Assignment 6

Name – Aastha Jain

Roll No – 2018IMT 004

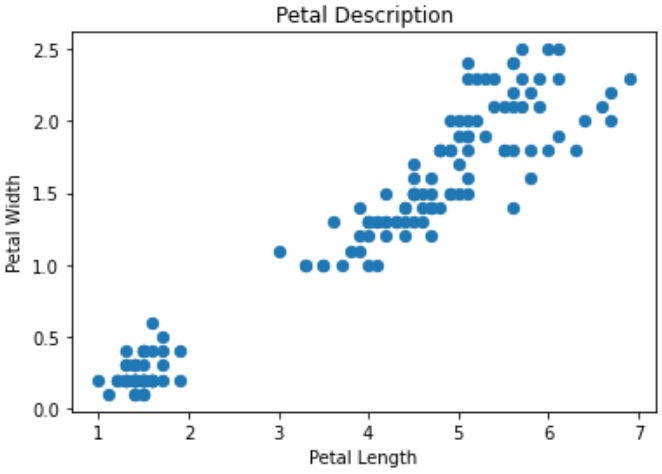
Course – ITIT-4103

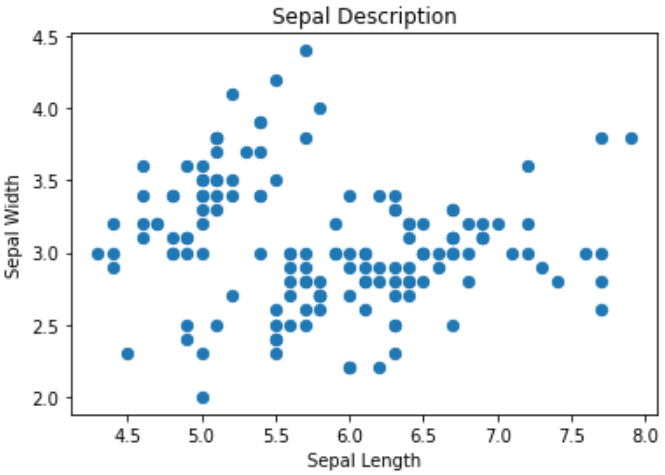
Aim

1. Considering the IRIS dataset discussed in previous assignment, apply EM algorithm to cluster the data (without considering the output labels) Use the same dataset for clustering using K-means algorithm. Compare the results of these two algorithms.

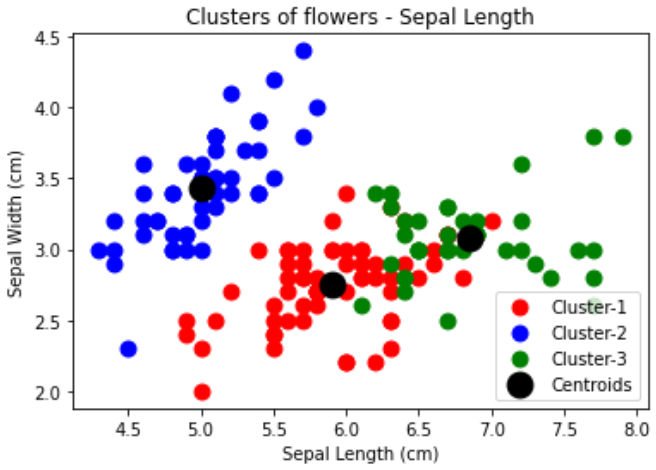
2. Apply PCA algorithm to obtain first two principal components and perform the clustering using both algorithms on the resultant data. Compare the results of these two algorithms.

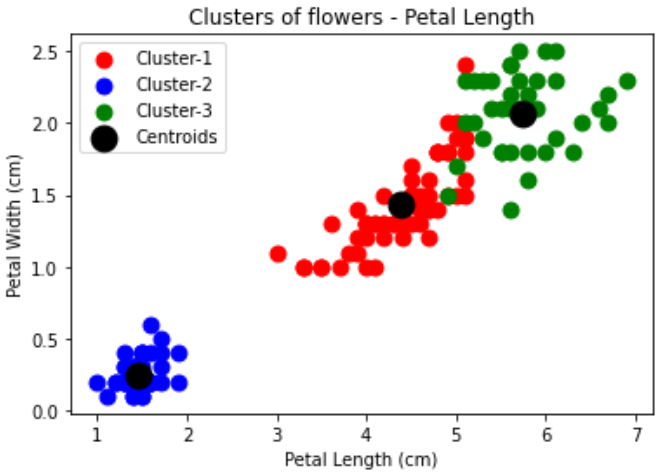
- Data Description

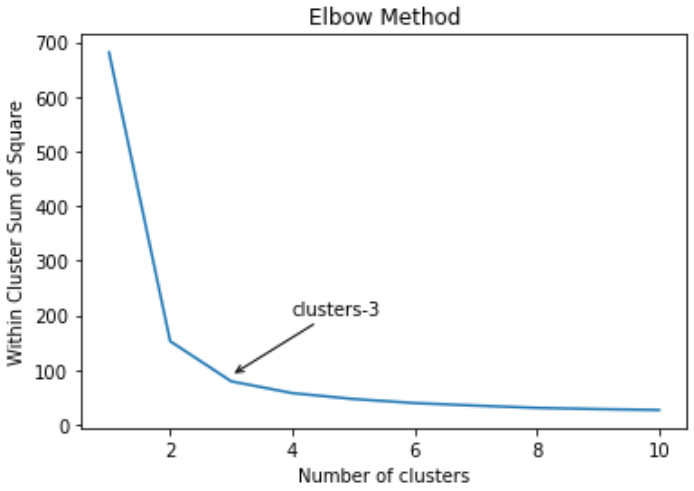




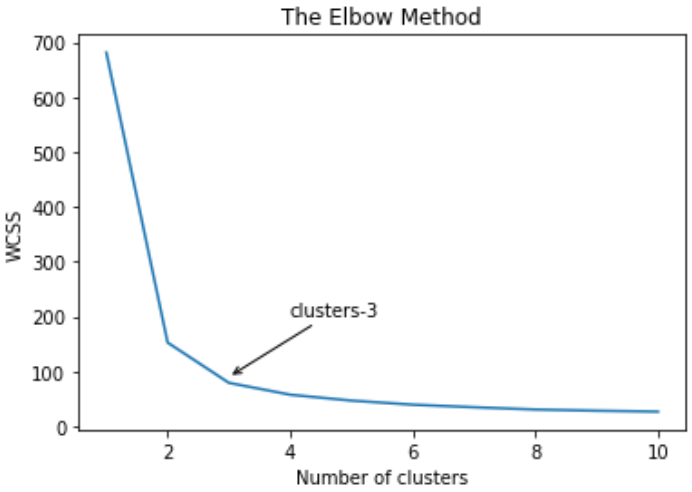
- K-Means

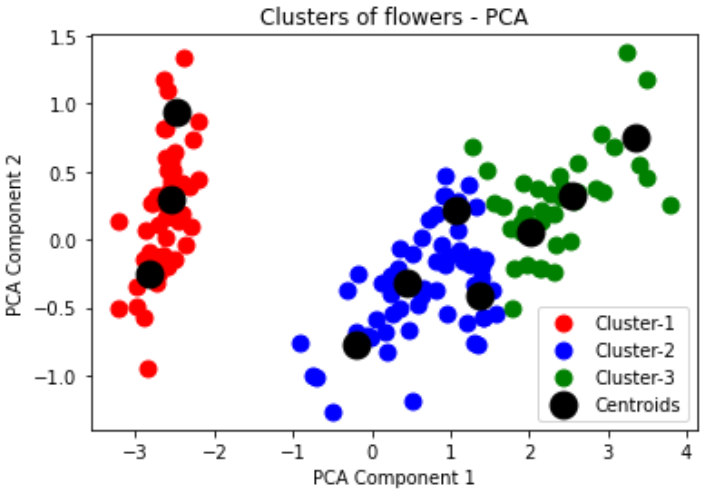






- K-Means PCA





- Results

  Accuracy of K-means and EM models

  1. K-Means model: 24 %
  2. Elbow Method: 41.333333333333336% %

  Accuracy of K-means and EM models with PCA

  1. K-Means model with PCA: 88.66666666666667 %
  2. Elbow Method with PCA: 64.0%

- Conclusions

  It can be seen that the EM method behaves and performs better than the K-means model in both raw data and PCA data (dimensionally reduced data).

  On semi-supervised learning, the EM Algorithm provides a viable alternative to classic k-means clustering.

  It finds multivariate Gaussian distributions for each cluster to offer stable solutions.