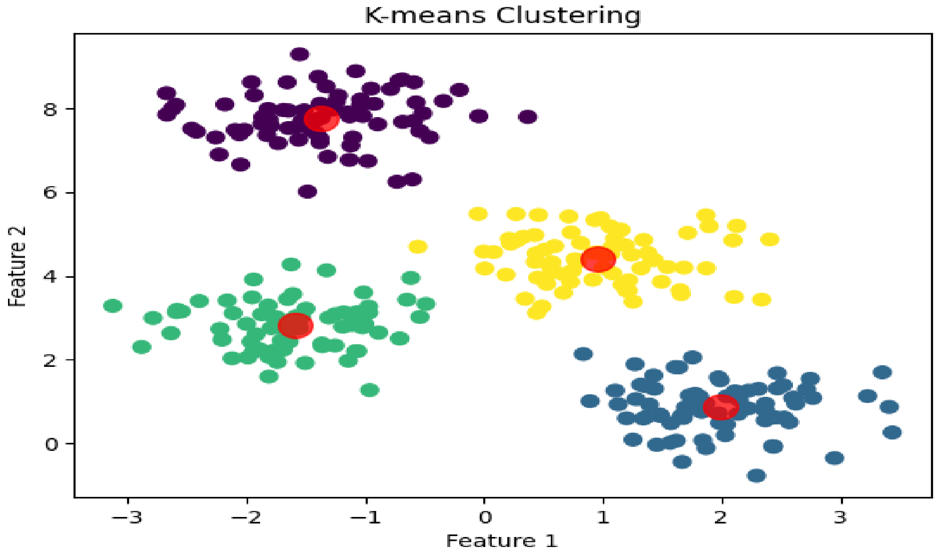
```
plt.scatter(X[:, 0], X[:, 1], c=y_kmeans, s=50, cmap='viridis')
centers = kmeans.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='red', s=200, alpha=0.75)
plt.title('K-means Clustering')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.show()
```

Output



This program does the following:

1. Generates synthetic data using make_blobs from sklearn.datasets.
2. Applies K-means clustering with n_clusters=4.
3. Plots the data points colored by their cluster assignments and shows the centroids as red circles.

You can adjust the parameters like the number of clusters, standard deviation, and number of samples in make_blobs to observe different clustering scenarios.

REFERENCES

Video Session

Videos: Clustering: https://www.youtube.com/watch?v=wk2ylI1qgU0

K-means clustering: https://www.youtube.com/watch?v=4b5d3muPQmA&t=119s

EXERCISES

A. Multiple Choice Questions

1. Which of the following are the types of correlation?
   a. Positive correlation
   b. Negative Correlation
   c. No correlation
   d. All of the above
2. Which of the following techniques is an analysis of the relationship between two variables to provide the prediction mechanism?
   a. Standard error
   b. Correlation
   c. Regression

   d. None of the above

3. Which of the given plots is suitable for testing the linear relationship between a dependent and independent variable?
   a. Bar chart
   b. Scatter plot
   c. Histograms
   d. All of the above
4. Which of the following scatter plots represents a positive correlation?
   a. points scattered randomly with no apparent trend
   b. points forming a diagonal line and bottom left to top right
   c. points forming a diagonal line from top left to bottom right
   d. points clustered around a central point
5. Which regression technique is used when there is only one independent variable?
   a. logistic regression
   b. multiple linear regression
   c. simple linear regression
   d. polynomial regression
6. What is one advantage of linear regression analysis?
   a. it is robust to outliers
   b. it can capture nonlinear relationships between variables
   c. it is simple and easy to interpret
   d. it is suitable for classification tasks
7. What is supervised learning in Artificial Intelligence?
   a. training a computer algorithm on input data that is not labelled.
   b. training a computer algorithm on input data that has been labelled for a specific output.
   c. training a computer algorithm without any input data
   d. training a computer algorithm to perform unsupervised tasks.
8. Which type of classification involves categorizing data into two distinct classes?
   a. multi-class classification
   b. binary classification
   c. unsupervised classification
   d. regression classification
9. What is logistic regression commonly used for in binary classification?
   a. categorizing observations into multiple classes
   b. predicting continuous values for input data
   c. categorizing observations into two distinct classes
   d. identifying unstructured data patterns
10. What is the primary goal of classification in AI?
    a. categorizing data into random groups
    b. locating and classifying things or concepts into predefined groups
    c. predicting continuous values for input data
    d. identifying unstructured data patterns

11. Which algorithm is commonly used for binary classification?
    a. Decision trees
    b. Support Vector Machine
    c. Logistic Regression
    d. k-Nearest Neighbors
12. The K-Nearest Neighbors (KNN) algorithm assigns a class to new data point by considering:
    a. Distance from the data point to a predefined decision boundary
    b. Majority vote of its K nearest neighbors in the training data
    c. Similarity of the data point to a cluster centroid
    d. probability of each class given the data point’s features.
13. What does a classification model in AI ultimately want to achieve?
    a. to identify patterns and associations in data
    b. to predict continuous numerical values
    c. to categorize input data into predefined classes or labels
    d. to optimize decision-making processes
14. What are some challenges in applying classification models to real-world problems?
    a. Data bias and fairness
    b. Interpretability and explainability
    c. overfitting and underfitting
    d. All of the above
15. What is clustering?
    a. Grouping labeled dataset
    b. Dividing data into different clusters
    c. Finding linear association between variables
    d. Predicting future behaviors of a dependent variable
16. Which type of learning does clustering belong to?
    a. Supervised learning
    b. Unsupervised learning
    c. Semi-supervised learning
    d. Reinforcement learning
17. Which method is used to group highly dense areas into clusters?
    a. Partitioning clustering
    b. Density-based clustering
    c. Distribution model-based clustering
    d. Hierarchical clustering
18. Which algorithm is an example of partitioning clustering?
    a. Mean-shift algorithm
    b. DBSCAN algorithm
    c. K-Means algorithm
    d. Fuzzy clustering algorithm

19. Which clustering method allows data objects to belong to more than one group or cluster?
    a. Partitioning clustering
    b. Density-based clustering
    c. Distribution model-based clustering
    d. Fuzzy clustering
20. Which clustering algorithm is sensitive to outliers?
    a. K-Means algorithm
    b. Mean-shift algorithm
    c. DBSCAN algorithm
    d. Hierarchical clustering

B. Fill in the blanks

1. In ___________ type of ML, the models are not trained in labeled data sets.
2. The ___________________ measures the linear relationship between the independent and dependent variables.
3. _________________predicts continuous numerical values, while Logistic regression predicts discrete categories.
4. _____________ are data points on the scatterplot that do not follow the pattern of the dataset.
5. _______________ algorithm operates based on the principle of proximity, making predictions by considering the similarity between data points.
6. Clustering is a machine learning technique used to group ____________ dataset.
7. Partitioning clustering divides the data into non-hierarchical groups, also known as ____________ method.
8. Density-based clustering connects highly dense areas into clusters, separated by areas of ____________.
9. The primary requirement for the number of clusters in K-Means algorithm is ____________ beforehand.
10. Clustering is widely used in applications such as market segmentation and ____________.

C. True or False:

1. Clustering is a supervised learning technique.
2. Hierarchical Clustering requires pre-specifying the number of clusters.
3. Fuzzy clustering is a hard clustering method.
4. Classification is an unsupervised learning technique.
5. In k-NN algorithm, k is the number of nearest data points.
6. K-Means algorithm requires specifying the number of clusters.