STA 3180 Statistical Modelling: Clustering

Topic: Clustering

I. Introduction to Clustering

A. Definition of Clustering

B. Types of Clustering

1. Hierarchical Clustering

2. K-Means Clustering

3. Density-Based Clustering

II. Clustering Algorithms

A. K-Means Algorithm

1. Main Things to Study:

a. Steps of the K-Means Algorithm

b. Distance Measures

c. Choosing the Number of Clusters

2. Problem Solving Strategies:

a. Visualize the data to identify clusters

b. Use the elbow method to determine the optimal number of clusters

c. Use the silhouette coefficient to evaluate the quality of the clusters

B. Hierarchical Clustering Algorithm

1. Main Things to Study:

a. Steps of the Hierarchical Clustering Algorithm

b. Linkage Criteria

2. Problem Solving Strategies:

a. Visualize the data to identify clusters

b. Use the dendrogram to determine the optimal number of clusters

c. Use the cophenetic correlation coefficient to evaluate the quality of the clusters

C. Density-Based Clustering Algorithm

1. Main Things to Study:

a. Steps of the Density-Based Clustering Algorithm

b. Density Estimation

2. Problem Solving Strategies:

a. Visualize the data to identify clusters

b. Use the density-based clustering algorithm to identify clusters in high-dimensional data

c. Use the Davies-Bouldin index to evaluate the quality of the clusters