Курагбедиани Г.К.
382003-1

№ 18

$\exists$ $\sigma(z) = \frac{1}{1+e^{-z}}$ (сигмоидальная ф-я). Проверить, что $\sigma' = \sigma(1-\sigma)$

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

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

$$\Rightarrow \sigma' = \sigma(1-\sigma)$$

№ 19

$\exists$ в задаче классификации на K классов $\{1, 2, \dots, K\}$ последний слой нейронной сети вычисляет softmax-функцию: $g_k(s_1, s_2, \dots, s_K) = \frac{e^{s_k}}{\sum_{l=1}^{K} e^{s_l}}$

В качестве штрафа используется кросс-энтропия (logloss-функция): $R^{(i)} = -\sum_{k=1}^{K} I(y^{(i)} = k) \ln g_k(s_1, s_2, \dots, s_K)$,

где $g_k(s_1, s_2, \dots, s_K)$ - softmax-функция. Доказать, что

1) $\frac{\partial g_k}{\partial s_l} = g_k(I(k=l) - g_l)$

2) $\frac{\partial R^{(i)}}{\partial g_k} = -\frac{I(y^{(i)})}{g_k}$

3) $\frac{\partial R^{(i)}}{\partial s_l} = g_l - I(l = y^{(i)})$.

1) $l = k$:

$$\frac{\partial g_k}{\partial s_l} = \frac{\partial}{\partial s_k}\left(e^{s_k}\right)\cdot\frac{1}{\sum_{i=1}^{K} e^{s_i}} + \frac{\partial}{\partial s_k}\left(\frac{1}{\sum_{i=1}^{K} e^{s_i}}\right) e^{s_k} =$$

$$= \frac{e^{s_k}}{\sum_{i=1}^{K} e^{s_i}} - \left(\frac{e^{s_k}}{\sum_{i=1}^{K} e^{s_i}}\right)^2 = g_k - g_k^2 = g_k(1-g_k)$$

$l \neq k$:

$$\frac{\partial g_k}{\partial s_l} = \frac{\partial}{\partial s_l}\left(\frac{1}{\sum_{i=1}^{K} e^{s_i}}\right) e^{s_k} = -\frac{e^{s_k}}{\sum_{i=1}^{K} e^{s_i}}\cdot\frac{e^{s_l}}{\sum_{i=1}^{K} e^{s_i}} = -g_k g_l$$

2) $$\frac{\partial R^{(i)}}{\partial g_k} = -\sum_{j=1}^{K} I(y^{(i)} = j)\cdot\frac{\partial}{\partial g_k}\left(\ln g_j(s_1, s_2, \dots, s_K)\right) =$$

$$= -\frac{I(y^{(i)} = k)}{g_k}$$

3) $$\frac{\partial R^{(i)}}{\partial s_l} = \sum_{j=1}^{K}\frac{\partial R^{(i)}}{\partial g_j}\cdot\frac{\partial g_j}{\partial s_l} = \sum_{j=1}^{K}\left(-\frac{I(y^{(i)} = j)}{g_j}\right) g_j\left(I(j = l) - g_l\right) =$$

$$= \sum_{j=1}^{K} I(y^{(i)} = j)\cdot\left(g_l - I(j = l)\right) =$$

$$= g_l\sum_{j=1}^{K} I(y^{(i)} = j) - \sum_{j=1}^{K} I(y^{(i)} = j)\, I(j = l) =$$

$$= g_l - I(y^{(i)} = l)$$