CSCI 567 Machine Learning

Homework #2

Name : Yi Zhao

Question 1.1 Answer :

$$\frac{\partial l}{\partial \boldsymbol{u}} = \frac{\partial l}{\partial \boldsymbol{a}} \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{h}} \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{u}}$$

$$\frac{\partial l}{\partial \boldsymbol{u}} = \boldsymbol{W}^{(2)\mathrm{T}} \cdot \frac{\partial l}{\partial \boldsymbol{a}} \cdot * H(\boldsymbol{u})$$

$$l = -\sum_k y_k \log \frac{e^{a_k}}{\sum_{k'} e^{a_{k'}}}$$

$$so, l = -\sum_k y_k a_k + \log \sum_{k'} e^{a_{k'}}$$

$$\frac{\partial l}{\partial \boldsymbol{a}} = \boldsymbol{z} - \boldsymbol{y}$$

$$\frac{\partial l}{\partial \boldsymbol{W}^{(1)}} = \frac{\partial l}{\partial \boldsymbol{u}} \cdot \boldsymbol{x}^{\mathrm{T}}$$

$$\frac{\partial l}{\partial \boldsymbol{b}^{(1)}} = \frac{\partial l}{\partial \boldsymbol{u}}$$

$$\frac{\partial l}{\partial \boldsymbol{W}^{(2)}} = \frac{\partial l}{\partial \boldsymbol{a}} \cdot \boldsymbol{h}^{\mathrm{T}}$$

Question 1.2 Answer :

Because the gradients of $\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)}, \boldsymbol{b}^{(1)}$ are all zero matrices or vectors, in every iterations, when update the parameters, there is no change for $\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)}, \boldsymbol{b}^{(1)}$.

Question 1.3 Answer :

$$\boldsymbol{a} = \boldsymbol{W}^{(2)}[\boldsymbol{W}^{(1)}\boldsymbol{x} + \boldsymbol{b}^{(1)}] + \boldsymbol{b}^{(2)}$$

$$\boldsymbol{a} = \boldsymbol{W}^{(2)}\boldsymbol{W}^{(1)}\boldsymbol{x} + \boldsymbol{W}^{(2)}\boldsymbol{b}^{(1)} + \boldsymbol{b}^{(2)}$$

$$So, \boldsymbol{U} = \boldsymbol{W}^{(2)}\boldsymbol{W}^{(1)}$$

$$\boldsymbol{v} = \boldsymbol{W}^{(2)}\boldsymbol{b}^{(1)} + \boldsymbol{b}^{(2)}$$

Question 2.1 Answer :

$$\boldsymbol{J}(\boldsymbol{w}) = \sum_n l(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x_n}), y_n) + \frac{\lambda}{2}||\boldsymbol{w}||_2^2$$

$$\frac{\partial \boldsymbol{J}(\boldsymbol{w})}{\partial \boldsymbol{w}} = \sum_n \frac{\partial l(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x_n}), y_n)}{\partial(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x_n}))}\boldsymbol{\phi}(\boldsymbol{x_n}) + \lambda\boldsymbol{w}$$

$$if,\ \frac{\partial \boldsymbol{J}(\boldsymbol{w})}{\partial \boldsymbol{w}} = 0$$

$$\boldsymbol{w}^* = -\frac{1}{\lambda}\sum_n \frac{\partial l(s_n, y_n)}{\partial s_n}\boldsymbol{\phi}(\boldsymbol{x_n}),\ where\ s_n = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x_n})$$

so, the optimal solution of w can be represented as a linear combination of $\boldsymbol{\phi}(\boldsymbol{x})$.

Question 2.2 Answer :

$$\min_{w} \sum_{j=1}^{N} l(\sum_{i=1}^{N} \alpha_i K_{ij}, y_j) + \frac{\lambda}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i K_{ij} \alpha_j$$