Question 1 You are using **KNN (K = 1)** on a dataset with **two features**:

- **Feature 1** is numeric (range: 0 to 1),
- **Feature 2** is numeric (range: 0 to 1000).

You try two approaches:

- **Model A**: Uses raw features
- **Model B**: Applies Min-Max scaling before training

You observe that Model B significantly outperforms Model A. Why? (b)

a) KNN performs worse after scaling due to loss of magnitude information
b)Model A overweighs Feature 2 due to scale imbalance
c)Model B fails to learn properly since scaling removes units

d) Scaling has no effect on KNN since it is non-parametric

Question 2 You are using KNN with K=5 on a dataset with **1000 features** and **only 200 training samples**. You observe poor classification accuracy.

Which of the following is the most plausible explanation for this behavior? ©

a) KNN suffers from underfitting due to a small K
b) KNN cannot work on high-dimensional data at all
c) Distance measures become less informative in high dimensions (curse of dimensionality)
d)PCA was not used, and KNN always requires PCA beforehand

Question 3 A dataset has 6 features. After applying PCA, the explained variance ratio is:

PC1: 48%
PC2: 26%
PC3: 14%
PC4: 8%
PC5: 3%
PC6: 1%

How many principal components do you need to retain ≥90% variance? (c)

a) 2
b) 3

c) 4
d)5

Question 4 You train an SVM classifier with a Gaussian (RBF) kernel on a dataset and get perfect training accuracy, but poor cross-validation accuracy. (b)

What should you try changing?

a) Increase C and decrease σ^2
b) Decrease C and increase σ^2
c) Increase both C and σ^2
d)Use a linear kernel instead

Question 5 You apply SVD on a large matrix M of shape (1000 x 500). The decomposition is:
$M = U\Sigma V^T$

What are the shapes of U, Σ, and V^T? (b)

a) U: (1000×1000), Σ: (1000×500), V^T: (500×500)
b) U: (1000×500), Σ: (500×500), V^T: (500×500)

c)U: (1000×500), Σ: (500×1000), V^T: (500×1000)

d) U: (1000×500), Σ: (500×500), V^T: (1000×500)

Question 6 Which of the following statements are TRUE? (multiple correct) (a,b,c)

a) KNN is a lazy learner and stores the entire training data
b) PCA creates new orthogonal features ordered by variance captured

c)SVM with RBF kernel can model non-linear boundaries
d) Decision Trees are more prone to overfitting than KNN on small datasets
e) Applying PCA always increases accuracy of any model
f)K-means works well when clusters are of different densities and shapes

Question 7 You have a fully grown decision tree with training accuracy 98% and test accuracy 72%. Which pruning method will most likely improve test accuracy? (b)
a) Pre-pruning by limiting max depth to 5
b) Post-pruning by cost complexity pruning
c) Increasing minimum samples per leaf to 1
d) Adding more features to split on

Question 8 Which of the following statements about decision tree ensembles are true? (a,b,c,d)

A. Bagging reduces variance by training multiple trees on bootstrap samples
B. Boosting trains trees sequentially to reduce bias
C. Random Forest selects a random subset of features at each split
D. Gradient Boosting trees use residuals to improve model performance
E. Ensembles eliminate the need for pruning individual trees

Question 9 Which statement best describes the eigenvector associated with the largest eigenvalue of a covariance matrix? (c)

a) It points in the direction of least variance in data.
b) It is orthogonal to the direction of maximum variance.
c) It represents the direction along which the data varies the most.
d) It minimizes the reconstruction error of PCA.

Question 10 **Scaling the dataset before PCA changes the eigenvectors but not the eigenvalues of the covariance matrix. (c)**

a) **True**
b) **False**
c) **Depends on data**
d) **Can't say**

Question 11 Suppose one feature ranges between 0 and 1, and another ranges between 0 and 1000. How would this affect K-means clustering, and what preprocessing step is recommended? (a)

A) The large scale of Feature 2 will dominate the distance calculations, so K-means may form clusters primarily based on Feature 2; standardize features to zero mean and unit variance before clustering.

B) Feature 1 will dominate the clustering because it is more precise; apply min-max scaling only to Feature 1.

C) K-means is scale-invariant, so no preprocessing is needed; it automatically balances feature influence.

D) Feature 2's scale difference has no effect if you use Euclidean distance; just normalize the final centroids after clustering.