

X


<https://swayam.gov.in>

https://swayam.gov.in/nc_details/NPTEL

reviewer4@nptel.iitm.ac.in ✓

NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » **Data Science for Engineers (course)**

 Announcements (announcements) **About the Course** (https://swayam.gov.in/nd1_noc20_cs28/preview)

Ask a Question (forum) Progress (student/home) Mentor (student/mentor)

Unit 10 - Week 8

Course outline

How does an NPTEL online course work?

Week 0

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

☐ K - Nearest Neighbors (kNN) (unit? unit=64&lesson=65)

☐ K - Nearest Neighbors implementation in R (unit? unit=64&lesson=66)

Assignment 8

The due date for submitting this assignment has passed. **Due on 2020-03-25, 23:59 IST.**
As per our records you have not submitted this assignment.

Click here (<https://drive.google.com/open?id=18GBZKzwbWTZbmb7K8a2T9dHve0ESDsvc>) to download the Data Set

1) For k-NN, which of the following distance measure is **NOT** valid for continuous variables? **1 point**

- ☐ Euclidean distance
- ☐ Manhattan distance
- ☐ Minkowski distance
- ☐ Hamming Distance

No, the answer is incorrect.

Score: 0

Accepted Answers:

Hamming Distance

2) Which of the following statement is/are **NOT** true about k-NN algorithm? **1 point**

- ☐ k-NN is a non-parametric machine learning algorithm
- ☐ It is also called instance-based learning
- ☐ Performance is not better when all the data have same order of magnitude
- ☐ k-NN is often termed as lazy learning algorithm

No, the answer is incorrect.

Score: 0

Accepted Answers:

Performance is not better when all the data have same order of magnitude

3) In K-means clustering, the clusters are compact and homogenous:- **1 point**

☐ K - means Clustering (unit? unit=64&lesson=67)

☐ K - means implementation in R (unit? unit=64&lesson=68)

☐ FAQ (unit? unit=64&lesson=69)

☐ Data Science for engineers - Summary (unit? unit=64&lesson=70)

☒ Week 8 data set (unit? unit=64&lesson=71)

☐ Quiz : Practice Assignment 8 (assessment? name=97)

☐ Quiz : Assignment 8 (assessment? name=127)

☐ Week 8 Feedback (unit? unit=64&lesson=129)

☒ Solution - Assignment 8 (unit? unit=64&lesson=131)

Text Transcripts

Download Videos

- ☐ When the WSS is lower and the BSS is higher
- ☐ When the BSS is lower and the WSS is higher
- ☐ When the TSS is lower and the BSS is higher
- ☐ When the TSS is higher and the BSS is lower

No, the answer is incorrect.

Score: 0

Accepted Answers:

When the WSS is lower and the BSS is higher

4) Which of the following statement(s) is/are **TRUE** with respect to elbow plot?

1 point

- ☐ Elbow plot is used to ascertain the optimal number of clusters
- ☐ Elbow plot is used to explain percentage variance as a function of number of clusters
- ☐ The value of K at which the percentage variance starts plateauing is called elbow
- ☐ None of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

Elbow plot is used to ascertain the optimal number of clusters

Elbow plot is used to explain percentage variance as a function of number of clusters

The value of K at which the percentage variance starts plateauing is called elbow

Based on the information given below answer the questions **5 and 6**.

Read the given dataset "**wine.csv**" in Rstudio.

Variables	Description
Alcohol	Alcohol
Malic	Malic acid
Ash	Ash
Alcalinity	Alcalinity of ash
Magnesium	Magnesium
Phenols	Total Phenols
Flavanoid	Flavanoids
Nonflavanoids	Non flavonoid phenols
Proanthocyanins	Proanthocyanins
Color	Color intensity
Hue	Hue
Dilution	D280/OD315 of diluted wines.
Proline	Proline

Normalize the data using scale function and build the K-means algorithm with the given conditions:

random number generator = 1234

number of clusters = 3

nstart=25

5) The within cluster sum of squares is

1 point

- ☐ 451.3, 284.6, 459.7
- ☐ 268.6, 290.0, 307.1

- ☐ 46.5, 203.3, 326.3
- ☐ 326.4, 385.7, 558.7

No, the answer is incorrect.

Score: 0

Accepted Answers:

326.4, 385.7, 558.7

6) The size of each of the clusters is

1 point

- ☐ 59, 65, 48
- ☐ 26, 57, 37
- ☐ 62, 65, 51
- ☐ 16, 30, 49

No, the answer is incorrect.

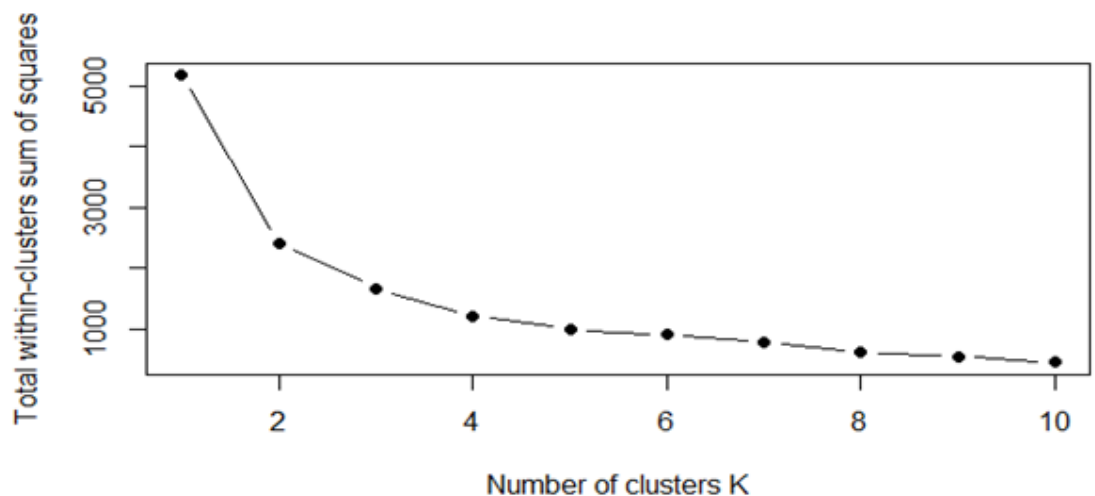
Score: 0

Accepted Answers:

62, 65, 51

7) From the elbow plot given below, the optimal numbers of clusters are

1 point



- ☐ 3 or 4
- ☐ 4 or 5
- ☐ 5 or 6
- ☐ 1 or 2

No, the answer is incorrect.

Score: 0

Accepted Answers:

3 or 4

8) Consider a kNN model built on a large set of data. Which of the following statements best describe the effect of choosing a lower k value when the data is considerably large?

1 point

- I - The model built with a lower k value makes it computationally expensive
- II- The model overfits

- ☐ I - True , II - True
- ☐ I- False , II - True

- ☐ I - True , II - False
- ☐ I- False , II – False

No, the answer is incorrect.

Score: 0

Accepted Answers:

I- False , II - True

9) The Euclidean distance between the two data points A (-3,3) and B (7,7) is _____

1 point

- ☐ $\sqrt{115}$
- ☐ $\sqrt{116}$
- ☐ $\sqrt{117}$
- ☐ $\sqrt{118}$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$\sqrt{116}$

10) When the k Nearest Neighbours algorithm is used for prediction, the values are calculated by: **1 point**

- ☐ minimizing the distance between the predictions
- ☐ averaging the values of the k nearest neighbours
- ☐ averaging the inter-cluster distances
- ☐ averaging the intra-cluster distances

No, the answer is incorrect.

Score: 0

Accepted Answers:

averaging the values of the k nearest neighbours

11) What is the advantage of using an odd value of k in binary classification problems using kNN algorithm? **1 point**

- ☐ Avoid tied votes on classes
- ☐ Makes decision boundary smoother
- ☐ Makes boundaries between classes more distinct
- ☐ Reduce the effect of noise

No, the answer is incorrect.

Score: 0

Accepted Answers:

Avoid tied votes on classes

12) kNN is used for both regression and classification problems

1 point

- ☐ True
- ☐ False

No, the answer is incorrect.

Score: 0

Accepted Answers:

True

13) Scaling is important in distance-based algorithms because

1 point

- ☐ variables with higher magnitude will influence the results more
- ☐ calculation of distances between points are affected by the magnitude of the variables
- ☐ the data is large and has many features
- ☐ it helps in speeding up the calculations in an algorithm

No, the answer is incorrect.

Score: 0

Accepted Answers:

variables with higher magnitude will influence the results more

calculation of distances between points are affected by the magnitude of the variables

it helps in speeding up the calculations in an algorithm