1

Mohamed Ehab Fathy 2002597

1) a) _ First layer

$$y_1^1 = w_1^{1T} x + b_1^1 = \begin{bmatrix} 2 & 2 \\ 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 9 \\ -2 \\ 5 \end{bmatrix}$$

- seCond layer

$$y^2 = w_1^{2T} y_1 + b_1^2 = [3 \quad 1 \quad 2] \begin{bmatrix} 9 \\ -2 \\ 5 \end{bmatrix} + 1 = 36$$

(b) using relu :

$$a_1 = \begin{bmatrix} 9 \\ 0 \\ 5 \end{bmatrix}$$

$$y_2 = [3 \quad 1 \quad 2] \begin{bmatrix} 9 \\ 0 \\ 5 \end{bmatrix} + 1 = 38$$

$$a_2 = 38$$

(c)

$$y = w_1^2 \overbrace{(w_1^T x + b_1)}^{y_1} + b_1^2 \quad , \quad J = (\hat{y} - y)^2$$

$$* \quad \frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial y} * \frac{\partial y}{\partial b_1^2}$$

$$= 2(\hat{y} - y) * 1 = 2 * (32 - 36) = -8$$

$$* \quad \frac{\partial J}{\partial w_{21}^2} = \frac{\partial J}{\partial y} * \frac{\partial y}{\partial w_{21}^2} = 2 * (\hat{y} - y) * y_2^1$$

$$= 2 * -4 * -2 = 16$$

$\boxed{2}$

* $$\frac{\partial J}{\partial b_2^1} = \frac{\partial J}{\partial y} * \frac{\partial y}{\partial y_2^1} * \frac{\partial y_2^1}{\partial b_2^1}$$

$$= 2 * (\hat{y} - y) * w_{21}^2 * 1$$

$$= 2 * (-4) * 1 * 1 = -8$$

* $$\frac{\partial J}{\partial w_{13}^{[1]}} = \frac{\partial J}{\partial y^2} * \frac{\partial y^2}{\partial y_3^1} * \frac{\partial y_3^1}{\partial w_{13}^{[1]}}$$

$$= 2(\hat{y} - y) * w_{31}^2 * x_1 = 2 * -4 * 2 * 1$$

$$= -16$$

1) d) $\cancel{*}$ $b_2^1(t+1) = b_2^1(t) - 2 * \frac{\partial J}{\partial b_1^2}$

$$= 1 - 2 * -8 = 17$$

* $w_{13}^{[1]} = 3 - 2 * -16 = 35$

e) yes, because using a sample that's not used to train the model is a good indicator of out-of-sample error

Scanned with CamScanner

$\boxed{3}$

(2)

(a) $$\frac{\partial F}{\partial x_1} = \frac{\partial F}{\partial g_1} * \frac{\partial g_1}{\partial x_1} + \frac{\partial F}{\partial g_2} * \frac{\partial g_2}{\partial x_1}$$

$$= \cos g_1 * e^{x_2} + 2 g_2 * 1$$

(b) $$\frac{\partial F}{\partial x_2} = \frac{\partial F}{\partial g_1} * \frac{\partial g_1}{\partial x_2} + \frac{\partial F}{\partial g_2} * \frac{\partial g_2}{\partial x_2}$$

$$= \cos g_1 * x_1 e^{x_2} + 2 g_2 * 2 x_2$$

(3) (1) $$\frac{\partial F}{\partial Z} = \frac{e^{-Z}}{(1+e^{-Z})^2} = \frac{1}{1+e^{-Z}}\left(1 - \frac{1}{1+e^{-Z}}\right)$$

$$= F(Z)(1-F(Z))$$

(2) let $Z = w^T x$

$$\frac{\partial F}{\partial w} = \frac{\partial F}{\partial Z} * \frac{\partial Z}{\partial w} = F(w)(1-F(w)) \cdot x$$

(3) $$\frac{\partial J(w)}{\partial w} = \frac{1}{2}\sum_{i=1}^{m} \frac{|w^T x^i - y^i|}{w^T x^i - y^i} \cdot x^i$$

Scanned with CamScanner

④ $$\frac{\partial J(w)}{\partial w} = \frac{1}{2} \sum_{i=1}^{m} 2(w^T x^i - y^i) \cdot x^i + 2\lambda w$$

⑤ let $z_i = w^T x^i$

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

$$J(z) = \sum_{i=1}^{m} \left[y^i \log(\sigma(z_i)) + (1-y^i) \log(1-\sigma(z_i)) \right]$$

$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial \sigma} * \frac{\partial \sigma}{\partial z} * \frac{\partial z}{\partial w}$$

$$= \sum_{i=1}^{m} \left[\frac{y_i}{\sigma(z_i)\ln(10)} + \frac{(1-y_i)}{(1-\sigma(z_i)) \times \ln 10} \right] * \sigma(z_i)(1-\sigma(z_i)) * x^i$$

6) $$\frac{\partial F(w)}{\partial w} = (1 - \tanh^2[w^T x])\, x$$