Q4 Write the expression for Negative log likelihood function for performing Multiclass Logistic Regression on data which has 512 features, 568454 rows, out of which 75·1. will be used as training data and there are 1000 categories in the data is balanced. Also Mention the dimentions of $\vec{\theta}$ matrix as well as $\vec{\theta}_0$ matrix as well as dimentions of a matrix for ground truth labels.

→ Before training the data, we need to define loss function.

So, the Negative log likelihood function or loss function for multiclass logistic regression is,

$$-\log_e L(\vec{\theta}_0, \vec{\theta}) = -\sum_{i=1}^{N_{train}} \sum_{j=1}^{n} \log_e \left(\frac{e^{(\vec{\theta}_0^j + \vec{\theta}^{jT} x_i)}}{\sum_{K=1}^{n} e^{(\vec{\theta}_0^K + \vec{\theta}^{KT} x_i)}} \right) \quad \text{—(1)}$$

Now we are going to use 75·1. of data as training data.

75·1. of N → 75·1. of 568454 Rows
≈ 426340 Rows.

∴ N_{train} = 426340
Categories = Θ = 1000
Feature vector x_i = 512

∴ In eqn (1),

$$-\log_e L(\vec{\theta}_0, \vec{\theta}) = -\sum_{i=1}^{426340\,999} \sum_{j=0} C_j^i \log_e \left(\frac{e^{(\vec{\theta}_0 j + \vec{\theta} i^T x_i)}}{\sum_{K=0}^{999} e^{(\vec{\theta}_0 K + \vec{\theta} K^T x_i)}} \right)$$

Where $x_i = \begin{bmatrix} x_1^1 \\ x_2^1 \\ \vdots \\ x_{512}^1 \end{bmatrix}$ 512×1 matrix

$\vec{\theta}_0$ is a row vector which is equal to

$[\vec{\theta}_0^0, \vec{\theta}_0^1, \ldots \vec{\theta}_0^{999}]_{1 \times 1000}$

& $\vec{\theta} = [\vec{\theta}^0 \; \vec{\theta}^1 \; \vec{\theta}^2 \ldots \vec{\theta}^{999}]_{512 \times 1000}$

Where every θ is a vector component

i.e $\vec{\theta}^0, \vec{\theta}^1, \ldots \vec{\theta}^{999}$ is 512×1 matrix

& $\vec{\theta} = 512 \times 1000$ matrix