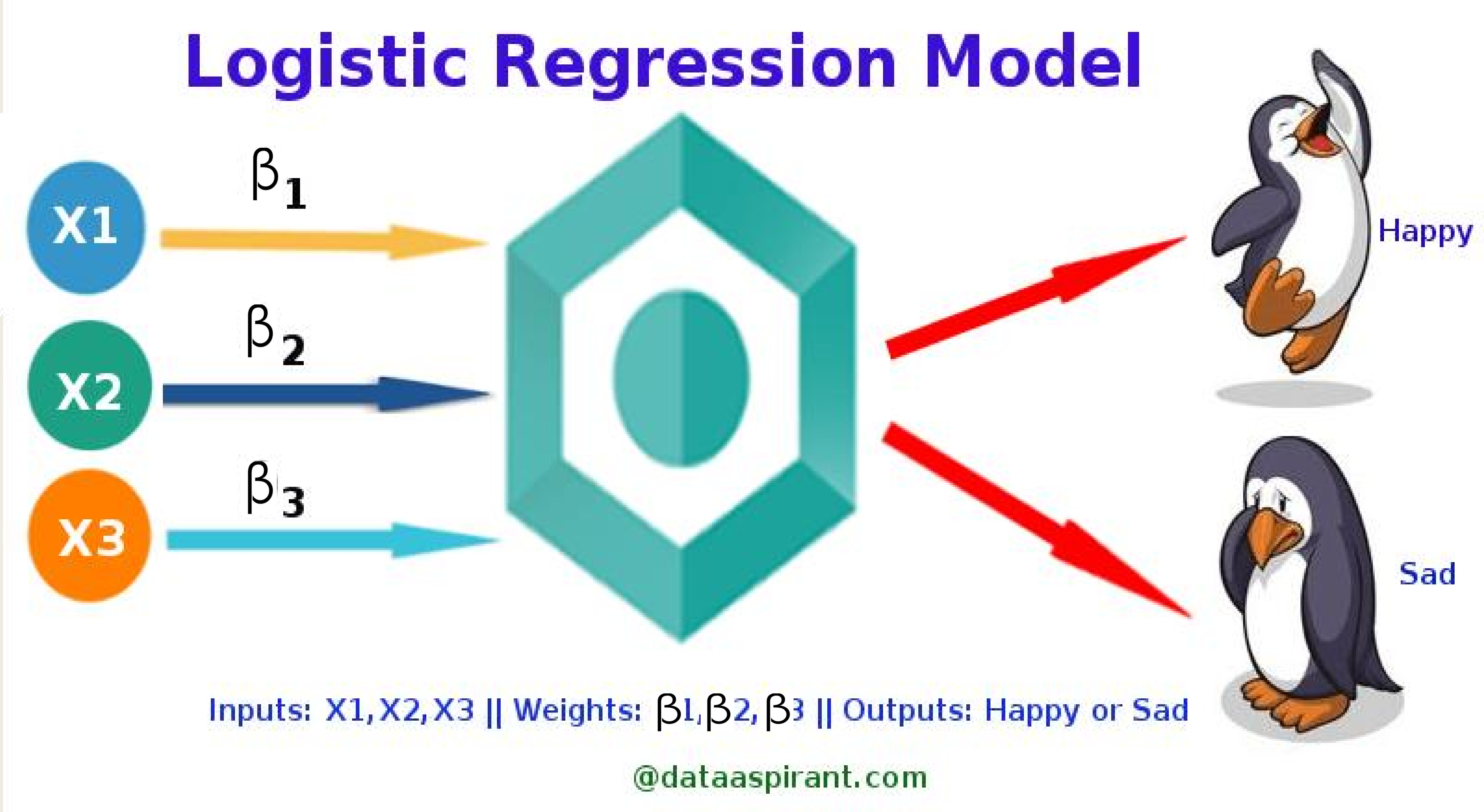
ML

Logistic Regression



What is Logistic Regression?

- Logistic regression is the appropriate regression analysis to conduct when the **dependent variable is dichotomous (binary).**

- Like all regression analyses, **logistic regression is a predictive analysis.**

- It 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.**

- This regression technique is similar to linear regression and can be used to predict the **Probabilities for classification problems.**

Types of Logistic Regression

- **Binary logistic regression**
- It is used to **predict the probability of a binary outcome, such as yes or no, true or false, or 0 or 1.** For example, it could be used to predict whether a patient has a disease or not, or whether a loan will be repaid or not.

- **Multinomial logistic regression**
- It is used to **predict the probability of one of three or more possible outcomes,** such as the type of product a customer will buy, the rating a customer will give a product, or the political party a person will vote for.

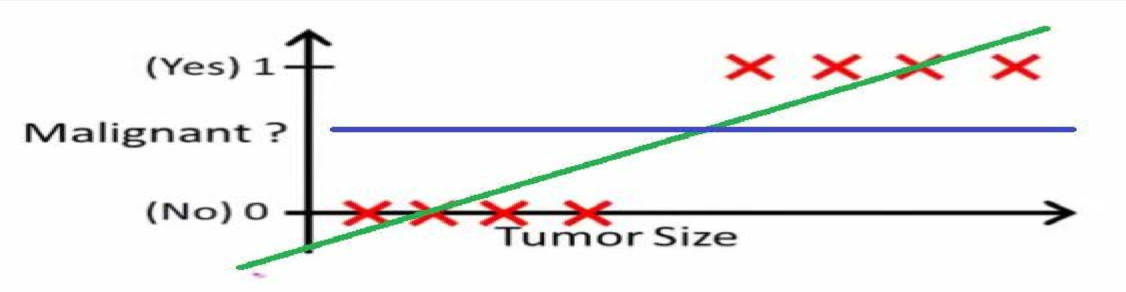
- **Ordinal logistic regression**
- It is used to **predict the probability of an outcome that falls into a predetermined order,** such as the level of customer satisfaction, the severity of a disease, or the stage of cancer.

Why do we use Logistic Regression rather than Linear Regression?

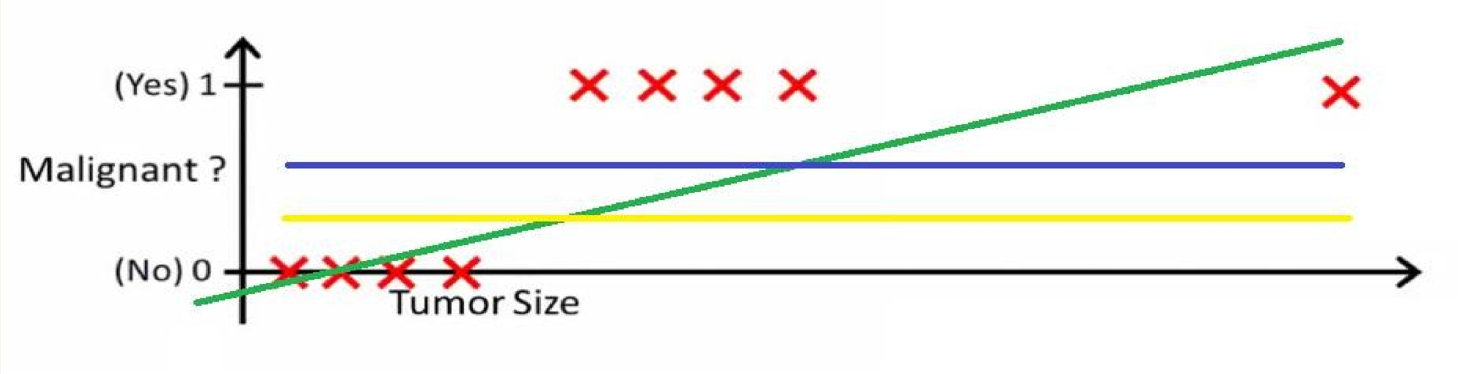
- **<u>Reason 1: -</u> Logistic Regression** is only used when our **dependent variable is binary** and in linear regression this **dependent variable is continuous.**

- **<u>Reason 2: -</u>** If we add an outlier in our dataset, the best fit line in linear regression shifts to fit that point.

- If we use linear regression to find the best fit line which aims at minimizing the distance between the predicted value and actual value, the line will be like this:



- Here the threshold value is 0.5, which means if the value of h(x) is greater than 0.5 then we predict malignant tumor (1) and if it is less than 0.5 then we predict benign tumor (0). Everything seems okay here but now let's change it a bit, we add some outliers in our dataset, now this best fit line will shift to that point. Hence the line will be somewhat like this:



Why do we use Logistic Regression rather than Linear Regression?

- The blue line represents the old threshold and the yellow line represents the new threshold which is maybe 0.2 here.
- To keep our predictions right we had to lower our threshold value.
- Hence we can say that linear regression is prone to outliers.
- Now here if h(x) is greater than 0.2 then only this regression will give correct outputs.
- Another problem with linear regression is that the predicted values may be out of range.
- We know that probability can be between 0 and 1, but if we use linear regression this probability may exceed 1 or go below 0.
- To overcome these problems we use Logistic Regression, **which converts this straight best fit line in linear regression to an S-curve using the sigmoid function, which will always give values between 0 and 1**.

Logistic Function

- Let’s start by mentioning the formula of logistic function:

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

- We all know the equation of the best fit line in linear regression is:

$$y = \beta_0 + \beta_1 x$$

- Let’s say instead of y we are taking probabilities (P). But there is an issue here, the value of (P) will exceed 1 or go below 0 and we know that range of Probability is (0-1). To overcome this issue we take “odds” of P:

$$\frac{P}{1 - P} = \beta_0 + \beta_1 x$$

- We know that odds can always be positive which means the range will always be (0,+∞). **Odds are nothing but the ratio of the probability of success and probability of failure.**

Logistic Function

- The problem here is that the range is restricted and we don't want a restricted range because if we do so then our correlation will decrease. By restricting the range we are actually decreasing the number of data points and of course, if we decrease our data points, our correlation will decrease. It is difficult to model a variable that has a restricted range. To control this we take the log of odds which has a range from (-∞,+∞).

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x$$

- Now we just want a function of P because we want to predict probability. To do so we will multiply by exponent on both sides and then solve for P.

$$\exp[\log(\frac{p}{1-p})] = \exp(\beta_0 + \beta_1 x)$$

$$e^{\ln[\frac{p}{1-p}]} = e^{(\beta_0 + \beta_1 x)}$$

$$\frac{p}{1-p} = e^{\left(\beta_0 + \beta_1 x\right)}$$

$$p = e^{\left(\beta_0 + \beta_1 x\right)} - pe^{\left(\beta_0 + \beta_1 x\right)}$$

Logistic Function

$$p = p\left[\frac{e^{\left(\beta_0 + \beta_1 x\right)}}{p} - e^{\left(\beta_0 + \beta_1 x\right)}\right]$$

$$1 = \frac{e^{\left(\beta_0 + \beta_1 x\right)}}{p} - e^{\left(\beta_0 + \beta_1 x\right)}$$

$$p\left[1 + e^{\left(\beta_0 + \beta_1 x\right)}\right] = e^{\left(\beta_0 + \beta_1 x\right)}$$

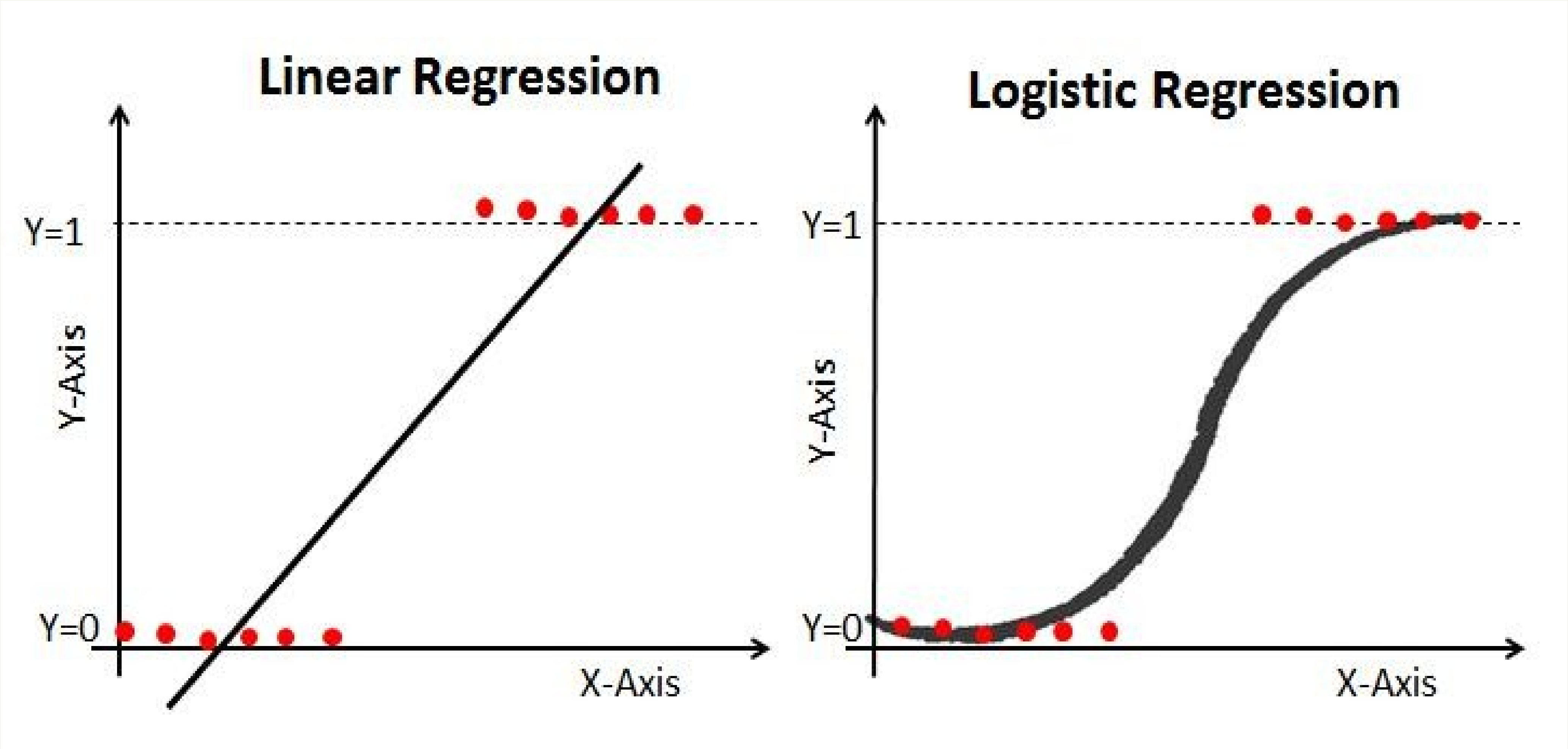
$$p = \frac{e^{\left(\beta_0 + \beta_1 x\right)}}{1 + e^{\left(\beta_0 + \beta_1 x\right)}}$$

$Now\ dividing\ by\ e^{\left(\beta_0 + \beta_1 x\right)},\ we\ will\ get$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad This\ is\ our\ sigmoid\ function.$$

Logistic Function

- Now we have our logistic function, also called a **sigmoid function.**
- The graph of a sigmoid function is as shown below. It squeezes a straight line into an S-curve.



Cost Function in Logistic Regression

- In linear regression, we use the Mean squared error as the cost function which can be shown as:

- In logistic regression Y_i is a non-linear function ($\hat{Y}=1/1+e^{-z}$). If we use this in the above MSE equation then it will give a non-convex graph with many local minima as shown

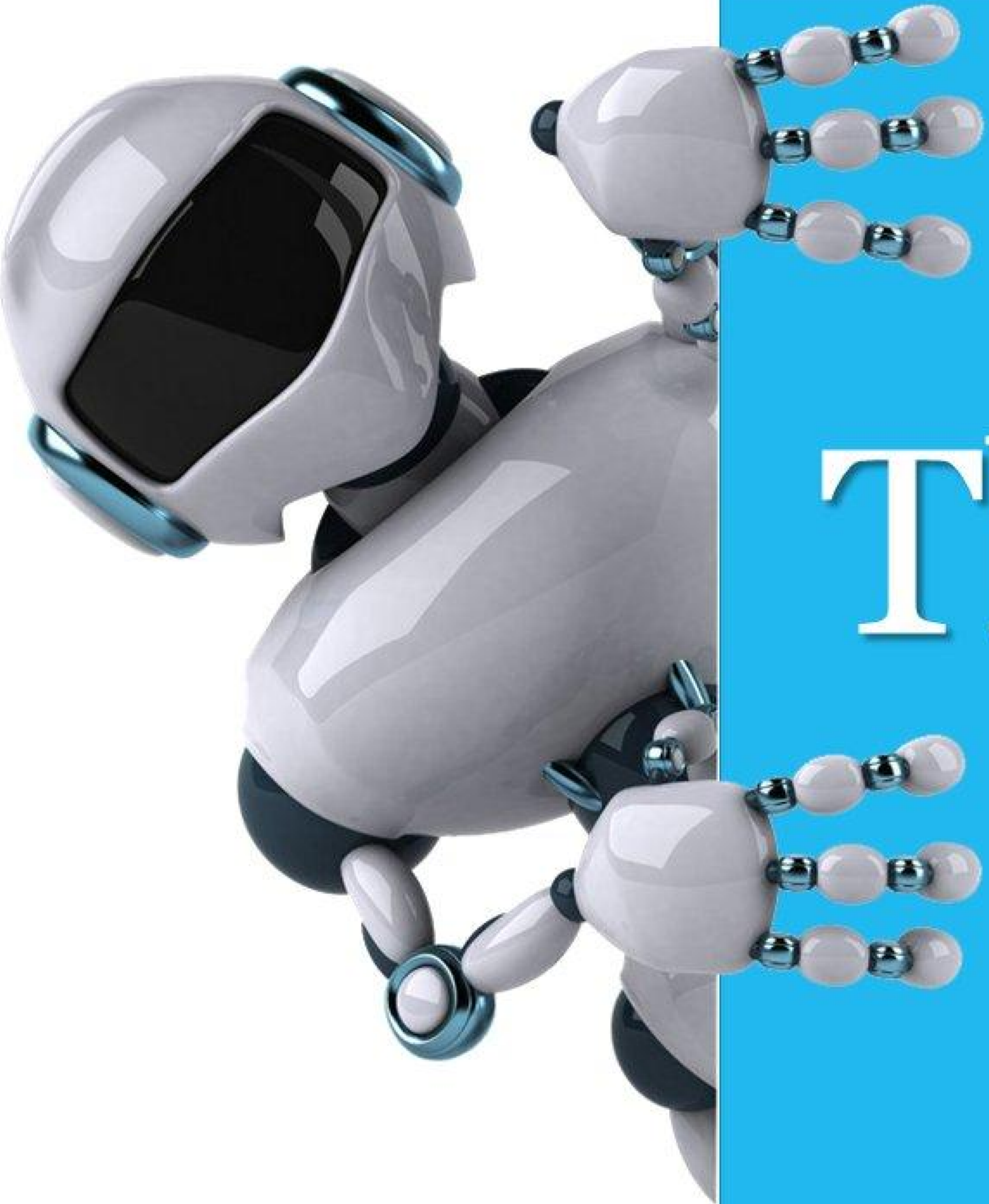
Cost Function in Logistic Regression

- The problem here is that this cost function will give results with local minima, which is a big problem because then we'll miss out on our global minima and our error will increase.
- In order to solve this problem, we derive a different cost function for logistic regression called **Log loss** which is derived from the maximum likelihood estimation method.

$$\text{Log loss} = \frac{1}{N} \sum_{i=1}^{N} -(y_i * \log(\hat{Y}_i) + (1 - y_i) * \log(1 - \hat{Y}_i))$$

- Above cost function can be solved using gradient descent as below:

$$\beta_{new} = \beta_{old} - \alpha\left[\sigma\left(\beta^T x\right) - y\right].x_j$$



Thank you