Com S 574	Spring Semester 2021
Machine Learning	
Final exam	
Instructor: Forrest Bao	Submitted by: Ibne Farabi Shihab

1 no question

Given that:

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/3 & 1/2 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.5 & 1 & 6 \\ 3 & -4 & 2 \end{bmatrix}$$

So,

$$A \circ B = \begin{bmatrix} 1 \times 0.5 & 1/2 \times 1 & 1/3 \times 6 \\ 1/3 \times 3 & 1/2 \times -4 & 1 \times 2 \end{bmatrix}$$

$$A \circ B = \begin{bmatrix} 0.5 & 0.5 & 2 \\ 1 & -2 & 2 \end{bmatrix}$$

2 no Question

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/3 & 1/2 & 1 \end{bmatrix}$$

$$B^T = \begin{bmatrix} 0.5 & 3 \\ 1 & -4 \\ 6 & 2 \end{bmatrix}$$

So,

$$AB^T = \begin{bmatrix} 3 & 5/3 \\ 20/3 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.5 & 1 & 6 \\ 3 & -4 & 2 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & 1/3 \\ 1/2 & 1/2 \\ 1/3 & 1 \end{bmatrix}$$

So, $BA^T = \begin{bmatrix} 3 & 20/3 \\ 5/3 & 1 \end{bmatrix}$

3 no Question

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/3 & 1/2 & 1 \end{bmatrix}$$

$$B^T = \begin{bmatrix} 0.5 & 3 \\ 1 & -4 \\ 6 & 2 \end{bmatrix}$$

$$AB^T = \begin{bmatrix} 3 & 5/3 \\ 20/3 & 1 \end{bmatrix}$$

$f(x) = x + 1$

So,

$f(AB^T) = AB^T + 1$

$$AB^T = \begin{bmatrix} 4 & 2.667 \\ 7.667 & 2 \end{bmatrix}$$

4 no Question

Given,

$x = \begin{bmatrix} x_0, x_1, x_2, x_3 \end{bmatrix}$

$= \begin{bmatrix} 1/2, 1/3, 1/4, 1/5 \end{bmatrix}$

$$w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

Prediction -

$\hat{y} = \phi(W^T X)$

$= \phi(W_0 X_0 + W_1 X_1 + W_2 X_2 + W_3 X_3))$

$= \phi(4)$

$= 4^2 = 16$

5 no Question

$\frac{\partial E}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}(\hat{y} + y)$

$= 1 + 0$

$= 1$

$\frac{\partial \hat{y}}{\partial W^T X} = \frac{\partial}{\partial W^T X}(W^T X)^2$

$= 2 * W^T X$

$= 2 * 4 = 8$

$\frac{\partial W^T X}{\partial X_1} = \frac{\partial}{\partial X_1} W_0X_0 + W_1X_1 + W_2X_2 + W_3X_3$

$= W_1 = 3$

$\frac{\partial E}{\partial X_1} = 1$*8*3 (From last 3 derivation)

$= 24$

6 no Question

$$\frac{\partial E}{\partial X} = \begin{bmatrix} \frac{\partial E}{\partial X_0} \\ \frac{\partial E}{\partial X_1} \\ \frac{\partial E}{\partial X_2} \\ \frac{\partial E}{\partial X_3} \end{bmatrix} = \begin{bmatrix} 8 \times 2 \\ 8 \times 3 \\ 8 \times 4 \\ 8 \times 5 \end{bmatrix} = \begin{bmatrix} 16 \\ 24 \\ 32 \\ 40 \end{bmatrix}$$

$$\frac{\partial E}{\partial W} = \begin{bmatrix} \frac{\partial E}{\partial W_0} \\ \frac{\partial E}{\partial W_1} \\ \frac{\partial E}{\partial W_2} \\ \frac{\partial E}{\partial W_3} \end{bmatrix} = \begin{bmatrix} 8 \times 1/2 \\ 8 \times 1/3 \\ 8 \times 1/4 \\ 8 \times 1/5 \end{bmatrix} = \begin{bmatrix} 4 \\ 8/3 \\ 2 \\ 8/5 \end{bmatrix}$$

7 No Question

$W^{(0)} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 0.1 & 0.1 & 0.1 \end{bmatrix}$ and

$x^{(0)} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$

$x^{(1)} = \phi(W^{(0)T} x^{(0)})$

$= \phi(\begin{bmatrix} 0.1 & 0.1 & 0.1 \end{bmatrix}^T)$

$$= \phi(\begin{bmatrix} 1 & 0.1 & 0.1 & 0.1 \end{bmatrix}) \text{ (Adding bias)}$$
$$x^{(2)} = \phi(W^{(1)T} x^{(1)})$$

$$= \phi(\begin{bmatrix} 0.65 & 0.65 \end{bmatrix}^T)$$
$$= \phi(\begin{bmatrix} 1 & 0.65 & 0.65 \end{bmatrix}) \text{ (Adding bias)}$$
$$x^{(3)} = \phi(W^{(2)T} x^{(2)})$$

$$= \phi(\begin{bmatrix} 0.575 & 0.575 \end{bmatrix}^T)$$
$$= \phi(\begin{bmatrix} 1 & 0.575 & 0.575 \end{bmatrix}) \text{ (Adding bias)}$$

8 No Question

$\delta^3 = (\hat{y} - y)^2$ is a 2*1 matrix where the prediction $\hat{y} = x^{(3)}$ and y is the target

$\delta^2 = x^{(2)} \circ (1 - x^{(2)}) \circ (w^{(2)} \delta^{(3)})$

$\delta^1 = x^{(1)} \circ (1 - x^{(1)}) \circ (w^{(1)} \delta^{(2)})$

$\delta^0 = x^{(0)} \circ (1 - x^{(0)}) \circ (w^{(0)} \delta^{(1)})$

here, $\delta^l - 1$ is the error from the previous layer

9 No Question

A model without regularization will have the chance of overfitting and this is because as the model might try to memorize the data instead of learning the generalization. For this a model with regularization might work better in general than model without regularization as it reduces the chances of learning a complex model.

10 No Question

For unbalanced dataset it is not wise to use the normal accuracy as model will predict the dominant most of the cases resulted in a high accuracy but poor performance as the model is not showing the overall picture.

11 No Question

Any optimal solution to K mean problem will have each element in M be the mean of some subset of the data points - hence, k-means. K means is a NP-hard problem.

For every iteration, the number of operations = [k*m*n] operations + [(k-1)*m*n] operations + [k*((m-1) + 1)*n] operations.

If the algorithm converges within ith iterations, then the operations = 6*[i*k*m*n] operations + [i*(k-1) *m*n] operations + [i*k*((m-1) + 1) *n] operations Therefore, the time complexity is O(i*k*m*n)