# DHANISH AHAMMED M V

**MS21218**


**ASSIGNMENT 02**

## QUESTION 3

Q3. ***Create a report investigating how different values of n and ϑ impact the ability for your logistic regression function to learn the coefficients, β, used to generate the output vector Y. Also include your derivation of the partial derivative of the cost function with respect to the parameters of the model.***

**Answer: -**

The output vector Y, which denotes the predicted class label for each input sample, is generated by the logistic regression function using the coefficients, that are learned during training. The values of n and, as well as other variables, affect the logistic regression function's capacity to learn the coefficients.

**n - size of data sets**

The number of training samples, denoted by n, is proportional to the amount of data available for training the model. A larger n value provides more data for the model to learn from, which can lead to improved model performance. A larger value of n, on the other hand, increases the computational cost of training the model.

**Θ - Threshold**

The value of determines the prediction threshold in logistic regression and influences the balance of precision and recall. Lowering it leads to more false positives, while increasing it leads to more false negatives. Thus, choosing the appropriate value of is critical in achieving the desired balance of precision and recall.

**Partial Derivative of Cost Function:**

3rd Question's Answer:

For logistic regression; the least squared errors result in a non-convex loss function with local minimum. Inorder to maintain the convex nature for loss function, a log loss error function can be created. The cost function is split into two cases:
$y=1$ and $y=0$

$$\Rightarrow Cos(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0 \end{cases}$$

$$\Rightarrow Cos(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

$\Rightarrow$ Combined Cost function

Derivative of cost function:
↳ Hypothesis function is sigmoid.
↳ Gradient of sigmoid functions :-

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

Now: ($\theta$) In terms of partial derivatives

$$\frac{d(K(\theta))}{\partial(\theta_k)} = \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)} \times \frac{1}{h_\theta(x^{(i)})} \times \frac{\partial(h_\theta(x^{(i)}))}{\partial(\theta_j)}\right]$$

$$+ \sum_{i=1}^{m}\left[(1-y^{i}) \times \frac{1}{(1-h_\theta(x^{(i)}))} \times \frac{\partial(1-h_\theta(x^{(i)}))}{\partial(\theta_k)}\right] \quad \text{—— (1)}$$

Evaluting the derivative of sigmoid function:-

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

$$\frac{\partial(\sigma(x))}{\partial x} = \frac{0 \times (1+e^{-x}) - 1 \times (e^{-x} \times (-1))}{(1+e^{-x})^2}$$

$$\frac{\partial(\sigma(x))}{\partial x} = \frac{(e^{-x})}{(1+e^{-x})^2} = \frac{1-1+(e^{-x})}{(1+e^{-x})^2}$$

$$= \frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2} = \sigma(x)(1-\sigma(x)) \quad \text{— (2)}$$

Applying (2) in (1):

$$\frac{\partial(K(\theta))}{\partial(\theta_k)} = \frac{-1}{m}\sum_{i=1}^{m}\left[y^i \times \frac{1}{h_\theta(x^i)} \times \sigma(z)(1-\sigma(z)) \times \frac{\partial(\theta^T x)}{\partial(\theta_k)}\right]$$

$$+ \sum_{i=1}^{m}\left[(1-y^i) \times \frac{1}{1-h_\theta(x^i)} \times -\sigma(z)(1-\sigma(z)) \times \frac{\partial(\theta^T x)}{\partial \theta_k}\right]$$

$$= \frac{-1}{m}\left[\sum_{i=1}^{m}\left[y^i \times \frac{1}{h_\theta(x^i)} h_\theta(x^i)(1-h_\theta(x^i)) \times x^i_k\right] + \right.$$

$$\left.\sum_{i=1}^{m}\left[(i-y) \times \frac{1}{1-h_\theta(x^i)} \times -h_\theta(x^i)(1-h_\theta(x^i)) \times x^i_k\right]\right.$$

$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y^i \times (1-h_\theta(x^i)) \times x^i_k - (1-y^i) \times h_\theta(x^i) \times x^i_z\right.$$

Upon simplification:

$$\Rightarrow \frac{\partial(K(\theta))}{\partial(\theta_k)} = \frac{-1}{m}\left(\sum_{i=1}^{m}\left[y^i - h_\theta(x^i)\right] x^i_k\right]$$

Removing the summation term: (convert to matrix form)

$$\therefore \frac{\partial(K(\theta))}{\partial(\theta)} = \frac{1}{m} \times X^T \left[h_\theta(x) - y\right]$$