

Midterm

Due Oct 28 at 11:59pm **Points** 48

Questions 16

Available Oct 26 at 12am - Oct 29 at 1am 3
days

Time Limit 100 Minutes

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	40 minutes	45 out of 48

Score for this quiz: **45** out of 48

Submitted Oct 27 at 6:34pm

This attempt took 40 minutes.

Question 1

3 / 3 pts

Select the correct sentence or sentences that define cosine and correlation measures.



The range of values that are possible for the cosine measure is [0, 1]



If two objects have a cosine similarity of 1, they are identical

Correct!



If X and Y have a mean of 0, $\text{corr}(X,Y) = \cos(X,Y)$



A, B



A, B and C

Question 2

3 / 3 pts

Select all possible use-cases of clustering



Credit card fraud detection



Market segmentation

orrect!

Summarize news

A, B, and C

A, B

Question 3

3 / 3 pts

Select all predictive techniques.

orrect!

Anomaly Detection

orrect!

Regression

orrect!

Classification

Clustering

Dimensionality Reduction

Question 4

3 / 3 pts

Select all unsupervised learning applications

orrect!



Marketing and sales promotion using Association Rule

Association Rule Mining is an unsupervised learning method.

orrect!



Document clustering

Clustering is an unsupervised learning method.



Face recognition



Fraud detection in credit card transactions

Question 5

3 / 3 pts

Select all correct sentences that define correlation.

orrect!



Correlation -1 means that the vectors have a perfect negative linear relationship.

Correct. If the correlation is -1, the vectors follow the perfect negative linear correlation.



Correlation 1 means that the vectors have a perfect negative linear relationship.

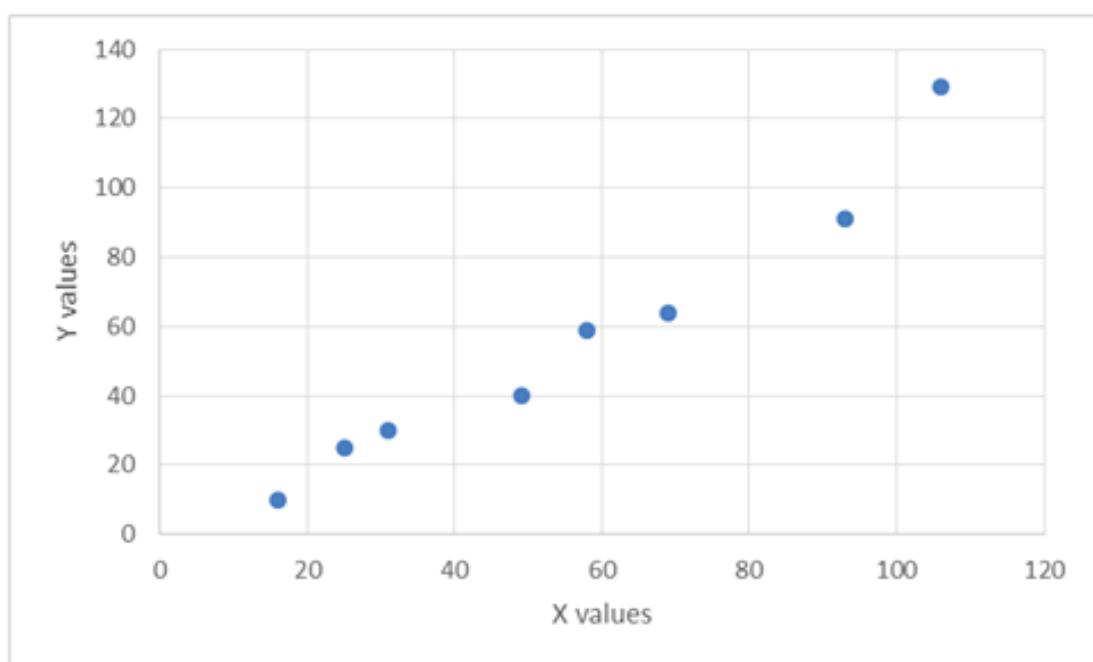


Correlation 0 means that the vectors have a perfect positive linear relationship.

Question 6

3 / 3 pts

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Compute cosine similarity between X and Y

orrect!

0.991

ect Answers

Between 0.95 and 1

Question 7

3 / 3 pts

Select all techniques used to measure node impurity.

orrect!

Gini Index

Gini Index is a measure of node impurity

average

orrect!

Entropy

Entropy is a measure of node impurity

split

Question 8

3 / 3 pts

Select all sentences that correctly define errors.

orrect!



Prediction errors consist of bias, variance, and irreducible error.

orrect!



Bias error results in high training error.



Variance error results in high training error.

orrect!



Irreducible error cannot be optimized.

Question 9

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about the decision tree model?

orrect!



The ID is the best attribute to predict Earlier Adopter



The Carrier is the best attribute to predict Earlier Adopter



The Device OS is the best attribute to predict Earlier Adopter



The Device Color is the best attribute to predict Earlier Adopter

Question 10

3 / 3 pts

Select Methods for good model evaluation.

orrect!



Maximize values of both precision and recall

we need to maximize both precision and recall

Maximize accuracy

Maximize precision

Minimize recall

Question 11

3 / 3 pts

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute F1 for 'pear' class

orrect!

0

ect Answers

0 (with margin: 0)

Question 12

3 / 3 pts

Select all sentences that are correct about cross-validation.

orrect!

Leave-one-out is a type of cross-validation.



Cross-validation guarantees reduction in test error.



K-fold cross-validation requires K repetitions of the same train set on K different test sets..

orrect!



In cross validation the same training record can be selected multiple times.

Question 13

3 / 3 pts

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute the accuracy as a fraction

orrect!

0.5

ect Answers

0.5 (with margin: 0)

Question 14

3 / 3 pts

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the Gini index? What is the Gini index of the split attribute?

orrect!



Attribute 1 is the best split and Gini index is 0.3444



Attribute 1 is the best split and Gini index is 0.4889



Attribute 2 is the best split and Gini index is 0.3444



Attribute 2 is the best split and Gini index is 0.4889

Question 15

3 / 3 pts

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the best split among attribute 1, attribute 2, and attribute 3, according to the information gain.

orrect!

Attribute 1

Attribute 2

Attribute 3

Question 16

0 / 3 pts

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

Answered

0.492

Correct Answers

Between 0.49 and 0.4915

Quiz Score: **45** out of 48

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	13 minutes	47 out of 48

Score for this quiz: **47** out of 48

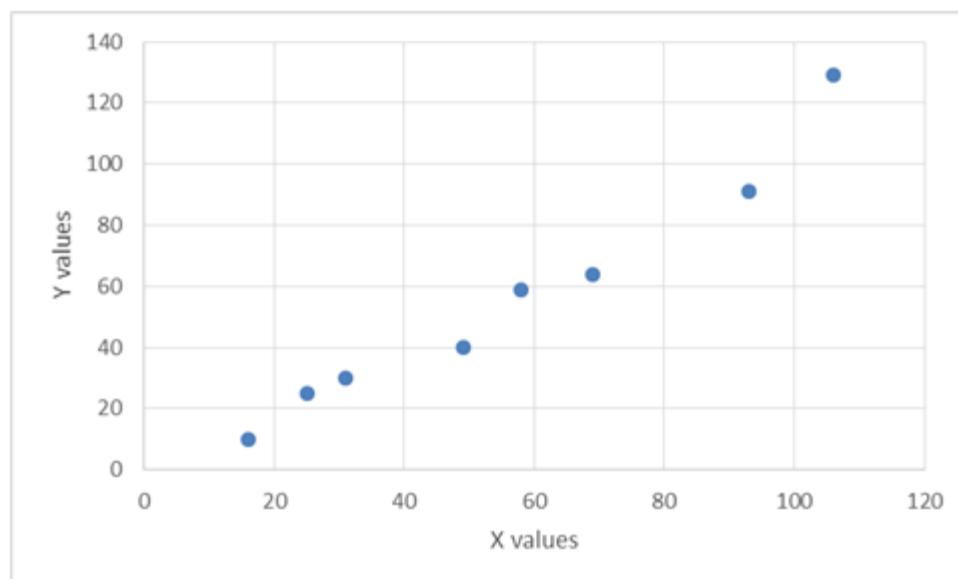
Submitted Oct 27 at 3:23pm

This attempt took 13 minutes.

Question 1

3 / 3 pts

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Compute cosine similarity between X and Y

Correct!

0.973

Incorrect Answers

Between 0.95 and 1

Question 2

3 / 3 pts

Select all correct sentences which make up the clustering.

- K-Nearest Neighbor is a typical clustering technique
- Association rule discovery is a typical technique of clustering
- Clustering techniques do not require any parameters
- A, B
- A, B, and C

Correct!

Question 3**3 / 3 pts**

Select all possible use-cases of clustering

- Credit card fraud detection
- Market segmentation
- Summarize news

Correct!

- A, B, and C
- A, B

Question 4**2 / 3 pts**

Select all predictive techniques.

Incorrect Answer

- Anomaly Detection

Correct!

- Regression

Correct!

- Classification

- Clustering

- Dimensionality Reduction

Question 5**3 / 3 pts**

Select the correct sentence or sentences that define cosine and correlation

measures.

The range of values that are possible for the cosine measure is [0, 1]

If two objects have a cosine similarity of 1, they are identical

Correct!

If X and Y have a mean of 0, $\text{corr}(X,Y) = \cos(X,Y)$

A, B

A, B and C

Question 6

3 / 3 pts

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

Tid

Correct!

Refund

Correct!

Marital Status

Correct! Taxable Income**Correct!** Cheat**Question 7****3 / 3 pts**

Select the statements below that are computed as part of the performance metric.

Correct! Precision can be computed using True Positive and False Positive. Recall can be computed using True Positive and False Positive.**Correct!** F1 can be computed using True Positive and False Positive, and False Negative. Accuracy can be computed using True Positive, False Positive, and False Negative.**Question 8****3 / 3 pts**

Select all sentences that correctly define errors.

Correct! Errors committed on training records are training errors.**Correct!** Generalization error are expected errors of the model on previously unseen records. Low training errors guarantee low generalization errors.

Generalization errors can be reduced by increasing the model complexity.

Question 9**3 / 3 pts**

True or False? The tree induction algorithm C4.5 is suitable for large dataset.

 True False

C4.5 is suitable for small data set.

Correct!**Question 10****3 / 3 pts**

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

Correct!

0.491

Incorrect Answers

Between 0.49 and 0.4915

Question 11

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about the decision tree model?

Correct!

-
- The ID is the best attribute to predict Earlier Adopter
 - The Carrier is the best attribute to predict Earlier Adopter

 - The Device OS is the best attribute to predict Earlier Adopter

 - The Device Color is the best attribute to predict Earlier Adopter

Question 12

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about a decision tree model?

Correct!

- Gini index for each ID value is 0.

Correct!

- Gini for Android in Device OS is the same with Gini for IOS in Device OS.

Correct!

- Gini index for the carrier is 0.1625.

Correct!

- Gini index for the overall collection of the example is 0.5.

Question 13**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the Gini index? What is the Gini index of the split attribute?

Correct!

- Attribute 1 is the best split and Gini index is 0.3444
- Attribute 1 is the best split and Gini index is 0.4889
- Attribute 2 is the best split and Gini index is 0.3444
- Attribute 2 is the best split and Gini index is 0.4889

Question 14**3 / 3 pts**

Select all classification techniques.

Correct!

- Decision tree

Correct!

- Naïive Bayes

- Clustering

Correct!

- Neural Networks

Decision tree, Naïive Bayes and Neural Network are commonly used classification techniques.

Question 15**3 / 3 pts**

A good classification model has:

- Low training error
- Low estimated generalization error
- A and B are correct

We need both low training error and low generalization error.

Correct!

- None of the above

Question 16**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the correct information gain pair of Attribute 1 and Attribute 2.

Correct!

- Attribute 1- 0.229, Attribute 2 – 0.007
- Attribute 1- 0.991, Attribute 2 – 0.007
- Attribute 1- 0.007, Attribute 2 – 0.229
- Attribute 1- 0.229, Attribute 2 – 0.991

Quiz Score: **47** out of 48

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

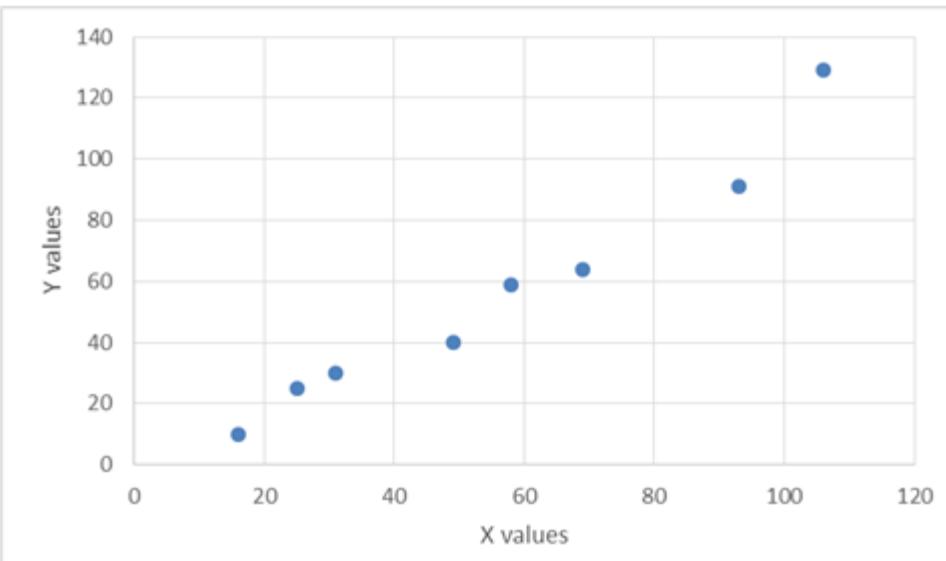
	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	10 minutes	45 out of 48

Score for this quiz: **45** out of 48

Submitted Oct 27 at 3:15pm

This attempt took 10 minutes.

Question 1			3 / 3 pts
	X	Y	
A	16	10	
B	25	25	
C	31	30	
D	49	40	
E	58	59	
F	69	64	
G	93	91	
H	106	129	



Compute cosine similarity between X and Y

Correct!

0.965

Incorrect Answers

Between 0.95 and 1

Question 2

3 / 3 pts

Select all correct sentences that define correlation.

Correct!



Correlation -1 means that the vectors have a perfect negative linear relationship.



Correct. If the correlation is -1, the vectors follow the perfect negative linear correlation.



Correlation 1 means that the vectors have a perfect negative linear relationship.



Correlation 0 means that the vectors have a perfect positive linear relationship.

Question 3**3 / 3 pts**

Select all unsupervised learning applications

Correct!

- Marketing and sales promotion using Association Rule

Association Rule Mining is an unsupervised learning method.

Correct!

- Document clustering

Clustering is an unsupervised learning method.

- Face recognition

- Fraud detection in credit card transactions

Question 4**3 / 3 pts**

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

Tid

Correct!

Refund

Correct!

Marital Status

Correct!

Taxable Income

Correct!

Cheat

Question 5

3 / 3 pts

Select all correct sentences which make up the clustering.

K-Nearest Neighbor is a typical clustering technique

Correct!

Association rule discovery is a typical technique of clustering

Clustering techniques do not require any parameters

A, B A, B, and C**Question 6****3 / 3 pts**

Select the correct sentence or sentences that define cosine and correlation measures.

 The range of values that are possible for the cosine measure is [0, 1] If two objects have a cosine similarity of 1, they are identical If X and Y have a mean of 0, $\text{corr}(X,Y) = \cos(X,Y)$ A, B A, B and C**Correct!****Question 7****3 / 3 pts**

Select all sentences that correctly define errors.

Correct! Errors committed on training records are training errors.**Correct!**
Generalization error are expected errors of the model on previously unseen records. Low training errors guarantee low generalization errors.

Generalization errors can be reduced by increasing the model complexity.

Question 8**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

When is the best split for Attribute 3 according to the information gain?

Correct!

- The split point equal to 2.0
- The split point equal to 3.5
- The split point equal to 5.5
- The split point equal to 6.5
- The split point equal to 7.5

Question 9**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the best split among attribute 1, attribute 2, and attribute 3, according to the information gain.

Correct! Attribute 1 Attribute 2 Attribute 3**Question 10****3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the classification rate and what is the classification error of the split attribute?

Correct!

- Attribute 1 is the best split, 2/9 error rate
- Attribute 1 is the best split, 4/9 error rate
- Attribute 2 is the best split, 2/9 error rate
- Attribute 2 is the best split, 4/9 error rate

Question 11

3 / 3 pts

A good classification model has:

- Low training error
- Low estimated generalization error

Correct!

- A and B are correct

We need both low training error and low generalization error.

- None of the above

Question 12**3 / 3 pts**

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

Correct!

0.491

Incorrect Answers

Between 0.49 and 0.4915

Question 13**3 / 3 pts**

Select all statements that correctly address underfitting and overfitting.

Correct! Underfitting can be reduced by reducing bias error. Overfitting can be reduced by reducing bias error. Underfitting can be reduced by reducing variance error.**Correct!** Overfitting can be reduced by reducing variance error.**Question 14****3 / 3 pts**

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute the accuracy as a fraction

Correct!

0.5

Incorrect Answers

0.5 (with margin: 0)

Question 15**3 / 3 pts**

Select the statements below that are computed as part of the performance metric.

Correct!

Precision can be computed using True Positive and False Positive.

Recall can be computed using True Positive and False Positive.

Correct!

F1 can be computed using True Positive and False Positive, and False Negative.

Accuracy can be computed using True Positive, False Positive, and False Negative.

Question 16**0 / 3 pts**

Select all sentences that are correct about cross-validation.

Correct!

Leave-one-out is a type of cross-validation.

You Answered

Cross-validation guarantees reduction in test error.

K-fold cross-validation requires K repetitions of the same train set on K different test sets..

Correct Answer

In cross validation the same training record can be selected multiple times.

Quiz Score: 45 out of 48

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	40 minutes	44.25 out of 48

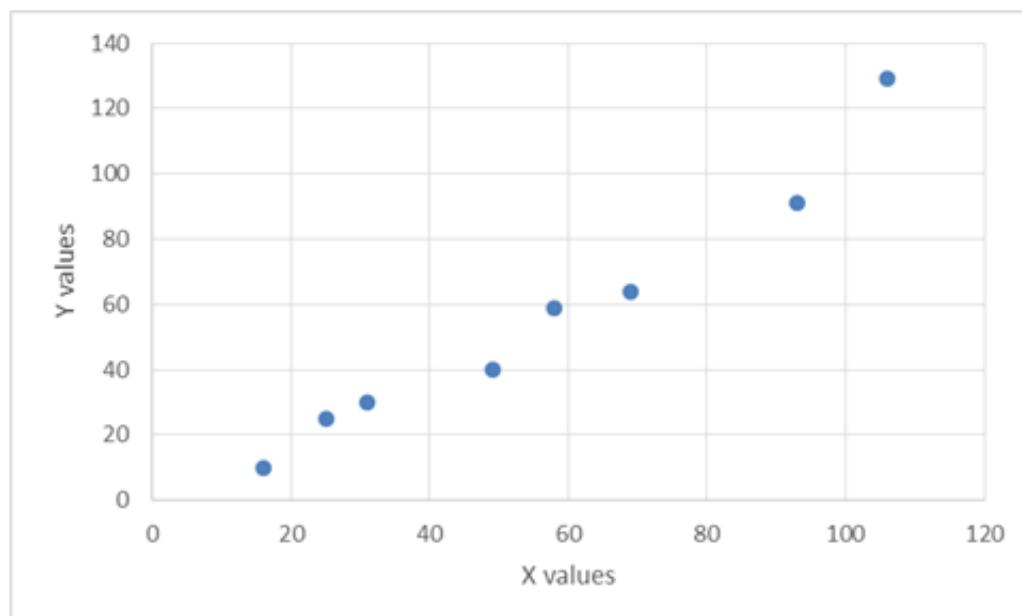
Score for this quiz: **44.25** out of 48

Submitted Oct 26 at 8:23pm

This attempt took 40 minutes.

Question 1	3 / 3 pts

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Select the closes answer for the correlation between X and Y

0.5 -0.5 -1 1 0**Correct!****Question 2****3 / 3 pts**

Select all correct sentences that define correlation.

Correct!

Correlation -1 means that the vectors have a perfect negative linear relationship.

Correct. If the correlation is -1, the vectors follow the perfect negative linear correlation.



Correlation 1 means that the vectors have a perfect negative linear relationship.



Correlation 0 means that the vectors have a perfect positive linear relationship.

Question 3**2.25 / 3 pts**

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

 Tid**Correct!** Refund**Correct!** Marital Status**Correct Answer** Taxable Income**Correct!** Cheat**Question 4****3 / 3 pts**

Select all unsupervised learning applications

Correct!

- Marketing and sales promotion using Association Rule

Association Rule Mining is an unsupervised learning method.

Correct!

- Document clustering

Clustering is an unsupervised learning method.

- Face recognition

- Fraud detection in credit card transactions

Question 5

3 / 3 pts

Select all correct sentences which make up the clustering.

- K-Nearest Neighbor is a typical clustering technique

- Association rule discovery is a typical technique of clustering

- Clustering techniques do not require any parameters

- A, B

- A, B, and C

Question 6

0 / 3 pts

Select the correct sentence or sentences that define cosine and correlation measures.



The range of values that are possible for the cosine measure is [0, 1]



If two objects have a cosine similarity of 1, they are identical



If X and Y have a mean of 0, $\text{corr}(X, Y) = \cos(X, Y)$



A, B



A, B and C

Correct Answer**You Answered****Question 7**

3 / 3 pts

Select all classification techniques.

Correct! Decision tree**Correct!** Naïive Bayes**Correct!** Clustering Neural Networks

Decision tree, Naïve Bayes and Neural Network are commonly used classification techniques.

Question 8**3 / 3 pts**

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about a decision tree model?

Correct!

- Gini index for each ID value is 0.

Correct!



Gini for Android in Device OS is the same with Gini for IOS in Device OS.

Correct!

- Gini index for the carrier is 0.1625.

Correct!

- Gini index for the overall collection of the example is 0.5.

Question 9

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about the decision tree model?

-
- The ID is the best attribute to predict Earlier Adopter
 - The Carrier is the best attribute to predict Earlier Adopter
 - The Device OS is the best attribute to predict Earlier Adopter
 - The Device Color is the best attribute to predict Earlier Adopter

Correct!

Question 10**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the classification rate and what is the classification error of the split attribute?

Correct!

- Attribute 1 is the best split, 2/9 error rate
- Attribute 1 is the best split, 4/9 error rate
- Attribute 2 is the best split, 2/9 error rate
- Attribute 2 is the best split, 4/9 error rate

Question 11**3 / 3 pts**

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

Correct!

0.491

Correct Answers

Between 0.49 and 0.4915

Question 12

3 / 3 pts

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute F1 for 'pear' class

Correct! 0**Correct Answers**

0 (with margin: 0)

Question 13

3 / 3 pts

Review the table below.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Select the correct precision and recall for 'apple' class.

-
- Precision – 0.5 and recall – 0.5

Correct!

- Precision – 1 and recall – 0.5
- Precision – 0.5 and recall – 1
- Precision – 1 and recall – 1

Question 14**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the correct information gain pair of Attribute 1 and Attribute 2.

Correct!

- Attribute 1- 0.229, Attribute 2 – 0.007

- Attribute 1- 0.991, Attribute 2 – 0.007
- Attribute 1- 0.007, Attribute 2 – 0.229
- Attribute 1- 0.229, Attribute 2 – 0.991

Question 15**3 / 3 pts**

Select all sentences that correctly define errors.

Correct!

- Prediction errors consist of bias, variance, and irreducible error.

Correct!

- Bias error results in high training error.

- Variance error results in high training error.

Correct!

- Irreducible error cannot be optimized.

Question 16**3 / 3 pts**

Select Methods for good model evaluation.

Correct!

- Maximize values of both precision and recall

we need to maximize both precision and recall

- Maximize accuracy
- Maximize precision
- Minimize recall

Quiz Score: **44.25** out of 48

Module 2 Practice Quiz

Due No due date **Points** 10 **Questions** 10
Time Limit None **Allowed Attempts** Unlimited

Instructions

10 questions!

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	less than 1 minute	0.58 out of 10

Submitted Oct 28 at 12:52pm

Question 1

0.25 / 1 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following builds the decision tree model?

Correct Answer

- Gini index for each ID value is 0.

Correct Answer

- Gini for Android in Device OS is the same with Gini for IOS in Device OS.

Correct!

- Gini index for the carrier is 0.1625.

Correct Answer

- Gini index for the overall collection of the example is 0.5.

Question 2

0 / 1 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following builds the decision tree model?

-
- The ID is the best attribute to predict Earlier Adopter
 - The Carrier is the best attribute to predict Earlier Adopter
 - The Device OS is the best attribute to predict Earlier Adopter
 - The Device Color is the best attribute to predict Earlier Adopter
-

Correct Answer

You Answered

Unanswered

Question 3

0 / 1 pts

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

DU Answered

orrect Answers

Between 0.49 and 0.4915

Inanswered

Question 4

0 / 1 pts

Select all sentences that correctly define errors.

orrect Answer

- Errors committed on training records are training errors.

Correct Answer

Generalization error are expected errors of the model on previously unseen records.

Low training errors guarantee low generalization errors.

Generalization errors can be reduced by increasing the model complexity.

Question 5**0.33 / 1 pts**

Select all sentences that correctly define errors.

Correct!

Prediction errors consist of bias, variance, and irreducible error.

Correct Answer

Errors due to bias are primarily caused by models that underfit.

Correct Answer

Errors due to variance are caused by models that overfit.

Irreducible errors are caused by complex models.

Inanswered**Question 6****0 / 1 pts**

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute F1 for 'pear' class

You Answered**Correct Answers**

0 (with margin: 0)

Inanswered**Question 7**

0 / 1 pts

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute the accuracy

You Answered**Correct Answers**

0.5 (with margin: 0)

Inanswered**Question 8**

0 / 1 pts

Review the following table.

i	ii	iii	# of instances	
			+	-
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

According to the classification error rate, which attribute would be chosen as the first splitting attribute?

Correct Answer

'i'

'ii'

'iii'

anyone

Inanswered

Question 9

0 / 1 pts

Review the following table.

i	ii	iii	# of instances	
			+	-
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

After the first splitting and 'i' = F child node, which attribute would be chosen as the second splitting attribute?

'i'

'ii'

'iii'

anyone

Correct Answer

In answered

Question 10

0 / 1 pts

Review the following table.

I	II	III	# of instances	
			+	-
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

Compute the accuracy by the resulting decision tree.

You Answered

Correct Answers

0.2 (with margin: 0)

20 (with margin: 0)

Knowledge Check: Classification Issues

Due No due date	Points 4	Questions 4
Time Limit None	Allowed Attempts Unlimited	

Instructions

4 questions!

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	1 minute	1.5 out of 4

Submitted Oct 24 at 2:15am

Question 1 0 / 1 pts

A good classification model has:

A) Low training error

B) Low generalization error

B) Low generalization error alone is not sufficient for a good classification model.

Answered

Correct Answer

C) A and B are correct

D) None of the above

Question 2**1 / 1 pts**

Choose the best method for good model evaluation.

Correct!

- Maximize values of both precision and recall

we need to maximize both precision and recall

- Maximize accuracy
- Maximize precision
- Minimize recall

Question 3**0.5 / 1 pts**

Select all that are true about estimating generalization errors.

Correct!

- Generalization error is always minimized by minimizing training error



Given two models of similar generalization errors, one should prefer the simpler model over the more complex model

Complex model may not perform well on unseen data.

Correct Answer

- Complex models can accidentally fit errors in data
- All of the above

Question 4**0 / 1 pts**

Which of the following are true regarding bias and variance tradeoff?

You Answered

- Underfitting is caused by low variance alone.

We need to consider also bias when evaluating the model.

Correct Answer

- Simple model tend to have High variance and low bias.
- Complex models overfit because of low variance
- None of the above.

Knowledge Check: Introduction to Classification Tasks

Due No due date	Points 4	Questions 4
Time Limit None	Allowed Attempts Unlimited	

Instructions

4 questions!

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	less than 1 minute	0 out of 4

Submitted Oct 28 at 12:54pm

In answered

Question 1 0 / 1 pts

Select all techniques used to measure node impurity.

Correct Answer

Gini Index

average

Correct Answer

Entropy

split

average and split do not give you information about node's impurity

Inanswered

Question 2

0 / 1 pts

True or False? The tree induction algorithm C4.5 is suitable for large dataset.

 True False

Correct Answer

Inanswered

Question 3

0 / 1 pts

Which are not issues of decision trees?

 Specifying the attribute test condition. Determining the best split. Determine when to stop splitting.

Correct Answer

 Splitting on discrete attributes. All of the above

Inanswered

Question 4

0 / 1 pts

Select all that is true about best split of node of a decision tree.

Correct Answer

- Non-homogeneous nodes have high degree of impurity

Correct Answer

- Nodes with homogeneous class distribution are preferred

Correct Answer

- Homogeneous nodes have low degree of impurity.

- None of the above

Knowledge Check: Introduction to Classification

Due No due date **Points** 2 **Questions** 2
Time Limit None **Allowed Attempts** Unlimited

Instructions

2 questions!

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	less than 1 minute	0 out of 2

Submitted Oct 28 at 12:54pm

In answered	Question 1	0 / 1 pts
<p>True or False? A test set is used to find a model for class attribute as a function of the values of other attributes</p> <p><input type="radio"/> True</p> <p><input checked="" type="radio"/> False</p>		

In answered	Question 2	0 / 1 pts
<p>Select all classification techniques.</p>		

orrect Answer Decision tree**orrect Answer** Rule-based systems Clustering**orrect Answer** Neural Networks

Clustering is an unsupervised learning method.

Knowledge Check: Review of Initial Data Exploration Techniques

Due No due date	Points 2	Questions 2
Time Limit None	Allowed Attempts Unlimited	

Instructions

2 questions.

LO: Review and summarize data exploration techniques for use in initial data analysis

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	less than 1 minute	0 out of 2

Submitted Oct 28 at 1:02pm

In answered	Question 1	0 / 1 pts
	Select all correct sentences that make up the similarity property.	
Correct Answer	<input type="checkbox"/> If $p = q$, similarity, $s(p, q)$ is one (or maximum similarity)	
	<input type="checkbox"/> Similarity is always in the range 0 to 1	
Correct Answer	<input type="checkbox"/> Similarity is usually in the range -1 to 1	

Correct Answer

Cosine and Jaccard-Coefficient are typical similarity measurement methods

In case of non-negative inputs, similarity is in the range 0 to 1. So it is not ALWAYS in range -1 to 1. However, correlation is in range -1 to 1

Inanswered**Question 2**

0 / 1 pts

Select all correct sentences that define correlation.



Correlation 0 means that the vectors have a perfect positive linear relationship.



Correlation -1 means that the vectors have a perfect negative linear relationship.



Correlation 1 means that the vectors have a perfect negative linear relationship.

Correct Answer

Module 1 Practice Quiz

Due No due date **Points** 10 **Questions** 10
Time Limit None **Allowed Attempts** Unlimited

Instructions

10 questions!

LO:

- 1.1 Explain the history and purpose of data mining across multiple disciplines
- 1.2 Differentiate what is and what is not data mining
- 1.3 Describe different data mining tasks
- 1.4 Recognize attributes of data needed for data mining
- 1.5 Review and summarize data exploration techniques for use in initial data analysis

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	less than 1 minute	0 out of 10

Submitted Oct 28 at 12:54pm

In answered	Question 1	0 / 1 pts
	Select all directly relevant fields to data mining that have influenced the breakthroughs in the field of data mining.	
Correct Answer	<input type="checkbox"/> Machine Learning / Pattern Recognition	

orrect Answer

-
- Statistics / Artificial Intelligence

orrect Answer

-
- Database System

-
- Computer Network

Inanswered**Question 2**

0 / 1 pts

What factor or factors have influenced the rise of the use of data mining breakthroughs?



- Services like Youtube become popular and allowed for the collection of large amounts of data.



- Sensor data collection with pervasive manner became more common



- Hardware became cheaper and more powerful



- A, B



- A, B and C

orrect Answer**Inanswered****Question 3**

0 / 1 pts

Select all correct pairs of data attributes type and the appropriate example.

orrect Answer

-
- Nominal – hair colors

orrect Answer

-
- Interval – centigrade scale

orrect Answer

- Ratio – Degree Kelvin or absolute

orrect Answer

- Ordinal – order of finishing a race

Inanswered**Question 4****0 / 1 pts**

Select all correct sentences that make up the classification

- Classification is an unsupervised learning method

- Feature extraction is a mandatory preprocessing

orrect Answer

- Classification is a supervised learning method

orrect Answer

- Spam filtering is a typical classification application

- Dimensionality reduction is a typical classification technique

Inanswered**Question 5****0 / 1 pts**

Select all possible use-cases of clustering

- Credit card fraud detection

- Market segmentation

- Summarize news

orrect Answer

- A, B, and C

- A, B

Inanswered**Question 6****0 / 1 pts**

Select all sentences that are true.

K-Nearest Neighbor is a typical clustering technique

Association rule discovery is a typical technique for supervised learning

Clustering techniques do not require any parameters

A, B

None of the above

Correct Answer**Inanswered****Question 7****0 / 1 pts**

Select all predictive techniques.

Regression

Correct Answer

Classification

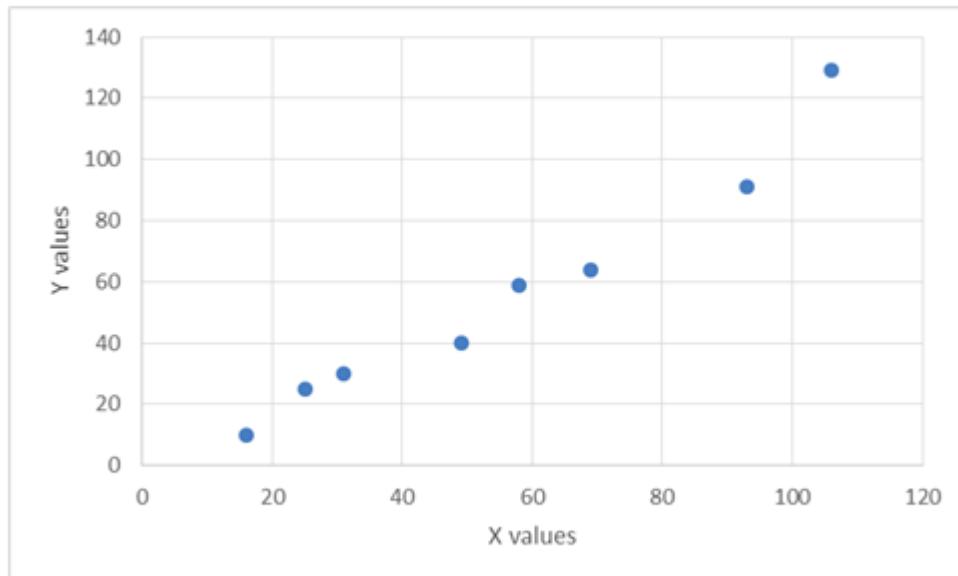
Correct Answer

Clustering

Dimensionality Reduction

Inanswered**Question 8****0 / 1 pts**

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Select the correct answer of Euclidean distance between (A and B), (D and E), and (G and H)

7.81(A and B), 39.62(D and E), 50.99(G and H)

17.49(A and B), 21.02(D and E), 50.99(G and H)

- 24(A and B), 28(D and E), 51(G and H)

orrect Answer

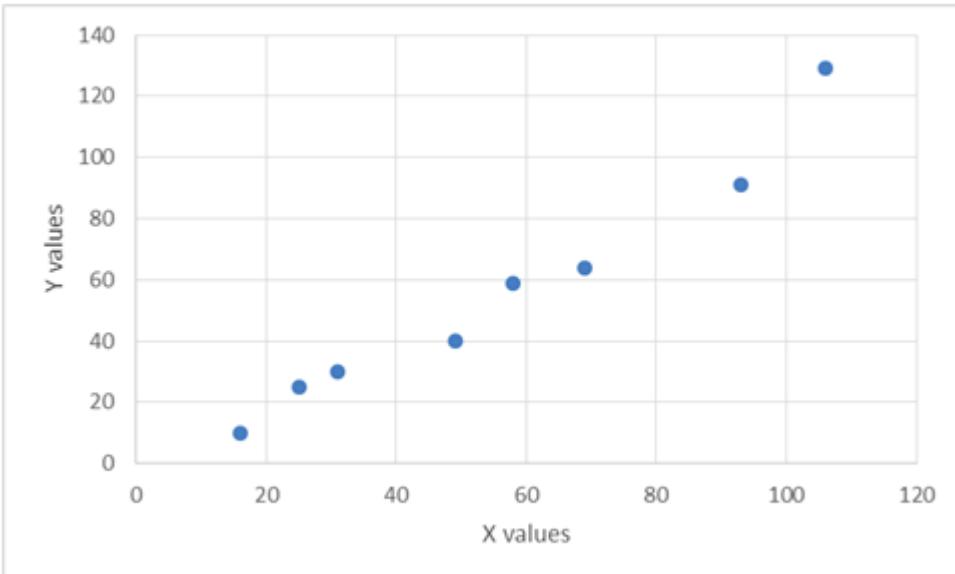
- None

Inanswered

Question 9

0 / 1 pts

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Compute cosine similarity between X and Y

DU Answered

Correct Answers

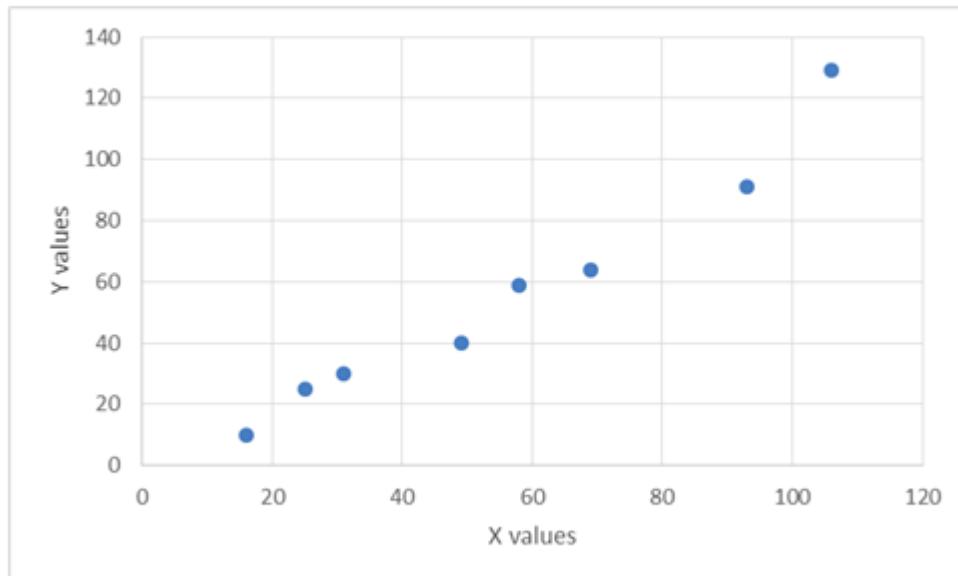
Between 0.99 and 0.992

Inanswered

Question 10

0 / 1 pts

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Select the most similar correlation between X and Y

0.5

-0.5

-1

Correct Answer 1 0

Knowledge Check: Data Attributes Needed for Data Mining

Due No due date **Points** 2 **Questions** 2
Time Limit None **Allowed Attempts** Unlimited

Instructions

2 questions!

LO: Recognize attributes of data needed for data mining

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	1 minute	0 out of 2

Submitted Oct 28 at 1:01pm

In answered	Question 1	0 / 1 pts
	Select all correct sentences that define attributes.	
	<input type="checkbox"/> The continuous attributes are often represented as integer variables.	
	<input type="checkbox"/> The binary attributes are discrete attributes.	
Correct Answer	<input checked="" type="checkbox"/>	The discrete attribute has only a finite or countably infinite set of values.

Unanswered

Question 2

0 / 1 pts

Select from the list below the attributes that could contribute to data quality problems.

Correct Answer Outliers**Correct Answer** Missing values**Correct Answer** Noise**Correct Answer** Duplicate Data

Noise refers to the modification of original values. Outliers are data objects with characteristics that are considerably different than most of the other data objects in the data set. Missing values and duplicate data are intuitively critical to data quality.

Knowledge Check: Introduction to Data Mining Tasks

Due No due date	Points 2	Questions 2
Time Limit None	Allowed Attempts Unlimited	

Instructions

2 questions!

LO: Describe different data mining tasks

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	less than 1 minute	0 out of 2

Submitted Oct 28 at 1pm

In answered	Question 1	0 / 1 pts
	Select all supervised learning methodologies	
	<input type="checkbox"/> Clustering	
	<input type="checkbox"/> Dimensionality Reduction	
Correct Answer	<input checked="" type="checkbox"/> Classification or categorization	
Correct Answer	<input checked="" type="checkbox"/> Regression	

Unanswered

Question 2

0 / 1 pts

Select all unsupervised learning applications

Correct Answer

- Marketing and sales promotion using Association Rule

Correct Answer

- Document clustering

- Face recognition

- Fraud detection in credit card transactions

Knowledge Check: History and Purpose of Data Mining

Due No due date **Points** 2 **Questions** 2
Time Limit None **Allowed Attempts** Unlimited

Instructions

2 questions!

LO: Explain the history and purpose of data mining across multiple disciplines

[Take the Quiz Again](#)

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	less than 1 minute	0 out of 2

Submitted Oct 28 at 12:59pm

In answered	Question 1	0 / 1 pts
	Discovering a student name in the class roster is an example of data mining function.	
	<input type="radio"/> True	
Correct Answer	<input checked="" type="radio"/> False	

Since the student name in the class roster can be obtained by looking up, the student name is already known information (or knowledge).

Unanswered

Question 2

0 / 1 pts

Big data is required for data mining.

 True False

Correct Answer

Although the increase in collected data size necessitates data mining, there is no specific data size needed to perform data mining.

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	87 minutes	44 out of 48

Score for this quiz: **44** out of 48

Submitted Oct 27 at 7:37pm

This attempt took 87 minutes.

Question 1	3 / 3 pts
Calculate Jaccard similar between X and Y X = 0101010001 Y = 0100011000	
Correct! 0.4	
Correct Answers 0.4 (with margin: 0)	

Question 2	3 / 3 pts

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

Tid

Correct!

Refund

Correct!

Marital Status

Correct!

Taxable Income

Correct!

Cheat

Question 3

2 / 3 pts

Select all predictive techniques.

Anomaly Detection

Correct Answer

Correct! Regression**Correct!** Classification Clustering Dimensionality Reduction**Question 4****3 / 3 pts**

Select all unsupervised learning applications

Correct! Marketing and sales promotion using Association Rule

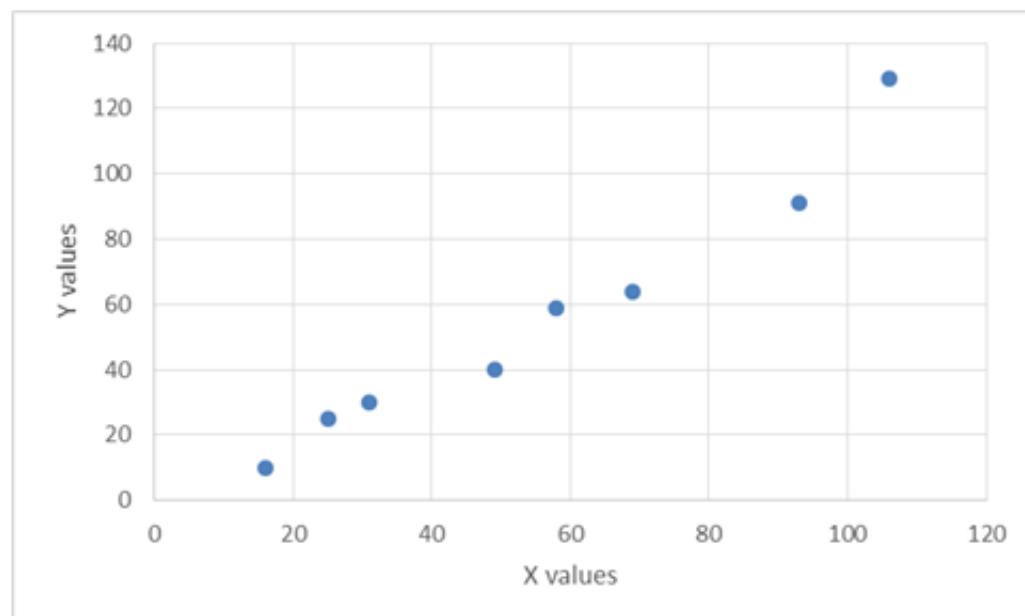
Association Rule Mining is an unsupervised learning method.

Correct! Document clustering

Clustering is an unsupervised learning method.

 Face recognition Fraud detection in credit card transactions**Question 5****3 / 3 pts**

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Select the closes answer for the correlation between X and Y

Correct! 0.5 -0.5 -1 1 0**Question 6****3 / 3 pts**

Select all correct sentences that define correlation.

Correct!

Correlation -1 means that the vectors have a perfect negative linear relationship.

Correct. If the correlation is -1, the vectors follow the perfect negative linear correlation.



Correlation 1 means that the vectors have a perfect negative linear relationship.



Correlation 0 means that the vectors have a perfect positive linear relationship.

Question 7**3 / 3 pts**

Select all sentences that correctly define errors.

Correct!

- Prediction errors consist of bias, variance, and irreducible error.

Correct!

- Bias error results in high training error.

- Variance error results in high training error.

Correct!

- Irreducible error cannot be optimized.

Question 8**0 / 3 pts**

Select all sentences that are correct about cross-validation.

Correct!

- Leave-one-out is a type of cross-validation.

You Answered

- Cross-validation guarantees reduction in test error.

-

K-fold cross-validation requires K repetitions of the same train set on K different test sets..

Correct Answer

-

In cross validation the same training record can be selected multiple times.

Question 9**3 / 3 pts**

Select all statements that correctly address underfitting and overfitting.

Correct!

- Underfitting can be reduced by reducing bias error.
- Overfitting can be reduced by reducing bias error.
- Underfitting can be reduced by reducing variance error.

Correct!

- Overfitting can be reduced by reducing variance error.

Question 10**3 / 3 pts**

Review the table below.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Select the correct precision and recall for 'apple' class.

Correct!

- Precision – 0.5 and recall – 0.5
- Precision – 1 and recall – 0.5
- Precision – 0.5 and recall – 1

- Precision – 1 and recall – 1

Question 11

3 / 3 pts

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute the accuracy as a fraction

Correct!

0.5

Correct Answers

0.5 (with margin: 0)

Question 12

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about the decision tree model?

Correct!

-
- The ID is the best attribute to predict Earlier Adopter
 - The Carrier is the best attribute to predict Earlier Adopter
 - The Device OS is the best attribute to predict Earlier Adopter
 - The Device Color is the best attribute to predict Earlier Adopter
-

Question 13

3 / 3 pts

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute F1 for 'pear' class

Correct!

Correct Answers

0 (with margin: 0)

Question 14

3 / 3 pts

Select the statements below that are computed as part of the performance metric.

Correct!

 Precision can be computed using True Positive and False Positive. Recall can be computed using True Positive and False Positive.

Correct!

F1 can be computed using True Positive and False Positive, and False Negative.



Accuracy can be computed using True Positive, False Positive, and False Negative.

Question 15

3 / 3 pts

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the Gini index? What is the Gini index of the split attribute?

Correct!

- Attribute 1 is the best split and Gini index is 0.3444

- Attribute 1 is the best split and Gini index is 0.4889
- Attribute 2 is the best split and Gini index is 0.3444
- Attribute 2 is the best split and Gini index is 0.4889

Question 16**3 / 3 pts**

Select all classification techniques.

Correct!

- Decision tree

Correct!

- Naïive Bayes

- Clustering

Correct!

- Neural Networks

Decision tree, Naïive Bayes and Neural Network are commonly used classification techniques.

Quiz Score: 44 out of 48

Midterm

Due Oct 28 at 11:59pm

Points 48

Questions 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	71 minutes	47 out of 48

Score for this quiz: **47** out of 48

Submitted Oct 27 at 7:47pm

This attempt took 71 minutes.

Question 1					3 / 3 pts
<hr/>					
Tid	Refund	Marital Status	Taxable Income	Cheat	
1	Yes	Single	125K	No	
2	No	Married	100K	No	
3	No	Single	70K	No	
4	Yes	Married	120K	No	
5	No	Divorced	95K	Yes	
6	No	Married	60K	No	
7	Yes	Divorced	220K	No	
8	No	Single	85K	Yes	
9	No	Married	75K	No	
10	No	Single	90K	Yes	

Select all possible columns which can be set as the class columns.

Tid

Correct! Refund**Correct!** Marital Status**Correct!** Taxable Income**Correct!** Cheat**Question 2****3 / 3 pts**

Select the correct sentence or sentences that define cosine and correlation measures.



The range of values that are possible for the cosine measure is [0, 1]



If two objects have a cosine similarity of 1, they are identical



If X and Y have a mean of 0, $\text{corr}(X,Y) = \cos(X,Y)$



A, B



A, B and C

Question 3**3 / 3 pts**

Calculate Jaccard similar between X and Y

X = 0101010001

Y = 0100011000

Correct!

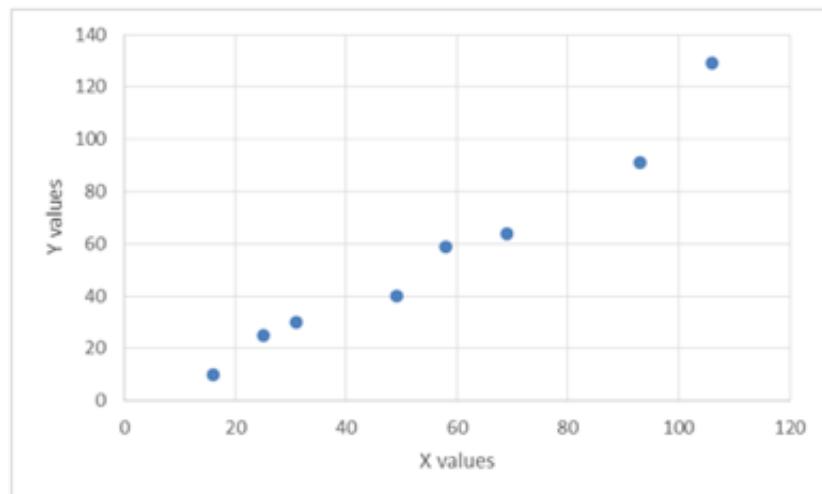
0.4

Correct Answers

0.4 (with margin: 0)

Question 4**3 / 3 pts**

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Compute cosine similarity between X and Y

Correct!

0.991

Correct Answers

Between 0.95 and 1

Question 5

3 / 3 pts

Select all unsupervised learning applications

Correct!

Marketing and sales promotion using Association Rule

Association Rule Mining is an unsupervised learning method.

Correct!

Document clustering

Clustering is an unsupervised learning method.

Face recognition

- Fraud detection in credit card transactions

Question 6**2 / 3 pts**

Select all predictive techniques.

Incorrect Answer

- Anomaly Detection

Correct!

- Regression

Correct!

- Classification

- Clustering

- Dimensionality Reduction

Question 7**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the classification rate and what is the classification error of the split attribute?

Correct!

Attribute 1 is the best split, 2/9 error rate

Attribute 1 is the best split, 4/9 error rate

Attribute 2 is the best split, 2/9 error rate

Attribute 2 is the best split, 4/9 error rate

Question 8

3 / 3 pts

Select all sentences that are correct about cross-validation.

Correct!

Leave-one-out is a type of cross-validation.

- Cross-validation guarantees reduction in test error.
- K-fold cross-validation requires K repetitions of the same train set on K different test sets..

Correct!

- In cross validation the same training record can be selected multiple times.

Question 9

3 / 3 pts

Select Methods for good model evaluation.

Correct!

- Maximize values of both precision and recall
 - we need to maximize both precision and recall
- Maximize accuracy
- Maximize precision
- Minimize recall

Question 10

3 / 3 pts

True or False? A test set is used to find a model for class attribute as a function of the values of other attributes

True False

Correct. A test set is used to determine the accuracy of the model.

Question 11

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about a decision tree model?

Correct!

- Gini index for each ID value is 0.

Correct!

- Gini for Android in Device OS is the same with Gini for IOS in Device OS.

Correct!

- Gini index for the carrier is 0.1625.

Correct!

- Gini index for the overall collection of the example is 0.5.

Question 12

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about the decision tree model?

-
- The ID is the best attribute to predict Earlier Adopter
 - The Carrier is the best attribute to predict Earlier Adopter
 - The Device OS is the best attribute to predict Earlier Adopter
 - The Device Color is the best attribute to predict Earlier Adopter

Correct!

Question 13**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the best split among attribute 1, attribute 2, and attribute 3, according to the information gain.

Correct!

 Attribute 1

 Attribute 2

 Attribute 3

Question 14**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the correct information gain pair of Attribute 1 and Attribute 2.

Correct!

- Attribute 1- 0.229, Attribute 2 – 0.007
- Attribute 1- 0.991, Attribute 2 – 0.007
- Attribute 1- 0.007, Attribute 2 – 0.229
- Attribute 1- 0.229, Attribute 2 – 0.991

Question 15

3 / 3 pts

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute the accuracy as a fraction

Correct!

0.5

Correct Answers

0.5 (with margin: 0)

Question 16

3 / 3 pts

Select the statements below that are computed as part of the performance metric.

Correct!



Precision can be computed using True Positive and False Positive.



Recall can be computed using True Positive and False Positive.

Correct!



F1 can be computed using True Positive and False Positive, and False Negative.



Accuracy can be computed using True Positive, False Positive, and False Negative.

Quiz Score: 47 out of 48

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	13 minutes	47 out of 48

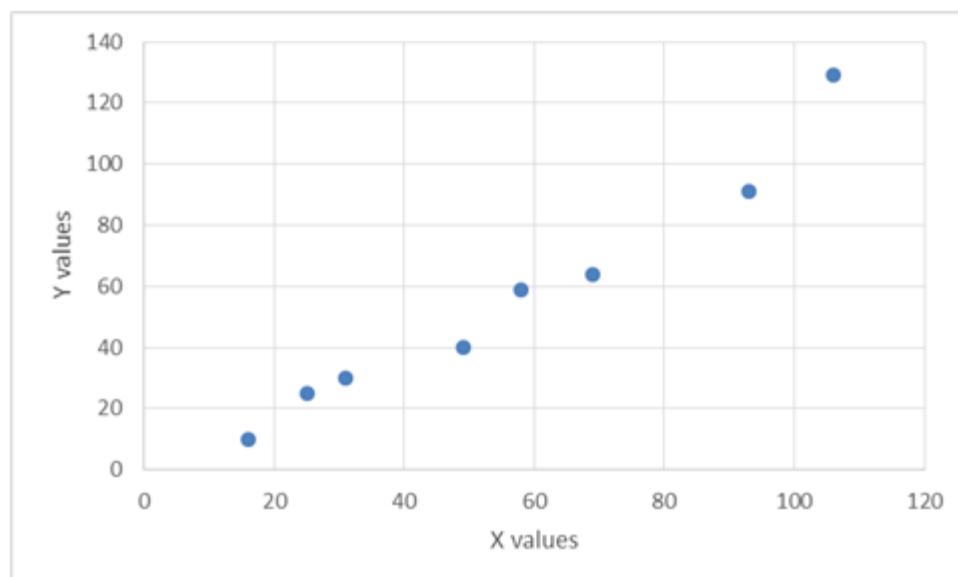
Score for this quiz: **47** out of 48

Submitted Oct 27 at 3:23pm

This attempt took 13 minutes.

Question 1 3 / 3 pts

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Compute cosine similarity between X and Y

Correct!

0.973

Incorrect Answers

Between 0.95 and 1

Question 2

3 / 3 pts

Select all correct sentences which make up the clustering.

- K-Nearest Neighbor is a typical clustering technique
- Association rule discovery is a typical technique of clustering
- Clustering techniques do not require any parameters
- A, B
- A, B, and C

Correct!

Question 3**3 / 3 pts**

Select all possible use-cases of clustering

- Credit card fraud detection
- Market segmentation
- Summarize news

Correct!

- A, B, and C
- A, B

Question 4**2 / 3 pts**

Select all predictive techniques.

Incorrect Answer

- Anomaly Detection

Correct!

- Regression

Correct!

- Classification

- Clustering

- Dimensionality Reduction

Question 5**3 / 3 pts**

Select the correct sentence or sentences that define cosine and correlation

measures.

The range of values that are possible for the cosine measure is [0, 1]

If two objects have a cosine similarity of 1, they are identical

Correct!

If X and Y have a mean of 0, $\text{corr}(X,Y) = \cos(X,Y)$

A, B

A, B and C

Question 6

3 / 3 pts

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

Tid

Correct!

Refund

Correct!

Marital Status

Correct! Taxable Income**Correct!** Cheat**Question 7****3 / 3 pts**

Select the statements below that are computed as part of the performance metric.

Correct! Precision can be computed using True Positive and False Positive. Recall can be computed using True Positive and False Positive.**Correct!** F1 can be computed using True Positive and False Positive, and False Negative. Accuracy can be computed using True Positive, False Positive, and False Negative.**Question 8****3 / 3 pts**

Select all sentences that correctly define errors.

Correct! Errors committed on training records are training errors.**Correct!** Generalization error are expected errors of the model on previously unseen records. Low training errors guarantee low generalization errors.

Generalization errors can be reduced by increasing the model complexity.

Question 9**3 / 3 pts**

True or False? The tree induction algorithm C4.5 is suitable for large dataset.

 True False

C4.5 is suitable for small data set.

Correct!**Question 10****3 / 3 pts**

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

Correct!

0.491

Incorrect Answers

Between 0.49 and 0.4915

Question 11

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about the decision tree model?

Correct!

-
- The ID is the best attribute to predict Earlier Adopter
 - The Carrier is the best attribute to predict Earlier Adopter

 - The Device OS is the best attribute to predict Earlier Adopter

 - The Device Color is the best attribute to predict Earlier Adopter

Question 12

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about a decision tree model?

Correct!

- Gini index for each ID value is 0.

Correct!

- Gini for Android in Device OS is the same with Gini for IOS in Device OS.

Correct!

- Gini index for the carrier is 0.1625.

Correct!

- Gini index for the overall collection of the example is 0.5.

Question 13**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the Gini index? What is the Gini index of the split attribute?

Correct!

- Attribute 1 is the best split and Gini index is 0.3444
- Attribute 1 is the best split and Gini index is 0.4889
- Attribute 2 is the best split and Gini index is 0.3444
- Attribute 2 is the best split and Gini index is 0.4889

Question 14**3 / 3 pts**

Select all classification techniques.

Correct!

- Decision tree

Correct!

- Naïve Bayes

- Clustering

Correct!

- Neural Networks

Decision tree, Naïve Bayes and Neural Network are commonly used classification techniques.

Question 15**3 / 3 pts**

A good classification model has:

- Low training error
- Low estimated generalization error
- A and B are correct

We need both low training error and low generalization error.

Correct!

- None of the above

Question 16**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the correct information gain pair of Attribute 1 and Attribute 2.

Correct!

- Attribute 1- 0.229, Attribute 2 – 0.007
- Attribute 1- 0.991, Attribute 2 – 0.007
- Attribute 1- 0.007, Attribute 2 – 0.229
- Attribute 1- 0.229, Attribute 2 – 0.991

Quiz Score: **47** out of 48

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

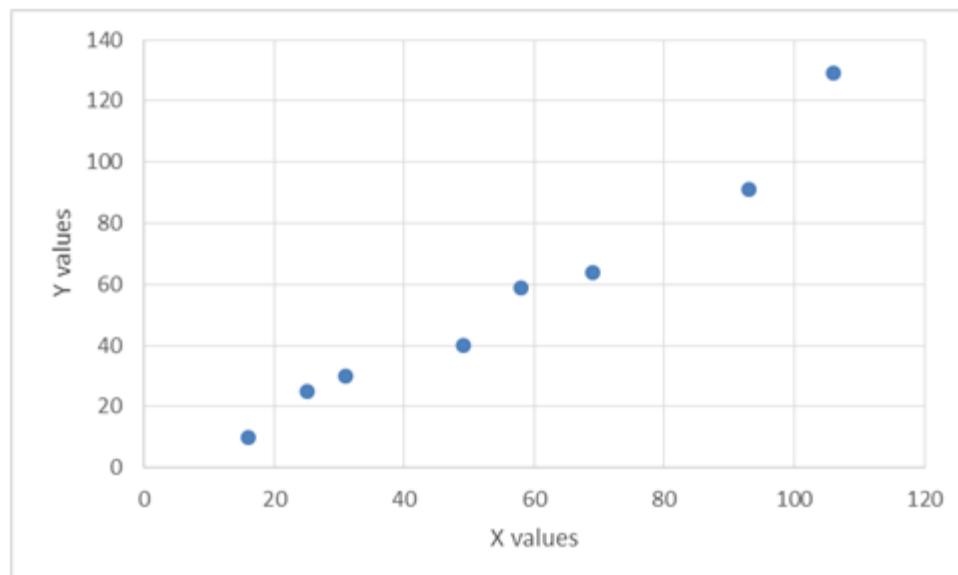
	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	10 minutes	45 out of 48

Score for this quiz: **45** out of 48

Submitted Oct 27 at 3:15pm

This attempt took 10 minutes.

Question 1			3 / 3 pts																											
<table border="1"><thead><tr><th></th><th>X</th><th>Y</th></tr></thead><tbody><tr><td>A</td><td>16</td><td>10</td></tr><tr><td>B</td><td>25</td><td>25</td></tr><tr><td>C</td><td>31</td><td>30</td></tr><tr><td>D</td><td>49</td><td>40</td></tr><tr><td>E</td><td>58</td><td>59</td></tr><tr><td>F</td><td>69</td><td>64</td></tr><tr><td>G</td><td>93</td><td>91</td></tr><tr><td>H</td><td>106</td><td>129</td></tr></tbody></table>					X	Y	A	16	10	B	25	25	C	31	30	D	49	40	E	58	59	F	69	64	G	93	91	H	106	129
	X	Y																												
A	16	10																												
B	25	25																												
C	31	30																												
D	49	40																												
E	58	59																												
F	69	64																												
G	93	91																												
H	106	129																												



Compute cosine similarity between X and Y

Correct!

0.965

Incorrect Answers

Between 0.95 and 1

Question 2

3 / 3 pts

Select all correct sentences that define correlation.

Correct!



Correlation -1 means that the vectors have a perfect negative linear relationship.



Correct. If the correlation is -1, the vectors follow the perfect negative linear correlation.



Correlation 1 means that the vectors have a perfect negative linear relationship.



Correlation 0 means that the vectors have a perfect positive linear relationship.

Question 3**3 / 3 pts**

Select all unsupervised learning applications

Correct!

- Marketing and sales promotion using Association Rule

Association Rule Mining is an unsupervised learning method.

Correct!

- Document clustering

Clustering is an unsupervised learning method.

- Face recognition

- Fraud detection in credit card transactions

Question 4**3 / 3 pts**

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

Tid

Correct!

Refund

Correct!

Marital Status

Correct!

Taxable Income

Correct!

Cheat

Question 5

3 / 3 pts

Select all correct sentences which make up the clustering.

K-Nearest Neighbor is a typical clustering technique

Correct!

Association rule discovery is a typical technique of clustering

Clustering techniques do not require any parameters

A, B A, B, and C**Question 6****3 / 3 pts**

Select the correct sentence or sentences that define cosine and correlation measures.

 The range of values that are possible for the cosine measure is [0, 1] If two objects have a cosine similarity of 1, they are identical If X and Y have a mean of 0, $\text{corr}(X,Y) = \cos(X,Y)$ A, B A, B and C**Correct!****Question 7****3 / 3 pts**

Select all sentences that correctly define errors.

Correct! Errors committed on training records are training errors.**Correct!**
Generalization error are expected errors of the model on previously unseen records. Low training errors guarantee low generalization errors.

Generalization errors can be reduced by increasing the model complexity.

Question 8**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

When is the best split for Attribute 3 according to the information gain?

Correct!

- The split point equal to 2.0
- The split point equal to 3.5
- The split point equal to 5.5
- The split point equal to 6.5
- The split point equal to 7.5

Question 9**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the best split among attribute 1, attribute 2, and attribute 3, according to the information gain.

Correct! Attribute 1 Attribute 2 Attribute 3**Question 10****3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the classification rate and what is the classification error of the split attribute?

Correct!

- Attribute 1 is the best split, 2/9 error rate
- Attribute 1 is the best split, 4/9 error rate
- Attribute 2 is the best split, 2/9 error rate
- Attribute 2 is the best split, 4/9 error rate

Question 11

3 / 3 pts

A good classification model has:

- Low training error
- Low estimated generalization error

Correct!

- A and B are correct

We need both low training error and low generalization error.

- None of the above

Question 12**3 / 3 pts**

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

Correct!

0.491

Incorrect Answers

Between 0.49 and 0.4915

Question 13**3 / 3 pts**

Select all statements that correctly address underfitting and overfitting.

Correct! Underfitting can be reduced by reducing bias error. Overfitting can be reduced by reducing bias error. Underfitting can be reduced by reducing variance error.**Correct!** Overfitting can be reduced by reducing variance error.**Question 14****3 / 3 pts**

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute the accuracy as a fraction

Correct!

0.5

Incorrect Answers

0.5 (with margin: 0)

Question 15**3 / 3 pts**

Select the statements below that are computed as part of the performance metric.

Correct!

Precision can be computed using True Positive and False Positive.

Recall can be computed using True Positive and False Positive.

Correct!

F1 can be computed using True Positive and False Positive, and False Negative.

Accuracy can be computed using True Positive, False Positive, and False Negative.

Question 16**0 / 3 pts**

Select all sentences that are correct about cross-validation.

Correct!

Leave-one-out is a type of cross-validation.

You Answered

Cross-validation guarantees reduction in test error.

K-fold cross-validation requires K repetitions of the same train set on K different test sets..

Correct Answer

In cross validation the same training record can be selected multiple times.

Quiz Score: 45 out of 48

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	40 minutes	44.25 out of 48

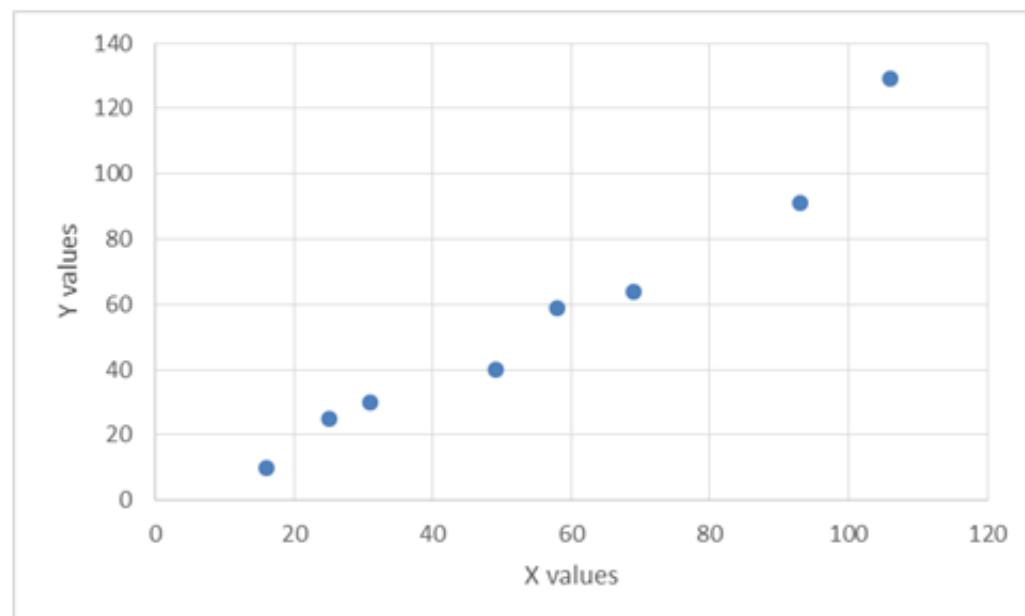
Score for this quiz: **44.25** out of 48

Submitted Oct 26 at 8:23pm

This attempt took 40 minutes.

Question 1	3 / 3 pts
<div style="height: 400px; width: 100%;"></div>	

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Select the closes answer for the correlation between X and Y

0.5 -0.5 -1 1 0**Correct!****Question 2****3 / 3 pts**

Select all correct sentences that define correlation.

Correct!

Correlation -1 means that the vectors have a perfect negative linear relationship.

Correct. If the correlation is -1, the vectors follow the perfect negative linear correlation.



Correlation 1 means that the vectors have a perfect negative linear relationship.



Correlation 0 means that the vectors have a perfect positive linear relationship.

Question 3**2.25 / 3 pts**

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

 Tid**Correct!** Refund**Correct!** Marital Status**Correct Answer** Taxable Income**Correct!** Cheat**Question 4****3 / 3 pts**

Select all unsupervised learning applications

Correct!

- Marketing and sales promotion using Association Rule

Association Rule Mining is an unsupervised learning method.

Correct!

- Document clustering

Clustering is an unsupervised learning method.

- Face recognition

- Fraud detection in credit card transactions

Question 5

3 / 3 pts

Select all correct sentences which make up the clustering.

- K-Nearest Neighbor is a typical clustering technique

- Association rule discovery is a typical technique of clustering

- Clustering techniques do not require any parameters

- A, B

- A, B, and C

Question 6

0 / 3 pts

Select the correct sentence or sentences that define cosine and correlation measures.



The range of values that are possible for the cosine measure is [0, 1]



If two objects have a cosine similarity of 1, they are identical



If X and Y have a mean of 0, $\text{corr}(X, Y) = \cos(X, Y)$



A, B



A, B and C

Correct Answer**You Answered****Question 7**

3 / 3 pts

Select all classification techniques.

Correct! Decision tree**Correct!** Naïive Bayes**Correct!** Clustering Neural Networks

Decision tree, Naïve Bayes and Neural Network are commonly used classification techniques.

Question 8**3 / 3 pts**

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about a decision tree model?

Correct!

- Gini index for each ID value is 0.

Correct!



Gini for Android in Device OS is the same with Gini for IOS in Device OS.

Correct!

- Gini index for the carrier is 0.1625.

Correct!

- Gini index for the overall collection of the example is 0.5.

Question 9

3 / 3 pts

Review the table below.

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Based on the table, which of the following statements are correct about the decision tree model?

-
- The ID is the best attribute to predict Earlier Adopter
 - The Carrier is the best attribute to predict Earlier Adopter
 - The Device OS is the best attribute to predict Earlier Adopter
 - The Device Color is the best attribute to predict Earlier Adopter

Correct!

Question 10**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the classification rate and what is the classification error of the split attribute?

Correct!

- Attribute 1 is the best split, 2/9 error rate
- Attribute 1 is the best split, 4/9 error rate
- Attribute 2 is the best split, 2/9 error rate
- Attribute 2 is the best split, 4/9 error rate

Question 11**3 / 3 pts**

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

Correct!

0.491

Correct Answers

Between 0.49 and 0.4915

Question 12

3 / 3 pts

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute F1 for 'pear' class

Correct! 0**Correct Answers**

0 (with margin: 0)

Question 13

3 / 3 pts

Review the table below.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Select the correct precision and recall for 'apple' class.

-
- Precision – 0.5 and recall – 0.5

Correct!

- Precision – 1 and recall – 0.5
- Precision – 0.5 and recall – 1
- Precision – 1 and recall – 1

Question 14**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the correct information gain pair of Attribute 1 and Attribute 2.

Correct!

- Attribute 1- 0.229, Attribute 2 – 0.007

- Attribute 1- 0.991, Attribute 2 – 0.007
- Attribute 1- 0.007, Attribute 2 – 0.229
- Attribute 1- 0.229, Attribute 2 – 0.991

Question 15**3 / 3 pts**

Select all sentences that correctly define errors.

Correct!

- Prediction errors consist of bias, variance, and irreducible error.

Correct!

- Bias error results in high training error.

- Variance error results in high training error.

Correct!

- Irreducible error cannot be optimized.

Question 16**3 / 3 pts**

Select Methods for good model evaluation.

Correct!

- Maximize values of both precision and recall

we need to maximize both precision and recall

- Maximize accuracy
- Maximize precision
- Minimize recall

Quiz Score: **44.25** out of 48

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

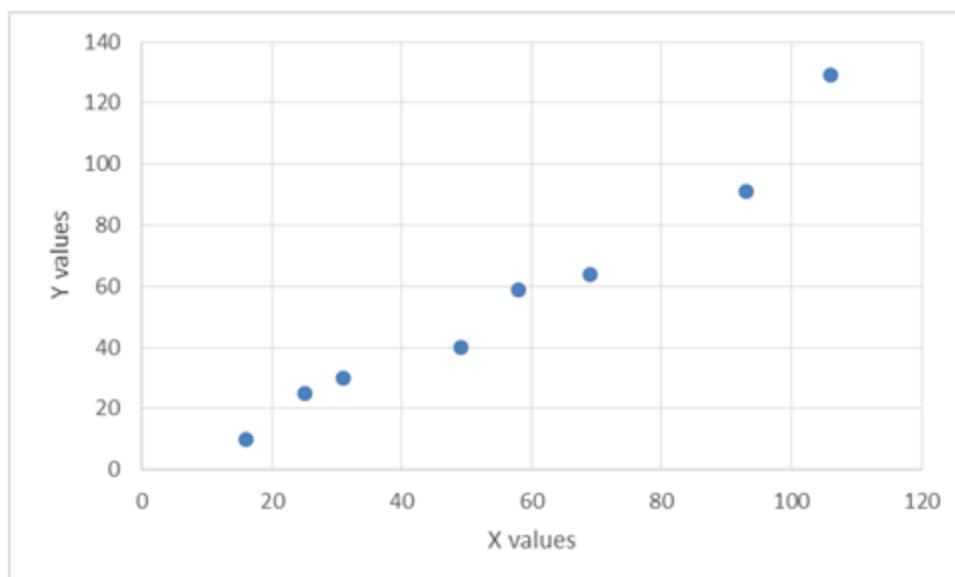
	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	43 minutes	46.5 out of 48

Score for this quiz: **46.5** out of 48

Submitted Oct 26 at 9:23pm

This attempt took 43 minutes.

Question 1			3 / 3 pts
	X	Y	
A	16	10	
B	25	25	
C	31	30	
D	49	40	
E	58	59	
F	69	64	
G	93	91	
H	106	129	



Compute cosine similarity between X and Y

Correct!

0.991

Correct Answers

Between 0.95 and 1

Question 2

1.5 / 3 pts

Select all predictive techniques.

You Answered

- Anomaly Detection

Correct!

- Regression

Correct!

- Classification

- Clustering

- Dimensionality Reduction

Question 3

3 / 3 pts

Select all correct sentences which make up the clustering.

Correct!

- K-Nearest Neighbor is a typical clustering technique

- Association rule discovery is a typical technique of clustering

- Clustering techniques do not require any parameters

- A, B

- A, B, and C

Question 4

3 / 3 pts

Select the correct sentence or sentences that define cosine and correlation measures.

Correct!

The range of values that are possible for the cosine measure is [0, 1]

If two objects have a cosine similarity of 1, they are identical

If X and Y have a mean of 0, $\text{corr}(X,Y) = \cos(X,Y)$

A, B

A, B and C

Correct!**Question 5****3 / 3 pts**

Select all supervised learning methodologies

Clustering

Dimensionality Reduction

Classification or categorization

Classification is a discrete data based supervised learning method.

Correct!

Regression

Regression is a continuous data based supervised learning method.

Correct!**Question 6****3 / 3 pts**

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

Tid

Correct!

Refund

Correct!

Marital Status

Correct!

Taxable Income

Correct!

Cheat

Question 7

3 / 3 pts

ID	Device OS	Carrier	Device Color	Earlier Adopter (Class)
1	Android	T-Mobile	White	Y
2	Android	AT&T	Gray	Y
3	Android	AT&T	Gray	Y
4	Android	AT&T	Black	Y
5	Android	AT&T	Pink	Y
6	Android	AT&T	Pink	Y
7	IOS	AT&T	White	Y
8	IOS	AT&T	White	Y
9	IOS	AT&T	Gray	Y
10	IOS	Verizon	Black	Y
11	Android	T-Mobile	Black	N
12	Android	T-Mobile	Pink	N
13	Android	T-Mobile	Gray	N
14	Android	Verizon	Pink	N
15	IOS	Verizon	White	N
16	IOS	Verizon	White	N
17	IOS	Verizon	Gray	N
18	IOS	Verizon	Gray	N
19	IOS	Verizon	Gray	N
20	IOS	Verizon	Black	N

Compute the Gini index for Device Color attribute using multiway split

Correct!

0.491

Correct Answers

Between 0.49 and 0.4915

Question 8

3 / 3 pts

Select the statements below that are computed as part of the performance metric.

Correct!

Precision can be computed using True Positive and False Positive.

Recall can be computed using True Positive and False Positive.

Correct!

F1 can be computed using True Positive and False Positive, and False Negative.

Accuracy can be computed using True Positive, False Positive, and False Negative.

Question 9

3 / 3 pts

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

When is the best split for Attribute 3 according to the information gain?

Correct!

The split point equal to 2.0

- The split point equal to 3.5
- The split point equal to 5.5
- The split point equal to 6.5
- The split point equal to 7.5

Question 10**3 / 3 pts**

Select all classification techniques.

Correct!

- Decision tree

Correct!

- Naïive Bayes

- Clustering

Correct!

- Neural Networks

Decision tree, Naïive Bayes and Neural Network are commonly used classification techniques.

Question 11**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

Select the best split among attribute 1, attribute 2, and attribute 3, according to the information gain.

Correct!

Attribute 1

Attribute 2

Attribute 3

Question 12

3 / 3 pts

Select all sentences that correctly define errors.

Correct!

Errors committed on training records are training errors.

Correct!

Generalization error are expected errors of the model on previously unseen records.

Low training errors guarantee low generalization errors.

Generalization errors can be reduced by increasing the model complexity.

Question 13**3 / 3 pts**

True or False? The tree induction algorithm C4.5 is suitable for large dataset.

True

False

C4.5 is suitable for small data set.

Correct!**Question 14****3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the Gini index? What is the Gini index of the split attribute?

Correct!

Attribute 1 is the best split and Gini index is 0.3444

Attribute 1 is the best split and Gini index is 0.4889

Attribute 2 is the best split and Gini index is 0.3444

Attribute 2 is the best split and Gini index is 0.4889

Question 15

3 / 3 pts

Review the table below.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Select the correct precision and recall for 'apple' class.

Correct!

- Precision – 0.5 and recall – 0.5
- Precision – 1 and recall – 0.5
- Precision – 0.5 and recall – 1
- Precision – 1 and recall – 1

Question 16

3 / 3 pts

Select all statements that correctly address underfitting and overfitting.

Correct!

- Underfitting can be reduced by reducing bias error.
- Overfitting can be reduced by reducing bias error.
- Underfitting can be reduced by reducing variance error.
- Overfitting can be reduced by reducing variance error.

Correct!

Quiz Score: **46.5** out of 48

Midterm

Due Oct 28 at 11:59pm **Points** 48 **Questions** 16

Available Oct 26 at 12am - Oct 29 at 1am 3 days

Time Limit 100 Minutes

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	54 minutes	48 out of 48

Score for this quiz: **48** out of 48

Submitted Oct 28 at 12:03am

This attempt took 54 minutes.

Question 1

3 / 3 pts

Select all possible use-cases of clustering

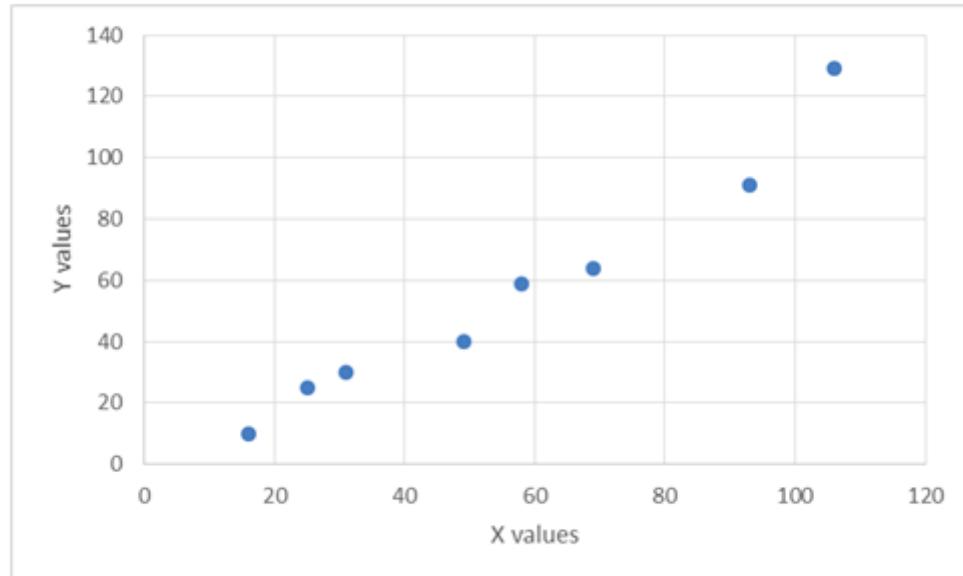
- Credit card fraud detection
- Market segmentation
- Summarize news
- A, B, and C
- A, B

Correct!

Question 2

3 / 3 pts

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Select the closest answer for the correlation between X and Y

0.5

-0.5

-1

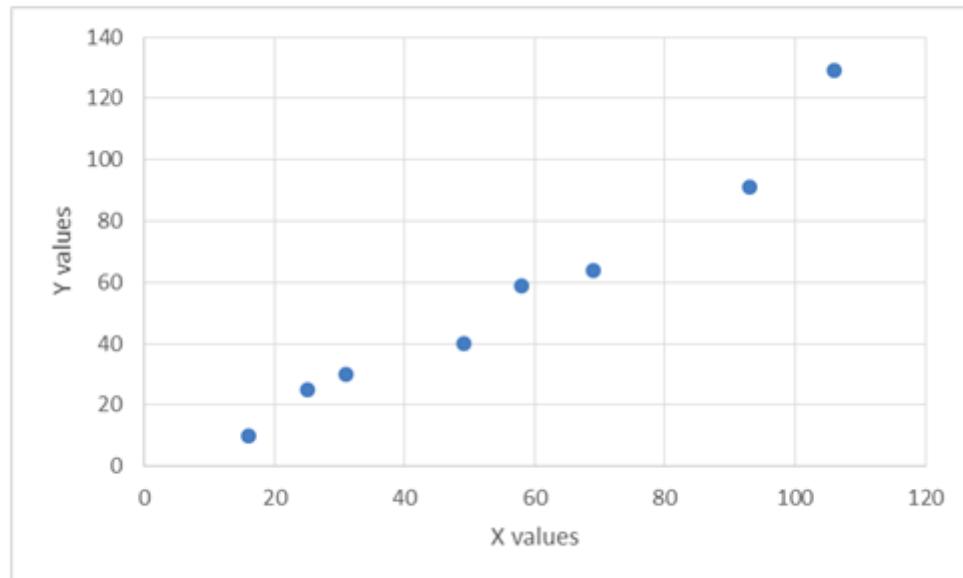
Correct! 1 0**Question 3****3 / 3 pts**

Select all correct sentences which make up the clustering.

- K-Nearest Neighbor is a typical clustering technique
- Association rule discovery is a typical technique of clustering
- Clustering techniques do not require any parameters
- A, B
- A, B, and C

Question 4**3 / 3 pts**

	X	Y
A	16	10
B	25	25
C	31	30
D	49	40
E	58	59
F	69	64
G	93	91
H	106	129



Compute cosine similarity between X and Y

Correct!

0.991

Correct Answers

Between 0.95 and 1

Question 5**3 / 3 pts**

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Select all possible columns which can be set as the class columns.

 Tid**Correct!** Refund**Correct!** Marital Status**Correct!** Taxable Income**Correct!** Cheat**Question 6****3 / 3 pts**

Select all predictive techniques.

 Anomaly Detection**Correct!**

Correct! Regression**Correct!** Classification Clustering Dimensionality Reduction**Question 7****3 / 3 pts**

True or False? The tree induction algorithm C4.5 is suitable for large dataset.

 True False

C4.5 is suitable for small data set.

Correct!**Question 8****3 / 3 pts**

Select all sentences that are correct about cross-validation.

Correct! Leave-one-out is a type of cross-validation. Cross-validation guarantees reduction in test error.

K-fold cross-validation requires K repetitions of the same train set on K different test sets..

Correct!

In cross validation the same training record can be selected multiple times.

Question 9**3 / 3 pts**

Select the statements below that are computed as part of the performance metric.

Correct!

Precision can be computed using True Positive and False Positive.



Recall can be computed using True Positive and False Positive.

Correct!

F1 can be computed using True Positive and False Positive, and False Negative.



Accuracy can be computed using True Positive, False Positive, and False Negative.

Question 10**3 / 3 pts**

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute the accuracy as a fraction

Correct!

0.5

orrect Answers

0.5 (with margin: 0)

Question 11**3 / 3 pts**

Select all statements that correctly address underfitting and overfitting.

Correct! Underfitting can be reduced by reducing bias error. Overfitting can be reduced by reducing bias error. Underfitting can be reduced by reducing variance error.**Correct!** Overfitting can be reduced by reducing variance error.**Question 12****3 / 3 pts**

Review the following table.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute F1 for 'pear' class

Correct!

0

orrect Answers

0 (with margin: 0)

Question 13**3 / 3 pts**

Review the following table.

Instance	Attribute 1	Attribute 2	Attribute 3	class
1	T	T	1	Y
2	T	T	6	Y
3	T	F	5	N
4	F	F	4	Y
5	F	T	7	N
6	F	T	3	N
7	F	F	8	N
8	T	F	7	Y
9	F	T	5	N

What is the best split between Attribute 1 and Attribute 2 according to the Gini index? What is the Gini index of the split attribute?

Correct!

- Attribute 1 is the best split and Gini index is 0.3444
- Attribute 1 is the best split and Gini index is 0.4889
- Attribute 2 is the best split and Gini index is 0.3444
- Attribute 2 is the best split and Gini index is 0.4889

Question 14**3 / 3 pts**

Select all techniques used to measure node impurity.

Correct! Gini Index

Gini Index is a measure of node impurity

 average**Correct!** Entropy

Entropy is a measure of node impurity

 split**Question 15**

3 / 3 pts

A good classification model has:

 Low training error Low estimated generalization error A and B are correct

We need both low training error and low generalization error.

 None of the above**Question 16**

3 / 3 pts

Review the table below.

Ground Truth	apple									
Predict	apple	pear	pear	apple	apple	pear	pear	pear	apple	apple

Compute F1 for 'apple' class (the result should be rounded to the thousandths place)

Correct!

0.667

Correct Answers Between 0.6 and 0.7

Quiz Score: **48** out of 48