**Editorial: Assignment-W7A2: From Groups to Grayscale: Exploring Clustering and CNNs**

**Question 1 (MCQ)**

A store manager at "FreshMart Groceries" wants to group similar products for optimized shelf placement using K-Means.

1. Assign each data point to the nearest cluster center.
2. Re-assign each point to the nearest cluster center.
3. Assign cluster centers randomly.
4. Re-compute cluster centers.
5. Specify the number of clusters.

**Question:**  
Arrange the steps of the K-Means algorithm in the correct sequence:  
A) 3, 5, 1, 4, 2  
B) 5, 3, 1, 2, 4  
C) 5, 3, 1, 4, 2  
D) 3, 5, 2, 4, 1  
E) None of these

**Correct Answer:** C)

**Explanation:**  
The correct sequence is:

1. Specify the number of clusters (5)
2. Assign cluster centers randomly (3)
3. Assign each data point to the nearest cluster center (1)
4. Re-compute cluster centers (4)
5. Re-assign each point to the nearest cluster center. (2)

**Question 2 (MSQ)**

A data analyst at "Shopify Insights" runs K-Means multiple times on customer purchase data to identify shopping patterns.

**Question:** (Select all that apply)  
Which of the following statements are correct?  
A) All cluster initializations lead to the same result  
B) Different initializations might find different local/global minima  
C) Some initializations may fail to converge  
D) Initialization affects iteration count for convergence

**Correct Answer:** B), D)

**Explanation:**  
K-Means can converge to **local minima** depending on initialization, and different initial centers may require varying iterations to stabilize. Convergence is guaranteed but not necessarily to the global optimum.

**Question 3 (MCQ)**

A data scientist at "HealthTrack" uses K-Means clustering with k=2 and 1 centroid remains to be initialized. The squared distances from existing centroids for data points x₁, .. ,x₅ are 25, 67, 89, 24, and 56 respectively.

**Question:**  
Which will be the next centroid selected?  
A) Any point is chosen uniformly at random  
B) x₃ will definitely be chosen.  
C) x₃ has highest probability but isn't guaranteed  
D) x4 will be chosen.

**Correct Answer:** A)  
**Explanation:**  
In **standard K-Means**, centroids are initialized randomly from all data points without considering distance weights. All points have equal selection probability regardless of their distances.

**Question 4 (MCQ)**

The same data scientist now uses **K-Means++** with k=2 and 1 centroid remains to be initialized. Squared distances for x₁, .., x₅ are 25, 67, 89, 24, and 56 respectively.

**Question:**  
Which will be the next centroid selected?  
A) Uniform random selection  
B) x₃ is guaranteed selection  
C) x₃ has highest probability but isn't guaranteed D) x4 will be chosen. **Correct Answer:** C)  
**Explanation:**  
K-Means++ assigns selection probabilities **proportional to squared distances**, making x₃ (distance=89) most likely but not certain. This balances exploration and optimization.

**Question 5 (MCQ)**

A K-Means clustering problem has three clusters with final centroids at **(1, 1)**, **(6, 3)**, and **(2, 7)**. A new data point is added at **(2, 0)**.

**Which cluster center’s cluster does the new data point belong to?**  
A) (1, 1)  
B) (6, 3)  
C) (2, 7)  
D) Could be any one of the three.

**Correct Answer:** A) (1, 1)

**Explanation:**

* Distance to (1, 1): √[(2−1)² + (0−1)²] = √2 ≈ 1.41
* Distance to (6, 3): √[(2−6)² + (0−3)²] = √25 = 5
* Distance to (2, 7): √[(2−2)² + (0−7)²] = 7

The smallest distance is to **(1, 1)**, so the new point is assigned to that cluster.

**Question 6 (MCQ)**

 In the attached image, there is a dataset. We are performing K-Means ++ Clustering technique. K = 3, 2 centroids A and A1 are intialized. Which datapoint will be the 3rd centroid?

A) I most likely  
B) S most likely  
C) R most likely  
D) Z most likely

**Correct Answer:** D)

**Explanation:**  
The 3rd centroid chosen will be based on sum of squared distances from existing centroids. And, then probability is applied therefore, most likely Z will be chosen as it is farthest from both of the existing centroids.

**Question 7 (MCQ)**

A team at "MedImaging AI" builds a CNN to detect tumors in X-rays. Which activation function is commonly used in CNNs to introduce non-linearity in hidden layers?

A) Sigmoid  
B) Softmax  
C) ReLU  
D) Tanh

**Correct Answer:** C)

**Explanation:**  
ReLU (Rectified Linear Unit) is the most commonly used activation function in CNNs because it efficiently introduces non-linearity, is simple to compute, and helps mitigate the vanishing gradient problem that can occur with sigmoid or tanh activations. This makes ReLU especially effective for training deep convolutional networks.

**Question 8 (MCQ)**

During training, you face CNN struggles with slow convergence. Its due to vanishing gradients. Which activation function most commonly causes vanishing gradients?

A) ReLU  
B) Leaky ReLU  
C) Sigmoid  
D) Swish

**Correct Answer:** C)

**Explanation:**  
The sigmoid activation function compresses outputs to the range (0, 1), causing gradients to become very small for inputs far from zero-especially in deep networks-leading to the vanishing gradient problem. This makes it difficult for the network to update weights in earlier layers, slowing or even halting learning. In contrast, activation functions like ReLU are designed to mitigate this issue.

**Question 9 (MCQ)**

The team needs to preprocess 8-bit grayscale X-rays. **Question:** Using 8-bit encoding, what is the highest decimal number we can represent?  
A) 128  
B) 256  
C) 255  
D) 512

**Correct Answer:** C)  
**Explanation:**  
The maximum value for an 8-bit unsigned integer is calculated as: When all 8 bits are set to 1 (11111111₂), the decimal equivalent is 1 + 2 + 4 + .... + 128 = 255. This represents the upper limit for 8-bit binary encoding systems.

**Question 10**

To enhance tumor visibility, you apply a transformation. **Question:** What does applying a negative transformation to an image do?

A) Increases brightness  
B) Converts the image into negative space  
C) Inverts the color values  
D) Applies blur

**Correct Answer:** C)

**Explanation:**  
Applying a negative transformation to an image inverts the color or intensity values of each pixel, effectively reversing the brightness levels. This operation converts dark areas to light and light areas to dark, producing a photographic negative of the original image.

**Question 11 (MSQ)**

A computer vision engineer at "AutoDrive" processes road images. **Question:** (Select all that apply) Which statements about RGB channels are correct?  
A) (0,0,0) represents white  
B) 8-bit encoding allows 256 values per channel  
C) OpenCV uses BGR format by default  
D) Channel isolation preserves color information

**Correct Answer:** B), C)  
**Explanation:**  
8-bit channels (0-255 values) and OpenCV's BGR default are fundamental concepts in image processing.

**Question 12 (MCQ)**

They use DBSCAN to detect road obstacles in LIDAR data.

**Question:** Which parameter controls the neighborhood radius in DBSCAN?  
A) min\_samples  
B) eps  
C) linkage  
D) kernel\_size

**Correct Answer:** B)  
**Explanation:**  
eps (ε) defines the radius for neighborhood calculations in density-based clustering

**Question 13 (MCQ)**

A retail chain uses hierarchical clustering to group stores.

**Question:** Which linkage method minimizes intra-cluster variance?  
A) Single  
B) Complete  
C) Average  
D) Ward

**Correct Answer:** D)  
**Explanation:**  
Ward's method merges clusters based on minimal variance increase

**Question 14 (MCQ)**

A geospatial firm analyzes irregular-shaped earthquake zones.

**Question:** Which algorithm handles non-convex clusters best?  
A) K-Means  
B) DBSCAN  
C) Hierarchical  
D) PCA

**Correct Answer:** B)  
**Explanation:**  
DBSCAN identifies arbitrary-shaped clusters through density analysis

**Question 15 (MSQ)**

"TrendAnalytics" evaluates customer segmentation quality.

**Question:** (Select all that apply) Which metrics are valid?  
A) Silhouette Score  
B) Davies-Bouldin Index  
C) F1 Score  
D) Mean Squared Error

**Correct Answer:** A), B)  
**Explanation:**  
Silhouette Score and Davies-Bouldin Index are standard clustering metrics