QUIZ NAVIGATION

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15

Show one page at a time

Finish review

Started on	Thursday, 7 October 2021, 12:10 PM
State	Finished
Completed on	Thursday, 7 October 2021, 1:10 PM
Time taken	59 mins 59 secs
Grade	**53.00** out of 64.00 (**83**%)

Question 1

Complete

Mark 4.00 out of 4.00

Flag question

Suppose, you are clustering the following eight points into three clusters: $A_1(2, 10)$, $A_2(2, 5)$, $A_3(8, 4)$, $A_4(5, 8)$, $A_5(7, 5)$, $A_6(6, 4)$, $A_7(1, 2)$, $A_8(4, 9)$. Initial cluster centers are: $A_1(2, 10)$, $A_4(5, 8)$ and $A_7(1, 2)$. The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as

$d(a,b)=|x_2 - x_1|+|y_2 - y_1|$

After the first iteration of the k-means algorithm, how many points will belong to the second cluster?

Answer: 5

Question 2

Complete

Mark 4.00 out of 4.00

Flag question

Suppose you are running a k-means clustering algorithm to cluster N data points into K clusters. Each data point is D dimensional. If you run the algorithm for I iterations, the running time complexity of the algorithm will be

Select one:

a. $O(NKDI)$

b. $O(NKD^2I)$

c. $O(NKI)$

d. $O(NK^2DI)$

Question 3

Complete

Mark 0.00 out of 4.00

Flag question

Suppose we want to train a Bayes classifier for a dataset where the input is a 6-dimensional real vector and output is a binary variable. Suppose, the 6-dimensional feature vector follow Gaussian distribution. How many independent parameters need to be estimated to train the classifier?

Answer: 12

Question 4

Complete

Mark 4.00 out of 4.00

Flag question

In a random survey of 1000 U.S. college students, 220 identified themselves as smokers. Construct a 90% confidence interval for the proportion of all U.S. college students who would identify themselves as smokers.

Select one:

a.

(0.199, 0.220)

b.

(0.207, 0.254)

c.

(0.199, 0.242)

d.

(0.220, 0.242)

Question 5

Complete

Mark 4.00 out of 4.00

Apply *DBSCAN* to the dataset D. Consider ε = 4 and *MinPts* = 3. Definition of core point: if a point *x* has *Min Pts* neighbors around it within distance ε, then *x* is a core point. Identify border point(s). Use Euclidean distance for your computation. For example, *x* has neighbors *p*, *q* and *r* within ε-distance. We say *x* has 3 neighbors.

Flag question

Table 1: Dataset D

Feature-1	Feature-2
13	28
13	24
13	29
13	27
17	12
15	12
15	11
16	12
8	3
8	4

Select one:

a.
(13,24) and (17,12)

b.

(8,3) and (15,11)

c.
(13,24) and (13,29)

d.
(13,24) and (17,12)

Question 6

Complete

Mark 4.00 out of 4.00

Flag question

You are given a dataset {0, 1, 1, 2, 3, 4, 4, 4, 5}. Consider the following kernel

$$k(d) = \begin{cases} 1 - |d| & \text{if } |d| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

where d is a function of the distance of sample x_i to the value in question x divided by the bandwidth. Compute the kernel density estimate for $x = 2$ when the bandwidth is 2. The answer should be correct upto 4 decimal places.

Answer: 0.13889

Question 7

Complete

Mark 4.00 out of 4.00

Based on the data what is P(Favourite Sport= Cricket | Income= Medium)?

Flag question

Table 2: Training examples.

Age	Income	Favourite Sport	Time spent watching sport
30-40	High	Cricket	Medium
20-30	Medium	Football	Low
30-40	High	Hockey	Low
30-40	Low	Cricket	High
20-30	High	Basketball	Medium
40-50	Medium	Cricket	High
20-30	Low	Football	Low
40-50	High	Cricket	High

Answer: 0.5

Question 8

Complete

Mark 4.00 out of 4.00

Flag question

Consider a data set which provides the value of an attribute A and corresponding class labels, either 0 or 1. Given a test sample with the attribute value 50 compute the posterior probability for class 0 using k-NN classification method, where the value of $k = 6$.

Value of *A* for an item	class
-23	0
35	1
-12	1
33	0
-45	1
22	0
-16	0
44	0
-77	0
-17	1
62	0

Use Euclidean distance to calculate nearest neighbor

Answer: 0.6667

Question 9

Complete

Mark 6.00 out of 6.00

Flag question

Which of the following statements are true in the context of Apriori algorithm for mining association rules from relational databases.

Select one or more:

- [ ] a.
  It uses the measured lift to reduce the number of candidate association rules.

- [x] b.
  It finds associations among values of attributes of relational tables.

- [ ] c.
  It finds associations among attributes of relational tables.

- [x] d.
  It uses the measure support in a greedy approach.

Question 10

Complete

Mark 3.00 out of 6.00

Flag question

Given N data points which of the following statements are true in the context of Hierarchical clustering algorithm.

Select one or more:

- [x] a.
  It has *N* clusters at its bottom-most level of hierarchy.

- [ ] b.

The ordering of data points affect the outcome.

c.

Given the distances between a pair of data points, explicit values of data points are also
required in computation.

d.

There is exactly N levels in the hierarchy.

Question 11

Complete

Mark 4.00 out of 4.00

Flag question

Consider a naive Bayes classifier with 3 boolean input variables, X_1, X_2 and X_3, and one boolean output, Y. How many independent parameters would have to be estimated to learn the above classifier if we do not make the naive Bayes conditional independence assumption?

Answer: 15

Question 12

Complete

Mark 4.00 out of 4.00

Flag question

Consider a naive Bayes classifier with 3 boolean input variables, X_1, X_2 and X_3, and one boolean output, Y. How many independent parameters need to be estimated to train such a naive Bayes classifier?

Answer: 7

Question 13

Complete

Mark 4.00 out of 4.00

Flag question

Consider the following 10 2D data points: $x_1(1, 5)$, $x_2(1, 7)$, $x_3(3, 3)$, $x_4(0, 0)$, $x_5(2, 7)$, $x_6(0, 3)$, $x_7(1, 1)$, $x_8(5, 5)$, $x_9(1, 3)$, $x_{10}(2, 9)$. Given the parameters of the DBSCAN algorithm as *Distance Threshold*(ε)= 3, and *Minimum Number Of Neighboring Points*(*Min Pnt*)=3. Find all core points, border points and noisy points, if exist.

NOTE: While you count number of neighbors, do not count the point itself. For example if a point x finds y and z as her neighbor, then x's |neighborhood|=2

Use Euclidean distance

Select one:

a.

Core={$x1$,$x2$,$x3$,$x6$,$x7$,$x9$}
Border={$x5$}
Noise= {$x4$, $x8$, $x10$}

b.

Core={$x1$,$x2$,$x3$,$x5$,$x6$,$x7$,$x9$}
Border={$x8$, $x10$}
Noise= {$x4$}

c.

Core={$x1$,$x2$,$x3$,$x4$,$x6$,$x7$}
Border={$x8$, $x9$, $x10$}
Noise= {$x5$}

d.

Core={$x1$,$x2$,$x3$,$x5$,$x6$,$x7$,$x9$}

Border={x4, $x8$, $x10$}

Noise= {}

e.

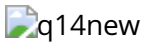
Core={$x1$,$x2$,$x3$,$x5$,$x6$,$x7$,$x9$}
Border={$x4$, $x8$}
Noise= {$x10$}

Question 14

Complete

Mark 4.00 out of 4.00

Flag question



Answer should be correct up to 4 decimal places.

Answer: 0.0144

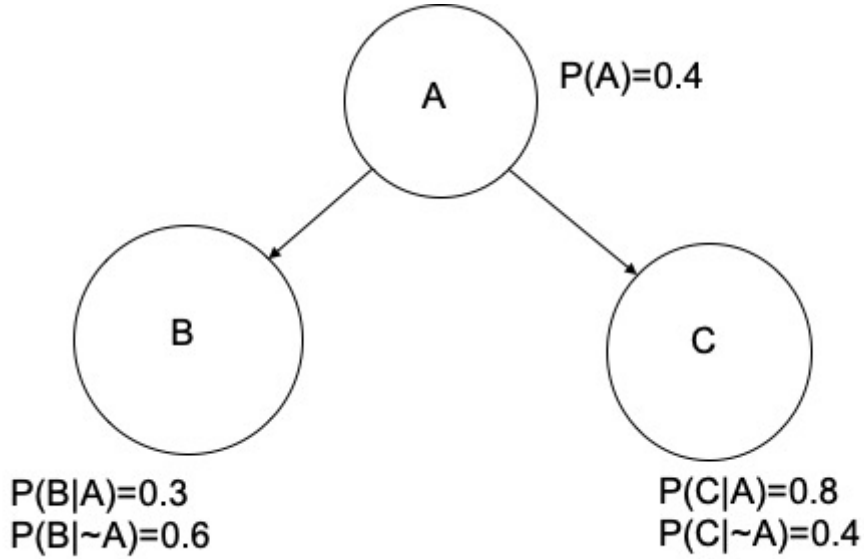
Question 15

Complete

Mark 0.00 out of 4.00

Flag question

Consider the following figure depicting a Bayesian Network of three nodes, each representing a boolean variable. Corresponding probabilities associated with nodes are also shown in the figure. Compute $P(A| \sim B, C)$.



Answer: 0.769

Finish review