Que-3 $-\log_e L(\theta_0, \theta) =$

$$-\sum_{i=1}^{N}\left[y_i \log_e\left(\frac{1}{1+e^{-(\theta_0+\theta^T x_i)}}\right) + (1-y_i)\log\left(1-\frac{1}{1+e^{-(\theta_0+\theta^T x)}}\right)\right]$$

Suppose $\frac{1}{1+e^{-(\theta_0+\theta^T x)}} = \sigma$

$$\frac{d\sigma}{d\theta} = \frac{1}{1+e^{-(\theta_0+\theta^T x)}}\cdot\theta^T x$$

$$\frac{d\sigma}{d\theta} = \sigma(\theta^T x)$$

$$\frac{-\log_e L(\theta_0, \theta)}{d\theta} = -\sum_{i=1}^{N}\left[\frac{d}{d\sigma} y_i \log_e \sigma + \frac{d}{d\sigma}(1-y_i)\log(1-\sigma)\right]\frac{d\sigma}{d\theta}$$

$$= -\sum_{i=1}^{N}\left[\frac{y_i}{\sigma} + \frac{1-y_i}{1-\sigma}\right]\sigma(\theta^T x)$$

$$= -\sum_{i=1}^{N}\left[\frac{y_i(1+e^{-(\theta_0+\theta^T x)})}{1} + \frac{(1-y_i)(1+e^{-(\theta_0+\theta^T x)})}{-e^{-(\theta_0+\theta^T x)}}\right]$$

$$\sigma(\theta^T x)$$

$$= -\sum_{i=1}^{N}\left[\frac{y_i(1+e^{-(\theta_0+\theta^T x)})^2 + (1-y_i)(1+e^{-\theta-\theta_0}}{e^{-(\theta_0+\theta^T x)}}\right.$$

$$+ \frac{d}{d\theta}\frac{1}{1+e^{-(\theta_0+\theta^T x)}}\theta^T x$$

Date : / / 20

Que-3

$$= -\sum_{i=1}^{N} \left[1 + e^{-(\hat{\theta} + \theta^T x)} \left[y_i (1 + e^{-(\hat{\theta} + \theta^T x)}) - (1 - y_i) \right] \right.$$

$$\left. + (1 + e^{-(\hat{\theta} + \theta^T x)})^{-2} (-x) \right]$$

$$= -\sum_{i=1}^{N} \left[1 + e^{-(\hat{\theta} + \theta^T x)} (y_i (1 + e^{-(\hat{\theta} + \theta^T x)}) - (1 - y_i)) \right]$$

$$- x_i (1 + e^{-(\hat{\theta} + \theta^T x)})^{-2}$$

$$= -\sum_{i=1}^{N} (1 + e^{-(\hat{\theta} + \theta^T x)})^{-1} (y_i (1 + e^{-(\hat{\theta} + \theta^T x)}) - (1 - y_i))$$

$$\frac{-\log \theta}{d\theta} = -\sum_{i=1}^{N} \frac{y_i (1 + e^{-(\hat{\theta} + \theta^T x)}) - (1 - y_i)}{1 + e^{-(\hat{\theta} + \theta^T x)}} \cdot x_i$$

Que-4

Negative logliklihood Function For multiclass logistic regression

$$-\log_e L(\hat{\theta}_0, \hat{\theta}) = -\sum_{i=1}^{N} \sum_{j=0}^{m} c_i^j \log_e \left(\frac{e^{(\theta_0^j + \hat{\theta}^{jT} x_i)}}{\sum_{k=0}^{m} e^{(\theta_0^k + \hat{\theta}^{kT} x_i)}} \right)$$

categories $= \theta = 1000$
features vector $x_i = 512$
$N = 568454 \times 0.75 = 426340$

$$-\log_e L(\hat{\theta}_0 . \hat{\theta}) = -\sum_{i=1}^{N} \sum_{j=0}^{999} c_i^j \log_e \left(\frac{e^{(\theta_0^j + \hat{\theta}^{jT} x_i)}}{\sum_{k=0}^{999} e^{(\theta_0^k + \hat{\theta}^{kT} x_i)}} \right)$$

where $N = 426340$

$$x_i = \begin{bmatrix} x_1^i \\ x_2^i \\ \vdots \\ x_{512}^i \end{bmatrix}_{512 \times 1}$$

$\hat{\theta}_0$ is row vector

$$\theta_0 = [\hat{\theta}_0^0, \theta_0^1 \cdots \hat{\theta}_0^{999}]$$

$$\hat{\theta}^* = [\hat{\theta}^0, \hat{\theta}^1 \ldots \hat{\theta}^{999}]_{512 \times 1000}$$

where

$$\hat{\theta}^0 = \begin{bmatrix} \hat{\theta}_1^0 \\ \theta_2^0 \\ \vdots \\ \theta_{512}^0 \end{bmatrix}$$