

Logistic Regression

Logistic Regression is a supervised learning algorithm.
Logistic Regression is of Multinomial(Multiple Outcomes) and Binary(2 Outcomes).

Basic Idea :

Logistic Regression is similar to Linear regression such that the Probability is scaled down to (0 to 1). A threshold value is set and the outcome is classified with reference to the threshold.

Hypothesis :

The function Z is similar to the Hypothesis of Linear Regression which is,

$$Z = \sum_{i=0}^N (\theta_i)(X_i) \quad (\text{Where, } X_0 = 1)$$

θ – Weight/Parameter/Regression Coefficients

N – Number of Features/Inputs

[Note : H has N + 1 terms where θ_0 is Y-Intercept]

Logistic(Sigmoid) Function :

Here the probability/Hypothesis (H) varies from $-\infty$ to ∞ so sigmoid function (S) is used in order to limit the Values between 0 to 1.

$$S(x) = \frac{1}{1+e^{-x}}$$

$$H \text{ (Hypothesis)} = \frac{1}{1+e^{-Z}}$$

Cost Function :

The Cost function (J) here is,

$$J = \begin{cases} -\text{Log}(H) & \text{for } Y = 1 \\ -\text{Log}(1 - H) & \text{for } Y = 0 \end{cases}$$

Which can be brought into single function as,

$$J = \left(\frac{-1}{m}\right)(Y(\text{Log}(H)) + (1 - Y)(\text{Log}(1 - H)))$$

Gradient Descent :

Gradient Descent is used as an Optimisation method to reduce the value of Cost function (J).

Basic Idea behind the Gradient Descent algorithm is to reach the Global minimum of the Cost Function.

Let $Y(\text{Log}(H)) = P$ and $Q = (1 - Y)(\text{Log}(1 - H))$

Derivative of P,

$$\frac{\partial P}{\partial \theta} = \left(\frac{Y}{H}\right)\left(\frac{\partial H}{\partial \theta}\right) = \left(\frac{Y}{H}\right)\left(\frac{\partial H}{\partial Z}\right)\left(\frac{\partial Z}{\partial \theta}\right)$$

$$\frac{\partial P}{\partial \theta} = \left(\frac{Y}{H}\right)H(1 - H)X \quad (\text{Since } \frac{\partial H}{\partial Z} = H(1 - H) \text{ and } \frac{\partial Z}{\partial \theta} = X)$$

$$\frac{\partial P}{\partial \theta} = Y(1 - H)X$$

Derivative of Q,

$$\frac{\partial P}{\partial \theta} = \left(\frac{1 - Y}{1 - H}\right)\left(-\frac{\partial H}{\partial \theta}\right) = \left(\frac{1 - Y}{1 - H}\right)\left(-\frac{\partial H}{\partial Z}\right)\left(\frac{\partial Z}{\partial \theta}\right)$$

$$\frac{\partial P}{\partial \theta} = -\left(\frac{1 - Y}{1 - H}\right)H(1 - H)X$$

$$\frac{\partial P}{\partial \theta} = -H(1 - Y)X$$

On adding the derivatives of P and Q,

$$\frac{\partial J}{\partial \theta} = \left(\frac{-1}{m}\right)(Y - H)X$$

Here the Updation of θ is same as the Linear Regression.

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j}(J)$$

Question :

1) Why Mean Squared error can't be used as Cost function in Logistic Regression?

Mean Square Error can't be used for Logistic Regression as the Function is Non-Convex for Logistic Regression. Which has multiple Local minimum.

2) How is Learning Rate α related with the accuracy of the model?

Lower the Learning Rate the higher the accuracy of the model (This relation holds true till specific value of α) as having higher value of α might skip the Global Minimum.

3) Is the Cost Function (J) Convex or Non-Convex?

The Cost Function here is Convex such that it has single Global minimum.

4) How do we use Logistic regression for Multinomial?

One of the technique is to split Multi-Class Classification into multiple Binary Classification.

5) Is there any other algorithm that can be used instead of Gradient Descent to Minimise the Cost Function?

Alternatives for Gradient Descent are Evolutionary Algorithm (EA) and Particle Swarm Optimisation (PSO).