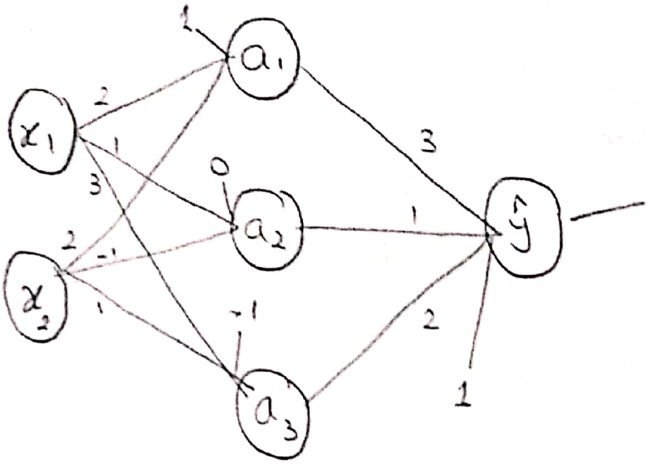

- $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ • $y = 32$

- $\begin{pmatrix} w_{11}^{[1]} & w_{12}^{[1]} & w_{13}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} & w_{23}^{[1]} \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ 2 & -1 & 1 \end{pmatrix}$ • $\begin{pmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

- $\begin{pmatrix} w_{11}^{[2]} \\ w_{21}^{[2]} \\ w_{31}^{[2]} \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$ • $b_1^{[2]} = 1$

1a)

$$\rightarrow W^{[1]T} \cdot X = \begin{bmatrix} 2 & 2 \\ 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 8 \\ -2 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 9 \\ -2 \\ 5 \end{bmatrix}$$

$$\hat{y} = [3 \quad 1 \quad 2] \begin{bmatrix} 9 \\ -2 \\ 5 \end{bmatrix} + 1 = 35 + 1 = \boxed{36}$$

1b) $\mathrm{Relu}(W^{[1]T} X) = \mathrm{Relu}\left(\begin{bmatrix} 9 \\ -2 \\ 5 \end{bmatrix} \right) = \begin{pmatrix} 9 \\ 0 \\ 5 \end{pmatrix}$

$$\hat{y} = \mathrm{Relu}\left([3 \quad 1 \quad 2] \begin{bmatrix} 9 \\ 0 \\ 5 \end{bmatrix} + 1 \right) = \boxed{38}$$

1c) $J = (\hat{y} - y)^2$

$\rightarrow \hat{y} = w_{11}^{[2]} a_1 + w_{21}^{[2]} a_2 + w_{31}^{[2]} a_3 + b_1^{[2]}$

$\rightarrow \frac{\partial J}{\partial b_1^{[2]}} = 2(\hat{y} - y) * 1 = 2(36-32)(1) = \boxed{8}$

$\rightarrow \frac{\partial J}{\partial w_{21}^{[2]}} = 2(\hat{y} - y) * a_2 = 2(36-32)(-2) = \boxed{-16}$

$\rightarrow \hat{y} = w_{11}^{[2]} a_1 + w_{21}^{[2]} a_2 + w_{31}^{[2]} a_3 + b_1^{[2]}$

$= w_{11}^{[2]} a_1 + w_{21}^{[2]} [w_{12}^{[1]} x_1 + w_{22}^{[1]} x_2 + b_2^{[1]}] + w_{31}^{[2]} a_2 + b_1^{[2]}$

$\rightarrow \frac{\partial J}{\partial b_2^{[1]}} = 2(\hat{y} - y) * w_{21}^{[2]} = 2(36-32)(1) = \boxed{8}$

$\rightarrow \hat{y} = w_{11}^{[2]} a_1 + w_{21}^{[2]} a_2 + w_{31}^{[2]} [w_{13}^{[1]} x_1 + w_{23}^{[1]} x_2 + b_3^{[1]}] + b_1^{[2]}$

$\rightarrow \frac{\partial J}{\partial w_{13}^{[1]}} = 2(\hat{y} - y) * w_{31}^{[2]} * x_1 = 2(36-32)(2)(1) = \boxed{16}$

1d) $\eta = 2$

$\rightarrow b_2^{[1]} = b_2^{[1]} - \eta \frac{\partial J}{\partial b_2^{[1]}} = 0 - 2(8) = \boxed{-16}$

$\rightarrow w_{13}^{[1]} = w_{13}^{[1]} - \eta \frac{\partial J}{\partial w_{13}^{[1]}} = 3 - 2(16) = \boxed{-29}$

1e) → The performance on the unseen Test set will be good estimate of out-of-sample error

→ However, This may lead to overfitting on the test set, or choosing a model that performs well just on this subset

→ So, The test set should be diverse and covers big distribution

→ It is also useful To combine both test & Train accuracies when choosing the model

Scanned by CamScanner

2)

$f(g(x)) = f(g_1(x_1 \dots x_n) \dots g_m(x_1 \dots x_n))$

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

$\rightarrow \dfrac{\partial f}{\partial g_1} = \cos g_1$

$\rightarrow \dfrac{\partial f}{\partial g_2} = 2g_2$

$\rightarrow \dfrac{\partial g_1}{\partial x_1} = e^{x_2} \quad \rightarrow \dfrac{\partial g_1}{\partial x_2} = x_1 e^{x_2}$

$\rightarrow \dfrac{\partial g_2}{\partial x_1} = 1 \quad \rightarrow \dfrac{\partial g_2}{\partial x_2} = 2x_2$

$\rightarrow \dfrac{\partial f}{\partial x_1} = \dfrac{\partial f}{\partial g_1}\dfrac{\partial g_1}{\partial x_1} + \dfrac{\partial f}{\partial g_2}\dfrac{\partial g_2}{\partial x_1}$

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

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

$\rightarrow \dfrac{\partial f}{\partial x_2} = \dfrac{\partial f}{\partial g_1}\dfrac{\partial g_1}{\partial x_2} + \dfrac{\partial f}{\partial g_2}\dfrac{\partial g_2}{\partial x_2}$

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

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

3)

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

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

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

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

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

$$\frac{dJ}{dw} = \frac{1}{2}\sum_{i=1}^{m} x^{(i)}$$

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

$$\frac{dJ}{dw} = \sum_{i=1}^{m}(w^T x^{(i)} - y^{(i)})x^{(i)} + 2\lambda \|w\|_2$$

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

$$\frac{dJ}{dw} = y^{(i)}(1+e^{-w^T x^{(i)}})\left(\frac{x^{(i)} e^{-w^T x^{(i)}}}{(1+e^{-w^T x^{(i)}})^2}\right)$$

$$+(1-y^{(i)})\left(\frac{1+e^{-w^T x^{(i)}}}{e^{-w^T x^{(i)}}}\right)\left(\frac{-x^{(i)} e^{-w^T x^{(i)}}}{(1+e^{-w^T x^{(i)}})^2}\right)$$