Editorial: Assignment - W8A2: Foundations of Deep Learning and Clustering: Concept Checkpoint

Question 1 (MCQ)

A machine learning engineer at "NeuroNet Solutions" is building a deep learning model to predict housing prices. She wants to justify using multi-layer perceptrons for this regression task.

Which theorem supports her choice by stating that stacked artificial neurons can approximate any continuous function?

A) Central Limit Theorem

B) Universal Approximation Theorem

C) Bayes' Theorem

D) Fourier Theorem

Correct Answer: B)

Explanation: The Universal Approximation Theorem justifies using multi-layer perceptrons for complex problems by demonstrating their ability to approximate any continuous function.

Question 2 (MSQ)

A data scientist at "VisionAI" tries to use a Multi-Layer Perceptron (MLP) for image classification but finds the model performs poorly compared to CNNs.

Which of the following are valid reasons for the MLP's struggles?

A) Destruction of spatial relationships during flattening

B) Inability to handle translation invariance

C) Limited capacity for non-linear combinations

D) Excessive computational requirements

Correct Answer: A), B)

Explanation: MLPs flatten 2D images into 1D vectors, losing spatial patterns and making position-specific learning difficult.

Question 3 (MCQ)

A retail analyst at "ClusterMart" uses K-means to segment customer data. During the clustering process, she wants to know what happens during the E-step.

What is computed during the E-step of K-means clustering?

A) New centroids

B) Distance matrix

C) Cluster reassignment

D) Optimal k value

Correct Answer: C)

Explanation: The Expectation step assigns points to nearest centroids using Euclidean distance.

Question 4 (MCQ)

At your company you are building a neural network for multi-class text classification. The model's output layer must convert logits to probabilities for each class.

Which activation function should be used for this purpose?

A) Sigmoid

B) Tanh

C) ReLU

D) Softmax

Correct Answer: D)

Explanation: Softmax normalizes outputs into probability distributions for multi-class classification.

Question 5 (MCQ)

A data scientist at "ClusterEval" wants to compare the separation between clusters formed by K-means.

Which evaluation metric should she use?

A) Inertia

B) Silhouette score

C) Purity

D) Rand index

Correct Answer: B)

Explanation: Combines intra-cluster cohesion and inter-cluster separation.

Question 6 (MCQ)

A data scientist at "ClusterSmart" is initializing centroids for K-means clustering on a new dataset. She wants to improve cluster quality by using a method that considers distances from existing centroids.

Which initialization method uses distance-weighted probabilities for centroid selection?

A) Random

B) K-means++

C) Hierarchical

D) PCA-based

Correct Answer: B)

Explanation: K-means++ selects subsequent centroids proportional to squared distances from existing centers.

Question 7 (MCQ)

Working at an AI startup you are comparing activation functions for a CNN designed for X-ray image analysis.

What is the primary advantage of ReLU over Sigmoid in CNNs for this application?

A) Output normalization

B) Mitigating vanishing gradients

C) Handling negative values

D) Probability interpretation

Correct Answer: B)

Explanation: ReLU's linear region maintains stronger gradients during backpropagation.

Question 8 (MCQ)

As an engineer at an AI company, you are implementing forward propagation in an MLP.

Which matrix operation is used to calculate neuron outputs in each layer?

A) Convolution

B) Dot product

C) Transpose

D) Determinant

Correct Answer: B)

Explanation: Layer outputs are computed via weight matrix multiplication followed by activation.

Question 9 (MCQ)

A business analyst at a company is evaluating the results of K-means clustering on sales data.

Which metric should she use to measure how compact the clusters are?

A) Silhouette score

B) Rand index C) Within-cluster sum of squares

D) Purity score

Correct Answer: C)

Explanation: WCSS (inertia) measures total squared distance from points to their centroid.

Question 10 (MCQ)

A product team at a company is building a CNN to identify product defects in images.

Which architectural component enables their CNN to detect local patterns such as edges and textures?

A) Fully connected layers

B) Pooling layers

C) Convolutional filters

D) Dropout layers

Correct Answer: C)

Explanation: Convolutional kernels slide across input space to identify local features.

Question 11 (MCQ)

A data scientist at a company attempts to use K-means on a dataset with categorical features.

How does K-means handle categorical data?

A) Native support through distance adaptation

B) Requires one-hot encoding

C) Uses cosine similarity

D) Not directly applicable

Correct Answer: D)

Explanation: K-means relies on Euclidean distance which isn't suitable for categorical variables.

Question 12 (MSQ)

You are working at a company and using One-vs-Rest approach for multi-class classification on a dataset. What could be the limitations of this approach that you must keep in mind?

A) Computational complexity

B) Class imbalance

C) Non-linear separability

D) Feature scaling

Correct Answer: A), B)

Explanation:

1. **Computational Complexity (A):** OvR trains one classifier per class, becoming resource-intensive for large class counts or complex models.

2. **Class Imbalance (B):** Each classifier faces severe imbalance (one class vs all others), diluting minority class signals. Other options (C, D) relate to model/data properties, not OvR-specific constraints.

Question 13 (MCQ)

A deep learning researcher at a company observes unstable training in her CNN.

Which normalization technique can help stabilize training by normalizing layer inputs?

A) Batch normalization

B) Min-max scaling

C) Z-score

D) L2 normalization

Correct Answer: A)

Explanation: Stabilizes training by normalizing layer inputs.

Question 14 (MCQ)

A developer at "ImageScale" wants to reduce the spatial dimensions of feature maps in her CNN for faster computation.

Which type of layer should she use?

A) Convolutional

B) Pooling

C) Fully connected

D) Dropout

Correct Answer: B)

Explanation: Pooling (e.g., max pooling) downsamples feature maps.

Question 15 (MCQ)

A data scientist at a company is explaining why MLPs can model complex relationships.

Which component enables MLPs to learn non-linear relationships?

A) Hidden layers

B) Output normalization

C) Linear activation

D) Batch processing

Correct Answer: A)

Explanation: Hidden layers with non-linear activations enable function approximation.