ASSIGNMENT 5

Dataset Overview: Wine Chemical Analysis

This dataset is adapted from the Wine Data Set originally available on the UCI Machine Learning Repository. It has been modified for unsupervised learning by omitting the information regarding the specific types of wine. The dataset contains results from a chemical analysis of wines produced in a specific region of Italy, derived from three different cultivars. Each record represents a unique wine sample, analyzed for 13 distinct chemical constituents that contribute to the wine's characteristics and quality.

Attributes

Alcohol: The percentage of alcohol in the wine, indicating its strength and flavor profile.
Malic Acid: A natural organic acid found in wine, contributing to its tartness and overall acidity level.
Ash: The residue remaining after the combustion of the wine sample, representing the inorganic material present, which can affect taste and quality.
Alcalinity of Ash: A measure of the alkaline content of the ash, influencing the wine's pH and taste profile.
Magnesium: The concentration of magnesium in the wine, which can impact fermentation and flavor complexity.
Total Phenols: The total amount of phenolic compounds, which contribute to the wine's color, flavor, and mouthfeel.
Flavanoids: A subset of phenolic compounds known for their antioxidant properties and influence on flavor and color.
Nonflavanoid Phenols: Phenolic compounds that are not classified as flavonoids, contributing to the wine's overall profile.
Proanthocyanins: Compounds responsible for astringency and color in red wines, affecting both flavor and aging potential.
Color Intensity: A measure of the depth of color in the wine, indicating its richness and potential aging ability.
Hue: The overall color quality of the wine, influenced by its chemical composition.
OD280/OD315 of Diluted Wines: The optical density measured at two wavelengths (280 nm and 315 nm), used to assess the concentration of phenolic compounds in the wine.
Proline: An amino acid present in wine that can influence flavor and the wine's aromatic profile.

Relevant libraries

tidyverse, factoextra, caret, e1071, cluster, dplyr, tinytex, dbscan, fpc.

Answer the following 5 questions

K-Means Clustering

Q1-A) [30 Points] Explain how normalization and the number of clusters (k) can affect the clustering outcome?

Q1-B) [40 Points] Using the wine dataset, perform K-Means clustering to group the wines into clusters based on their chemical attributes and your optimal *k*. Make sure to discuss the significance of the clusters formed and any insights you can draw from the results, i.e., interpret each cluster.

Hint: Partial code view

```
wines.norm <- scale(wines)
#Finding the value of "k" by using "wss" and "silhouette"
fviz_nbclust()
kmeans()
fviz_cluster()#
kmeans.wines$centers
kmeans.wines$size
kmeans.wines$withinss
```

~~~~~~

DBSCAN Clustering

Q2-A) [30 Points] Explain how the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm differs from K-Means clustering, particularly in terms of handling noise and outliers.

Q2-B) [40 Points] Assume k=3, apply DBSCAN to the wine dataset, and analyze the clusters formed. How do the parameters epsilon (ε) and minPts influence the clustering results?

Hint: Partial code view:

```
#Finding the epsilon Value
dbscan::kNNdistplot()
abline()
#Clustering using DBSCAN
DBscan.wine <- fpc::dbscan()
plot(DBscan.wine,…)
```

~~~~~~

Compare K-Means and DBSCAN

Q3) [30 Points] Explain pros and cons of each of these two methods based on the outputs observed in the previous questions.

Note that Explanations can be included in html as comments using #. Also, there is no unique way to develop your code; provided partial codes (pseudocodes) just serve you as hints and you don't need to use them.

- ❖ Submit your report in HTML format using R-Markdown. In your answers, please include your questions numbers, namely, Q1-A, Q1-B, etc.