# KNN & PCA - Assignment

## K-Nearest Neighbors (KNN)

1️⃣ What is K-Nearest Neighbors (KNN) and how does it work?  
KNN ek simple supervised learning algorithm hai. Prediction ke liye, yeh nearest (sabse kareeb) ‘k’ data points dekhta hai. Jo class ya value in nearest points me zyada hoti hai, woh result hota hai.

2️⃣ What is the difference between KNN Classification and KNN Regression?  
KNN classification me majority voting hota hai (maximum nearest neighbors ki category). KNN regression me average nikalte hain nearest neighbors ki values ka.

3️⃣ What is the role of distance metric in KNN?  
Distance metric (jaise Euclidean ya Manhattan) define karta hai ki do points ke beech distance kaise calculate hoga. Isse ‘nearest’ neighbors identify kiye jaate hain.

4️⃣ What is the Curse of Dimensionality in KNN?  
Jab features ki sankhya badh jati hai, distance meaningful nahi rehta. Data sparse ho jata hai, jisse accuracy girti hai.

5️⃣ How can we choose the best value of k in KNN?  
Try karte hain alag-alag values of k aur unka accuracy dekhte hain (cross-validation ke through). Jo k best accuracy deta hai, use select karte hain.

6️⃣ What are KD Tree and Ball Tree in KNN?  
KD Tree aur Ball Tree efficient data structures hote hain jo KNN ki speed badhate hain by organizing data points.

7️⃣ When should you use KD Tree vs Ball Tree?  
KD Tree low-dimensional data ke liye achha hai, Ball Tree high-dimensional data ke liye better perform karta hai.

8️⃣ What are the disadvantages of KNN?  
- Slow prediction time (large dataset me).  
- Sensitive to irrelevant features.  
- Curse of dimensionality ka problem.

9️⃣ How does feature scaling affect KNN?  
Feature scaling important hota hai because distance metrics feature values pe depend karte hain. Agar scaling na ho to bade features dominate karte hain.

🔟 How does KNN handle missing values in a dataset?  
Generally KNN missing values ko handle nahi karta. Missing values ko pehle impute karna hota hai ya aise rows remove karne padte hain.

## Principal Component Analysis (PCA)

1️⃣1️⃣ What is PCA (Principal Component Analysis)?  
PCA ek technique hai jo data ke dimensions reduce karta hai by creating new features (principal components) jo maximum variance capture karte hain.

1️⃣2️⃣ How does PCA work?  
PCA covariance matrix bana kar eigenvectors aur eigenvalues nikalta hai. Sabse bade eigenvalues ke eigenvectors principal components hote hain.

1️⃣3️⃣ What is the geometric intuition behind PCA?  
PCA data ke direction of maximum spread (variance) pe focus karta hai aur nayi axes define karta hai.

1️⃣4️⃣ What is the difference between Feature Selection and Feature Extraction?  
- Feature Selection me existing features me se best features select kiye jaate hain.  
- Feature Extraction (jaise PCA) me naye features banaye jaate hain.

1️⃣5️⃣ What are eigenvalues and eigenvectors in PCA?  
Eigenvectors nayi directions (principal components) ko represent karte hain, aur eigenvalues unme variance ka amount batate hain.

1️⃣6️⃣ How do you decide the number of components to keep in PCA?  
Variance explained ka percentage dekhte hain (jaise 95%), ya scree plot use karte hain.

1️⃣7️⃣ Can PCA be used for classification?  
Haan, PCA se dimensionality reduction karke uske baad classification kar sakte hain.

1️⃣8️⃣ What are the limitations of PCA?  
- Linear relationships hi capture karta hai.  
- Data ka interpretation thoda mushkil ho sakta hai.

1️⃣9️⃣ How do KNN and PCA complement each other?  
PCA se dimensions kam karke KNN ki performance better ho sakti hai. High-dimensional data me KNN slow hota hai, PCA se speed badh sakti hai.

2️⃣0️⃣ What are the key differences between PCA and Linear Discriminant Analysis (LDA)?  
- PCA maximum variance capture karta hai, bina label ke.  
- LDA class separation ke liye dimensions reduce karta hai (labels ka use karta hai).