Subject Code: AML 5102

Subject Name: Deep Learning Principles and Applications

Quiz

1. What does $x_2^{(3)}$ represent in the following data matrix:

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

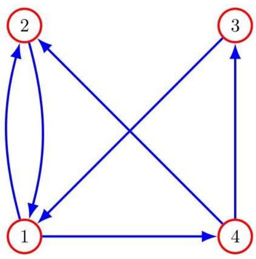
   a. BP of the 3rd patient
   b. BP of the 2nd patient
   c. TEMP of the 3rd patient
   d. TEMP of the 2nd patient

2. What does $\frac{x^{(1)} + x^{(2)} + x^{(3)} + x^{(4)}}{4}$ represent for the following data matrix:

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

   a. Average patient
   b. Average HR
   c. Average BP
   d. Average TEMP

3. Suppose we have four stations that are connected by train services as shown in the following graph:



The adjacency matrix A associated with this graph has entries defined as

$$a_{ij} = \begin{cases} 1 & \text{if direct train service exists from station } j \text{ to station } i, \\ 0 & \text{otherwise.} \end{cases}$$

Choose the correct adjacency matrix from the options below:

a. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$

b. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$

c. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$

d. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

4. A train network comprising *20* stations and *50* paths is represented by the 50×20-matrix P whose entries are defined as

$$p_{ij} = \begin{cases} 1, & \text{if station } j \text{ is on route } i, \\ 0, & \text{otherwise;} \end{cases}$$

Suppose the 5th column of P is sparse; that is, only a few entries of that column are nonzero values while most of the other entries are zeros. Based on this information, which one of the following statement is correct:

a. Station-5 shows up only on a few paths
b. Path-5 has only a few stations
c. Station-5 never shows up on Path-5
d. Path-5 does never contains Station-5

5. Suppose we measured for 100 patients at 24 hourly timestamps (starting from 00:00 hours) 3 features (heart rate, blood pressure, and temperature). The resulting 3D tensor P has the shape structure *patients* × *timestamps* × *features*. What does the entry $p_{2,3,3}$ of the tensor represent?

a. Temperature of 2nd patient at 2PM
b. Temperature of 2nd patient at 2AM
c. Temperature of 3rd patient at 2AM

d. Tempetaure of 3rd patient at 2PM

6. The MAHE registrar has the complete list of courses taken by each graduating student in a program. This data is represented as an $m \times n$-matrix X such that

$$x_{ij} = \begin{cases} 1 & \text{if student } i \text{ has taken course } j, \\ 0 & \text{otherwise.} \end{cases}$$

The dot product $x^{(4)} \cdot \mathbf{1}$, where $\mathbf{1}$ is the vector whose entries are all equal to 1, gives the

a. number of students in the 4th course
b. number of courses taken by the 4th student
c. number of students who have not taken the 4th course
d. number of courses not taken by the 4th student

7. The MAHE registrar has the complete list of courses taken by each graduating student in a program. This data is represented as an $m \times n$-matrix X such that

$$x_{ij} = \begin{cases} 1 & \text{if student } i \text{ has taken course } j, \\ 0 & \text{otherwise.} \end{cases}$$

The number of students who have taken both the 5th and 6th courses is the dot product

a. $x_5 \cdot x^{(6)}$
b. $x^{(5)} \cdot x^{(6)}$
c. $x^{(5)} \cdot x_6$
d. $x_5 \cdot x_6$

8. The result of the matrix-vector product $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}$ is

a. $\begin{bmatrix} x_5 \\ x_3 \\ x_6 \end{bmatrix}$

b. $\begin{bmatrix} x_3 \\ x_5 \\ x_6 \end{bmatrix}$

c. $\begin{bmatrix} x_5 \\ x_6 \\ x_3 \end{bmatrix}$

d. $\begin{bmatrix} x_4 \\ x_5 \\ x_6 \end{bmatrix}$

9. What is the effect of multiplying a vector x by the matrix $\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$?
   a. Cycle the components of x downward by 1 step
   b. Cycle the components of x upward by 1 step
   c. Switch the 2nd and 5th components of x
   d. Switch the 1st and 5th components of x

10. Suppose we have the following patient data matrix X:

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

where HR is measured in *beats per minute*. Suppose we want to convert the HR values in X to *beats per second*. Which one of the following matrix-matrix product achieves that?

   a. DX, where D = $\begin{bmatrix} 1/60 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
   b. DX, where D = $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/60 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
   c. XD, where D = $\begin{bmatrix} 1/60 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
   d. XD, where D = $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/60 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

11. Suppose we have 10^4 samples corresponding to 10 output labels and that each sample is a 32 x 32 grayscale image. What will be the shape of the weights matrix used for calculating the sample's raw scores as $z = Wx + b$?
    a. 11 x 1024
    b. 1024 x 11
    c. 1024 x 10
    d. 10 x 1024

12. Suppose we have a sample that can correspond to 5 possible output classes. If the raw scores for that sample using some weights and biases for all the classes are equal, what is the multiclass SVM hinge loss for this sample?
    a. 4
    b. 5
    c. 0
    d. 6

13. Consider a dataset in which a sample has 2 output classes (survived/not survived) and 5 features:

    (1) heart rate (2) blood pressure (3) temperature (4) age (5) gender.

    The 3rd column of the weights matrix corresponds to what?

    a. Weights for the "survived" output class
    b. Weights for the "not survived" output class
    c. Weights for the feature "Blood Pressure"
    d. Weights for the feature "Temperature"

14. The sensitivity of the loss with respect to some weight parameter evaluated at its current value is -10. This means that decreasing this weight (by a tiny amount) would ________ the loss
    a. increase
    b. decrease
    c. not change
    d. negate

15. Suppose $L(\mathbf{w}) = ||\mathbf{w}||^2$. What is $\nabla_{\mathbf{w}}(L)$?
    a. $\mathbf{w}$
    b. $2\mathbf{w}$

c. $\mathbf{w}/2$
d. $4\mathbf{w}$

16. When performing the bias trick, the bias feature takes the value
    a. 1
    b. 0
    c. -1
    d. any fixed positive value

17. Suppose we have 10^3 samples corresponding to 10 output classes and that each sample is a 32 x 32 grayscale image. Suppose the samples are arranged columnwise in the data matrix and that the bias trick has been applied. What will be the shape of the data matrix?
    a. 1024 x 10
    b. 1025 x 1000
    c. 1000 x 1025
    d. 1024 x 1000

18. Consider a dataset in which a sample x can belong to either the "survived" class (labeled as 1) or the "not survived" class (labeled as 0). Suppose there are 4 patients in the dataset in which the first two survived and the last two did not. Using a particular weight vector w (bias trick applied), we get:

$$\begin{aligned} \sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(1)}\right) &= 0.1, \\ 1-\sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(2)}\right) &= 0.2, \\ \sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(3)}\right) &= 0.95, \\ 1-\sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(4)}\right) &= 0.15. \end{aligned}$$

    Without calculation, we can say that the loss is the highest for patient
    a. 1
    b. 2
    c. 3
    d. 4

19. Which if the following is <u>not</u> true about regularization?
    a. It shrinks the weight values closer to zero
    b. It helps prevent model overfitting
    c. It is applied to both weights and biases
    d. It helps the model to generalize well for unseen test data

20. Suppose we apply batch processing for a dataset with 1024 samples using a batch size 16. How many times will the weights be updated in one epoch?
    a. 64
    b. 16
    c. 32
    d. 8
21. Which one of the following statements is <u>not true</u> about a softmax activation layer in a zero hidden layer neural network architecture?
    a. It converts raw scores to probabilities
    b. It is fully connected to the dense layer before
    c. It is node-wise connected to the dense layer before
    d. It has the same number of nodes as the dense layer before

22. Which one of the following activation functions does not have the vanishing gradient problem?
    a. Sigmoid
    b. tanh
    c. ReLU
    d. Leaky ReLU

23. Which one of the following activation functions clips all negative raw scores to zeros?
    a. Sigmoid
    b. tanh
    c. ReLU
    d. Leaky ReLU

24. Which one of the following activation functions results in mean-centered activated values?
    a. Sigmoid
    b. tanh
    c. ReLU
    d. Leaky ReLU

25. Suppose we are dealing with a classification problem in which a sample can belong to one of 5 possible output categories labeled from 0 through 4 (Python indexing

style). Which one of the following is the correct one-hot encoded representation for a sample with output label 3?

a. [0, 0, 1, 0, 1]
b. [0, 0, 0, 0, 1]
c. [0, 0, 1, 0, 0]
d. [0, 0, 0, 1, 0]

26. The CCE loss for a sample with correct probability vector [0, 1, 0] and predicted probability vector [0.75, 0.15, 0.1] is

a. -log(0.75)
b. -log(0.1)
c. -log(0.15)
d. -log(0.85)

27. Suppose we apply a softmax classifier for a classification problem with 10 possible output categories. What is the shape of the gradient $\nabla_{\mathbf{z}} (\mathrm{softmax}(\mathbf{z}))$?

a. 11 x 11
b. 10 x 10
c. 9 x 9
d. 10 x 11

28. Suppose we use a single hidden layer neural network for a classification problem with 10 features per sample. If there are 6 nodes in the dense layer of the hidden layer, what is the shape of the weights matrix $\mathbf{W}^{[1]}$ if we ignore the bias feature?

a. 6 x 10
b. 6 x 11
c. 10 x 6
d. 11 x 6

29. What is the shape of the gradient $\nabla_{\mathbf{W}}(\mathbf{W}\mathbf{x})$ for a 5 x 10-matrix $\mathbf{W}$ and a 10-vector $\mathbf{x}$?

a. 5 x 10 x10
b. 10 x 5 x 10
c. 5 x 10
d. 10 x 10 x 5

30. Which one of the following is not a hyperparameter for a neural network?
    a. Weights
    b. Learning rate
    c. Regularization strength
    d. Number of nodes in the hidden layer

31. The number of hidden layers in a 5-layer neural network is
    a. 6
    b. 5
    c. 4
    d. 3

32. What does $z_5^{[3]}$ represent in a 6-layer deep neural network (layer indexing starts from 0 and node indexing starts from 1)?
    a. The raw score calculated by the 3rd neuron in hidden layer 5
    b. The raw score calculated by the 5th neuron in hidden layer 3
    c. The raw score calculated by the 5th neuron in hidden layer 2
    d. The raw score calculated by the 4th neuron in hidden layer 5

33. In a 3-layer neural network with $n^{[0]} = 10, n^{[1]} = 8, n^{[2]} = 3$, the shape of the matrix $\mathbf{W}^{[1]}$ is
    a. 8 x 11
    b. 11 x 8
    c. 10 x 8
    d. 8 x 10

34. When running a batch of size 32 through an L-layer deep neural network, where each sample could possibly belong to one of 3 output categories, the shape of the raw scores matrix $\mathbf{Z}^{[L]}$ is
    a. 3 x 32
    b. 32 x 3
    c. 32 x 32
    d. 3 x 3

35. A 2-layer neural network with 5 neurons in each layer has a total of __ parameters (i.e. weights and biases).

a. 59
b. 60
c. 61
d. 62

36. Which one of the following is the correct categorical cross-entropy loss expression for a sample with correct one-hot encoded output label vector $\mathbf{y}$ when using a 5-layer neural network?
a. $-\sum_k a_k^{[5]} \log(y_k)$
b. $-\sum_k y_k \log\left(a_k^{[5]}\right)$
c. $-\sum_k y_k \log\left(z_k^{[5]}\right)$
d. $-\sum_k y_k \log\left(a_k^{[6]}\right)$

37. The gradient flowing backward from the output direction through an activation layer with layer index 3 of a deep neural network is
a. $\nabla_{\mathbf{z}^{[3]}}(L)$
b. $\nabla_{\mathbf{a}^{[2]}}(L)$
c. $\nabla_{\mathbf{z}^{[3]}}\left(\mathbf{a}^{[3]}\right)$
d. $\nabla_{\mathbf{a}^{[3]}}(L)$

38. The local gradient of an activation layer with layer index 3 of a deep neural network is
a. $\nabla_{\mathbf{a}^{[3]}}\left(\mathbf{z}^{[3]}\right)$
b. $\nabla_{\mathbf{z}^{[2]}}\left(\mathbf{a}^{[3]}\right)$
c. $\nabla_{\mathbf{z}^{[3]}}\left(\mathbf{a}^{[3]}\right)$
d. $\nabla_{\mathbf{a}^{[3]}}\left(\mathbf{z}^{[2]}\right)$

39. The local gradient of an activation layer (ReLU) with layer index 3 of a deep neural network with $\mathbf{z}^{[3]} = \begin{bmatrix} -1 \\ -10 \\ 10 \\ 4 \end{bmatrix}$ is

e. $$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

f. $$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

g. $$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

h. $$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

40. In a 2-layer neural network with 5 neurons in each layer, the shape of the gradient $\nabla_{\mathbf{W}^{[1]}}\left(\mathbf{z}^{[1]}\right)$ (local weight gradient of dense layer 1) is
   a. 5 x 6 x 5
   b. 6 x 5 x 5
   c. 5 x 5 x 6
   d. 6 x 6 x 5