PCA Assignment

support@intellipaat.com

+91-7022374614

US: 1-800-216-8930 (Toll-Free)

Problem Statement:

You have been provided with a multi dimensional data that contains information on certain images. Using machine learning, you should be able to predict the images on the new set of data using the model that you have trained on the existing data.

Dataset Information:

Each point in the data is an 8x8 image.

Classes	10
Samples per class	~180
Samples total	1797
Dimensionality	64
Features	integers 0-16

1.PCA is _____?

A. Unsupervised
B. Supervised
C. Semisupervised
D. None of the above

2. Why is PCA needed in the Data Science field?

A. Data Visualization
B. In Dimensionality Reduction
C. Data Manipulation
D. All of the Above

3. Extract only features and scale the data using StandardScaler() , compute min covariance?

A. -0.5705
B. -1
C. 0.1
D. None of the above

4. Extract only features and scale the data using StandardScaler() , how many eigen_values will be calculated?

A. 1
B. 64
C. 1797
D. 0

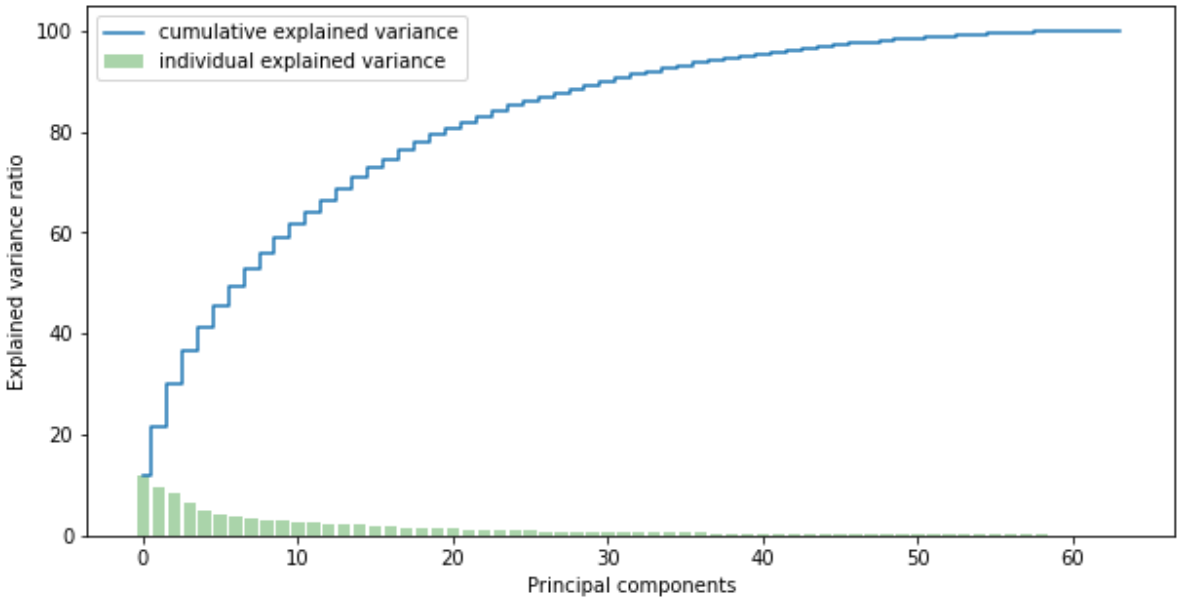
5. from the above eigenvalues and eigenvectors, create eigenpair and calculate maximum cumulative explained variance?

A. 100
B. 90
C. 80
D. 69

6. Why is cumulative explained variance required before applying PCA?

A. To check the maximum components to be selected
B. To check the distribution of the data
C. To check the minimum components to be selected
D. All of the above

7. from the below plot, select the range of components can be selected?



A. 0-10
B. 10-20
C. 40-60
D. 30-40

8. Fit the data to PCA and compute the maximum explained_variance_ratio?

A. 0.120
B. 0.9
C. 0.012
D. 0.011

9. fit the KMeans model on top of PCA fitted data and find the optimal number of clusters(k-value)?

A. 5-6

B. 6
C. 7-8
D. 4

10. Calculate the optimal number of clusters for the data without applying PCA?
A. 5
B. 6
C. 7
D. 8

With the above PCA data your decision tree model with train test ratio as 80:20, and random_state=40 .

11. With the help of the above model how many images have been misclassified by the model.
A. 40-50
B. 60-70
C. 50-60
D. 70-80

12.With the above model how many images are classified as 8.
A. 38
B. 36
C. 35
D. 32

13. Without applying PCA, how much misclassified data are there?
A. 45
B. 63
C. 23
D. 22

14. Which of the following non-zero vector stays parallel after matrix multiplication
A. EigenValue
B. EigenVector
C. Covariance vector
D. Explained variance ratio

15. The Output of PCA is always a new representation of data with a lower dimension than the original data representation.
A. True
B. False