

If already registered, click to check your payment status

Course outline

About NPTEL

How does an NPTEL online course work?

Setup Guide

Pre Course Material

Week 0

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

- ☐ K - Nearest Neighbors (kNN)
- ☐ K - Nearest Neighbors implementation in R
- ☐ K - means Clustering
- ☐ K - means implementation in R
- ☐ FAQ
- ☐ Data Science for engineers - Summary
- ☐ Datasets
- ☒ Practice: Week 8: Assignment 8 (Non Graded)

Quiz: Week 8: Assignment 8

- ☐ Week 8 Feedback Form : Data Science For Engineers

Text Transcripts

Download Videos

Books

Problem Solving Session - Jan 2024

Week 8: Assignment 8

The due date for submitting this assignment has passed.

Due on 2024-03-20, 23:59 IST.

Assignment submitted on 2024-03-20, 20:33 IST

Consider the dataset “USArrests.csv”. Answer questions 1 to 4 based on the information given below:

This data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

Variables	Description
States	The state where the incident occurred
Murder	No. of arrests for murder (per 100,000 residents)
Assault	No. of arrests for assault (per 100,000 residents)
UrbanPop	Percentage of urban population
Rape	Rape arrests (per 100,000 residents)

- Set the column "States" as index of the data frame while reading the data
 - Set the random number generator to set.seed(123)
 - Normalize the data using scale function and build the K-means algorithm with the given conditions:
 - number of clusters = 4
 - nstart=20

1) According to the built model, the within cluster sum of squares for each cluster is _____ (the order of values in each option could be different):- **1 point**

- ☒ 8.316061 11.952463 16.212213 19.922437

☐ 7.453059 12.158682 13.212213 21.158766

☐ 8.316061 13.952463 15.212213 19.922437

☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

8.316061 11.952463 16.212213 19.922437

2) According to the built model, the size of each cluster is _____ (the order of values in each option could be different):- **1 point**

- ☐ 13 13 7 14

☐ 11 18 25 24

☒ 8 13 16 13

☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

8 13 16 13

3) The Between Cluster Sum-of-Squares (BCSS) value of the built K-means model is _____ (Choose the appropriate range) **1 point**

- ☒ 100 - 200

☐ 200 - 300

☐ 300 – 350

☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

100 - 200

4) The Total Sum-of-Squares value of the built k-means model is _____ (Choose the appropriate range) **1 point**

- ☒ 100 - 200

☐ 200 - 300

☐ 300 – 350

☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

100 - 200

5) Which of the statement is INCORRECT about KNN algorithm? **1 point**

- ☒ KNN works ONLY for binary classification problems

☐ If k=1, then the algorithm is simply called the nearest neighbour algorithm

☐ Number of neighbours (K) will influence classification output

☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

KNN works ONLY for binary classification problems

6) K means clustering algorithm clusters the data points based on:- **1 point**

- ☐ Dependent and independent variables

☐ The eigen values

☒ Distance between the points and a cluster centre

☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

Distance between the points and a cluster centre

7) The method / metric which is NOT useful to determine the optimal number of clusters in unsupervised clustering algorithms is **1 point**

- ☒ Scatter plot

☐ Elbow method

☐ Dendrogram

☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

Scatter plot

8) The unsupervised learning algorithm which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest centroid is **1 point**

- ☐ Hierarchical clustering

☒ K-means clustering

☐ KNN

☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

K-means clustering

