Page No.

Date

Q-3 Find the derivative of Negative log Likelihood Function formulated for performing binary logistic regression:

$$-\log_e L(\theta_0, \hat{\theta}) = -\sum_{i=1}^{N_{train}} \left[y_i \log_e \left(\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}} \right) + (1-y_i) \log_e \left(1 - \frac{1}{1+(e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}} \right) \right]$$

Where, $\hat{\theta}_0$ is scalar and $\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$

→ Sigmoid function $6(z) = \frac{1}{1+e^{-z}}$

Where $z = (\hat{\theta}_0 + \hat{\theta}^T x_i)$

We have Negative log likelihood function for binary logistic regression:

$$-\log L(\theta_0, \hat{\theta}) = -\sum_{i=1}^{N_{train}} \left[y_i \log_e \left(\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i}} \right) + (1-y_i) \log_e \left(1 - \frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}} \right) \right]$$

$$= -\sum_{i=1}^{N_{train}} \left[y_i \log_e 6(\theta^T x_i) + (1-y_i) \log_e [1 - 6(\theta^T x_i)] \right]$$

Date

Now we have function for negative log likelihood we need to chose the value of theta that minimize it.

To minimize the Negative log likelihood function, we have to use gradient descent Algorithm

let's take derivative with respect to each theta-hat

$$-\frac{\partial}{\partial\hat{\theta}} \log_e L(\theta_0, \hat{\theta}) = -\frac{\partial}{\partial\hat{\theta}} y \log \sigma(\hat{\theta}^T x_i) + \frac{\partial}{\partial\hat{\theta}} (1-y) \log[1-\sigma(\hat{\theta}^T x_i)]$$

$$= \left[\frac{y}{\sigma(\hat{\theta}^T x_i)} - \frac{1-y}{1-\sigma(\hat{\theta}^T x_i)}\right] \frac{\partial}{\partial\hat{\theta}} \sigma(\hat{\theta}^T x_i)$$

$$= \left[\frac{y}{\sigma(\hat{\theta}^T x_i)} - \frac{1-y}{1-\sigma(\hat{\theta}^T x_i)}\right] \sigma(\hat{\theta}^T x_i) \cdot [1-\sigma(\hat{\theta}^T x_i)] x_i$$

$$= \left[\frac{y-\sigma(\hat{\theta}^T x_i)}{\sigma(\hat{\theta}^T x)(1-\sigma(\hat{\theta}^T x_i))}\right] \sigma(\hat{\theta}^T x_i)[1-\sigma(\hat{\theta}^T x_i)] x_i$$

$$= [y - \sigma(\hat{\theta}^T x_i)] x_i$$

where $\hat{\theta}$ is a $\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$ matrix.