Editorial: End Module B Quiz - A1 -> June 24

B4 Module B - End Exam - Attempt 1

Question 1 (MCQ)

Ravi is building a pricing model for a ride-sharing app called "SwiftRyde." After analyzing historical trip data, he notices a consistent trend: for every additional kilometer traveled, the fare increases at a steady rate, unaffected by traffic or surge pricing. The relationship looks clean and predictable.

Which regression technique should Ravi use to best model this direct, linear relationship between distance and fare?

Options: A) Linear Regression B) Logistic Regression C) Polynomial Regression D) Ridge Regression

Answer: A

Question 2 (MCQ)

Priya is designing a system that predicts whether a bank customer will default on a loan. The data includes information like income, credit history, and previous loan defaults. The system needs to predict one of two outcomes: “Default” or “No Default.”

Which type of machine learning algorithm should she choose for this binary prediction task?

Options: A) K-Means Clustering B) Logistic Regression C) Linear Regression D) Principal Component Analysis

Answer: B

Question 3 (NAT)

Consider the directed graph with 5 nodes: P, Q, R, S, and T, and the following edges:

P → Q P → R Q → R R → S S → T T → P

What is the outdegree of node P?

Answer: 2

Question 4 (MSQ)

A machine learning engineer is evaluating different regularization techniques to control overfitting in a linear regression model with many correlated features.

Which of the following statements correctly describe characteristics of L1 (Lasso) and L2 (Ridge) regularization?

Options: A) L1 regularization can completely eliminate some features by assigning their coefficients to zero. B) L2 regularization uses the square of the coefficients in its penalty term. C) L1 regularization tends to work better when most input features are highly correlated. D) L2 regularization typically retains all features but shrinks their coefficients.

Answers: A, B, D

Question 5 (NAT)

A dataset originally has 15 input features. After applying Lasso regression, 5 features retain non-zero coefficients. However, when Ridge regression is applied on the same dataset, all 15 features retain non-zero coefficients.

How many features were effectively eliminated (set to zero) by Lasso but not by Ridge?

Answer: 10

Question 6 (MCQ)

During the execution of the K-Means clustering algorithm, each data point is grouped based on the closest cluster center using a distance metric like Euclidean distance.

Which step of the K-Means algorithm does this describe?

Options: A) Initialization – where initial centroids are chosen randomly B) Assignment – where each data point is assigned to its nearest centroid C) Update – where centroids are recalculated based on assigned points D) Convergence – where the algorithm stops when centroids no longer change

Answer: B

Question 7 (MCQ)

What will be the output of the following Python code snippet, which implements a basic ReLU activation function?

```
import numpy as np


def relu(x):
    return np.maximum(0, x)


data = np.array([-3, -8, -5, -1, -2])
activated_data = relu(data)
```

print(activated_data)

Options: A) [3 8 5 1 2] B) [-3 -8 -5 -1 -2] C) [0 0 0 0 0] D) [1 1 1 1 1]

Answer: C

Question 8 (MCQ)

A researcher is working with an image dataset where each image is represented by thousands of pixel values. To reduce computational load and extract meaningful patterns without using all raw pixel values, they decide to apply a dimensionality reduction technique.

Which method is most suitable for this task?

Options: A) Batch Normalization B) One hot Label Encoding C) Principal Component Analysis D) K-Means Clustering

Answer: C

Question 9 (MCQ)

A machine learning engineer is designing a neural network (NN) to classify customer reviews as positive, negative, or neutral. They are encountering issues with model convergence and generalization. Which of the following is a crucial practice or component for building an effective NN in this multi-class classification scenario?

Options: A) The number of hidden layers and neurons should always be maximized to prevent underfitting. B) The activation function for the output layer should be softmax for multi-class classification problems to output probabilities for each class. C) L1 or L2 regularization should only be applied to the bias terms, not the weights, to reduce overfitting. D) Dropout layers should be used at the input layer to randomly set input features to zero during training.

Answer: B

Question 10 (MCQ)

An AI researcher is developing a Convolutional Neural Network (CNN) for autonomous vehicles to detect pedestrians and traffic signs from camera feeds. They are struggling with the model's performance on diverse real-world images and computational efficiency. Which of the following design principles is commonly applied in effective CNN architectures?

Options: A) Using larger filter sizes in the convolutional layers will always improve the model's ability to capture fine-grained details. B) Pooling layers (e.g., max pooling or average pooling) are primarily used to increase the spatial dimensions of the feature maps,

enhancing detail. C) The number of filters in a convolutional layer typically increases with depth to allow the network to learn more complex and abstract features from the input. D) Batch normalization should only be applied to the output layer to stabilize predictions, not hidden layers.

Answer: C

Question 11 (MCQ)

A researcher is developing a machine learning model to classify different species of plants from images, with 10 distinct species. The final layer of their neural network outputs raw scores for each species. Which of the following statements accurately describes the primary purpose and characteristic of the Softmax activation function in this multi-class classification context?

Options: A) The Softmax function is primarily used to convert a single raw score into a binary classification outcome (0 or 1). B) Softmax outputs a probability distribution over multiple classes, ensuring that all probabilities are non-negative and sum to 1, making it suitable for multi-class classification. C) When applied in a neural network, Softmax prevents the vanishing gradient problem by scaling large output values. D) The Softmax function requires integer-only inputs and outputs only integer probabilities.

Answer:* B

Question 12 (MSQ)

A machine learning engineer is building a house price prediction model using a linear regression algorithm. During training, the model's loss is decreasing very slowly, and training is taking longer than expected.

Which of the following actions can be valid solutions or insights regarding Gradient Descent in this case?

Options: A) Reduce the learning rate to make smaller updates and avoid overshooting. B) Increase the learning rate to escape local minima and speed up convergence. C) Use stochastic or mini-batch gradient descent to speed up training. D) Normalize input features to ensure faster convergence of gradient descent.

Answers: B, C, D

Question 13 (MSQ)

A startup is using a neural network to classify email messages as spam or not spam. They notice the model is underfitting even after multiple epochs of training.

Which of the following might help improve the performance of their neural network?

Options: A) Increasing the number of hidden layers and neurons. B) Adding dropout layers to reduce overfitting. C) Using a more complex activation function like ReLU or tanh instead of sigmoid. D) Gathering more training data to better capture patterns.

Answers: A, C, D

Question 14 (MSQ)

A data analyst is training a recommendation system on a large dataset of user-movie ratings. The training set consists of 10 million data points. They experiment with different optimization strategies to balance speed and convergence quality.

Which of the following statements correctly describe the characteristics or trade-offs of the different types of gradient descent?

Options: A) Batch Gradient Descent computes the gradient using the entire dataset, which can be slow but gives a stable descent path. B) Stochastic Gradient Descent updates weights for each individual training example, often resulting in noisy convergence but faster updates. C) Mini-batch Gradient Descent offers a compromise between speed and convergence stability by using small subsets of data per update. D) Batch Gradient Descent is more suitable when training data doesn't fit in memory, as it handles smaller chunks at a time.

Answers: A, B, C

Question 15 (NAT)

In Support Vector Machines, support vectors are the data points that lie closest to the decision boundary and influence the position and orientation of the hyperplane. True or False? If this statement is true, enter 1. If it is false, enter 0.

Answer: 1

Question 16 (NAT)

Below is part of a CNN model written for the MNIST dataset (which contains grayscale images of handwritten digits, 28×28 in size).

```
class DigitCNN(nn.Module):
  def __init__(self):
    super(DigitCNN, self).__init__()
    self.features = nn.Sequential(
      nn.Conv2d(1, 16, 3, padding=1),
```

```
        nn.ReLU(),
        nn.MaxPool2d(2, 2),

        nn.Conv2d(16, 32, 3, padding=1),
        nn.ReLU(),
        nn.MaxPool2d(2, 2)
    )
    self.classifier = nn.Sequential(
        nn.Flatten(),
        nn.Linear(32*7*7, _____________________),
    )
```

What value should replace the blank in the final Linear layer to correctly output class probabilities?

Answer: 10

Question 17 (MSQ)

A hospital uses a machine learning model to predict whether a patient has a rare disease. The following confusion matrix summarizes the model's performance on 100 test cases:

7(TP) 3(FN) 1(FP) 89(TN)

The vlalues are as per the standard format entry of a confusion matrix. Which of the following statements are true based on this confusion matrix?

Options: A) The model's accuracy is greater than 95%. B) There are more false negatives than false positives. C) The model's recall is 70%. D) The precision of the model is 80%.

Answers: A, B, C

Question 18 (MCQ)

A 7×7 image is passed through a 3×3 convolution filter with no padding and a stride of 2. What will be the width and height of the output feature map?

Options: A) 5×5 B) 3×3 C) 2×2 D) 4×4

Answer: B

Question 19 (MCQ)

Which of the following statements correctly describes a key difference between Max Pooling and Average Pooling in Convolutional Neural Networks?

Options: A) Max Pooling tends to preserve the most dominant features, while Average Pooling smooths the feature map. B) Max Pooling reduces overfitting more than Average Pooling because it uses more parameters. C) Average Pooling introduces non-linearity, while Max Pooling is a linear operation. D) Max Pooling and Average Pooling always produce identical results on the same input.

Answer: A

Question 20 (NAT)

Apply 2×2 max pooling on the following input matrix:

[[4, 1, 3, 2], [8, 6, 5, 7], [0, 9,11,10], [2,13,14,12]]

What is the sum of the pooled output?

Answer: 42

Question 21 (MCQ)

Which of the following hyperplanes does an SVM aim to select during training?

Options: A) The one that minimizes the classification error on the training data. B) The one that maximizes the margin between the support vectors. C) The one that passes through the maximum number of data points. D) The one that minimizes the variance of the data.

Answer: B

Question 22 (MSQ)

A data scientist is applying DBSCAN to a dataset with irregularly shaped clusters. Which of the following are valid characteristics or behaviors of DBSCAN?

Options: A) It can detect noise points (outliers) in the dataset. B) It requires specifying the number of clusters beforehand. C) It can handle clusters of varying shapes and densities. D) It uses two parameters: epsilon and Minimum neighbors.

Answers: A, C, D

Question 23 (MCQ)

In the K-Means algorithm, what does the “K” refer to?

Options: A) The number of iterations B) The number of dimensions in the input space C) The number of clusters to form D) The number of features selected

Answer: C

Question 24 (MCQ)

What type of problems can a single-layer perceptron solve?

Options: A) Any non-linear classification problem B) Only linearly separable classification problems C) Regression problems D) Clustering problems

Answer: B

Question 25 (MCQ)

Which of the following best describes the goal of Principal Component Analysis (PCA)?

Options: A) Transforming categorical data into numerical form B) Increasing the number of features in the dataset C) Reducing dimensionality while retaining most variance D) Normalizing feature scales

Answer: C

Question 26 (MSQ)

During forward and backward propagation in a neural network, which of the following happen?

Options: A) Forward propagation computes activations layer by layer B) Backward propagation updates weights using gradients C) Gradients are calculated using chain rule (backpropagation) D) Forward propagation minimizes the loss function

Answers: A, B, C

Question 27 (MCQ)

Which of the following code snippets correctly adds a Dense (fully connected) layer with 64 units and ReLU activation to a Keras Sequential model?

A)

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

model = Sequential()

model.add(Dense(64, activation='relu')))

B)

from tensorflow.keras.models import Model

from tensorflow.keras.layers import Dense

model = Model()

model.add(Dense(64, activation='relu'))

C) ```python from tensorflow.keras.models import Sequential from tensorflow.keras.layers import Layer

model = Sequential() model.compile(Dense(64, activation='relu'))

D) ```python

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

model = Sequential()

Dense(64, activation='relu')(model)

Answer: A

Question 28 (MSQ)

Which of the following statements are true about the architectures: LeNet-5, AlexNet, and VGGNet?

Options: A) LeNet-5 was one of the earliest CNN models, designed for digit recognition. B) AlexNet popularized the use of ReLU activation and dropout for deep learning. C) VGGNet uses very large convolution filters (e.g., 11×11) to capture spatial features. D) VGGNet increases depth using stacked 3×3 convolution layers.

Answers: A, B, D

Question 29 (MCQ)

A data analyst clusters customer data using K-Means with different values of K. When K = 3, the silhouette score is 0.82. When K = 6, the silhouette score drops to 0.41.

What can be inferred from this result?

Options: A) The clustering is better when K = 6. B) K = 3 gives well-separated clusters. C) A higher silhouette score means the clusters overlap more. D) The silhouette score is unrelated to cluster quality.

Answer: B

Question 30 (MSQ)

A student builds the following CNN model using Keras:

```
model = keras.Sequential([
   layers.Conv2D(32, (3, 3), activation='relu', input_shape=(28, 28, 1)),
   layers.MaxPooling2D((2, 2)),
   layers.Flatten(),
   layers.Dense(64, activation='relu'),
   layers.Dense(10, activation='softmax')
])
```

Which of the following statements are true about this model?

Options: A) The model starts with a convolution layer that extracts features. B) The MaxPooling layer reduces spatial dimensions. C) The Flatten layer increases the number of parameters. D) The final Dense layer is suitable for multi-class classification.

Answers: A, B, D

Explanation

Question 1 Answer: A Justification: Linear Regression is ideal for modeling direct, constant-rate relationships between variables, as described for distance and fare.

Question 2 Answer: B Justification: Logistic Regression is a classification algorithm specifically designed for binary outcomes, making it suitable for predicting "Default" or "No Default."

Question 3 Answer: 2 Justification: The outdegree of node P is the count of edges starting from P. The edges are P → Q and P → R, totaling two.

Question 4 Answer: A, B, D Justification: • A: L1 (Lasso) regularization adds a penalty that can shrink coefficients exactly to zero, effectively eliminating features. • B: L2 (Ridge) regularization adds a penalty proportional to the square of the coefficients. • D: L2

regularization shrinks coefficients towards zero but generally retains all features, rather than eliminating them entirely.

Question 5 Answer: 10 Justification: Lasso regression set 15−5=10 features to zero. Ridge regression retained all 15. Therefore, 10 features were eliminated by Lasso but not by Ridge.

Question 6 Answer: B Justification: The "Assignment" step in K-Means involves assigning each data point to the cluster whose centroid is closest to it based on a distance metric.

Question 7 Answer: C Justification: The ReLU function f(x)=max(0,x) outputs 0 for any negative input. Since all values in data are negative, the output activated_data will be [0 0 0 0 0].

Question 8 Answer: C Justification: Principal Component Analysis (PCA) is a powerful dimensionality reduction technique that reduces the number of features (like pixel values) while retaining most of the data's variance, thereby reducing computational load.

Question 9 Answer: B Justification: For multi-class classification, the Softmax activation function in the output layer is essential as it converts raw scores into a probability distribution over all classes, with probabilities summing to 1.

Question 10 Answer: C Justification: In deep CNN architectures, the number of filters typically increases with depth. This allows the network to learn more complex and abstract features as it processes the input through successive layers.

Question 11 Answer: B Justification: Softmax is specifically designed for multi-class classification. It transforms raw output scores into a probability distribution, ensuring all probabilities are non-negative and sum to 1.

Question 12 Answer: B, C, D Justification: • B: Increasing the learning rate can help speed up convergence if the loss is decreasing slowly, preventing it from getting stuck in shallow regions. • C: Stochastic or mini-batch gradient descent update weights more frequently than batch gradient descent, leading to faster training on large datasets. • D: Normalizing input features ensures that all features are on a similar scale, which helps gradient descent converge more quickly and efficiently without oscillations.

Question 13 Answer: A, C, D Justification: • A: Increasing model complexity (more layers/neurons) can help capture more intricate patterns, addressing underfitting. • C: More complex activation functions like ReLU or tanh can enable the network to learn non-linear relationships more effectively, mitigating underfitting often caused by limited model capacity. • D: More training data can provide the model with a broader representation of patterns, helping it generalize better and reduce underfitting, especially if the current data is insufficient or biased.

Question 14 Answer: A, B, C Justification: • A: Batch Gradient Descent uses the entire dataset for each update, leading to stable convergence but is slow for large datasets. • B: Stochastic Gradient Descent updates weights for each example, offering faster updates but

with noisy convergence due to high variance gradients. • C: Mini-batch Gradient Descent provides a balance by using small subsets of data for updates, compromising between speed and stability.

Question 15 Answer: 1 Justification: Support vectors are the data points closest to the decision boundary in an SVM. They are critical because they directly define the margin and the optimal position of the hyperplane.

Question 16 Answer: 10 Justification: The MNIST dataset contains 10 distinct classes (digits 0-9). The final nn.Linear layer in a classification model must output a value for each class to represent class probabilities.

Question 17 Answer: A, B, C Justification: • A: Accuracy = (7+89)/100=96%, which is greater than 95%. (True) • B: False Negatives (3) are more than False Positives (1). (True) • C: Recall = TP/(TP+FN)=7/(7+3)=7/10=70%. (True)

Question 18 Answer: B Justification: The output feature map will be 3x3 from the given data

Question 19 Answer: A Justification: Max Pooling extracts the most prominent feature (maximum activation) within a region, preserving strong signals. Average Pooling, by averaging, smooths the feature map and retains more general information.

Question 20 Answer: 42 Justification: Applying 2x2 max pooling yields the maximums from each block: 8 (from top-left), 7 (from top-right), 13 (from bottom-left), and 14 (from bottom-right). Their sum is 8+7+13+14=42.

Question 21 Answer: B Justification: The core principle of an SVM is to find the hyperplane that maximizes the margin (the distance) between the closest data points (support vectors) of different classes.

Question 22 Answer: A, C, D Justification: • A: DBSCAN identifies "noise" points that do not belong to any cluster. • C: It is effective at discovering arbitrarily shaped clusters and can handle varying densities. • D: DBSCAN requires two parameters: epsilon (maximum distance for neighborhood) and min_samples (minimum number of points to form a dense region).

Question 23 Answer: C Justification: In the K-Means algorithm, 'K' explicitly represents the pre-determined number of clusters that the algorithm will attempt to partition the data into.

Question 24 Answer: B Justification: A single-layer perceptron can only learn to classify data that is linearly separable, meaning a single straight line (or hyperplane in higher dimensions) can separate the classes.

Question 25 Answer: C Justification: PCA's main objective is to reduce the dimensionality of a dataset by transforming it into a smaller set of principal components that capture the maximum possible variance of the original data.

Question 26 Answer: A, B, C Justification: • A: Forward propagation computes the output of each layer sequentially. • B: Backward propagation uses calculated gradients to adjust (update) the network's weights. • C: Gradients for weight updates are computed efficiently

through the network using the chain rule, which is the essence of the backpropagation algorithm.

Question 27 Answer: A Justification: This snippet correctly initializes a Keras Sequential model and uses model.add() to append a Dense (fully connected) layer with specified units and activation function, which is the standard Keras API.

Question 28 Answer: A, B, D Justification: • A: LeNet-5 was indeed an early and influential CNN, notably for digit recognition. • B: AlexNet's success in ImageNet significantly popularized ReLU activation and dropout. • D: VGGNet is characterized by its use of stacked small 3x3 convolutional filters to increase network depth and receptive field effectively.

Question 29 Answer: B Justification: A higher silhouette score (0.82 for K=3) indicates better-defined and more separated clusters. A lower score (0.41 for K=6) suggests less distinct or overlapping clusters.

Question 30 Answer: A, B, D Justification: • A: The initial Conv2D layer is responsible for extracting hierarchical features from the input image. • B: MaxPooling2D layers reduce the spatial dimensions (height and width) of the feature maps, which helps in reducing computation and increasing translation invariance. • D: A Dense layer with 10 units and a softmax activation is the standard output configuration for a multi-class classification problem with 10 classes.