

X

reviewer2@nptel.iitm.ac.in ▼

Courses » Introduction to Data Analytics

Announcements

Course

Ask a Question

Progress



Unit 9 - Week 8 - Clustering Analysis and Prescriptive Analytics

Course outline

[How to access the portal](#)[Week 1 - Course Overview and Descriptive Statistics](#)[Week 2 - Probability Distributions & Inferential Statistics](#)[Week 3 - Inferential Statistics](#)[Week 4 - Machine Learning](#)[Week 5 - Supervised Learning \(Regression and Classification Techniques\) - I](#)[Week 6 : Supervised Learning \(Regression and Classification Techniques\)-II](#)[Week 7 - Association Rule Mining and Big Data](#)[Week 8 - Clustering Analysis and Prescriptive Analytics](#) Clustering Analysis Clustering Analysis (cont'd) Introduction to Experimentation and Active Learning Introduction to Experimentation and Active Learning(cont'd) An Introduction to Online Learning - Reinforcement Learning An Introduction to Online Learning - Reinforcement Learning (cont'd) Quiz : Assignment 8 Feedback for week 8 Assignment 8: Solution[Course Summary+ Insight into the Final Exam](#)

Assignment 8

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2017-09-20, 23:55 IST**1 point**

1) In solving a n-armed bandit problem, we decide to use the ϵ -greedy algorithm. We have the following two choices for the ϵ parameter, $\epsilon = 0.1$ and $\epsilon = 0.01$. If we run the algorithm for a very long time, and want to maximize the cumulative reward, which value of ϵ should we use?

Suppose we want to find an optimal arm in the fastest time. In this situation, which value of ϵ should we use?

- for maximum reward use $\epsilon = 0.1$; for minimising the time required to find an optimal arm, use $\epsilon = 0.01$
- for maximum reward use $\epsilon = 0.01$; for minimizing the time required to find an optimal arm, use $\epsilon = 0.1$
- for both cases, use $\epsilon = 0.1$
- for both cases, use $\epsilon = 0.01$

No, the answer is incorrect.

Score: 0

Accepted Answers:

for maximum reward use $\epsilon = 0.01$; for minimizing the time required to find an optimal arm, use $\epsilon = 0.1$

2) Suppose that you have been given a number of different drug formulations to treat a particular disease and your job is to identify one **1 point** among them that best meets certain criteria with regards to its efficacy in treating the disease. Before you run the experiments, you need to provision for the samples that would be required. Treating this as a n-armed bandit problem, which kind of solution method would you prefer for identifying the best option?

- asymptotic correctness
- regret optimality
- PAC optimality

No, the answer is incorrect.

Score: 0

Accepted Answers:

PAC optimality

3) After 12 iterations of the UCB algorithm applied on a 4-arm bandit problem, we have $n_1 = 3, n_2 = 4, n_3 = 3, n_4 = 2$ and $\bar{x}_1 = 0.55, \bar{x}_2 = 0.63, \bar{x}_3 = 0.61, \bar{x}_4 = 0.40$. Which arm should be played next? **2 points**

- 1
- 2
- 3
- 4

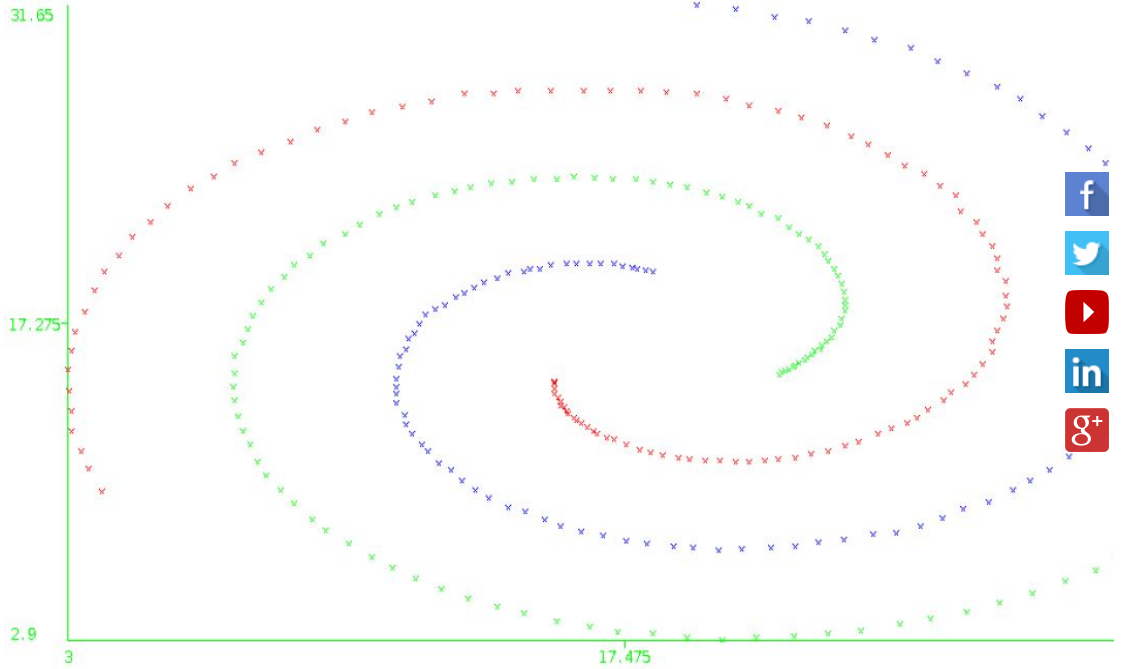
No, the answer is incorrect.

Score: 0

Accepted Answers:

4

4) Consider the following image showing data points belonging to three different clusters (indicated by the colors of the points). 1 point



If we run the k-means clustering algorithm with $k = 3$, do you think the algorithm will be able to correctly cluster the data points belonging to the three clusters?

- no
- yes

No, the answer is incorrect.

Score: 0

Accepted Answers:

no

5) Consider the following one dimensional data set: 12, 22, 2, 3, 33, 27, 5, 16, 6, 31, 20, 37, 8, 18. Given $k = 3$ and initial cluster centers to be 5, 6 and 31, what are the final cluster centers obtained on applying the k-means algorithm? 1 point

- 5, 18, 30
- 5, 18, 32
- 6, 19, 32
- 4.8, 17.6, 32

No, the answer is incorrect.

Score: 0

Accepted Answers:

4.8, 17.6, 32

6) In outlier mining, we try to detect data points that do not fit a particular distribution. Which among the following are disadvantages of using DBSCAN for performing outlier mining? 1 point

(Note that more than one statement may be correct)

- No need to make any changes to the DBSCAN algorithm.
- DBSCAN does not work well when there is large variation in the density of the data point groupings.
- DBSCAN can detect outlier points which lie between groupings of data points and not just points far away from all other data points.
- Slight change in the DBSCAN parameters can lead to large variation in the resultant outlier points detected.

No, the answer is incorrect.

Score: 0

Accepted Answers:

DBSCAN does not work well when there is large variation in the density of the data point groupings.

Slight change in the DBSCAN parameters can lead to large variation in the resultant outlier points detected.

7) What would, in general, be the effect of increasing MinPts in DBSCAN while retaining the same Eps parameter? (Note that more than one statement may be correct) 1 point

- Increase in the sizes of individual clusters
- Decrease in the sizes of individual clusters
- Increase in the number of clusters
- Decrease in the number of clusters

No, the answer is incorrect.

Score: 0

Accepted Answers:

Decrease in the sizes of individual clusters

Increase in the number of clusters

8) Given a two-class training data set with 100 unlabelled data points, suppose we randomly select 10 data points and query for their labels. We supply these 10 labelled data points to a SVM, and obtain a decision boundary. Assuming a limit on the number of additional points that we can select to improve this classifier, in general, would you prefer to query the labels of points lying close to the decision surface or those that are far from the decision surface? 1 point

- close to the decision surface

far from the decision surface

No, the answer is incorrect.

Score: 0

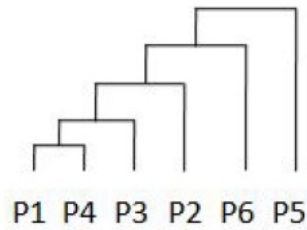
Accepted Answers:

close to the decision surface

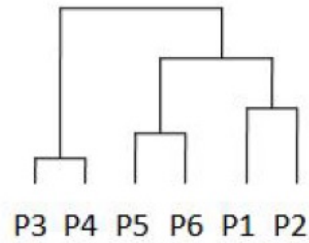
9) For the similarity matrix given below, show the hierarchy of clusters created by the single link clustering algorithm.

1 point

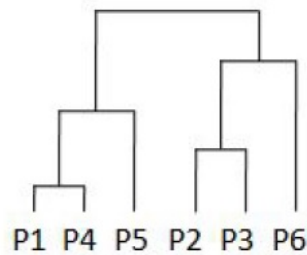
	P1	P2	P3	P4	P5	P6
P1	1.0000	0.7895	0.1579	0.0100	0.5292	0.3542
P2	0.7895	1.0000	0.3684	0.2105	0.7023	0.5480
P3	0.1579	0.3684	1.0000	0.8421	0.5292	0.6870
P4	0.0100	0.2105	0.8421	1.0000	0.3840	0.5573
P5	0.5292	0.7023	0.5292	0.3840	1.0000	0.8105
P6	0.3542	0.5480	0.6870	0.5573	0.8105	1.0000



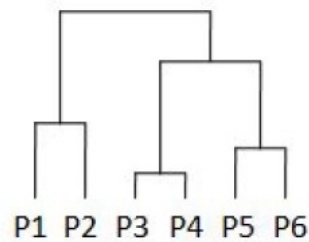
(a)



(b)



(c)



(d)

- a
- b
- c
- d

No, the answer is incorrect.

Score: 0

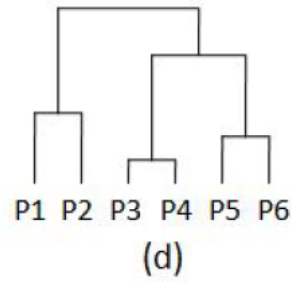
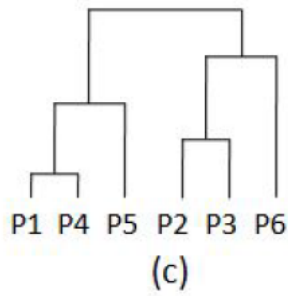
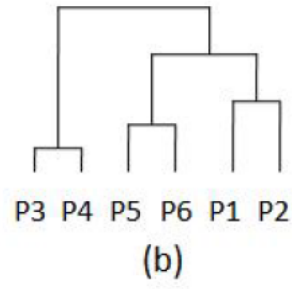
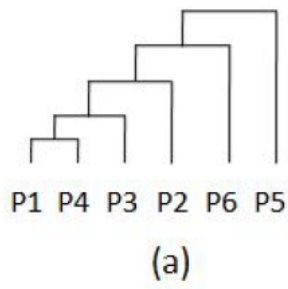
Accepted Answers:

b

10) For the similarity matrix given in the previous question, show the hierarchy of clusters created by the complete link clustering algorithm.

1 point





- a
- b
- c
- d

No, the answer is incorrect.

Score: 0

Accepted Answers:

d

Previous Page

End

