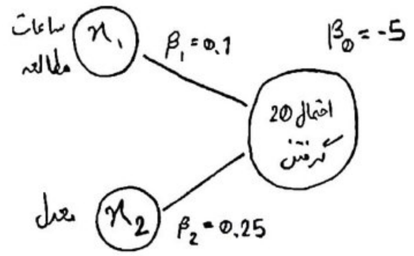
402211723 سید احسان حسن‌بیگی

پرسش 1



$\boxed{1}$ $\sigma\left(80 \times 0.1 + 18 \times 0.25 - 5\right) = \sigma(7.5) = \dfrac{1}{1+e^{-7.5}} = 0.99944$

$\boxed{2}$ $\sigma\left(x_1 \times 0.1 + 16 \times 0.25 - 5\right) = \dfrac{1}{1+e^{-\left(\frac{x_1}{10}-1\right)}} = 0.9 \longrightarrow e^{-\left(\frac{x_1}{10}-1\right)} = \dfrac{1}{9}$

$\longrightarrow x_1 = 10 \ln 9 + 10 \simeq 31.97$

پرسش 2

$s_k^{(i)}$ یعنی احتمالی که مدل ما به کلاس kاُم برای سمپل iاُم نسبت داده است و داریم $s_k^{(i)} = w_k \cdot x^{(i)}$

$\boxed{1}$ $\lg p\left(Y = y_k^{(i)} \mid x^{(i)}\right) = \lg \operatorname{softmax}\left(s^{(i)}_{y_k^{(i)}}\right) \xrightarrow{\text{one-hot form}} \sum_{k=1}^{k} \underbrace{y_k^{(i)}}_{\text{one-hot}} \lg \operatorname{softmax}\left(s_k^{(i)}\right)$

$\longrightarrow \lg p(Y \mid X) = \sum_{i=1}^{N} \sum_{k=1}^{k} y_k^{(i)} \lg \operatorname{softmax}\left(s_k^{(i)}\right)$

$= \sum_{i=1}^{N} \sum_{k=1}^{k} y_k^{(i)} \lg \left(\dfrac{e^{w_k \cdot x^{(i)}}}{\sum_{j=1}^{k} e^{w_j \cdot x^{(i)}}} \right)$

$= \sum_{i=1}^{N} \sum_{k=1}^{k} y_k^{(i)}\, w_k \cdot x^{(i)} - \sum_{i=1}^{N} \sum_{k=1}^{k} y_k^{(i)} \lg \left(\sum_{j=1}^{k} e^{w_j \cdot x^{(i)}} \right)$

می‌دانیم $y_k^{(i)}$ به فرم one-hot است و داریم: $\sum_{k=1}^{k} y_k^{(i)} = 1$

$= \sum_{i=1}^{N} \sum_{k=1}^{k} y_k^{(i)}\, w_k \cdot x^{(i)} - \sum_{i=1}^{N} \lg \left(\sum_{j=1}^{k} e^{w_j \cdot x^{(i)}} \right)$

$\boxed{\text{پرسش } 2}$ ادامه

$\boxed{2}$

$$\frac{\partial}{\partial w_\ell} \sum_{i=1}^{N}\sum_{k=1}^{k} y_k^{(i)}\, w_k \cdot x^{(i)} - \sum_{i=1}^{N}\sum_{k=1}^{k} \lg\left(\sum_{j=1}^{k} e^{w_j \cdot x^{(i)}}\right) - \lambda \sum_{k=1}^{k} \|w_k\|^2$$

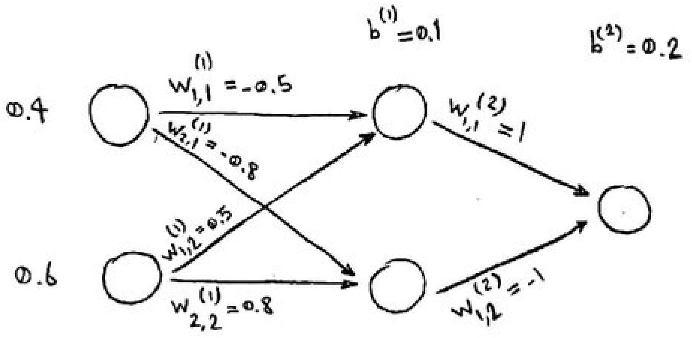
$$= \sum_{i=1}^{N} y_\ell^{(i)}\, x^{(i)} - \sum_{i=1}^{N} \frac{x^{(i)}\, e^{w_\ell \cdot x^{(i)}}}{\sum_{j=1}^{k} e^{w_j \cdot x^{(i)}}} - 2\lambda\, w_\ell$$

$\boxed{3}$

$$w_k^{(t+1)} = w_k^{(t)} + \eta \left(\sum_{i=1}^{N} y_k^{(i)}\, x^{(i)} - \sum_{i=1}^{N} \frac{x^{(i)}\, e^{w_k^{(t)} \cdot x^{(i)}}}{\sum_{j=1}^{k} e^{w_j^{(t)} \cdot x^{(i)}}} - 2\lambda\, w_k^{(t)} \right)$$

gradient ascent

$\boxed{\text{پرسش } 3}$



در صورت سوال گفته شده $w_{2,1}^{(2)} = -1$ که با توجه به تعریف صورت سوال وجود ندارد و ما فرض کرده ایم:

$$w_{1,2}^{(2)} = -1$$

$\boxed{1}$

$$a_1^{(1)} = \sigma\left(\underbrace{0.4\times -0.5}_{-0.2} + \underbrace{0.6\times 0.5}_{0.3} + 0.1\right) = \ln\left(1+e^{0.2}\right) = 0.798$$

$$a_2^{(1)} = \sigma\left(\underbrace{0.4\times -0.8}_{-0.32} + \underbrace{0.6\times 0.8}_{0.48} + 0.1\right) = \ln\left(1+e^{0.26}\right) = 0.831$$

$$a_1^{(2)} = \sigma\left(0.798\times 1 + 0.831\times -1 + 0.2\right) = \ln\left(1+e^{0.167}\right) = 0.780$$

$\boxed{2}$

$$\ln(1+e^z) = \ln(1+e^z) - \ln(e^z) + z = \boxed{\ln\left(1+e^{-z}\right) + z}$$

در حالت قبل e^z می توانست بسیار بزرگ شود، حال آنکه قرار بود به ورودی ln آن گرفته شود و دوباره کوچک شود.

در فرمول جایگزین این مشکل وجود ندارد و اگر z بزرگ باشد $\ln(1+e^{-z})$ نزدیک صفر می شود و به مقدار z می رسیم که مطلوب ماست

$\boxed{\text{پرسش 3}}$ ادامه

در صورت سوال گفته شده $w^{(2)}_{j,1}$ که با توجه به تعریف صورت سوال وجود ندارد و ما فرض کرده ایم:

$$w^{(2)}_{1,j}$$

$\boxed{3}$

$$\frac{\partial loss}{\partial a_1^{(2)}} = \frac{\partial \frac{1}{2}\|y-a_1^{(2)}\|^2}{\partial a_1^{(2)}} = -1\left(y-a_1^{(2)}\right) = a_1^{(2)} - y$$

$$\frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} = \frac{\partial \ln\left(1+e^{z_1^{(2)}}\right)}{\partial z_1^{(2)}} = \frac{e^{z_1^{(2)}}}{1+e^{z_1^{(2)}}} = \frac{1}{1+e^{-z_1^{(2)}}}$$

$$\left.\begin{aligned} \frac{\partial z_1^{(2)}}{\partial w_{1,1}^{(2)}} &= \frac{\partial a_1^{(1)} w_{1,1}^{(2)} + a_2^{(1)} w_{1,2}^{(2)} + b^{(2)}}{\partial w_{1,1}^{(2)}} = a_1^{(1)} \\ \frac{\partial z_1^{(2)}}{\partial w_{1,2}^{(2)}} &= \frac{\partial a_1^{(1)} w_{1,1}^{(2)} + a_2^{(1)} w_{1,2}^{(2)} + b^{(2)}}{\partial w_{1,2}^{(2)}} = a_2^{(1)} \end{aligned}\right\} \frac{\partial z_1^{(2)}}{\partial w_{1,j}^{(2)}} = a_j^{(1)}$$

$\boxed{4}$

$$\frac{\partial loss}{\partial w_{1,1}^{(2)}} = \frac{\partial loss}{\partial a_1^{(2)}} \times \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \times \frac{\partial z_1^{(2)}}{\partial w_{1,1}^{(2)}} = \left(a_1^{(2)} - y\right)\left(\frac{1}{1+e^{-z_1^{(2)}}}\right)\left(a_1^{(1)}\right)$$

$$= \left(0.78-1\right)\left(\frac{1}{1+e^{-0.167}}\right)\left(0.798\right) = -0.095$$

$$\rightarrow w_{1,1}^{(2)} = w_{1,1}^{(2)} - \eta\left(-0.095\right) = 1 + 0.095 = 1.095$$

$\boxed{5}$

$$\frac{\partial loss}{\partial w_{1,1}^{(1)}} = \frac{\partial loss}{\partial a_1^{(2)}} \times \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \times \frac{\partial z_1^{(2)}}{\partial a_1^{(1)}} \times \frac{\partial a_1^{(1)}}{\partial z_1^{(1)}} \times \frac{\partial z_1^{(1)}}{\partial w_{1,1}^{(1)}}$$

$$= \left(a_1^{(2)} - y\right)\left(\frac{1}{1+e^{-z_1^{(2)}}}\right)\left(w_{1,1}^{(2)}\right)\left(\frac{1}{1+e^{-z_1^{(1)}}}\right)\left(a_1^{(0)}\right)$$

$$= \left(0.78-1\right)\left(\frac{1}{1+e^{-0.167}}\right)\left(1\right)\left(\frac{1}{1+e^{-0.2}}\right)\left(0.4\right) = -0.026$$

$$\rightarrow w_{1,1}^{(1)} = w_{1,1}^{(1)} - \eta\left(-0.026\right) = -0.5 + 0.026 = -0.474$$