Editorial: Assignment-W6A2: Core Concepts in Neural Networks and K-Means Clustering

Question 1 (MSQ)

A neural network is used to classify handwritten digits from the MNIST dataset.

Question: (Select all that apply)
Which of the following statements about the input representation for an FNN in this task are correct?
A) The image is used as a 2D array of pixels
B) The image is flattened into a 1D vector
C) Each input neuron represents one pixel's brightness
D) The input is a sequence of pixel values over time

Correct Answer: B), C)

Explanation:
The 28x28 image is flattened into a 784-length vector, with each input neuron representing a pixel's brightness.

Question 2 (MCQ)

You are working in a startup and a building a neural network for prediction. You observed that sometimes simpler neural network models were outperforming deeper ones.

Question: What is a key reason that simpler neural network models can sometimes outperform deeper ones? A) They always have higher accuracy
B) They require more computation
C) They are less likely to overfit with limited data
D) They ignore regularization

Correct Answer: C)

Explanation:
Simpler models often generalize better, especially when data is limited, as deeper models may overfit without enough data or regularization

Question 3 (MCQ)

A research team at an AI company is training a neural network for a regression task. During backpropagation, they calculate the gradient of the loss with respect to the output as 0.4 and the gradient of the output with respect to a weight as 0.3.

Question:
What is the gradient of the loss with respect to the weight?
A) 1.33
B) 0.75
C) 0.4
D) 0.12

Correct Answer: D) 0.12

Explanation:

By the chain rule, the gradient of the loss with respect to the weight is the product of the two given gradients:

$\partial L/\partial w = (\partial L/\partial y) \times (\partial y/\partial w) = 0.4 \times 0.3 = 0.12$

Question 4 (MSQ)

A data scientist at a fruit-sorting company is designing a neural network to classify fruit images captured by a conveyor belt camera. The camera outputs RGB images where pixel values range from 0 to 255. Before feeding this data to the model, the scientist normalizes the pixel values to be in the range [0, 1].

Question:

Why is normalization of pixel values important when training a neural network on image data?

Select all that apply.

A) It helps ensure all features contribute equally during learning
B) It improves model convergence during gradient descent
C) It removes background noise from the images
D) It increases the resolution of the image data

Correct Answer:

✅ A), B)

Explanation:

Normalization ensures consistent scale across input features, which avoids one feature dominating the learning process. It also speeds up convergence by improving gradient behavior. It does **not** remove noise or change the image resolution.

Question 5 (MSQ)

You were assigned to modify an existing feedforward neural network (FNN) which is trained on the MNIST dataset. You were wondering...

Question: (Select all that apply)
What is the role of the hidden layers in the FNN?
A) They directly output the class prediction
B) They learn some pattern like basic shapes and digit parts
C) They combine features into higher-level representations
D) They only store the input data

Correct Answer: B), C)

Explanation:
Hidden layers in FNNs learn to detect some pattern like shapes and combine them into higher-level representations like digit parts, which are then used for classification.

Question 6 (MCQ)

A data scientist is designing a feedforward neural network to classify handwritten digits and must choose an activation function for the hidden layers.

Question:
Which activation function is used in the hidden layers of the FNN for MNIST, and why?
A) Sigmoid, for non-linearity
B) ReLU, for efficient non-linearity and sparse activations
C) Softmax, to output probabilities
D) Tanh, to center activations

Correct Answer: B)

Explanation:
ReLU is used because it efficiently introduces non-linearity and allows for sparse, fast computations by zeroing out negative values.

Question 7 (MSQ)

A team is developing a Recurrent Neural Network (RNN) to predict the next word in a sentence based on previous words.

Question: (Select all that apply)
In the context of RNNs, what does the hidden state represent?
A) The current input only
B) The memory of the network at a given time step
C) Accumulated information from previous time steps
D) The output probabilities

Correct Answer: B), C)

Explanation:
The hidden state in an RNN acts as memory, carrying information from previous time steps and integrating it with the current input1.

Question 8 (MSQ)

A data scientist wants to segment customers into groups based on their purchasing behavior using K-means clustering.

Question: (Select all that apply)
Which of the following statements about K-means clustering are correct?
A) It is a supervised learning algorithm
B) It assigns each data point to the nearest centroid

C) The number of clusters (K) must be specified in advance

D) It guarantees finding the global minimum of the objective function

Correct Answer: B), C)

Explanation:

K-means is an unsupervised algorithm that assigns each data point to the nearest centroid, and the number of clusters K must be set before running the algorithm. It does not guarantee a global minimum due to random initialization.

Question 9 (MCQ)

A marketing analyst applies K-means clustering to customer data and notices that the resulting clusters are different each time the algorithm is run.

Question:
What is the most likely cause?

A) The data is not normalized

B) The number of clusters is too large

C) The dataset is too small

D) The algorithm uses random initialization of centroids

Correct Answer: D)

Explanation:

K-means uses random initialization for centroids, which can lead to different results on each run.

Question 10 (MSQ)

You are clustering data points using K-means. After several iterations, the assignment of points to clusters no longer changes.

Question: (Select all that apply)
What does this indicate?

A) The algorithm has converged

B) The centroids are still moving

C) The algorithm has reached a local minimum.

D) Further iterations will not change the clusters

Correct Answer: A), D)

Explanation:

When assignments stop changing, K-means has converged and further iterations will not alter the clusters.

Question 11 (MCQ)

A data scientist at a retail company has applied K-means clustering to group customer profiles based on their purchasing behavior. She now wants to assess how well the customers are grouped internally within each cluster.

Question:
Which metric should she use to evaluate how tightly grouped the customers are within each cluster?

A) Inter-cluster distance
B) Between-cluster variance
C) Number of clusters
D) Within-cluster sum of squares (WCSS)

Correct Answer: D)

Explanation:
K-means minimizes the within-cluster sum of squares, which measures how closely related the data points are within each cluster — a key metric for evaluating clustering quality.

Question 12 (MSQ)

You were working on a dataset with many outliers. You apply K-means clustering.

Question: (Select all that apply)
How might outliers affect the clustering result?
A) Outliers can distort the position of centroids
B) K-means is robust to outliers
C) Outliers may form their own clusters
D) The algorithm may converge more slowly

Correct Answer: A), C), D)

Explanation:
Outliers can pull centroids away from dense regions, may form their own clusters, and slow convergence. K-means is not robust to outliers.

Question 13 (MSQ)

A data analyst is using the elbow method to decide the number of clusters for K-means on a sales dataset.

Question: (Select all that apply)
What does the "elbow" point on the plot indicate?
A) The point where adding more clusters yields diminishing returns
B) The lowest inertia value
C) The optimal number of clusters
D) The maximum number of clusters possible

Correct Answer: A), C)

Explanation:
The elbow point shows where increasing K further does not significantly reduce inertia, suggesting the optimal K.

Question 14 (MSQ)

A machine learning engineer is reviewing the limitations of K-means clustering before applying it to a new dataset.

Question: (Select all that apply)
Which of the following are limitations of the standard K-means algorithm?
A) Sensitive to initial centroid positions
B) Can handle non-spherical clusters well
C) Sensitive to outliers
D) Requires specification of K in advance

Correct Answer: A), C), D)

Explanation:
K-means is sensitive to initial centroids and outliers, and the number of clusters K must be specified.

Question 15 (MSQ)

A researcher is considering K-means clustering for a dataset with mixed data types and several outliers.

Question: (Select all that apply)
Which situations are not suitable for K-means clustering? A) Data with non-spherical clusters
B) Data with many outliers
C) Data with categorical variables only
D) Data that is well-separated and spherical

Correct Answer: A), B), C)

Explanation:
K-means works best for numerical, spherical, well-separated clusters and struggles with non-spherical clusters, outliers, and categorical data.