6 11

Curso: Tópicos de Ciencia de la Computación 2
Nombre: Cynthia Lizeth Calixtro Aires. Código: 20150456D
Test: 3

★ Resolución de la pregunta 2:

Exercise 8.3:

a) Sea $\sigma(a) = \dfrac{1}{1+e^{-a}}$

$$\Rightarrow \frac{\partial \sigma(a)}{\partial a} = \cancel{-}\frac{1}{(1+e^{-a})^2}\cdot(\cancel{-}e^{-a}) = \underbrace{\frac{1}{(1+e^{-a})}}\cdot\frac{(1+e^{-a}-1)}{(1+e^{-a})}$$

$$= \sigma(a)\cdot(1-\sigma(a))$$

1

b) Sea $NLL(w) = -\sum_{i=1}^{N}\left[y_i \log \mu_i + (1-y_i)\log(1-\mu_i)\right]$

donde:

- w es el vector de pesos $\Rightarrow w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$
- n es el número de features.
- N es el número de ejemplos.
- y_i es el i-ésimo target.
- $x^{(i)}$ es el i-ésimo ejemplo $\Rightarrow x^{(i)} = \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}$
- $\mu_i = \sigma(w^T x^{(i)}) = \dfrac{1}{1+e^{-w^T x^{(i)}}}$, con $w^T x^{(i)} = w_0 x_0^{(i)} + w_1 x_1^{(i)} + \ldots + w_j x_j^{(i)} + \ldots + w_n x_n^{(i)}$

1

Hallemos $\nabla NLL(w)$:

$$\Rightarrow \frac{\partial}{\partial w_j} NLL(w) = -\sum_{i=1}^{N}\left[\frac{\partial}{\partial w_j} y_i \log \mu_i + \frac{\partial}{\partial w_j}(1-y_i)\log(1-\mu_i)\right]$$

$$= -\sum_{i=1}^{N}\left[\left(\frac{y_i}{\mu_i} - \frac{(1-y_i)}{1-\mu_i}\right)\cdot\frac{\partial \mu_i}{\partial w_j}\right] = \sum_{i=1}^{N}\frac{(\mu_i - y_i)}{\mu_i(1-\mu_i)}\cdot\frac{\partial \mu_i}{\partial w_j}$$

Usando el resultado obtenido en (a) y la regla de la cadena:

$$\frac{\partial \mu_i}{\partial w_j} = \sigma(w^T x^{(i)}) \cdot (1 - \sigma(w^T x^{(i)})) \cdot \underbrace{\frac{\partial}{\partial w_j} w^T x^{(i)}}_{x_j^{(i)}} = \mu_i (1 - \mu_i) x_j^{(i)} \quad \text{—} \; (*)$$

Luego:

$$\Rightarrow \frac{\partial NLL(w)}{\partial w_j} = \sum_{i=1}^{N} \left[\frac{(\mu_i - y_i)}{\cancel{\mu_i (1-\mu_i)}} \cdot \cancel{\mu_i (1-\mu_i)} x_j^{(i)} \right] = \sum_{i=1}^{N} (\mu_i - y_i) x_j^{(i)}$$

Vemos que:

$$\frac{\partial}{\partial w_j} NLL(w) = \left[x_j^{(1)} \; \cdots \; x_j^{(i)} \; \cdots \; x_j^{(N)} \right] \begin{bmatrix} \mu_1 - y_1 \\ \vdots \\ \mu_i - y_i \\ \vdots \\ \mu_N - y_N \end{bmatrix}$$

Así:

$$\Rightarrow \nabla NLL(w) = \underbrace{\begin{bmatrix} | & & | & & | \\ x^{(1)} & \cdots & x^{(i)} & \cdots & x^{(N)} \\ | & & | & & | \end{bmatrix}}_{X^T} \cdot \left(\underbrace{\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_N \end{bmatrix}}_{\mu} - \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix}}_{y} \right)$$

$$= X^T (\mu - y)$$

c) Hallemos H:

$$H_{kj} = \frac{\partial \, NLL(w)}{\partial w_k \partial w_j} = \sum_{i=1}^{N} \frac{\partial}{\partial w_k}(\mu_i - y_i)\, x_j^{(i)} = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial w_k} x_j^{(i)} \overset{(*)}{=} \sum_{i=1}^{N} \mu_i(1-\mu_i)\, x_k^{(i)} x_j^{(i)}$$

Luego, sea a un vector de dimensión $n+1$:

$$a^T H a = \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k H_{kj} = \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k \sum_{i=1}^{N} \mu_i(1-\mu_i)\, x_k^{(i)} x_j^{(i)}$$

$$= \sum_{i=1}^{N} \mu_i(1-\mu_i) \sum_{j=0}^{n} a_j x_j^{(i)} \sum_{k=0}^{n} a_k x_k^{(i)}$$

Como: $\sum_{l=0}^{n} a_l x_l^{(i)} = a^T x^{(i)}$

$$\Rightarrow a^T H a = \sum_{i=1}^{N} \mu_i(1-\mu_i)\left(a^T x^{(i)}\right)^2$$

Como $0 < \mu_i < 1$, entonces $a^T H a \geq 0$ y así H es PSD.

Para que H sea PD se debe cumplir $a^T H a > 0$. Veamos que cuando X es full rank H es PD:

Si $\underset{N\times(n+1)}{X}$ es full rank entonces $\operatorname{rank}(X) = n+1$ (Asumiendo que $N \gg n+1$). Además, como para cualquier matriz A se cumple que:

$$\operatorname{rank}(A) + \dim(\text{Null-Space}(A)) = \#\text{ de columnas de } A$$

$$\Rightarrow \dim(\text{Null-Space}(X)) = 0.$$

Por lo cual, $a \notin \text{Null-Space}(X)$, es decir, $Xa \neq \bar{0}$ y así nunca se cumpliría:

$$a^T H a = 0.$$

EXERCISE 8.4:

a) Sean $\eta_i = \begin{bmatrix} \eta_{i1} \\ \eta_{i2} \\ \vdots \\ \eta_{iC} \end{bmatrix}_{C\times 1} = W^T x^{(i)} \wedge \mu_{ik} = S(\eta_i)_k = \dfrac{e^{\eta_{ik}}}{\sum_{c'=1}^{C} e^{\eta_{ic'}}}$

Luego:

$$\frac{\partial \mu_{ik}}{\partial \eta_{ij}} = \frac{\partial}{\partial \eta_{ij}}\left(\frac{e^{\eta_{ik}}}{\sum_{c'=1}^{C} e^{\eta_{ic'}}}\right)$$

Si $k \neq j$:

$$\Rightarrow \frac{\partial \mu_{ik}}{\partial \eta_{ij}} = -\frac{e^{\eta_{ik}}}{\left(\sum_{c'=1}^{C} e^{\eta_{ic'}}\right)^2} \cdot \underbrace{\frac{\partial}{\partial \eta_{ij}}\left(\sum_{c'=1}^{C} e^{\eta_{ic'}}\right)}_{e^{\eta_{ij}}} = -\underbrace{\frac{e^{\eta_{ik}}}{\sum_{c'=1}^{C} e^{\eta_{ic'}}}}\cdot\underbrace{\frac{e^{\eta_{ij}}}{\sum_{c'=1}^{C} e^{\eta_{ic'}}}}$$

$$= -\mu_{ik} \cdot \mu_{ij}$$

Si $k = j$:

$$\Rightarrow \frac{\partial \mu_{ik}}{\partial \eta_{ij}} = \frac{\frac{\partial}{\partial \eta_{ij}} e^{\eta_{ik}} \cdot \sum_{c'=1}^{C} e^{\eta_{ic'}} - e^{\eta_{ik}} \cdot \frac{\partial}{\partial \eta_{ij}}\left(\sum_{c'=1}^{C} e^{\eta_{ic'}}\right)}{\left(\sum_{c'=1}^{C} e^{\eta_{ic'}}\right)^2} = \underbrace{\frac{e^{\eta_{ik}}}{\sum_{c'=1}^{C} e^{\eta_{ic'}}}} - \underbrace{\frac{e^{\eta_{ik}}}{\sum_{c'=1}^{C} e^{\eta_{ic'}}}}\cdot\underbrace{\frac{e^{\eta_{ij}}}{\sum_{c'=1}^{C} e^{\eta_{ic'}}}}$$

$$= \mu_{ik} - \mu_{ik} \cdot \mu_{ij}$$

$$= \mu_{ik}(1 - \mu_{ij})$$

En general:

$$\Rightarrow \frac{\partial \mu_{ik}}{\partial \eta_{ij}} = \mu_{ik}(\delta_{kj} - \mu_{ij}), \quad \text{con} \quad \delta_{kj} = \begin{cases} 0, & k \neq j \\ 1, & k = j \end{cases}$$

b) Sea

$$\ell(W) = \sum_{i=1}^{N} \left[\sum_{c=1}^{C} y_{ic} w_c^T x^{(i)} - \log\left(\sum_{c'=1}^{C} \exp(w_{c'}^T x^{(i)}) \right) \right]$$

Luego:

$$\Rightarrow \nabla_{w_c} \ell = \frac{\partial}{\partial w_c} \ell(W) = \sum_{i=1}^{N} \left[\underbrace{\frac{\partial}{\partial w_c}\left(\sum_{c=1}^{C} y_{ic} w_c^T x^{(i)} \right)} - \underbrace{\frac{\partial}{\partial w_c}\left(\log\left(\sum_{c'=1}^{C} \exp(w_{c'}^T x^{(i)}) \right)\right)} \right]$$

$$= \sum_{i=1}^{N} \left[y_{ic} x^{(i)} - \frac{\frac{\partial}{\partial w_c}\left(\sum_{c'=1}^{C} \exp(w_{c'}^T x^{(i)}) \right)}{\sum_{c'=1}^{C} \exp(w_{c'}^T x^{(i)})} \right]$$

$$= \sum_{i=1}^{N} \left[y_{ic} x^{(i)} - \underbrace{\frac{\exp(w_c^T x^{(i)})}{\sum_{c'=1}^{C} \exp(w_{c'}^T x^{(i)})}} \cdot x^{(i)} \right] \quad \text{—} (**)$$

$$= \sum_{i=1}^{N} (y_{ic} - \mu_{ic})\, x^{(i)}$$

c) $$H_{c,c'} = \frac{\partial}{\partial w_c} \nabla_{w_{c'}} \ell \overset{(**)}{=} \sum_{i=1}^{N} \left[\cancelto{0}{\frac{\partial}{\partial w_c}(y_{ic'} x^{(i)})} - \frac{\partial}{\partial w_c}\left(\frac{\exp(w_{c'}^T x^{(i)}) \cdot x^{(i)}}{\sum_{c''=1}^{C} \exp(w_{c''}^T x^{(i)})} \right) \right]$$

Si $c \neq c'$:

$$\Rightarrow H_{c,c'} = -\sum_{i=1}^{N} \left[\frac{-\exp(w_{c'}^T x^{(i)}) . x^{(i)}}{\left(\sum_{c''=1}^{C} \exp(w_{c''}^T x^{(i)}) \right)^2} \cdot \underbrace{\frac{\partial}{\partial w_c}\left(\sum_{c''=1}^{C} \exp(w_{c''}^T x^{(i)}) \right)}_{\exp(w_c^T x^{(i)})\,(x^{(i)})^T} \right]$$

$$= \sum_{i=1}^{N} \underbrace{\frac{\exp(w_{c'}^T x^{(i)})}{\sum_{c''=1}^{C} \exp(w_{c''}^T x^{(i)})}}_{\mu_{ic'}} \cdot \underbrace{\frac{\exp(w_c^T x^{(i)})}{\sum_{c''=1}^{C} \exp(w_{c''}^T x^{(i)})}}_{\mu_{ic}} \cdot x^{(i)} . (x^{(i)})^T$$

$$= \sum_{i=1}^{N} \mu_{ic'} \mu_{ic} \cdot x^{(i)} . (x^{(i)})^T$$

Si $c = c'$:

$$H_{c,c'} = -\sum_{i=1}^{N}\left[\frac{\frac{\partial}{\partial w_c}\left(\exp(w_{c'}^T x^{(i)})\, x^{(i)}\right)\cdot\sum_{c''=1}^{C}\exp(w_{c''}^T x^{(i)}) - \exp(w_{c'}^T x^{(i)})\, x^{(i)}\cdot\frac{\partial}{\partial w_c}\left(\sum_{c''=1}^{C}\exp(w_{c''}^T x^{(i)})\right)}{\left(\sum_{c''=1}^{C}\exp(w_{c''}^T x^{(i)})\right)^2}\right.$$

$$= -\sum_{i=1}^{N}\left[\frac{\exp(w_c^T x^{(i)})\, x^{(i)}(x^{(i)})^T\cdot\sum_{c''=1}^{C}\exp(w_{c''}^T x^{(i)}) - \exp(w_{c'}^T x^{(i)})\, x^{(i)}\cdot\exp(w_c^T x^{(i)})(x^{(i)})^T}{\left(\sum_{c''=1}^{C}\exp(w_{c''}^T x^{(i)})\right)^2}\right]$$

$$= -\sum_{i=1}^{N}\left(\underbrace{\frac{\exp(w_c^T x^{(i)})}{\sum_{c''=1}^{C}\exp(w_{c''}^T x^{(i)})}}_{\mu_{ic}} - \underbrace{\frac{\exp(w_{c'}^T x^{(i)})}{\sum_{c''=1}^{C}\exp(w_{c''}^T x^{(i)})}}_{\mu_{ic'}}\cdot\underbrace{\frac{\exp(w_c^T x^{(i)})}{\sum_{c''=1}^{C}\exp(w_{c''}^T x^{(i)})}}_{\mu_{ic}}\right) x^{(i)}\cdot(x^{(i)})^T$$

$$= -\sum_{i=1}^{N}\mu_{ic}(1-\mu_{ic'})\, x^{(i)}\cdot(x^{(i)})^T$$

En general:

$$H_{c,c'} = -\sum_{i=1}^{N}\mu_{ic}(\delta_{cc'} - \mu_{ic'})\, x^{(i)}\cdot(x^{(i)})^T$$