Consider the following distance matrix. Assuming that C and D are initial centroids for K-means algorithm, what would be the clusters resulted after the first epoch?

	A	B	C	D	E	F	G
A	0						
B	45	0					
C	13	6	0				
D	21	21	9	0			
E	32	6	40	1	0		
F	15	16	9	28	11	0	
G	19	25	32	9	5	10	0

Select one:

- a. A, D, C and B, E, F, G
- b. B, C, G and A, D, E, F
- c. A, B, C, E and D, F, G
- d. B, C, F, G and A, D, E
- e. A, B, C, F and D, E, G

Next page

Online Exams

Sri Lanka Institute of Information Technology

Question 9

Not yet answered

Marked out of 1.0

Flag question

Which of the following can contain in the output obtained by executing K-means algorithm?

Select one or more:

- a. Centroids
- b. Predicated category label
- c. Groups of observations
- d. a trend line
- e. Predicted numerical value

Next page

1

9

17

25

Finish a

Time le

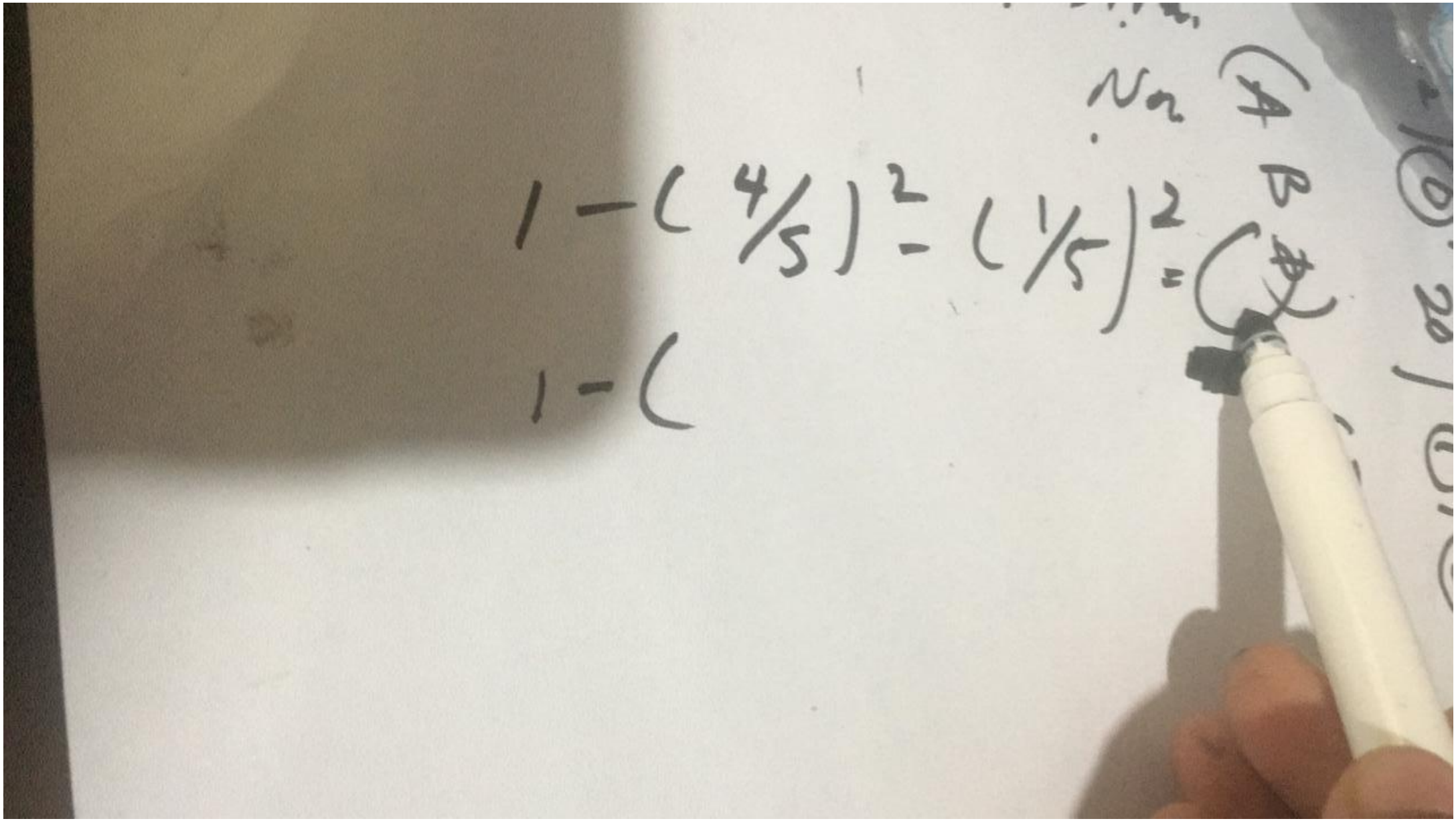
Sri Lanka Institute of Information Technology

Which of the following are correct?

Select one or more:

- a. Gini index is not suitable for the measurement of purity.
- b. In greedy method all attributes are lined up and different split points are tried and tested using a cost function
- c. In classification a partition is pure if all the tuples in it belong to the same class
- d. An attribute selection measure is a heuristic for selecting the splitting criterion that "best" separates a given data partition
- e. Attribute selection measures are also known as splitting rules

Next page


$1-\left(\frac{4}{5}\right)^2=\left(\frac{1}{5}\right)^2=$
$1-($
(A B #)

Sri Lanka Institute of Information Technology

2
swered
t of 1.0
estion

Consider the following distance matrix. What would be the second cluster resulted after applying complete-link agglomerative clustering?

	A	B	C	D
A	0			
B	45	0		
C	13	23	0	
D	20	21	9	0

Select one:

- a. B, C, D
- b. B, C, A
- c. A, C, D
- d. A, D, B
- e. None of the above

Next page

Sri Lanka Institute of Information Technology

Which of the following is not an application of clustering?

Select one or more:

- [ ] a. Identifying different user groups in Facebook
- [ ] b. Identifying most suitable airline to travel for a customer with a given profile
- [ ] c. Detecting spam emails
- [ ] d. Predicting price of houses
- [ ] e. Market segmentation

Sri Lanka Institute of Information Technology

Question 23

Not yet answered

Marked out of 1.0

Flag question

Consider the following output obtained by using multiple regression technique in python in order to predict the Employed people in a country as a percentage based on the Gross National product (GNP). Assuming alpha is 0.05 what would be the employed people as a percentage in a country with GNP is 258.

	coef	std err	t	P>|t|	[95.0% Conf. Int.]	
const	51.8436	0.681	76.087	0.000	50.382	53.305
GNP	0.0348	0.002	20.374	0.000	0.031	0.038

Select one:

- a. 60%
- b. 76 %
- c. 51 %
- d. 80 %
- e. 0.03 %

Next page

Quiz nav

1 2 3

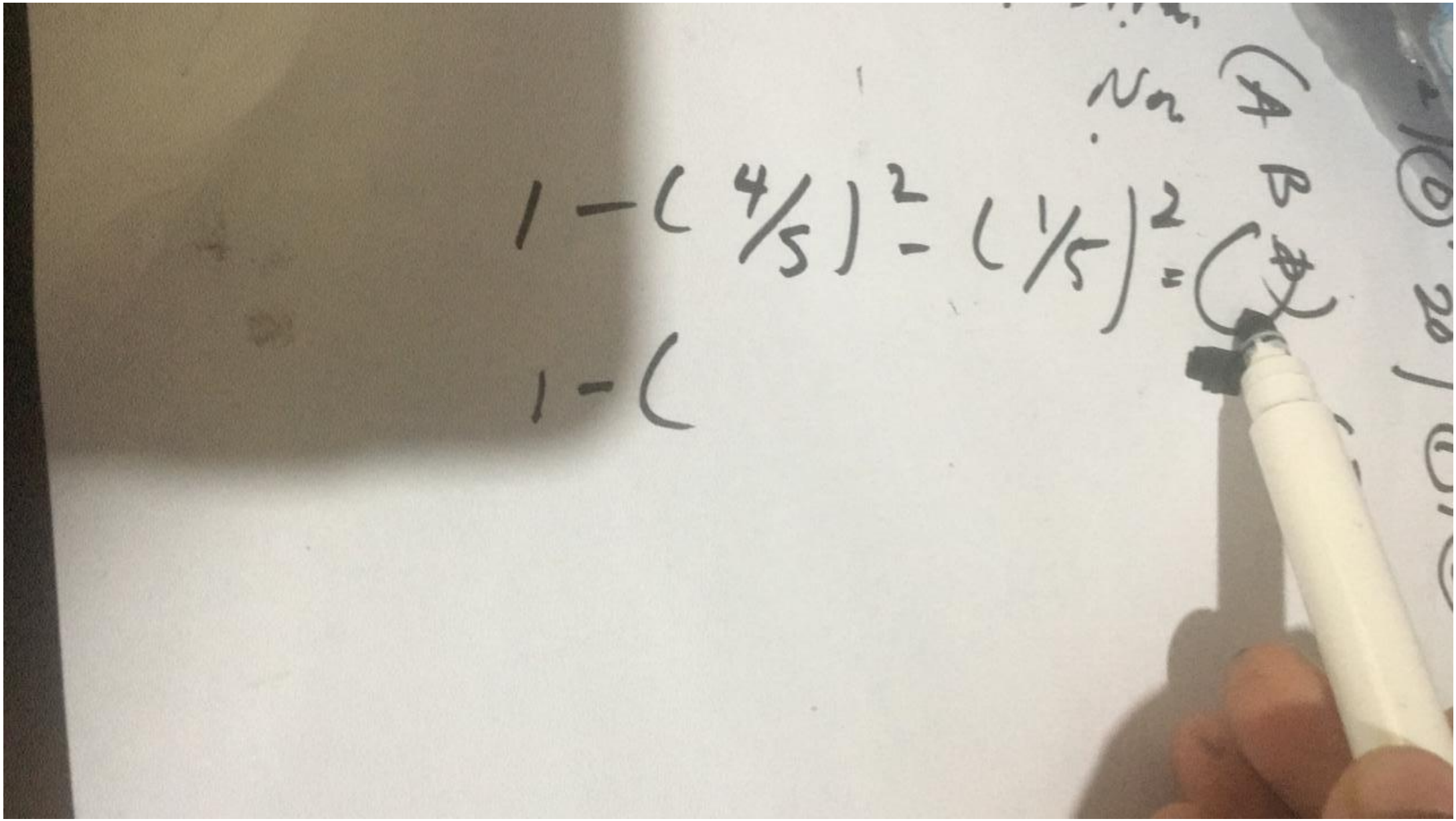
9 10 11

17 18 19

25

Finish attempt ...

Time left 0:05:03


$1-\left(\frac{4}{5}\right)^2=\left(\frac{1}{5}\right)^2=$
$1-($
(A B #)

The relationship between price and sales of an item is analyzed using simple linear regression with the following data. What would be the value for β0 in the resulted regression equation?

Price (k)	Sales (k)
3	20
6	9
11	0

Select one:

- a. 9.6
- b. 25
- c. 1.06
- d. 2.5
- e. -2.4

Next page

≡ Qu

1
9
17
25

Finish a

Time le

Online Exams

Sri Lanka Institute of Information Technology

Question 8

Not yet answered

Marked out of 1.0

Flag question

Select the most suitable answers to fill the blanks in the correct order

The ..... clustering, follows a ..... approach,which starts with all the objects in the same cluster whereas the ..... clustering, follows a ..... approach which starts with each object forming a separate group

Select one:

a.
divisive, agglomerative, top-down, bottom-up

b.
agglomerative, top-down , divisive , bottom-up

c.
divisive , top-down , agglomerative , bottom-up

d.
top-down, divisive , bottom-up, agglomerative

e. None of the above

Consider the following data set

TID	Items Bought
T001	{Milk, Orange, Bread, Eggs, Youghurt, Butter }
T002	{Noodles, Orange, Bread, Milk, Eggs, Youghurt}
T003	{Milk, Sugar, Bread, Egg}
T004	{Chocolate, Sugar , Tea bags, Butter, Youghurt, Biscuits}
T005	{Biscuits, Orange, Bread, Ice cream, Eggs }

What is the confidence for {Milk} -> Bread?

Select one:

- a. 0.5
- b. 0.25
- c. 1.0
- d. 0.2
- e. 0.75

Consider the following data set

TID	Items Bought
T001	{Milk, Orange, Bread, Eggs, Youghurt, Butter }
T002	{Noodles, Orange, Bread, Milk, Eggs, Youghurt}
T003	{Milk, Sugar, Bread, Egg}
T004	{Chocolate, Sugar , Tea bags, Butter, Youghurt, Biscuits}
T005	{Biscuits, Orange, Bread, Ice cream, Eggs }

What is the confidence for {Milk} -> Bread?

Select one:

- a. 0.5
- b. 0.25
- c. 1.0
- d. 0.2
- e. 0.75

Sri Lanka Institute of Information Technology

Question 13

Not yet answered

Marked out of 1.0

Flag question

Consider the following distance matrix. What would be the second cluster resulted after applying single-link agglomerative clustering?

	A	B	C	D
A	0			
B	9	0		
C	1	0	0	
D	16	4	8	0

Select one:

- a. B, C, D
- b. B, C, A
- c. A, C, D
- d. A, D, B
- e. None of the above

Next page

≡ Qui

1 2

9 10

17 18

25

Finish atte

Time left 0

Sri Lanka Institute of Information Technology

Question 13

Not yet answered

Marked out of 1.0

Flag question

Consider the following distance matrix. What would be the second cluster resulted after applying single-link agglomerative clustering?

	A	B	C	D
A	0			
B	9	0		
C	1	0	0	
D	16	4	8	0

Select one:

- a. B, C, D
- b. B, C, A
- c. A, C, D
- d. A, D, B
- e. None of the above

Next page

≡ Qui

1 2

9 10

17 18

25

Finish atte

Time left 0

Online Exams

Sri Lanka Institute of Information Technology

An emergency room in a hospital measures 17 variables to measure the medical condition of newly admitted patients to decide whether the patient should put in an intensive- care unit. Due to the high cost of ICU, those patients who may survive more a month are given higher priority. However, the problem is to predict high risk patients and discriminate them from low- risk patients.

Select the most suitable analytical method for the above scenario.

Select one:

- a. association rule mining
- b. clustering
- c. linear regression analysis
- d. classification
- e. time series analysis

Next page

ered
f 1.0
ion

Select the main characteristic of the Supervised learning

Select one:

- a. computationally complex
- b. Analysts are not aware about different categories of data
- c. Algorithms are used against data which is not labelled
- d. Algorithms are trained using labeled data.
- e. None of the above

ered
f 1.0
ion

Select the main characteristic of the Supervised learning

Select one:

- a. computationally complex
- b. Analysts are not aware about different categories of data
- c. Algorithms are used against data which is not labelled
- d. Algorithms are trained using labeled data.
- e. None of the above

Sri Lanka Institute of Information Technology

A credit card company typically receives hundreds of thousands of applications for new cards with several different attributes of applicant details. The company needs to identify the applicants who can be placed into good credit, bad credit or fall into a gray area.

Select the most suitable analytical technique for the above scenario.

Select one:

- a. K-Means analysis
- b. decision tree analysis
- c. linear regression analysis
- d. Time series analysis
- e. Apriori algorithm

Next page

A credit card company typically receives hundreds of thousands of applications for new cards with several different attributes of applicant details. The company needs to identify the applicants who can be placed into good credit, bad credit or fall into a gray area.

Select the most suitable analytical technique for the above scenario.

Select one:

- a. K-Means analysis
- b. decision tree analysis
- c. linear regression analysis
- d. Time series analysis
- e. Apriori algorithm

Next page

Which of the following is not true about classification?

Select one or more:

- [ ] a. The only input for the classifier are tuples
- [ ] b. Second step of classification involves testing the classification model and predicting
- [ ] c. Classification have two main steps namely learning and predicting
- [ ] d. Same set of rows are usually used in first and second phase of classification
- [ ] e. The class label for of the data set is a continuous variable

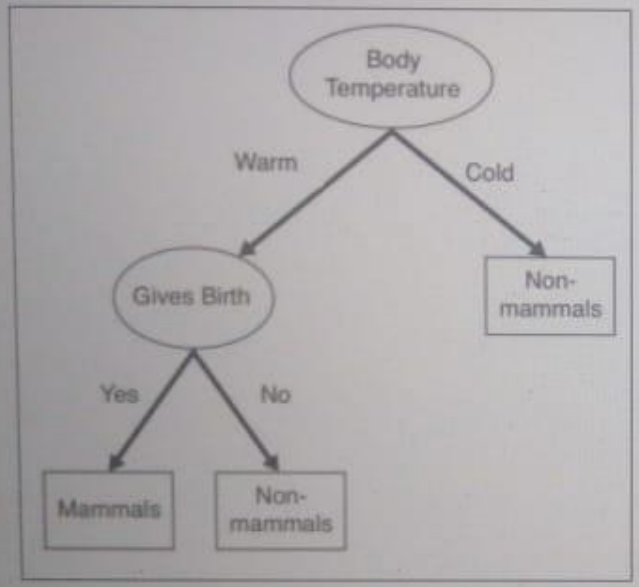
Which of the following is not true about classification?

Select one or more:

- [ ] a. The only input for the classifier are tuples
- [ ] b. Second step of classification involves testing the classification model and predicting
- [ ] c. Classification have two main steps namely learning and predicting
- [ ] d. Same set of rows are usually used in first and second phase of classification
- [x] e. The class label for of the data set is a continuous variable

Question 17
Not yet answered
Marked out of 1.0
Flag question

A decision tree has constructed to determine the class of animals according to the body temperature and birth way as shown in following figure.



Select the correct answer which contains the class of the unlabeled data shown in following table for the animals in order; <gibbon, flamingo, lion and cockroach>.

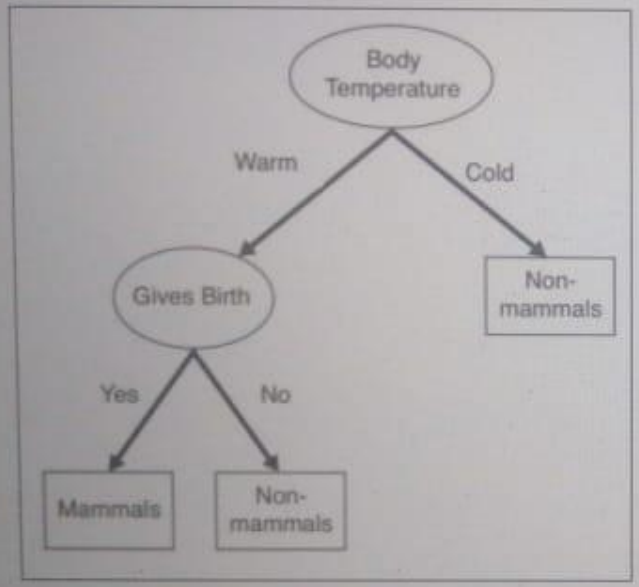
Name	Body Temperature	Gives Birth	Class
gibbon	warm	yes	?
flamingo	warm	no	?
lion	warm	yes	?

Select one:

a. Non - Mammals, Mammals, Mammals, Non-

Question 17
Not yet answered
Marked out of 1.0
Flag question

A decision tree has constructed to determine the class of animals according to the body temperature and birth way as shown in following figure.



Select the correct answer which contains the class of the unlabeled data shown in following table for the animals in order; <gibbon, flamingo, lion and cockroach>.

Name	Body Temperature	Gives Birth	Class
gibbon	warm	yes	?
flamingo	warm	no	?
lion	warm	yes	?

Select one:

a. Non - Mammals, Mammals, Mammals, Non-

uestion 20
lot yet answered
larked out of 1.0
Flag question

In a regression and correlation analysis if R squared = 1, then

Select one:

a. SSR = SST

b. SST=1

c. SSR = SSE

d. SSE = 1

e. SSE = SST

uestion 20
lot yet answered
larked out of 1.0
Flag question

In a regression and correlation analysis if R squared = 1, then

Select one:

a. SSR = SST

b. SST=1

c. SSR = SSE

d. SSE = 1

e. SSE = SST

Which of the following is not a technique for estimation of regression coefficients?

Select one or more:

- a. Bayesian regression
- b. Kernal regression
- c. Ordinary least square
- d. Minimal likelihood estimation
- e. Maximum least square

Next page

Sri Lanka Institute of Information Technology

Which of the following is not true with relevance to regression?

Select one or more:

- a. Regression relies on statistical techniques
- b. Regression could be applied to variables having causal relationships
- c. Regression could be used to predict categorical variables
- d. Regression is the process of finding a mathematical equation that best fits the noisy data
- e. All of the above

Sri Lanka Institute of Information Technology

Which of the following is not true with relevance to regression?

Select one or more:

- [ ] a. Regression relies on statistical techniques
- [ ] b. Regression could be applied to variables having causal relationships
- [ ] c. Regression could be used to predict categorical variables
- [ ] d. Regression is the process of finding a mathematical equation that best fits the noisy data
- [ ] e. All of the above

Sri Lanka Institute of Information Technology

on 10
t answered
d out of 1.0
g question

Consider the following data set. Assuming that A, B and C are initial centroids for K-means algorithm, what would be the resulted clusters after the first epoch?

	A	B	C	D	E	F	G
A	0						
B	9	0					
C	1	0	0				
D	16	4	8	0			
E	3	12	0	20	0		
F	11	26	13	12	1	0	
G	19	25	23	2	8	10	0

Select one:

- a. {A, D} {B, E} and { C, F, G}
- b. {A, E} {B, D,F} {C, G}
- c. {A, F, G}, {B, D} and {C, E}
- d. {A, F, G} { B, D} and {C, E}
- e. {A, D, E} { B, F} {C, G}

Next page

≡ Quiz na

1 2 3
9 10 1
17 18 19
25

Finish attempt ...

Time left 0:43:20

Sri Lanka Institute of Information Technology

on 10
t answered
d out of 1.0
g question

Consider the following data set. Assuming that A, B and C are initial centroids for K-means algorithm, what would be the resulted clusters after the first epoch?

	A	B	C	D	E	F	G
A	0						
B	9	0					
C	1	0	0				
D	16	4	8	0			
E	3	12	0	20	0		
F	11	26	13	12	1	0	
G	19	25	23	2	8	10	0

Select one:

- a. {A, D} {B, E} and { C, F, G}
- b. {A, E} {B, D,F} {C, G}
- c. {A, F, G}, {B, D} and {C, E}
- d. {A, F, G} { B, D} and {C, E}
- e. {A, D, E} { B, F} {C, G}

Next page

Quiz na

1 2
9 10
17 18
25

Finish attempt ...

Time left 0:43:20

Consider the following distance matrix. What would be the second cluster resulted after applying complete-link agglomerative clustering?

	A	B	C	D
A	0			
B	10	0		
C	13	6	0	
D	24	21	9	0

Select one:

- a. B, C, D
- b. B, C, A
- c. A, C, D
- d. A, D, B
- e. None of the above

Next page

Consider the following distance matrix. What would be the second cluster resulted after applying complete-link agglomerative clustering?

	A	B	C	D
A	0			
B	10	0		
C	13	6	0	
D	24	21	9	0

Select one:

- a. B, C, D
- b. B, C, A
- c. A, C, D
- d. A, D, B
- e. None of the above

Next page

2
swered
ut of 1.0
uestion

Which of the following is not true with relevance to regression?

Select one or more:

- a. Regression relies on statistical techniques
- b. Regression could be applied to variables having causal relationships
- c. Regression could be used to predict categorical variables
- d. Regression is the process of finding a mathematical equation that best fits the noisy data
- e. All of the above

Next page

2
swered
ut of 1.0
uestion

Which of the following is not true with relevance to regression?

Select one or more:

- a. Regression relies on statistical techniques
- b. Regression could be applied to variables having causal relationships
- c. Regression could be used to predict categorical variables
- d. Regression is the process of finding a mathematical equation that best fits the noisy data
- e. All of the above

Next page

mammals

Select the correct answer which contains the class of the unlabeled data shown in following table for the animals in order: <gibbon, flamingo, lion and cockroach>.

Name	Body Temperature	Gives Birth	Class
gibbon	warm	yes	?
flamingo	warm	no	?
lion	warm	yes	?
cockroach	cold	no	?

Select one:

a. Non - Mammals, Mammals, Mammals, Non-mammals

b. Mammals, Non-mammals, Non - Mammals, Non-mammals

c. Mammals, Non-mammals, Mammals, Non-mammals

d. Non - Mammals, Non- Mammals, Mammals, Non-mammals

e. Mammals, Mammals, Mammals, Non-mammals

Next page

mammals

Select the correct answer which contains the class of the unlabeled data shown in following table for the animals in order: <gibbon, flamingo, lion and cockroach>.

Name	Body Temperature	Gives Birth	Class
gibbon	warm	yes	?
flamingo	warm	no	?
lion	warm	yes	?
cockroach	cold	no	?

Select one:

a. Non - Mammals, Mammals, Mammals, Non-mammals

b. Mammals, Non-mammals, Non - Mammals, Non-mammals

c. Mammals, Non-mammals, Mammals, Non-mammals

d. Non - Mammals, Non- Mammals, Mammals, Non-mammals

e. Mammals, Mammals, Mammals, Non-mammals

Next page

Consider the following distance matrix. What would be the second cluster resulted after applying complete-link agglomerative clustering?

	A	B	C	D
A	0			
B	45	0		
C	13	23	0	
D	20	21	9	0

Select one:

- a. B, C, D
- b. B, C, A
- c. A, C, D
- d. A, D, B
- e. None of the above

Next page

uestion 18

ot yet answered

arked out of 1.0

Flag question

Consider the following count matrix developed for the prediction of students' grade using specialty.

Specialty	Grade	
	Yes	No
IT	4	1
Medicine	1	3
Engineering	0	2
Sociology	1	1

Which of the following is Gini(Specialty)?

Select one:

- a. 0.0
- b. 0.315
- c. 0.500
- d. 0.037
- e. 0.375

uestion 18

ot yet answered

arked out of 1.0

Flag question

Consider the following count matrix developed for the prediction of students' grade using specialty.

Specialty	Grade	
	Yes	No
IT	4	1
Medicine	1	3
Engineering	0	2
Sociology	1	1

Which of the following is Gini(Specialty)?

Select one:

- a. 0.0
- b. 0.315
- c. 0.500
- d. 0.037
- e. 0.375

Online Exams

Sri Lanka Institute of Information Technology

swered

t of 1.0

estion

Which of the following are imitations of the K-means algorithm

Select one or more:

- a. Not Sensitive to outliers
- b. Issues in clustering data of various sizes and density
- c. cannot find clusters with arbitrary shapes
- d. K- should be selected
- e. Initial centroids are selected systematically

Online Exams

Sri Lanka Institute of Information Technology

swered

t of 1.0

estion

Which of the following are imitations of the K-means algorithm

Select one or more:

- a. Not Sensitive to outliers
- b. Issues in clustering data of various sizes and density
- c. cannot find clusters with arbitrary shapes
- d. K- should be selected
- e. Initial centroids are selected systematically

Sri Lanka Institute of Information Technology

Which of the following is true regarding correlation between two variables x and y?

Select one or more:

- a. If there is a very strong correlation between x and y correlation coefficient must be zero
- b. Correlation coefficient could be used to determine a specific value of the y-variable given a specific value of the x-variable
- c. If there is a very strong correlation between x and y correlation coefficient must be any value larger than 1
- d. The strength of the relationship between the x and y could be determined by correlation
- e. If x and y have a strong negative correlation coefficient is a much smaller than 0

Next page

1
9
17
25

Finish

Time l

Sri Lanka Institute of Information Technology

Which of the following is true regarding correlation between two variables x and y?

Select one or more:

- a. If there is a very strong correlation between x and y correlation coefficient must be zero
- b. Correlation coefficient could be used to determine a specific value of the y-variable given a specific value of the x-variable
- c. If there is a very strong correlation between x and y correlation coefficient must be any value larger than 1
- d. The strength of the relationship between the x and y could be determined by correlation
- e. If x and y have a strong negative correlation coefficient is a much smaller than 0

Next page

1
9
17
25
Finish

Question 25
Not yet answered
Marked out of 1.0
Flag question

Consider the following data frame named data.

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	California	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	California	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12
6	134615.46	147198.87	127716.82	California	156122.51
7	130298.13	145530.06	323876.68	New York	155752.60
8	120542.52	148718.95	311613.29	New York	152211.77
9	123334.88	108679.17	304981.62	California	149759.96

which of the following is true?

Select one or more:

- a. data.iloc[, 2: 4] returns the content of Administration and Marketing Spend columns
- b. data['Administration'].describe() generates descriptive statistics for the data in column Administration.
- c. data[:, 4] command provide the same output as data['Profit']
- d. data.shape[1] returns 4
- e. data.head() returns all column values for the first five rows in the data frame

Quiz navigation

1 2 3 4 5
9 10 11 12 13
17 18 19 20 21
25

Finish attempt ...

Time left 0:01:50

Question 25
Not yet answered
Marked out of 1.0
Flag question

Consider the following data frame named data.

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	California	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	California	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12
6	134615.46	147198.87	127716.82	California	156122.51
7	130298.13	145530.06	323876.68	New York	155752.60
8	120542.52	148718.95	311613.29	New York	152211.77
9	123334.88	108679.17	304981.62	California	149759.96

which of the following is true?

Select one or more:

- a. data.iloc[, 2: 4] returns the content of Administration and Marketing Spend columns
- b. data['Administration'].describe() generates descriptive statistics for the data in column Administration.
- c. data[:, 4] command provide the same output as data['Profit']
- d. data.shape[1] returns 4
- e. data.head() returns all column values for the first five rows in the data frame

Quiz navigation

1 2 3 4 5
9 10 11 12 13
17 18 19 20 21
25

Finish attempt ...

Time left 0:01:50

Online Exams

Sri Lanka Institute of Information Technology

Consider the following distance matrix. Assuming A and E are are initial centroids for K-means algorithm, what would be the resulted clusters after the first epoch?

	A	B	C	D	E	F	G
A	0						
B	9	0					
C	1	0	0				
D	16	4	8	0			
E	3	12	0	20	0		
F	11	26	13	12	1	0	
G	19	25	23	2	8	10	0

Select one:

- a. A, C, D, G and B, E, F
- b. A, D and B, C, E, F, G
- c. A, B, G and C, D, E, F
- d. A, B, C and D, E, F, G
- e. A, B, D and C, E, F, G

Next page

Quiz navigation

1 2 3 4 5
9 10 11 12 13
17 18 19 20 21
25

Finish attempt ...

Time left 0:36:47

Online Exams

Sri Lanka Institute of Information Technology

Consider the following distance matrix. Assuming A and E are are initial centroids for K-means algorithm, what would be the resulted clusters after the first epoch?

	A	B	C	D	E	F	G
A	0						
B	9	0					
C	1	0	0				
D	16	4	8	0			
E	3	12	0	20	0		
F	11	26	13	12	1	0	
G	19	25	23	2	8	10	0

Select one:

- a. A, C, D, G and B, E, F
- b. A, D and B, C, E, F, G
- c. A, B, G and C, D, E, F
- d. A, B, C and D, E, F, G
- e. A, B, D and C, E, F, G

Next page

Quiz navigation

1 2 3 4 5
9 10 11 12 13
17 18 19 20 21
25

Finish attempt ...

Time left 0:36:47

Online Exams

Sri Lanka Institute of Information Technology

Consider the following distance matrix. Assuming A and E are are initial centroids for K-means algorithm, what would be the resulted clusters after the first epoch?

	A	B	C	D	E	F	G
A	0						
B	9	0					
C	1	0	0				
D	16	4	8	0			
E	3	12	0	20	0		
F	11	26	13	12	1	0	
G	19	25	23	2	8	10	0

Select one:

- a. A, C, D, G and B, E, F
- b. A, D and B, C, E, F, G
- c. A, B, G and C, D, E, F
- d. A, B, C and D, E, F, G
- e. A, B, D and C, E, F, G

Next page

Quiz navigation

1 2 3 4 5
9 10 11 12 13
17 18 19 20 21
25

Finish attempt ...

Time left 0:36:47

ered
of 1.0
tion

Which of the following is **not** a classification technique?

Select one:

- a. K-nearest neighbor
- b. Random forest
- c. Support vector machine
- d. K-means analysis
- e. Neutral networks

ered
of 1.0
tion

Which of the following is **not** a classification technique?

Select one:

- a. K-nearest neighbor
- b. Random forest
- c. Support vector machine
- d. K-means analysis
- e. Neutral networks

4
swered
ut of 1.0
uestion

Which of the following are strengths of decision tree methods?

Select one or more:

- a. Decision tree classifiers perform series of condition checking with one attribute at a time
- b. Decision trees provides a clear indication of order of variables to check to come to a conclusion
- c. Decision trees are able to generate set of rules which are easy to understand.
- d. Calculations in decision trees can become complex when there are many class labels
- e. Decision trees are able to handle both continuous and categorical variables.

Nex

4
swered
ut of 1.0
uestion

Which of the following are strengths of decision tree methods?

Select one or more:

- a. Decision tree classifiers perform series of condition checking with one attribute at a time
- b. Decision trees provides a clear indication of order of variables to check to come to a conclusion
- c. Decision trees are able to generate set of rules which are easy to understand.
- d. Calculations in decision trees can become complex when there are many class labels
- e. Decision trees are able to handle both continuous and categorical variables.

Nex