Logistic Regression-1

Why do we need Logistic Regression?

- Useful for binary classification

What are the assumptions of Logistic regression?

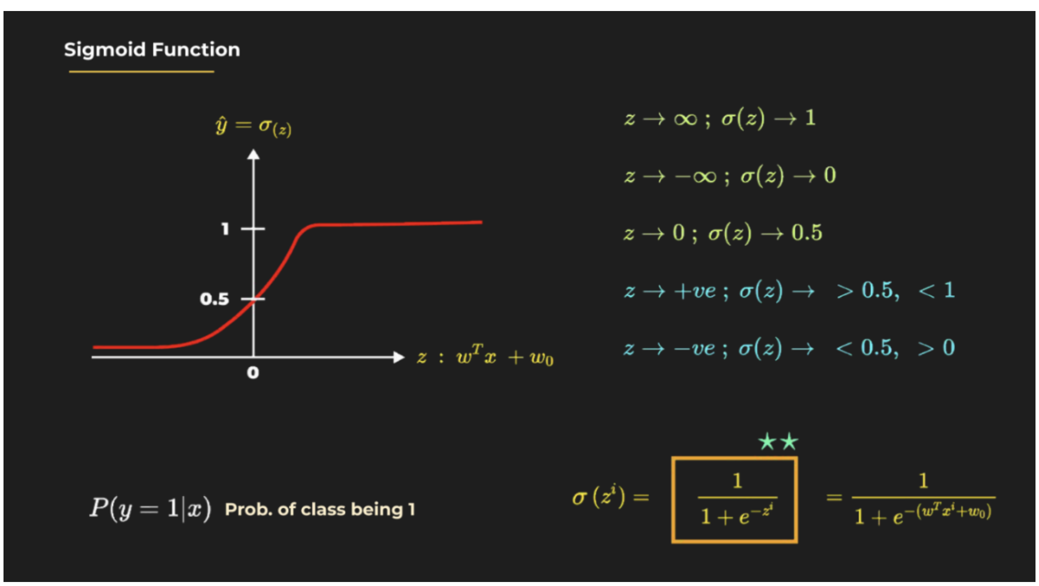
- Data should be linearly separable

What is the goal of this algorithm?

- To find a hyperplane π which accurately separates the data

How to perform prediction through logistic regression?

- Given labels -> $y_i \in \{0, 1\}$
- Compute a linear function of x -> $z_i = w^t x_i + w_0 \in \{-\infty, \infty\}$
- Compute Sigmoid(z_i) -> $\sigma(z_i) = \frac{1}{1+e^{-z_i}}$
- Predicted label -> $\hat{y}_i = 1\ if\ \sigma(z_i) > threshold\ else\ 0$



What are the properties of Sigmoid function?

- **Range** -> (0, 1)
- Smooth and differentiable at all points

What is the derivative of a logistic regression model?

$$\sigma'(z) = \sigma(z) \times [1 - \sigma(z)]$$

What is the significance of the sigmoid function?

- $\sigma(z_i)$ is the probability of x_i belonging to class 1

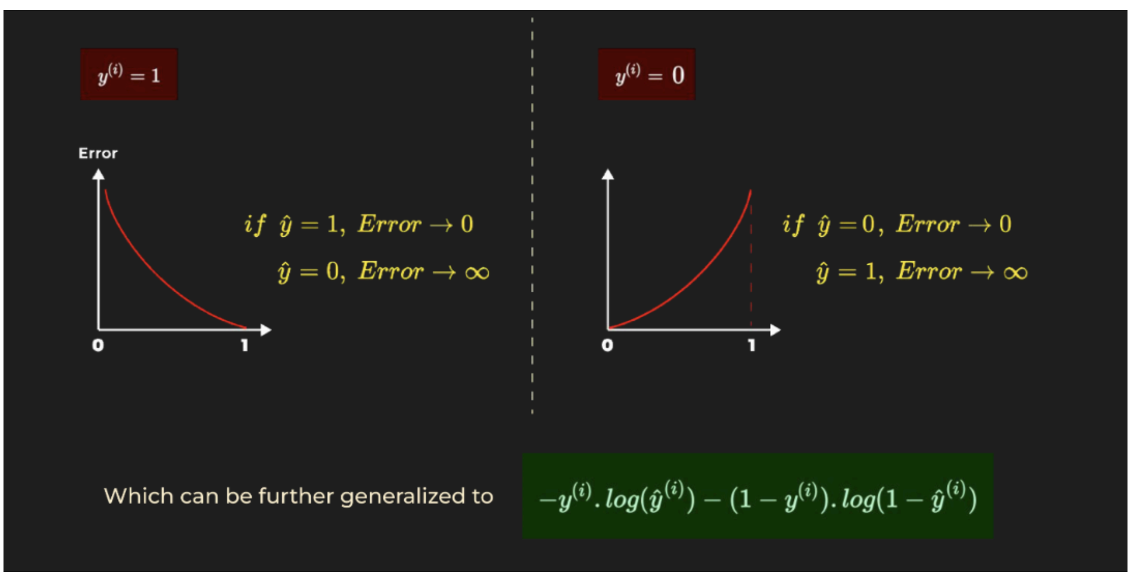
Which loss function is used for Logistic Regression?

- **Log loss ->** Combination of:
    - $-\log(\hat{y}_i)$ when $y_i = 1$
    - $-\log(1 - \hat{y}_i)$ when $y_i = 0$.

$$Log\ Loss = -y_i \log(\hat{y}_i) - (1 - y_i)\log(1 - \hat{y}_i)$$

Total loss becomes:

$$TotalLoss = Log\ Loss + \lambda \sum_{j=1}^{d} w_j^2$$



How does log loss help train the logistic regression model?

- $-\log(\hat{y}_i)$ -> High when $\hat{y}_i = 0$ and very low when $\hat{y}_i = 1$.
- $-\log(1-\hat{y}_i)$ ->High when $\hat{y}_i = 1$ and very low when $\hat{y}_i = 0.$

Thus both of these components help penalize the model most when making wrong predictions.

Also if $y_i \in \{-1, 1\}$, then $Log\ Loss = \sum_{i=1}^{n} log(1 + e^{-y_i z_i})$

But why can't we use Mean Square Error as in Linear regression?

- Non-convex curve: contains a lot of local minima
- Difficult for Gradient Descent to reach global minima

What does the derivative of Log Loss look like?

$$\frac{\partial Log\ Loss}{\partial w} = \frac{\partial(-y\log(\hat{y}) - (1-y)\log(1-\hat{y}))}{\partial w},\ \hat{y} = \sigma(z)\ and\ z = w^t x + w_0$$

Intermediate steps: [reference blog]

On solving:

$$\frac{\partial Log\ Loss}{\partial w} = -y\frac{\partial z}{\partial w} + y \times \hat{y}\frac{\partial z}{\partial w} + \hat{y}\frac{\partial z}{\partial w} - y \times \hat{y}\frac{\partial z}{\partial w}$$

$$\frac{\partial Log\ Loss}{\partial w} = -y\frac{\partial z}{\partial w} + \hat{y}\frac{\partial z}{\partial w}$$

Also since $\frac{\partial z}{\partial w} = x$; $\frac{\partial Log\ Loss}{\partial w} = (\hat{y} - y)\,x$

Can we use R2 score to evaluate Logistic Regression ?

Ans: No. R2 tells us how close we are far/close we are to actual result

Which metric can we use instead? → accuracy

Goal: Check how many correct values are predicted. → satisfied by **accuracy**

$$\text{Accuracy} = \frac{\text{no. of correct predictions}}{\text{Total number of predictions}}$$