

Assignment 3: Unsupervised Learning

The data set to be used in this lab assignment is WINE.txt, which is already uploaded into this folder. This data set contains 178 rows and 14 columns. Each row corresponds to an object. Column 1 represents the classes of the objects. Columns (2-14) correspond to the 13 features of the objects. The tasks in this assignment are unsupervised learning, so only the data in columns 2-14 (features of the objects) are used.

The tasks in this lab are: 1) principal component analysis (PCA) for dimension reduction; and 2) K-means clustering using the new features from PCA. They are concretely described in the following:

1) PCA for dimension reduction

- Obtain and save the principal component coefficient matrix for the WINE data
- Calculate and save the new feature values (according to PCA) for all objects in the WINE data
- Visualize the objects in the WINE data using the top two principal components

You can use a tool or open library in performing this task

2) K-means clustering (K=3)

- Divide the objects into 3 clusters using the top 2 principal components, store the centroid and number of objects in each cluster
- Divide the objects into 3 clusters using the top 5 principal components, store the centroid and number of objects in each cluster
- Divide the objects into 3 clusters using the top 8 principal components, store the centroid and number of objects in each cluster

You have to make your own program to implement the K-means algorithm

The report has to include the following:

- A short summary of the basic principle of PCA, with around 150-200 words
- A short summary of the K-means algorithm, with around 150-200 words
- A picture to visualize the objects of WINE data based on PCA
- A table to give the centroid and number of data points for each cluster, based on the three experiments of clustering (using 2, 5, 8 principal components respectively)

Before submitting the report, you have to present this assignment first to Ning Xiong. After it is approved, you can then send the report in CANVAS.