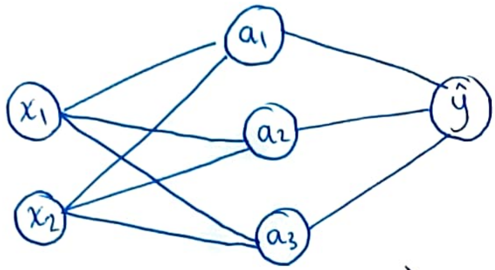
CSE616: Assignment 1

Name: Robeir Samir George
ID : 1901823



$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad y = 32$$

$$\begin{pmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ 2 & -1 & 1 \end{pmatrix}, \quad \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} w_{11}^{(2)} \\ w_{21}^{(2)} \\ w_{31}^{(2)} \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \quad b_1^{(2)} = 1$$

1a when activation functions are <u>identification</u> functions :-

$$a = w^{(1)T} x + b^{(1)} \Rightarrow \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ 2 & -1 & 1 \end{pmatrix}^T \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 2 \\ 1 & -1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 8 \\ -2 \\ 6 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix}$$

$$\hat{y} = w^{(2)T} a + b^{(2)} \Rightarrow \hat{y} = \begin{pmatrix} 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix} + 1 = 27 - 2 + 10 + 1 = \boxed{36} \ \#$$

1b when activation functions are <u>relu</u> functions :-

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 9 \\ 0 \\ 5 \end{pmatrix} \Rightarrow \hat{y} = \begin{pmatrix} 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} 9 \\ 0 \\ 5 \end{pmatrix} + 1$$

$$= 27 + 0 + 10 + 1 = \boxed{38} \ \#$$

1c $J = (\hat{y} - y)^2$ $\hat{y} = w^{T(2)} a + b_1^{(2)}$

- $\dfrac{\partial J}{\partial b_1^{(2)}} = \dfrac{\partial J}{\partial \hat{y}} \cdot \dfrac{\partial \hat{y}}{\partial b_1^{(2)}} = 2(\hat{y} - y) \cdot (1)$
  $= 2(36 - 32) = \boxed{8}$

- $\dfrac{\partial J}{\partial w_{21}^{(2)}} = \dfrac{\partial J}{\partial \hat{y}} \cdot \dfrac{\partial \hat{y}}{\partial w_{21}^{(2)}} = 2(\hat{y} - y) \cdot a_2^{(2)}$
  $= 2(36 - 32)(-2)$
  $= 2(4)(-2) = \boxed{-16}$

- $\dfrac{\partial J}{\partial b_2^{(1)}} = \dfrac{\partial J}{\partial \hat{y}} \cdot \dfrac{\partial \hat{y}}{\partial a_2^{(1)}} \cdot \dfrac{\partial a_2^{(1)}}{\partial b_2^{(1)}}$
  $= 2(\hat{y} - y) \cdot (1) \cdot (1) = 2(36 - 32) = \boxed{8}$

- $\dfrac{\partial J}{\partial w_{13}^{(1)}} = \dfrac{\partial J}{\partial \hat{y}} \cdot \dfrac{\partial \hat{y}}{\partial a_3^{(1)}} \cdot \dfrac{\partial a_3^{(1)}}{\partial w_{13}^{(1)}} = 2(\hat{y} - y)(2)(3)$
  $= 2(4)(2)(3) = \boxed{48}$

1d $b_2^{(1)} = b_2^{(1)} - \eta \dfrac{\partial J}{\partial b_2^{(1)}}$

$= 0 - (2)(8) = \boxed{-16}$

$w_{13}^{(1)} = w_{13}^{(1)} - \eta \dfrac{\partial J}{\partial w_{13}^{(1)}}$

$= 2 - (2) \cdot (48)$

$= 2 - 96$

$= \boxed{-94}$

1e Selecting the best model performing on the test set doesn't mean that it will be a good indicator of the out-of-sample error. As the out-of-sample error is called generalization error, and for the model to be good it has to have low error (high accuracy) on both train and test sets. Also, trying the best performing model on the current test set doesn't mean that the model will have the best preformance - in general - against other models, as the models were not experimented over other test sets having different features. So, the answer is NO, not always this will be an indicator, as we are searching for the best "general" performance (general error)

2
$$\frac{\partial f}{\partial x_i} = \sum_{i=1}^{m} \frac{\partial f}{\partial g_j} \cdot \frac{\partial g_j}{\partial x_i}$$

$$f = \sin g_1 + g_2^2$$
$$g_1 = x_1 e^{x_2}$$
$$g_2 = x_1 + x_2^2$$

- $$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial g_1} \cdot \frac{\partial g_1}{\partial x_1} + \frac{\partial f}{\partial g_2} \cdot \frac{\partial g_2}{\partial x_1}$$

$$= \cos g_1 \cdot e^{x_2} + 2 g_2 \cdot (1)$$

$$= \boxed{e^{x_2} \cos(x_1 e^{x_2}) + 2x_1 + 2x_2^2} \quad \#$$

- $$\frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial g_1} \cdot \frac{\partial g_1}{\partial x_2} + \frac{\partial f}{\partial g_2} \cdot \frac{\partial g_2}{\partial x_2}$$

$$= \cos g_1 \cdot x_1 e^{x_2} + (2 g_2) \cdot (2x_2)$$

$$= x_1 e^{x_2} \cos(x_1 e^{x_2}) + 4x_2(x_1 + x_2^2)$$

$$= \boxed{x_1 e^{x_2} \cos(x_1 e^{x_2}) + 4x_1 x_2 + 4x_2^3} \quad \#$$

3 (1)

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

$$\frac{df}{dz} = \frac{+e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}}$$

$$= f(z) \cdot \frac{1+e^{-z}-1}{1+e^{-z}} = f(z)\left[\frac{1+e^{-z}}{1+e^{-z}} - \frac{1}{1+e^{-z}}\right]$$

$$\frac{df}{dz} = f(z)\left[1 - f(z)\right] \quad \#$$

---

(2)

$$f(w) = \frac{1}{1+e^{-w^T x}} = (1+e^{-w^T x})^{-1}$$

$$\frac{df(w)}{dw} = (-1)(1+e^{-w^T x})^{-2} \cdot (e^{-w^T x})(-x)$$

$$= \frac{x\, e^{-w^T x}}{(1+e^{-w^T x})^2} = x\left[f(w)\,(1-f(w))\right] \quad \#$$

---

(3)

$$f(w) = \frac{1}{2}\sum_{i=1}^{m} \left| w^T x^{(i)} - y^{(i)} \right| = \frac{1}{2}(w^T X - y)^T (w^T X - y)$$

$$= \frac{1}{2}(Xw - y)^T (Xw - y) = \frac{1}{2}\left((Xw)^T - y^T\right)(Xw - y)$$

$$= \frac{1}{2}\left(w^T X^T - y^T\right)(Xw - y) = \frac{1}{2}\left(w^T X^T X w - y^T X w - w^T X^T y + y^T y\right)$$

$$= \frac{1}{2}\left(w^T X^T X w - 2w^T X^T y + y^T y\right)$$

$$\frac{df}{dw} = \frac{1}{2}\left(2X^T X w - 2X^T y\right)$$

$$= \boxed{X^T X w - X^T y} \quad \#$$

④ $J(w) = \frac{1}{2}\left[\sum_{i=1}^{m} (w^T x^{(i)} - y^{(i)})^2\right] + \lambda \|w\|_2^2$

$$\frac{\partial}{\partial w}(w^T x^{(i)} - y^{(i)})^2 = 2(w^T x^{(i)} - y^{(i)}) \cdot x^{(i)}$$

$$\frac{\partial}{\partial w} \lambda \|w\|_2^2 = 2\lambda \|w\|$$

$$\frac{\partial J}{\partial w} = \sum_{i=1}^{m} \left[(w^T x^{(i)} - y^{(i)})\, x^i\right] + 2\lambda \|w\| \quad \#$$

---

⑤ $J(w) = \sum_{i=1}^{m} \left[y^{(i)} \log\left(\frac{1}{1+e^{-w^T x(i)}}\right) + (1-y^{(i)}) \log\left(1 - \frac{1}{1+e^{-w^T x^{(i)}}}\right)\right]$

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

$$= \sum_{i=1}^{m} x^{(i)} \cdot y^{(i)} \cdot \frac{e^{-w^T x^{(i)}}}{1+e^{-w^T x^{(i)}}} + (-x^{(i)})(1-y^{(i)}) \frac{\cancel{1+e^{-w^T x^{(i)}}}}{\cancel{e^{-w^T x^{(i)}}}} \cdot \frac{e^{-w^T x^{(i)}}}{(1+e^{-w^T x^{(i)}})^2}$$

$$= \sum_{i=1}^{m} \frac{x\left(y\left(e^{-w^T x^{(i)}} + 1\right) - 1\right)}{1+e^{-w^T x^{(i)}}} \quad \#$$

---

⑥ $\frac{d}{dw} \tanh(w^T x)$ let $w^T x = z$ $\Rightarrow \frac{d}{dw} \tanh(z) = \underbrace{\frac{d \tanh(z)}{dz}}_{?} \cdot \underbrace{\frac{dz}{dw}}_{x}$

$$\frac{d \tanh(z)}{dz} = \frac{d}{dz} \frac{\sinh(z)}{\cosh(z)} = \frac{\frac{d}{dz}\sinh(z)\cosh(z) - \frac{d}{dz}\cosh(z)\sinh(z)}{\cosh^2(z)}$$

$$= \frac{\cosh^2(z) - \sinh^2(z)}{\cosh^2(z)} = 1 - \frac{\sinh^2(z)}{\cosh^2(z)} = 1 - \tanh^2(z)$$

$$\boxed{\frac{d}{dw} \tanh(w^T x) = x\left(1 - \tanh^2(w^T x)\right)} \quad \#$$

End of Assignment