➔ What is Logistic Regression ?

It is a supervised ML algorithm used for classification tasks. Despite its name, it is actually a classification problem not a regression.

e.g:

email apam or not spam

disease or no disease

- ## Hypothesis :

The hypothesis depends on our problem but mostly we use sigmoid function.

Its graph like S shape curve

➢ Sigmoid function :

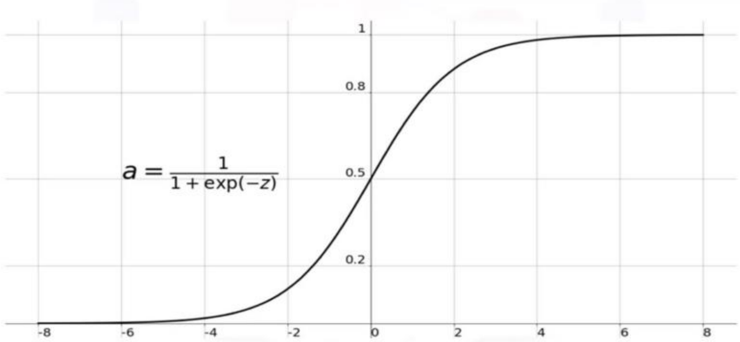
Hypothesis = 1/1+e^(-z)

Where,

e = exponent

z = weights.T*X

Sigmoid Function



➔Why we use sigmoid in logistic regression ?

It convert the linear output into probability

Helps the model make binary decision by adding threshold (usually 0.5)

➔ Decision boundary :

When the output is mapped to 0 and 1 based on a threshold 0.5.

If the output is greater than 0.5, then it is classify is a positive class (1)

If the output in less than 0.5, then it consider as negative class (0)

➢ What is the cost function in logistic regression :

In logistic regression we use log loss function as a cost function which measure the error between predicted probability and actual labels.

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$$

$$J(\theta) = \frac{1}{m}[\sum_{i=1}^{m} -y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))]$$

$$m = number\ of\ samples$$

➔ Why we use log loss :

- Non-linear relationships :
  MSE is not suitable because the sigmoid is non linear.
- Better gradient behavior :
  Faster convergence
- Probability interpretation :
  Make the output more natural for classification

Optimization:

➢ Gradient Descent :

It is an optimization technique use to minimize the cost function in machine learning models including logistic regression. It iteratively adjust the model parameters (W and b) to find the best set of parameter that minimize the cost function.

$$J(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\text{Cost}(h_\theta(x), y)$$

$$\boxed{\text{Cost}(h_\theta(\mathbf{x}), y) = -ylog(h_\theta(\mathbf{x})) - (1-y)log(1-h_\theta(\mathbf{x}))}$$

<u>Gradient descent for logistic regression:</u>

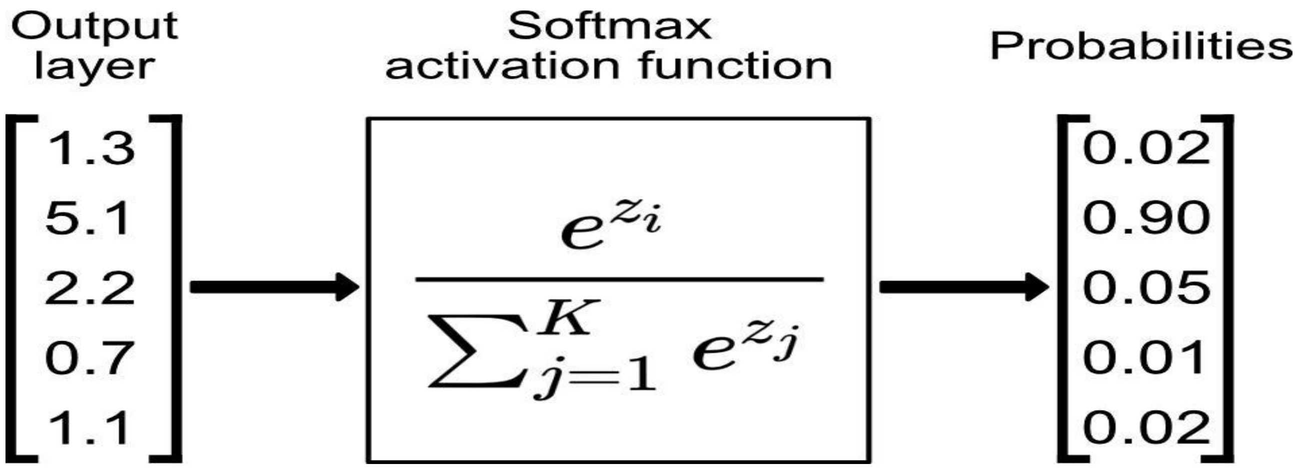
while not converged {

$$\theta_j^{\text{new}} = \theta_j^{\text{old}} - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \text{ for } j = 0, 1, \ldots, n$$

}

➢ Softmax regression :

This is use for multi class classification.
It same like sigmoid but there the probability output classify two classes and here it will classify two or more classes.
In softmax we use loss function called categorical cross entropy.
It divede the output into different categories depends on no of classes.

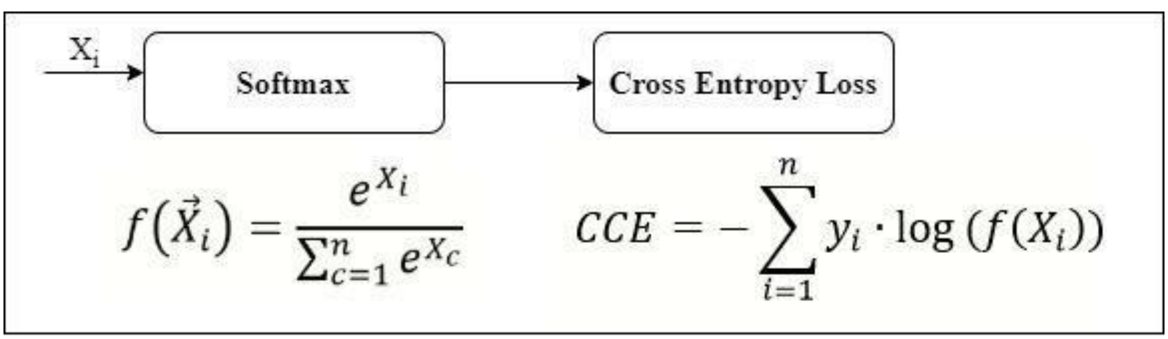


Where,

Z = wT*X

K = no of classes.

Graphical view of categorical cross entropy and softmax



We use these activation function in deep learning mostly the concept of perceptron is based on logistic regression. (Next lecture is deep learning)

End of lecture

Thank you