tomorrow [illegible] corp finance [illegible] 8-

Q1.

(a) $y^T z = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{matrix} 1 \times 2 + \\ 3 \times 3 \end{matrix}$

$= 2 + 9 = \underline{\underline{11}}$

(b) $Xy = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2+12 \\ 1+9 \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$

(c) $X^2 = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$

$= \begin{bmatrix} 4+4 & 8+12 \\ 2+3 & 4+9 \end{bmatrix} = \begin{bmatrix} 8 & 20 \\ 5 & 13 \end{bmatrix}$

(d)

$\det(X) = 2 \times 3 - 4 \times 1 = 6 - 4 = 2 \neq 0$

hence, inverse of X exists.

$X^{-1} = \frac{1}{\det(X)} \begin{bmatrix} 3 & -4 \\ -1 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 3 & -4 \\ -1 & 2 \end{bmatrix}$

(e) ~~rank(X) = 2 i.e. 2 linearly independent rows~~

(e) Solving row echlon form

$$\begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \xrightarrow{(2)-\frac{1}{2}(1)} \begin{bmatrix} 2 & 4 \\ 0 & 1 \end{bmatrix}$$

$$\xrightarrow{(1)-4(2)} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{\frac{1}{2}(1)+(2)}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$ we have 2

independent cols in matrix, hence

Rank $(x) = 2$

Q2

(a) $\frac{\partial y}{\partial x} = 3x^2 + 1$

(b) $\frac{\partial y}{\partial x}$ $y = (5x^3 - 2x)(2x)$

$10x^4 - 4x^2$

$\frac{\partial y}{\partial x} = 40x^3 - 8x = 4x(10x^2 - 2)$

(c) y

f'(

=

ⓒ $y = \frac{2x^2+3}{8x+1}$

$$f'(x) = \frac{u'(x)\,v(x) - u(x)\,v'(x)}{[v(x)]^2}$$

$u(x) = 2x^2 + 3$ $\qquad$ $v(x) = 8x+1$

$u'(x) = 4x$ $\qquad$ $v'(x) = 8$

now, substitute:

$$= f'(x) = \frac{(4x)(8x+1) - (2x^2+3)(8)}{(8x+1)^2}$$

$$f'(x) = \frac{32x^2 + 4x - 16x^2 - 24}{(8x+1)^2}$$

$$f'(x) = \frac{16x^2 + 4x - 24}{(8x+1)^2} \;//$$

ⓓ $y = (3x-2)^8$

$$\frac{dy}{dx} = 8(3x-2)^7 \cdot (3) = 24(3x-2)^7$$

2)

(c) $y = \log(x^2 + x)$

$$\frac{\partial y}{\partial x} = \frac{u'(x)}{u(x)}$$

$u(x) = x^2 + x$
$u'(x) = 2x + 1$

$$\boxed{f'(x) = \frac{2x+1}{x^2+x}}$$

Task 3

$$CE(y, \hat{y}) = -\sum_{\Delta} y_i \log(\hat{y}_i)$$

output $= \hat{y} = \text{softmax}(Q)$

$$\frac{\partial CE(y, \hat{y})}{\partial Q} \;?$$

softmax function:

$$S_i = \frac{e^{z_i}}{\sum_l e^{z_l}} \quad : R^N \to R^N$$

softmax function:

$$\sigma(\theta_j) = \frac{e^{\theta_j}}{\sum_{i=1}^{D} e^{\theta_i}} \qquad j \in \{1, 2, \ldots D\}$$

D-dimensional

computing

$$\frac{\partial \sigma(z_j)}{\partial z_k} = \frac{\partial}{\partial z_j} \cdot e$$

$$\frac{\partial \sigma(\theta_j)}{\partial \theta_k} = \frac{\partial}{\partial \theta_j} \cdot \frac{e^{\theta_j}}{\sum_D e^{\theta_i}}$$ when $j = K$

applying quotient rule

$$= \frac{\frac{\partial e^{\theta_j}}{\partial \theta_j} \cdot \sum_D e^{\theta_i} - e^{\theta_j} \cdot \frac{\partial \sum_D e^{\theta_i}}{\partial \theta_j}}{\left(\sum e^{\theta_i}\right)^2}$$

$$= \frac{e^{\theta_j}}{\sum e^{\theta_i}} - \left(\frac{e^{\theta_j}}{\sum e^{\theta_i}}\right)^2$$

$$\sigma(\theta_j) - (\sigma(\theta_j))^2$$

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

when $j \neq k$

$$\frac{\partial \sigma(\theta_j)}{\partial \theta_k} = \frac{\partial}{\partial \theta_k} \cdot \frac{e^{\theta_j}}{\sum_i e^{\theta_i}}$$

$$= \frac{\frac{\partial e^{\theta_j}}{\partial \theta_k} \cdot \sum_i e^{\theta_i} - e^{\theta_j} \cdot \frac{\partial \sum_i e^{\theta_i}}{\partial \theta_k}}{\left(\sum_i e^{\theta_i} \right)^2}$$

by quotient rule

$$= \frac{-e^{\theta_j}}{\sum e^{\theta_i}} \cdot \frac{e^{\theta_k}}{\sum e^{\theta_i}}$$

$$= -\sigma(\theta_j) \cdot \sigma(\theta_k)$$

Hence, we formulate the derivative of softmax (θ) as:

$$\frac{\partial \sigma(\theta_j)}{\partial(\theta_k)} = \begin{cases} \sigma(\theta_j)(1-\sigma(\theta_j)) & \text{when } j = k \\ -\sigma(\theta_j) \cdot \sigma(\theta_k) & \text{when } j \neq k \end{cases}$$

finally, to compute $\frac{\partial CE(y,\hat{y})}{\partial Q}$

$CE = \sum_{\partial D} (-y_j \log \sigma(z_j))$

we use chain rule:

$$\frac{\partial}{\partial Q_k} CE = \frac{\partial}{\partial Q_k} \cdot \sum_{j=i}^{D} (-y_j \log \sigma(Q_j))$$

$$= -\sum_{j=1}^{D} y_j \frac{\partial}{\partial Q_k} \log \sigma(Q_j)$$

$$= -\sum_{j=1}^{D} y_j \frac{1}{\sigma(Q_j)} \cdot \frac{\partial}{\partial Q_k} \cdot \sigma(Q_j)$$

$$= -y_k \cdot \frac{\sigma(Q_k)(1-\sigma(k))}{\sigma(Q_k)} - \sum_{j \neq k} y_j \cdot \frac{-\sigma(Q_j)\sigma(Q_k)}{\sigma(Q_j)}$$

case considering both $j=k$ and $j \neq k$

$$\cancel{= y_k \cdot (1 - \sigma(Q_k)) + \sum_{j \neq k} y_j \frac{-\sigma(Q_j)\sigma(Q_k)}{\sigma(Q_j)}}$$

$$-y_k \cdot (1 - \sigma(\theta_k)) + \sum_{j \neq k} y_j \sigma(\theta_k)$$

$$= -y_k + y_k \sigma(\theta_k) + \sum_{j \neq k} y_j \sigma(\theta_k)$$

$$\boxed{= -y_k + \sigma(\theta_k) \sum_j y_j}$$

Since y is normalised,

$$\boxed{\sum_j y_j = 1}$$

Here $\frac{\partial}{\partial \theta_k} CE = \sigma(\theta_k) - y_k$

Total parameters of the neural network

number of weights connecting input layer to hidden layer $= D_x \times D_h = D_x D_h$

Additionally, bias term 1 per neuron in hidden layer $= D_h$ total bias terms

$$= D_x D_h + D_h = D_h (D_x + 1).$$

$\Rightarrow$ Total parameters from Hidden to output layer:

$$D_h \cdot D_y + D_y \text{ (bias terms)}$$
$$D_y (D_h + 1).$$

total Parameters $= D_h (D_x + 1) + D_y (D_h + 1)$.