ECE 421 Homework 4

Name: Mark Qi

Student ID: 1006764645

Question 1,

$x \in \mathbb{R}^{d+1}$, $y \in \{1, 2\}$

$w(1)$, $w(2)$ weigh vectors for class 1 and 2

Probability of x belonging to class $i \in \{1, 2\}$

$$P^{SM}(y = i \mid x) = \frac{e^{w(i)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}}$$

Loss Function

$$e_n^{SM}(w(1), w(2)) = -\log(P^{SM}(y_n \mid x_n))$$

$$= -\log\left[\frac{e^{w(y_n)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}}\right]$$

(a),

Gradient With Respect To $w(1)$

$y_n = 1$

$$\nabla_{w(1)} e_n^{SM}(w(1), w(2)) = \nabla_{w(1)}\left[-\log\left[\frac{e^{w(1)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}}\right]\right]$$

$$= -\frac{e^{w(1)^T x} + e^{w(2)^T x}}{e^{w(1)^T x}} \times \nabla_{w(1)} \frac{e^{w(1)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}}$$

$$= -\cancel{\frac{e^{w(1)^T x} + e^{w(2)^T x}}{e^{w(1)^T x}}}$$

$$\frac{\underline{x}\, \cancel{e^{w(1)^T x}}\left(e^{w(1)^T x} + e^{w(2)^T x}\right) - \cancel{e^{w(1)^T x}} \times e^{w(1)^T x}}{\left(e^{w(1)^T x} + e^{w(2)^T x}\right)^{\cancel{2}}}$$

$$= \frac{-\underline{x}\, e^{w(2)^T x}}{\left(e^{w(1)^T x} + e^{w(2)^T x}\right)}$$

$y_n = 2$

$$\nabla_{w(1)} e_n^{SW}\left(w(1), w(2)\right) = \nabla_{w(1)} \left[-\log \left[\frac{e^{w(2)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}} \right] \right]$$

$$= -\frac{e^{w(1)^T x} + e^{w(2)^T x}}{e^{w(2)^T x}} \times \nabla_{w(1)} \frac{e^{w(2)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}}$$

$$= \cancel{-}\cancel{\frac{e^{w(1)^T x} + e^{w(2)^T x}}{e^{w(2)^T x}}}$$

$$\frac{\cancel{-}\cancel{e^{w(2)^T x}}\, \underline{x}\, e^{w(1)^T x}}{\left(e^{w(1)^T x} + e^{w(2)^T x}\right)^{\cancel{2}}}$$

$$= \frac{\underline{x}\, e^{w(1)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}}$$

Gradient With Respect To $\underline{w(2)}$

$y_n = 1$

$$\nabla_{w(2)} e_n^{SW}\left(w(1), w(2)\right) = \nabla_{w(2)}\left[-\log\left[\frac{e^{w(1)^T x}}{e^{w(1)^T x}+e^{w(2)^T x}}\right]\right]$$

$$= -\frac{e^{w(1)^T x}+e^{w(2)^T x}}{e^{w(1)^T x}} \times \nabla_{w(2)} \frac{e^{w(1)^T x}}{e^{w(1)^T x}+e^{w(2)^T x}}$$

$$= \cancel{-}\,\cancel{\frac{e^{w(1)^T x}+e^{w(2)^T x}}{e^{w(1)^T x}}}\;\frac{\cancel{-}\,\cancel{e^{w(1)^T x}}\; x\, e^{w(2)^T x}}{\left(e^{w(1)^T x}+e^{w(2)^T x}\right)^{\cancel{2}}}$$

$$= \frac{x\, e^{w(2)^T x}}{e^{w(1)^T x}+e^{w(2)^T x}}$$

$y_n = 2$

$$\nabla_{w(2)} e_n^{SW}\left(w(1), w(2)\right) = \nabla_{w(2)}\left[-\log\left[\frac{e^{w(2)^T x}}{e^{w(1)^T x}+e^{w(2)^T x}}\right]\right]$$

$$= -\frac{e^{w(1)^T x}+e^{w(2)^T x}}{e^{w(2)^T x}} \times \nabla_{w(2)} \frac{e^{w(2)^T x}}{e^{w(1)^T x}+e^{w(2)^T x}}$$

$$= -\cancel{\frac{e^{w(1)^T x}+e^{w(2)^T x}}{e^{w(2)^T x}}}\;\frac{x\, \cancel{e^{w(2)^T x}}\left(e^{w(1)^T x}+e^{w(2)^T x}\right)-\cancel{e^{w(2)^T x}}\; x\, e^{w(2)^T x}}{\left(e^{w(1)^T x}+e^{w(2)^T x}\right)^{\cancel{2}}}$$

$$= \frac{-x\, e^{w(1)^T x}}{\left(e^{w(1)^T x}+e^{w(2)^T x}\right)}$$

(b)

Binary Logistic Regression

$$p^{LR}(y=1 \mid x) = \frac{e^{w^T x}}{1+e^{w^T x}}$$

$$p^{LR}(y=2 \mid x) = 1 - p^{LR}(y=1 \mid x)$$

Loss Function

$$e_n^{LR}(w) = -\log(p^{LR}(y_n \mid x_n))$$

Find a relationship between $w(1)$, $w(2)$ and w

$$p^{SM}(y=1 \mid x) = \frac{e^{w(1)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}} = p^{LR}(y=1 \mid x)$$

$$= \frac{e^{(w(1)^T - w(2)^T)x}}{e^{(w(1)^T - w(2)^T)x} + 1} = \frac{e^{w^T x}}{1+e^{w^T x}}$$

$$p^{SM}(y=2 \mid x) = \frac{e^{w(2)^T x}}{e^{w(1)^T x} + e^{w(2)^T x}} = 1 - p^{LR}(y=1 \mid x)$$

$$= \frac{1}{e^{(w(1)^T - w(2)^T)x} + 1} = 1 - \frac{e^{w^T x}}{1+e^{w^T x}}$$

$$= \frac{1}{1 + e^{(w(1)^T - w(2)^T)x}} = \frac{1}{1+e^{w^T x}}$$

$$\boxed{w(1) - w(2) = w}$$

(c)

For Softmax Regression Model, Weight Vector Are Updated

$$\underline{W_{new}} = \underline{W_{old}} + \varepsilon^{SM} \nabla_{w(l)} e_n^{SM}(w(1), w(2)), \qquad l \in \{1,2\}$$

If $l = y_n$

$$\underline{W_{new}} = \underline{W_{old}} + \varepsilon^{SM} \left[-x_n + \frac{e^{w(y_n)^T x_n}}{e^{w(1)^T x_n} + e^{w(2)^T x_n}} \right]$$

If $l \neq y_n$

$$\underline{W_{new}} = \underline{W_{old}} + \varepsilon^{SM} \left[\frac{x_n e^{w(l)^T x_n}}{e^{w(1)^T x_n} + e^{w(2)^T x_n}} \right]$$

Calculating Gradient Of Binary Logistic Regression

$y_n = 1$

$$\nabla_{w} e_n^{LR}(w) = \nabla_{w} \left\{ -\log\left(\frac{e^{w^T x_n}}{1+e^{w^T x_n}} \right) \right\}$$

$$= \frac{\cancel{1+e^{w^T x_n}}}{\cancel{e^{w^T x_n}}} \times \frac{x_n \cancel{e^{w^T x_n}} - e^{w^T x_n} x_n \cancel{e^{w^T x_n}}}{(1+e^{w^T x_n})^{\cancel{2}}}$$

$$= \frac{x_n - x_n e^{w^T x_n}}{1+e^{w^T x_n}}$$