Devoir TP2

Nicholas Massad
MASN 2601

$y \sim$ Berouilli (ϕ_y) avec $\phi_y = p(y=1|x,w)$

$= \sigma(z) = \frac{1}{1+e^{-z}}$ où $z = w^T x = \sum_{j=1}^{d} w_j x_j$

$L_{CE}^{(i)}(\vec{w}) = -\log p(y^{(i)}|\vec{x}^{(i)}, \vec{w})$

$p(y^{(i)}|\vec{x}^{(i)}, \vec{w}) = \sigma(z^{(i)})^{y^{(i)}}(1-\sigma(z^{(i)}))^{1-y^{(i)}}, \quad y^{(i)} = 0, 1$

a) Montrez que

$L_{CE}^{(i)}(\vec{w}) = -y^{(i)}\log\sigma(z^{(i)}) - (1-y^{(i)})\log(1-\sigma(z^{(i)}))$

J'omet les i car c'est long à écrire

$L_{CE}(\vec{w}) = -\log p(y|\vec{x}, \vec{w}) = -\log\left(\sigma(z)^y(1-\sigma(z))^{1-y}\right)$

$= -\log(\sigma(z)^y) - \log((1-\sigma(z))^{1-y})$

loi des log

$= \boxed{-y\log\sigma(z) - (1-y)\log(1-\sigma(z))}$

B) $\frac{\partial}{\partial w_j} L_{CE}(\vec{w}) = \frac{\partial}{\partial\sigma(z)} L_{CE}(\vec{w}) \cdot \frac{\partial\sigma(z)}{\partial z} \cdot \frac{\partial z}{\partial w_j}$

$\frac{\partial}{\partial\sigma(z)}\left(-y\log\sigma(z) - (1-y)\log(1-\sigma(z))\right) \cdot \frac{\partial}{\partial z}\frac{1}{1+e^{-z}} \cdot \frac{\partial}{\partial w_j} w_j x_j$

$= \left(\frac{-y}{\sigma(z)} + \frac{(1-y)}{1-\sigma(z)}\right) \cdot \frac{e^{-z}}{(1+e^{-z})^2} \cdot x_j$

$= (-y(1-\sigma(z)) + (1-y)(\sigma(z)))x_j$

$= (\sigma(z) - y)x_j$

c)

$$L^{D}_{CE}(w) = \sum_{i=1}^{m} L_{CE}(w)$$

$$\frac{\partial}{\partial w_j} L^{(D)}_{CE}(w) = \frac{\partial}{\partial w_j} \sum_{i=1}^{m} L^{(i)}_{CE}(\vec{w}) = \sum_{i=1}^{m} (\sigma(z^{(i)}) - y^{(i)}) x_j^{(i)}$$

d) La valeur à back Propager pour chaque w_j, $j=1 \dots d$ est la somme des différence entre le sigmoïde du vecteur d'entrée x_i et la valeur attendu y. Tout ceci multiplier par x_j

#2

a) $\phi_k = P(y=k)$

$$\phi_y = P(y) = \prod_{i=1}^{k} \phi_i^{1\{y=i\}} = \prod_{i=1}^{k-1} \phi_i^{0} \phi_k^{1}$$

$$= (1) \cdot \phi_k = P(y=k) \Rightarrow P(y) \quad \text{Par évidence}$$

b) $\eta_k = \log \frac{\phi_k}{\phi_K} \quad k=1 \dots K$

Montrer $\phi_k = \frac{1}{\sum_{i=1}^{K} e^{\eta_i}}$

$$\frac{1}{\sum_{i=1}^{K} e^{\log \frac{\phi_i}{\phi_K}}} = \frac{1}{\sum_{i=1}^{K} \frac{\phi_i}{\phi_K}} = \frac{1}{\frac{1}{\phi_K} \sum_{i=1}^{K} \phi_i} = \phi_K \frac{1}{1} = \boxed{\phi_K}$$

c) $P(y \mid X, w_1, w_2, \dots w_k) = P(y \mid \eta_1, \eta_2, \dots \eta_k)$

si les η sont indépendant

$$= \prod_{i=1}^{k} P(y \mid \eta_i)$$

d) $L_{CE} = -y \log\left(\frac{1}{1+e^{-wx}}\right) - (1-y)\log\left(\frac{e^{-wx}}{1+e^{-wx}}\right)$

1 b)

$$\frac{\partial L_{CE}}{\partial w_i} = \left(\frac{1}{1+e^{-w_i x_i}} - y_i\right) x_i$$

#3 $\text{softmax}(z) = S = (S_1, \ldots S_k)$

$$S_k = \frac{e^{z_k}}{\sum_{l=1}^{k} e^{z_l}}$$

a)

Montrer $\qquad \vec{a}$ vecteur constant $a_i = a_{i+1} \quad \forall i$

$$S_k = \frac{e^{z+a_k}}{\sum_{l=1}^{k} e^{z+a_l}} = \frac{e^{z_k}}{\sum_{l=1}^{k} e^{z_l}}$$

$$\text{softmax}(z+a) = \frac{e^{z_k} e^{a_k}}{\sum_{l=1}^{k} e^{z_l} e^{a_l}} = \frac{e^{z_k}}{\sum_{l=1}^{k} e^{z_l}} \cdot \frac{e^{a_k}}{\sum_{l=1}^{k} e^{a_l}} \qquad e^{a_i} = e^{a_{i+1}} \quad \forall_i$$

$$= \frac{e^{z_k}}{\sum_{l=1}^{k} e^{z_l}} \cdot \frac{e^{a_k}}{e^{a_k} \sum_{l=1}^{k} 1}$$

$$= \frac{e^{z_k}}{\sum_{l=1}^{k} e^{z_l}} \cdot \frac{1}{\sum_{l=1}^{k} 1} = \frac{e^{z_k}}{\sum_{l=1}^{k} e^{z_l} \cdot 1} = \frac{e^{z_k}}{\sum_{l=1}^{k} e^{z_l}} = \text{softmax}(z)$$

$\square$

b) car nous cherchons à atteindre un minimum local d'où le "descente" dans descente de gradient

c) $\frac{\partial s_k}{\partial z_1}$

$$\frac{\partial s_k}{\partial z_1} = \frac{\partial}{\partial z_1} \frac{e^{z_k}}{\sum_{l=1}^{k} e^{z_l}}$$

$$= e^{z_k} \frac{\partial}{\partial z_1}\left(\sum_{l=1}^{k} e^{z_l}\right)^{-1} = \frac{-e^{z_k}}{\left(\sum_{l=1}^{k} e^{z_l}\right)^2} \frac{\partial}{\partial z_1}\left(\sum_{l=1}^{k} e^{z_l}\right)$$

$$= -\frac{e^{z_k} e^{z_1}}{\left(\sum_{l=1}^{k} e^{z_l}\right)^2} \quad \square$$