CMPUT 466 W7

Arun Woosaree

November 14, 2021

Problem 1

Derive the gradient in softmax regression $\frac{\partial J}{\partial w_{k,i}}, \frac{\partial J}{\partial b_k}$

$$y_k = \frac{e^{z_k}}{\sum_{k''} e^{z_{k''}}}$$

$$z_k = w_k^\top x + b$$

$$J = -\sum_{k'} t_{k'}^{(m)} \log y_{k'}^{(m)}$$

$$= -\sum_{k}^{\prime} t_{k'}^{(m)} \left(\log e^{z_{k'}^{(m)}} - \log \sum_{k''} e^{z_{k''}^{(m)}} \right)$$

$$= -\sum_{k}^{\prime} t_{k'}^{(m)} \left(z_{k'}^{(m)} - \log \sum_{k''} e^{z_{k''}^{(m)}} \right)$$

$$= -\left(\sum_{k'} t_{k'}^{(m)} z_{k'}^{(m)} - \sum_{k'} t_{k'}^{(m)} \log \sum_{k''} e^{z_{k''}^{(m)}} \right)$$

$\sum_{k'}$ is not needed for the second term because one t_k is one

$$J = -\left(\sum_{k'} t_{k'}^{(m)} z_{k'}^{(m)} - \log \sum_{k''} e^{z_{k''}^{(m)}} \right)$$

$$\frac{\partial J}{z_k} = \frac{\partial}{\partial z_k} - \left(t_k^{(m)} z_k^{(m)} - \log \sum_{k''} e^{z_{k''}^{(m)}} \right)$$

$$= -t_k^{(m)} + \frac{e^{z_k^{(m)}}}{\sum_{k''} e^{z_{k''}^{(m)}}}$$

$$= y_k^{(m)} - t_k^{(m)}$$

For one sample,

$$\frac{\partial J}{\partial w_{k,i}} = \frac{\partial J}{z_k}\frac{\partial z_k}{\partial w_{k,i}} = (y_k^{(m)} - t_k^{(m)})\frac{\partial z_k}{\partial w_{k,i}}$$
$$= (y_k^{(m)} - t_k^{(m)})x_i^{(m)}$$

$$\frac{\partial J}{\partial b_k} = \frac{\partial J}{z_k}\frac{\partial z_k}{\partial b_k} = y_k^{(m)} - t_k^{(m)}$$

Total loss of m samples:

$$\boxed{\begin{aligned}\frac{\partial J}{\partial w_{k,i}} &= \sum_{m=1}^{M}(y_k^{(m)} - t_k^{(m)})x_i^{(m)} \\ \frac{\partial J}{\partial b_k} &= \sum_{m=1}^{M} y_k^{(m)} - t_k^{(m)}\end{aligned}}$$

Problem 2

Show that logistic regression can also be reduced to 2-way softmax. i.e. for any parameter of the logistic regression model, there exists some parameter of the softmax regression model that does the same thing

$$y = \sigma(w^\top x + b) = \frac{1}{1 + e^{-(w^\top x + b)}}$$
$$= \frac{e^{w^\top x + b}}{1 + e^{w^\top x + b}}$$
$$= \frac{e^{w^\top x + b}}{e^{\mathbf{0}^\top x + \mathbf{0}} + e^{w^\top x + b}}$$

This is equivalent to a two-way softmax with weights $\begin{bmatrix} w^\top \\ \mathbf{0}^\top \end{bmatrix}$ and a bias of $\begin{bmatrix} b \\ 0 \end{bmatrix}$

Problem 3

Give a mapping from $\mathbf{y}$ to $\hat{t}$ that maximizes the total expected utility.

$$\mathbb{E}_{t\sim y}[u] = \sum_k y_k u_k \mathbb{1}\{\hat{t} = k\}$$

Choosing

$$\hat{t} = \arg\max_k y_k u_k$$

maximizes the utility