HW2_MVA_Fall2025

2025-12-05

Cluster Analysis and Discriminant Analysis

In the exercises below you will continue on working the player statistics of four teams taken part in Final Four of FIBA EuroBasket 2025.

The variables in the data set were defined as follows:

GP: Games Played

FG: Field Goals

2PT FG: 2pt Field Goal

3PT FG: 3pt Field Goal

FT: Free Throws

OREB: Offensive Rebounds

DREB: Defensive Rebounds

REB: Total Rebounds

AST: Assists

PF: Personal Fouls

TO: Turnovers

STL: Steals

BLK: Blocks

"+/-": Point differential of the score while player on the court

EFF: Efficiency

PTS: Points

You had already imported the data set as assigned player names as rows and assigned correct variable types for each variable in HW1. In the following exercises you will be using the data frame that you created in 1c of Hw1 which consists of the variables "Position", "MIN", "FG","2PT.FG","3PT.FG", "FT","OREB","DREB", "REB", "AST", "PF", "TO","STL","BLK","EFF" and "PTS".

1. Hierarchical Clustering.

Apply a hierarchical clustering on previously created data frame that includes variables "Position", "MIN", "FG","2PT.FG","3PT.FG", "FT","OREB","DREB", "REB", "AST", "PF", "TO","STL","BLK","EFF" and "PTS". Scale it if it is required.

    a) Decide the best method to draw dendrogram.(0.5p)

    b) How many clusters do you think exist in the data considering the dendrogram? (0.5p)

2. K means clustering.

a) Find the optimum number of clusters. (0.5p)

b) Apply k-means clustering using the number of clusters that you decided. Interpret each cluster by making a descriptive statistics table. (1.5p)

3. Hierarchical Clustering based on PCA and MCA.

a) Use HCPC() function and apply a PCA based clustering on the pca output that you had from HW1. Decide on number of clusters. (0.5)

b) Interpret each cluster by using profiling.(1p)

c) Use HCPC() function and apply a MCA based clustering on the multiple correspondence analysis output that you had from HW1.Decide on number of clusters. (0.5p)

d) Interpret each cluster by using profiling. (1p)

4. Conclusion

Compare the results of clustering methods that you applied and make a final comment on which clustering method would you prefer and why. (0.5p)

5. Discriminant Analysis

Our objective is to fit a prediction model to classify players according to their positions.

a) Consider only the variables that have approximately normal distribution and test the assumption of homogenity of variance. (1p)

b) Fit the proper discriminant model on the data set and explain why you use it. (0.5p)

c) Make a table cross classifying real and predicted values and interpret correct classification rates. (1p)

d) Test whether the classification is done by chance using Q statistic. (1p)

NOTE: You should change the name of the variables that includes number such "2PT.FG" and "3PT.FG" to "TwoPt.FG" and "ThreePt.FG" to avoid errors while running analyses.