Devoir TP3

MASN 2601
Nicholas Massad

x_1	x_2	y
0	0	1
1	1	1
1	0	0

a) $i=1$ $\quad w_1x_1 + w_2x_2 + b = 1$

$0 + 0 + b = 1$

$\boxed{b = 1}$

$i=2$ $\quad w_1(1) + w_2(1) + b = 1$

$\boxed{w_1 + w_2 + b = 1}$

$i=3$ $\quad w_1(1) + w_2(0) + b = 0$

$\boxed{w_1 + b = 0}$

b) $b = 1$

$w_1 + b = 0 \Rightarrow w_1 = -1$

$w_1 + w_2 + b = 1 \Rightarrow w_2 = 1$

$\boxed{w_1 = -1 \quad w_2 = 1 \quad b = 1}$

#2 $\quad F(x,y)$ où $x = x(t)$ et $y = y(t)$

a) $F(x,y) = x^2 + y^2 \quad x(t) = t \quad y(t) = e^t$

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial t}$$

$$= 2x\cdot(1) + 2y\cdot e^t = 2t + 2e^{2t} = \boxed{2(t + e^{2t})}$$

b) $z = wx + b$

$\hat{y} = \frac{1}{1+e^{-z}} = \sigma(z)$

$L = \frac{1}{2}(\hat{y} - y)^2$

$R = \frac{1}{2} w^2$

$L_{Reg} = L + \lambda R$

$\bar{L}_{Reg} = 1$

$\bar{L}, \bar{R} = \left(\frac{\partial L_{Reg}}{\partial L}, \frac{\partial L_{Reg}}{\partial R}\right) \cdot \frac{\partial L_{Reg}}{\partial L_{Reg}} = (1, \lambda) \cdot (1)$

$\bar{\hat{y}} = \frac{\partial L_{Reg}}{\partial L} \cdot \frac{\partial L}{\partial \hat{y}} = (1) \cdot 2(\hat{y} - y)$

$\bar{z} = \frac{\partial L_{Reg}}{\partial L} \cdot \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} = (1) \cdot 2(\hat{y} - y) \cdot \frac{1}{1+e^{-z}} \cdot \left(1 - \frac{1}{1+e^{-z}}\right)$

$\bar{w} = \frac{\partial L_{reg}}{\partial L} \cdot \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w} + \frac{\partial L_{Reg}}{\partial R} \cdot \frac{\partial R}{\partial w} = (1) \cdot (2\hat{y} - y) \cdot \sigma(z)(1 - \sigma(z))X + \lambda \cdot 2w$

$\bar{b} = \frac{\partial L_{reg}}{\partial L} \cdot \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial b} = (1) \cdot 2(\hat{y} - y) \cdot \sigma(z)(1 - \sigma(z)) \cdot (1)$

$\sigma(z) = \frac{1}{1+e^{-z}}$

c)

$\bar{L} = \frac{\partial L}{\partial L} = 1$

$\bar{y}_k = \frac{\partial L}{\partial \hat{y}_k} = (\hat{y}_k - y_k)$

$\bar{w}^{[2]}_{ki} = \frac{\partial L}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial w_{ki}} = (\hat{y}_k - y_k) \cdot h_i$

$\bar{b}^{[2]}_k = \frac{\partial L}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial b^{[2]}_k} = (\hat{y}_k - y_k) \cdot (1)$

$\sigma(z) = \frac{1}{1+e^{-z}}$

$\bar{h}_i = \frac{\partial L}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial h_i} = (\hat{y}_k - y_k) \cdot w_{ki}$

$\bar{z}_i = \frac{\partial L}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial h_i} \cdot \frac{\partial h_i}{\partial z_i} = (\hat{y}_k - y_k) \cdot w_{ki} \cdot \sigma(z_i)(1 - \sigma(z_i))$

$$\bar{w}_{ij}^{[1]} = \frac{\partial L}{\partial \hat{y}_K} \cdot \frac{\partial \hat{y}_K}{\partial h_i} \cdot \frac{\partial h_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_{ij}} = (\hat{y}_K - y) \cdot w_{Ki} \cdot \sigma(z_i)(1-\sigma(z_i)) \cdot x_j$$

$$\bar{b}_i^{[1]} = \frac{\partial L}{\partial \hat{y}_K} \cdot \frac{\partial \hat{y}_K}{\partial h_i} \cdot \frac{\partial h_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial b_i} = (\hat{y}_K - y) \cdot w_{Ki} \cdot \sigma(z_i)(1-\sigma(z_i)) \cdot (1)$$

3

a)

$$L_{SCE}(y, \hat{y}) = -\sum_{k=1}^{K} y_k \log(\hat{y}_k) \qquad y_k = \begin{Bmatrix} 1 & \text{if } k=i \\ 0 & \text{else} \end{Bmatrix} \quad i \in K$$

$$= -\underset{0}{\underset{\uparrow}{y_1}} \log(\hat{y}_1) - \underset{0}{\underset{\uparrow}{y_2}} \log(\hat{y}_2) - \dots - \underset{1}{\underset{\uparrow}{y_i}} \log(\hat{y}_i) - \dots \underset{0}{\underset{\uparrow}{y_K}} \log(\hat{y}_K)$$

$$= -\vec{y} \log(\hat{\vec{y}}) = -y_i \log(\hat{y}_i)$$

Puisque seulement un terme ne sera pas égale à 0 alors le produit scalaire ci-haut demeure vrai et est égale au L_{SCE}

b) $$\nabla_z L_{SCE} = \frac{\partial L_{SCE}}{\partial \hat{\vec{y}}} \cdot \frac{\partial \hat{\vec{y}}}{\partial \vec{z}}$$

$$\frac{\partial L_{SCE}}{\partial \hat{\vec{y}}} = -\frac{y}{\hat{y}} \qquad \frac{\partial \hat{\vec{y}}}{\partial \vec{z}} = \frac{\partial}{\partial \vec{z}} \frac{e^{z_k}}{\sum_{l=1}^{K} e^{z_l}} \quad k = 1, \dots, K$$

(Pour $k = i = l$) $$= \frac{e^{z_i} \sum_{l=1}^{K} e^{z_l} - e^{z_i} e^{z_i}}{\sum_{l=1}^{K} e^{z_l}}$$

Puisque seulement le terme i n'est pas égale à 0 alors $e^{z_k} = e^{z_i}$

$$= \frac{e^{z_i}}{\sum_{l=1}^{K} e^{z_l}} \cdot \frac{\sum_{l=1}^{K} e^{z_l} - e^{z_i}}{\sum_{l=1}^{K} e^{z_l}}$$

$$\hat{y}_i = S_K(z_i)$$

$$= \hat{y}_i \cdot (1 - \hat{y}_i)$$

$$\nabla_z L_{SCE} = -\frac{y}{\hat{y}_i} \cdot (\hat{y}_i (1 - \hat{y}_i)) = \boxed{\hat{y}_i - y} \quad \square$$

c) $\frac{\partial L_{SCE}}{\partial w_{1j}} = \frac{\partial L_{SCE}}{\hat{y}_1} \cdot \frac{\partial \hat{y}_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{1j}}$

comme vu Précedement

$$\frac{\partial L_{SCE}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} = \vec{\hat{y}} - \vec{y}$$

Puisqu'il y a $m \times j$ entreé et $m \times j$ sortie où m est le nombre de séquence d'entrée dans le mini-batch et j le nombre de Paramètre d'entré, et on sait que $z_1 = \vec{w_1} \vec{x} + b$

alors

$$\frac{\partial z_1}{\partial w_{1j}} = x_j$$

Finalement

(Plusieurs sortie et lable)

$$\frac{\partial L_{SCE}}{\partial w_{1j}} = \sum_{i=1}^{m} (\hat{y}_1^{[i]} - y_1^{[i]}) x_j$$

d) On voit que le résultat Précedent, dans sa Forme est très similaire à la régression Logistique Binaire mais l'erreur est cumuler sur le mini-batch complet. au lieu d'une seule donné. ceci est bien car cela Permettra de réduire la variabilité dans le Processus d'apprentissage