

Quiz 2

Due Aug 1 at 19:00

Points 5

Questions 20

Available Jul 31 at 19:00 - Aug 1 at 19:00 24 hours

Time Limit 60 Minutes

Instructions

- This Quiz is for **60 minutes** duration.
- Multiple attempts are **not** allowed.
- Choose the correct/most apt options from the given options.
- Correct Answers will be visible later. Please do not discuss answers.

All the Best!

Attempt History

| | Attempt | Time | Score |
|---------------|----------------------------------|-------------|--------------|
| LATEST | <u>Attempt 1</u> | 23 minutes | 5 out of 5 |

! Correct answers will be available on Aug 1 at 19:02.

Score for this quiz: **5 out of 5**

Submitted Jul 31 at 22:58

This attempt took 23 minutes.

Question 1

0.25 / 0.25 pts

Which of the following are signs of a High Variance Machine Learning model?

- P) Noise in dataset
- Q) Overfitting
- R) Underfitting
- S) Failure to capture data trends

R only

P and Q

P, Q and S

P only

Question 2

0.25 / 0.25 pts

Visualization techniques like box plots or scatter plots applied on given data is a good way to identify:

Outliers

Missing values

Errors against Codebook

Data Entry errors

Question 3**0.25 / 0.25 pts**

The Normal distribution is fully specified by the following two parameters

the mean and the median

the mean and the standard deviation

the mean and the upper bound

the mean and the lower bound

Question 4**0.25 / 0.25 pts**

Models trained in one domain can be easily generalized to other domains

True

 False**Question 5****0.25 / 0.25 pts**

Noise is a distortion of a value or addition of spurious values.

 True

 False**Question 6****0.25 / 0.25 pts**

Genomic Sequence data is an example of Ordered Data

 True

 False

Question 7**0.25 / 0.25 pts**

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes
- C) It preserves all details even after aggregation at all times
- D) It does not complement well with statistical analysis

 B, D

 C only

 A , B

 B only**Question 8****0.25 / 0.25 pts**

Choose the correct statements about association analysis.

- Association analysis requires symmetric binary attributes.
- For association problems, one binary attribute for each categorical value is introduced.
- In association analysis, only the attribute '1' is important.
- Association analysis requires asymmetric binary attributes.

Question 9

0.25 / 0.25 pts

The objective of Cross-Validation techniques is to improve _____ :

- Precision and Recall
- Model evaluation
- Quality of Fit
- ROC Curve

Question 10**0.25 / 0.25 pts**

Overfitting happens when:

- the model has low bias and low variance
- the model has high bias and high variance
- the model has high bias and low variance
- the model has low bias and high variance

Question 11**0.25 / 0.25 pts**

Suppose that you need to build model for 100000 records. You may partition the dataset into 10 subsets and the model is trained for the first 9 subsets and tested for the 10th subset. The technique followed in the above model is

- None of the mentioned

- Traditional validation
- Cross-Validation
- Root mean square

Question 12

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'Laptop Model Number.'

- Discrete, qualitative, nominal
- Continuous, qualitative, nominal
- Discrete, qualitative, ordinal

Question 13

0.25 / 0.25 pts

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. When data is converted to apply association analysis, what type of attributes would you have and how many?

- 100 symmetric binary attributes
- 20 asymmetric binary attributes
- 100 asymmetric binary attributes
- 20 symmetric binary attributes

Question 14**0.25 / 0.25 pts**

Which of the following is not a Classification model metric?

- Confusion Matrix
- Precision
- R Squared or Adjusted R Square
- Recall

Question 15**0.25 / 0.25 pts**

Match the following type of data

Eye color

Nominal

**Education Level**

Ordinal

**Number of planets around the Sun**

Discrete

**The height of a student**

Continuous

**Question 16****0.25 / 0.25 pts**

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness

as measured by a light meter. Continuous, quantitative, ratio.

'The opinion in terms: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree'

- Discrete, qualitative, nominal
- Continuous, qualitative, ordinal
- Discrete, qualitative, ordinal

Question 17

0.25 / 0.25 pts

These days purchasing used books through online store is becoming more popular. Assume that a user logs in to AbeBooks to do the same. Predicting whether he / she makes a purchase of a book is a _____ task.

- Classification
- Reinforcement
- Regression
- Association

Question 18**0.25 / 0.25 pts**

Suppose you need to record the temperature of furnace in a database. This type is an example for
_____ attribute

 Normal Continuous Discrete Asymmetric**Question 19****0.25 / 0.25 pts**

Quasi Structured data is considered to be more Structured than Semi-Structured data.

 True False

Question 20**0.25 / 0.25 pts**

Network event logs are the example of _____.

- Natural language
- Graph based
- Machine generated data
- Structured data

Quiz Score: 5 out of 5

Quiz 2

- **Due Aug 1 at 19:00**

- **Points 5**

- **Questions 20**

- **Available Jul 31 at 19:00 - Aug 1 at 19:00 24 hours**

- **Time Limit 60 Minutes**

Instructions

- This Quiz is for **60 minutes** duration.
- Multiple attempts are **not** allowed.
- Choose the correct/most apt options from the given options.
- Correct Answers will be visible later. Please do not discuss answers.

All the Best!

Attempt History

| | Attempt | Time | Score |
|---------------|---------------------------|------------|------------|
| LATEST | Attempt 1 | 22 minutes | 5 out of 5 |

Correct answers will be available on Aug 1 at 19:02.

Score for this quiz: 5 out of 5

Submitted Jul 31 at 22:54

This attempt took 22 minutes.

Question 1

0.25 / 0.25 pts

The objective of Cross-Validation techniques is to improve _____ :



ROC Curve



Model evaluation



Quality of Fit



Precision and Recall

Question 2

0.25 / 0.25 pts

Which of the following is not a Classification model metric?



Recall



Precision



R Squared or Adjusted R Square



Confusion Matrix

Question 3

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter.
Continuous, quantitative, ratio.

'The opinion in terms: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree'



Continuous, qualitative, ordinal



Discrete, qualitative, ordinal



Discrete, qualitative, nominal

Question 4

0.25 / 0.25 pts

Network event logs are the example of _____.



Natural language



Graph based



Structured data



Machine generated data

Question 5

0.25 / 0.25 pts

Inconsistencies between data sources or against your company's standardised values is called Interpretation Error.



True



False

Question 6

0.25 / 0.25 pts

Visualization techniques like box plots or scatter plots applied on given data is a good way to identify:

-
- Data Entry errors
-
- Errors against Codebook
-
- Missing values
-
- Outliers

Question 7

0.25 / 0.25 pts

Quasi Structured data is considered to be more Structured than Semi-Structured data.

-
- True
-
- False

Question 8

0.25 / 0.25 pts

Binary attributes are _____ type of attributes.

-
- Discrete
-
- Asymmetric
-
- Normal
-
- Continuous

Question 9

0.25 / 0.25 pts

Suppose that you need to build model for 100000 records. You may partition the dataset into 10 subsets and the model is trained for the first 9 subsets and tested for the 10th subset. The technique followed in the above model is



None of the mentioned



Traditional validation



Root mean square



Cross-Validation

Question 10

0.25 / 0.25 pts

Models trained in one domain can be easily generalized to other domains



True



False

Question 11

0.25 / 0.25 pts

Suppose you need to record the temperature of furnace in a database. This type is an example for _____ attribute



Normal



Discrete



Continuous



Asymmetric

Question 12

0.25 / 0.25 pts

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes
- C) It preserves all details even after aggregation at all times
- D) It does not complement well with statistical analysis



C only



B only



B, D



A , B

Question 13

0.25 / 0.25 pts

Genomic Sequence data is an example of Ordered Data



True



False

Question 14

0.25 / 0.25 pts

When we build models for the training sets in terms of functions say $f(X)$, the estimate for $f(X)$ should not vary too much between training sets. If a method has high variance then small changes in the training data can result in large changes in $f(X)$



True



False

Question 15

0.25 / 0.25 pts

Noise is a distortion of a value or addition of spurious values.



True



False

Question 16

0.25 / 0.25 pts

The Normal distribution is fully specified by the following two parameters



the mean and the upper bound



the mean and the standard deviation



the mean and the lower bound



the mean and the median

Question 17

0.25 / 0.25 pts

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. When data is converted to apply association analysis, what type of attributes would you have and how many?

- 20 asymmetric binary attributes
- 100 symmetric binary attributes
- 100 asymmetric binary attributes
- 20 symmetric binary attributes

Question 18

0.25 / 0.25 pts

Match the following set of job descriptions to their roles:

warehousing the data and defining database architecture

| | |
|----------------|---|
| Data Architect |  |
|----------------|---|

implements, test and maintains infrastructure components

| | |
|---------------|---|
| Data Engineer |  |
|---------------|---|

solves problems using ML and DM

| | |
|----------------|---|
| Data Scientist |  |
|----------------|---|

data collection and interpretation

| | |
|--------------|---|
| Data Analyst |  |
|--------------|---|

Question 19

0.25 / 0.25 pts

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. Can association analysis be used to analyze the results?

- No



Yes

Question 20

0.25 / 0.25 pts

What happens when a statistical method tries to match datapoints more closely or when a more flexible method is used?



the bias reduces, but variance increases



the bias reduces, and variance reduces



the bias increases, and variance increases



the bias increases, but variance reduces

Quiz Score: 5 out of 5

Quiz 2

Due Aug 1 at 19:00 **Points** 5 **Questions** 20
Available Jul 31 at 19:00 - Aug 1 at 19:00 24 hours **Time Limit** 60 Minutes

Instructions

- This Quiz is for **60 minutes** duration.
- Multiple attempts are **not** allowed.
- Choose the correct/most apt options from the given options.
- Correct Answers will be visible later. Please do not discuss answers.

All the Best!

Attempt History

| | Attempt | Time | Score |
|--------|---------------------------|------------|------------|
| LATEST | Attempt 1 | 22 minutes | 5 out of 5 |

❗ Correct answers will be available on Aug 1 at 19:02.

Score for this quiz: **5** out of 5

Submitted Jul 31 at 22:49

This attempt took 22 minutes.

| Question 1 | 0.25 / 0.25 pts |
|------------|-----------------|
| | |

The objective of Cross-Validation techniques is to improve

_____ :

Quality of Fit

Precision and Recall

Model evaluation

ROC Curve

Question 2

0.25 / 0.25 pts

Quasi Structured data is considered to be more Structured than
Semi-Structured data.

True

False

Question 3

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also
classify them as qualitative (nominal or ordinal) or quantitative
(interval or ratio). Some cases may have more than one
interpretation. Eg. Brightness as measured by a light meter.
Continuous, quantitative, ratio.

'Laptop Model Number.'

Discrete, qualitative, nominal

Discrete, qualitative, ordinal

Continuous, qualitative, nominal

Question 4

0.25 / 0.25 pts

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. Can association analysis be used to analyze the results?

No

Yes

Question 5

0.25 / 0.25 pts

Network event logs are the example of _____.

Machine generated data

Structured data

Graph based

Natural language

Question 6**0.25 / 0.25 pts**

Choose the correct statements about association analysis.

-
- Association analysis requires symmetric binary attributes.
-
- In association analysis, only the attribute '1' is important.
-
- For association problems, one binary attribute for each categorical value is introduced.
-
- Association analysis requires asymmetric binary attributes.

Question 7**0.25 / 0.25 pts**

When the model and the data don't match very well.

- A) Get a different model.
- B) Get different data.
- C) Do both or one of them.
- D) Only (A) or (B)

A

D

B

C

Question 8

0.25 / 0.25 pts

Binary attributes are _____ type of attributes.

- Continuous
- Normal
- Discrete
- Asymmetric

Question 9

0.25 / 0.25 pts

Match the following set of job descriptions to their roles:

**warehousing the data
and defining database
architecture**

Data Architect ▾

**implements, test and
maintains infrastructure
components**

Data Engineer ▾

**solves problems using
ML and DM**

Data Scientist ▾

**data collection and
interpretation**

Data Analyst ▾

Question 10

0.25 / 0.25 pts

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. When data is converted to apply association analysis, what type of attributes would you have and how many?

- 100 asymmetric binary attributes
- 20 asymmetric binary attributes
- 20 symmetric binary attributes
- 100 symmetric binary attributes

Question 11

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter.
Continuous, quantitative, ratio.

'Number of students in a school.'

- Discrete, quantitative, ratio
- Discrete, qualitative, nominal
- Continuous, quantitative, ratio

Question 12

0.25 / 0.25 pts

What happens when a statistical method tries to match datapoints

more closely or when a more flexible method is used?

- the bias increases, but variance reduces
- the bias reduces, and variance reduces
- the bias increases, and variance increases
- the bias reduces, but variance increases

Question 13

0.25 / 0.25 pts

Visualization techniques like box plots or scatter plots applied on given data is a good way to identify:

- Missing values
- Outliers
- Data Entry errors
- Errors against Codebook

Question 14

0.25 / 0.25 pts

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes
- C) It preserves all details even after aggregation at all times
- D) It does not complement well with statistical analysis

A , B

B only

C only

B, D

Question 15

0.25 / 0.25 pts

Noise is a distortion of a value or addition of spurious values.

True

False

Question 16

0.25 / 0.25 pts

Models trained in one domain can be easily generalized to other domains

True

False

Question 17

0.25 / 0.25 pts

Linked Webpages is an example of which type of data

Ordered Data

Internet Data

Graph Data

Record Data

Question 18

0.25 / 0.25 pts

Match the following type of data

Eye color

Nominal

Education Level

Ordinal

**Number of planets
around the Sun**

Discrete

The height of a student

Continuous

Question 19

0.25 / 0.25 pts

A statistical model serves which key purposes in a data analysis:

A) Quantitative summary of data.

B) Impose a specific structure on the population from which the data were sampled.

C) Both (a) and (b)

D) None of the above.

C

D

B

A

Question 20

0.25 / 0.25 pts

These days purchasing used books through online store is becoming more popular. Assume that a user logs in to AbeBooks to do the same. Predicting whether he / she makes a purchase of a book is a _____ task.

Association

Reinforcement

Classification

Regression

Quiz Score: **5** out of 5

Quiz 2

Due Aug 1 at 19:00

Points 5

Questions 20

Available Jul 31 at 19:00 - Aug 1 at 19:00 24 hours

Time Limit 60 Minutes

Instructions

- This Quiz is for **60 minutes** duration.
- Multiple attempts are **not** allowed.
- Choose the correct/most apt options from the given options.
- Correct Answers will be visible later. Please do not discuss answers.

All the Best!

Attempt History

| | Attempt | Time | Score |
|--------|---------------------------|------------|------------|
| LATEST | Attempt 1 | 11 minutes | 5 out of 5 |

! Correct answers will be available on Aug 1 at 19:02.

Score for this quiz: 5 out of 5

Submitted Jul 31 at 22:51

This attempt took 11 minutes.

Question 1

0.25 / 0.25 pts

Choose the correct statements about association analysis.

Association analysis requires asymmetric binary attributes.

In association analysis, only the attribute '1' is important.



For association problems, one binary attribute for each categorical value is introduced.

Association analysis requires symmetric binary attributes.

Question 2**0.25 / 0.25 pts**

Noise is a distortion of a value or addition of spurious values.

 True False**Question 3****0.25 / 0.25 pts**

When the model and the data don't match very well.

- A) Get a different model.
- B) Get different data.
- C) Do both or one of them.
- D) Only (A) or (B)

 B D C A**Question 4****0.25 / 0.25 pts**

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes
- C) It preserves all details even after aggregation at all times
- D) It does not complement well with statistical analysis

A , B

B only

C only

B, D

Question 5

0.25 / 0.25 pts

What happens when a statistical method tries to match datapoints more closely or when a more flexible method is used?

the bias increases, but variance reduces

the bias increases, and variance increases

the bias reduces, but variance increases

the bias reduces, and variance reduces

Question 6

0.25 / 0.25 pts

These days purchasing used books through online store is becoming more popular. Assume that a user logs in to AbeBooks to do the same. Predicting whether he / she makes a purchase of a book is a _____ task.

Regression

Classification

Association

Reinforcement

Question 7

0.25 / 0.25 pts

Models trained in one domain can be easily generalized to other domains

True

False

Question 8

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'The opinion in terms: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree'

Discrete, qualitative, nominal

Continuous, qualitative, ordinal

- Discrete, qualitative, ordinal

Question 9**0.25 / 0.25 pts**

Inconsistencies between data sources or against your company's standardised values is called Interpretation Error.

- True

- False

Question 10**0.25 / 0.25 pts**

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'Number of students in a school.'

- Discrete, quantitative, ratio

- Continuous, quantitative, ratio

- Discrete, qualitative, nominal

Question 11**0.25 / 0.25 pts**

Overfitting happens when:

- the model has high bias and low variance
- the model has high bias and high variance
- the model has low bias and low variance
- the model has low bias and high variance

Question 12

0.25 / 0.25 pts

Network event logs are the example of _____.

- Machine generated data
- Graph based
- Natural language
- Structured data

Question 13

0.25 / 0.25 pts

Match the following set of job descriptions to their roles:

warehousing the data and defining database architecture

Data Architect



**implements, test and
maintains infrastructure
components**

Data Engineer



**solves problems using
ML and DM**

Data Scientist



**data collection and
interpretation**

Data Analyst



Question 14

0.25 / 0.25 pts

Genomic Sequence data is an example of Ordered Data

True

False

Question 15

0.25 / 0.25 pts

Quasi Structured data is considered to be more Structured than Semi-Structured data.

True

False

Question 16

0.25 / 0.25 pts

When we build models for the training sets in terms of functions say $f(X)$, the estimate for $f(X)$ should not vary too much between training sets. If a method has high variance then small changes in the training data can result in large changes in $f(X)$

True

False

Question 17

0.25 / 0.25 pts

Visualization techniques like box plots or scatter plots applied on given data is a good way to identify:

Data Entry errors

Outliers

Errors against Codebook

Missing values

Question 18

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'Laptop Model Number.'

Discrete, qualitative, nominal

Discrete, qualitative, ordinal

Continuous, qualitative, nominal

Question 19

0.25 / 0.25 pts

Binary attributes are _____ type of attributes.

Continuous

Asymmetric

Normal

Discrete

Question 20

0.25 / 0.25 pts

Which of the following is not a Classification model metric?

Recall

Confusion Matrix

R Squared or Adjusted R Square

Precision

Quiz Score: 5 out of 5

Quiz 2

Due Aug 1 at 19:00

Points 5

Questions 20

Available Jul 31 at 19:00 - Aug 1 at 19:00 24 hours

Time Limit 60 Minutes

Instructions

- This Quiz is for **60 minutes** duration.
- Multiple attempts are **not** allowed.
- Choose the correct/most apt options from the given options.
- Correct Answers will be visible later. Please do not discuss answers.

All the Best!

Attempt History

| | Attempt | Time | Score |
|--------|---------------------------|-----------|------------|
| LATEST | Attempt 1 | 4 minutes | 5 out of 5 |

❗ Correct answers will be available on Aug 1 at 19:02.

Score for this quiz: 5 out of 5

Submitted Jul 31 at 22:45

This attempt took 4 minutes.

Question 1

0.25 / 0.25 pts

Which of the following is not a Classification model metric?

Recall

Precision

Confusion Matrix

R Squared or Adjusted R Square

Question 2**0.25 / 0.25 pts**

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'The opinion in terms: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree'

-
- Continuous, qualitative, ordinal

 - Discrete, qualitative, nominal

 - Discrete, qualitative, ordinal

Question 3**0.25 / 0.25 pts**

When we build models for the training sets in terms of functions say $f(X)$, the estimate for $f(X)$ should not vary too much between training sets. If a method has high variance then small changes in the training data can result in large changes in $f(X)$

-
- True

 - False

Question 4**0.25 / 0.25 pts**

Quasi Structured data is considered to be more Structured than Semi-Structured data.

True

False

Question 5

0.25 / 0.25 pts

Network event logs are the example of _____.

Natural language

Structured data

Machine generated data

Graph based

Question 6

0.25 / 0.25 pts

The Normal distribution is fully specified by the following two parameters

the mean and the lower bound

the mean and the upper bound

the mean and the median

- the mean and the standard deviation

Question 7

0.25 / 0.25 pts

Suppose you need to record the temperature of furnace in a database.
This type is an example for _____ attribute

- Continuous
- Normal
- Discrete
- Asymmetric

Question 8

0.25 / 0.25 pts

Choose the correct statements about association analysis.

- Association analysis requires symmetric binary attributes.

- In association analysis, only the attribute '1' is important.



For association problems, one binary attribute for each categorical value is introduced.

- Association analysis requires asymmetric binary attributes.

Question 9

0.25 / 0.25 pts

Noise is a distortion of a value or addition of spurious values.

True

False

Question 10

0.25 / 0.25 pts

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes
- C) It preserves all details even after aggregation at all times
- D) It does not complement well with statistical analysis

B, D

B only

A , B

C only

Question 11

0.25 / 0.25 pts

Genomic Sequence data is an example of Ordered Data

True

False

Question 12

0.25 / 0.25 pts

Match the following set of job descriptions to their roles:

warehousing the data and defining database architecture

Data Architect



implements, test and maintains infrastructure components

Data Engineer



solves problems using ML and DM

Data Scientist



data collection and interpretation

Data Analyst



Question 13

0.25 / 0.25 pts

Linked Webpages is an example of which type of data

Internet Data

Ordered Data

Record Data

Graph Data

Question 14**0.25 / 0.25 pts**

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. When data is converted to apply association analysis, what type of attributes would you have and how many?

- 100 symmetric binary attributes
- 100 asymmetric binary attributes
- 20 asymmetric binary attributes
- 20 symmetric binary attributes

Question 15**0.25 / 0.25 pts**

What happens when a statistical method tries to match datapoints more closely or when a more flexible method is used?

- the bias reduces, but variance increases
- the bias increases, and variance increases
- the bias reduces, and variance reduces
- the bias increases, but variance reduces

Question 16

0.25 / 0.25 pts

The objective of Cross-Validation techniques is to improve _____

:

 Model evaluation Precision and Recall Quality of Fit ROC Curve**Question 17**

0.25 / 0.25 pts

Match the following type of data

Eye color

Nominal

**Education Level**

Ordinal

**Number of planets around the Sun**

Discrete

**The height of a student**

Continuous

**Question 18**

0.25 / 0.25 pts

Suppose that you need to build model for 100000 records. You may partition the dataset into 10 subsets and the model is trained for the first 9 subsets and tested for the 10th subset. The technique followed in the above model is

-
- Cross-Validation
 - Traditional validation
 - None of the mentioned
 - Root mean square
-

Question 19

0.25 / 0.25 pts

Models trained in one domain can be easily generalized to other domains

-
- True
 - False
-

Question 20

0.25 / 0.25 pts

Which of the following are signs of a High Variance Machine Learning model?

- P) Noise in dataset
- Q) Overfitting
- R) Underfitting

S) Failure to capture data trends

P and Q

R only

P only

P, Q and S

Quiz Score: **5** out of 5

Quiz 2 Results for SANTOSHI NAYAK

! Correct answers will be available on Aug 1 at 19:02.

Score for this quiz: 5 out of 5

Submitted Jul 31 at 22:29

This attempt took 17 minutes.

Question 1 0.25 / 0.25 pts

Match the following set of job descriptions to their roles:

warehousing the data and defining database architecture

Data Architect

implements, test and maintains infrastructure components

Data Engineer

solves problems using ML and DM

Data Scientist

data collection and interpretation

Data Analyst

Question 2 0.25 / 0.25 pts

Models trained in one domain can be easily generalized to other domains

True

False

Question 3

0.25 / 0.25 pts

Suppose you need to record the temperature of furnace in a database.
This type is an example for _____ attribute

- Normal
- Asymmetric
- Discrete
- Continuous

Question 4

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg.
Brightness as measured by a light meter. Continuous, quantitative, ratio.

'The opinion in terms: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree'

- Discrete, qualitative, ordinal
- Discrete, qualitative, nominal
- Continuous, qualitative, ordinal

Question 5

0.25 / 0.25 pts

Genomic Sequence data is an example of Ordered Data

 True False**Question 6**

0.25 / 0.25 pts

The objective of Cross-Validation techniques is to improve _____ :

 Precision and Recall Model evaluation ROC Curve Quality of Fit**Question 7**

0.25 / 0.25 pts

Inconsistencies between data sources or against your company's standardised values is called Interpretation Error.

True False**Question 8**

0.25 / 0.25 pts

When the model and the data don't match very well.

- A) Get a different model.
- B) Get different data.
- C) Do both or one of them.
- D) Only (A) or (B)

 B C D A**Question 9**

0.25 / 0.25 pts

Overfitting happens when:

- the model has high bias and high variance
- the model has high bias and low variance

- the model has low bias and low variance
- the model has low bias and high variance

Question 10

0.25 / 0.25 pts

A statistical model serves which key purposes in a data analysis:

- A) Quantitative summary of data.
- B) Impose a specific structure on the population from which the data were sampled.
- C) Both (a) and (b)
- D) None of the above.

C

B

D

A

Question 11

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'Laptop Model Number.'

- Discrete, qualitative, ordinal
- Continuous, qualitative, nominal
- Discrete, qualitative, nominal

Question 12

0.25 / 0.25 pts

When we build models for the training sets in terms of functions say $f(X)$, the estimate for $f(X)$ should not vary too much between training sets. If a method has high variance then small changes in the training data can result in large changes in $f(X)$

- True
- False

Question 13

0.25 / 0.25 pts

Quasi Structured data is considered to be more Structured than Semi-Structured data.

- True
- False

Question 14

0.25 / 0.25 pts

What happens when a statistical method tries to match datapoints more closely or when a more flexible method is used?

-
- the bias reduces, but variance increases
 - the bias increases, and variance increases
 - the bias increases, but variance reduces
 - the bias reduces, and variance reduces
-

Question 15

0.25 / 0.25 pts

Suppose that you need to build model for 100000 records. You may partition the dataset into 10 subsets and the model is trained for the first 9 subsets and tested for the 10th subset. The technique followed in the above model is

-
- Traditional validation
 - Root mean square
 - None of the mentioned
 - Cross-Validation
-

Question 16

0.25 / 0.25 pts

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. Can association analysis be used to analyze the results?

Yes

No

Question 17

0.25 / 0.25 pts

Network event logs are the example of _____.

Natural language

Graph based

Structured data

Machine generated data

Question 18

0.25 / 0.25 pts

Choose the correct statements about association analysis.



For association problems, one binary attribute for each categorical value is introduced.

- In association analysis, only the attribute '1' is important.
-
- Association analysis requires asymmetric binary attributes.
-
- Association analysis requires symmetric binary attributes.

Question 19

0.25 / 0.25 pts

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes
- C) It preserves all details even after aggregation at all times
- D) It does not complement well with statistical analysis

B only

A , B

B, D

C only

Question 20

0.25 / 0.25 pts

Which of the following is not a Classification model metric?

Recall

- Precision
- R Squared or Adjusted R Square
- Confusion Matrix

Quiz Score: **5** out of 5

Quiz 2

Due Aug 1 at 19:00

Points 5

Questions 20

Available Jul 31 at 19:00 - Aug 1 at 19:00 24 hours

Time Limit 60 Minutes

Instructions

- This Quiz is for **60 minutes** duration.
- Multiple attempts are **not** allowed.
- Choose the correct/most apt options from the given options.
- Correct Answers will be visible later. Please do not discuss answers.

All the Best!

Attempt History

| | Attempt | Time | Score |
|--------|------------------|------------|------------|
| LATEST | <u>Attempt 1</u> | 13 minutes | 5 out of 5 |

Correct answers will be available on Aug 1 at 19:02.

Score for this quiz: 5 out of 5

Submitted Jul 31 at 22:21

This attempt took 13 minutes.

Question 1 0.25 / 0.25 pts

Which of the following are signs of a High Variance Machine Learning model?

- P) Noise in dataset
- Q) Overfitting
- R) Underfitting
- S) Failure to capture data trends

P only

P, Q and S

P and Q R only**Question 2****0.25 / 0.25 pts**

Quasi Structured data is considered to be more Structured than Semi-Structured data.

 True False**Question 3****0.25 / 0.25 pts**

Network event logs are the example of _____.

 Graph based Machine generated data Structured data Natural language**Question 4****0.25 / 0.25 pts**

Overfitting happens when:

- the model has high bias and low variance
- the model has low bias and high variance
- the model has high bias and high variance
- the model has low bias and low variance

Question 5**0.25 / 0.25 pts**

Inconsistencies between data sources or against your company's standardised values is called Interpretation Error.

- True
- False

Question 6**0.25 / 0.25 pts**

Suppose that you need to build model for 100000 records. You may partition the dataset into 10 subsets and the model is trained for the first 9 subsets and tested for the 10th subset. The technique followed in the above model is

- Root mean square
- None of the mentioned
- Traditional validation
- Cross-Validation

Question 7**0.25 / 0.25 pts**

A statistical model serves which key purposes in a data analysis:

- A) Quantitative summary of data.
- B) Impose a specific structure on the population from which the data were sampled.
- C) Both (a) and (b)
- D) None of the above.

 C

 D

 A

 B**Question 8****0.25 / 0.25 pts**

Binary attributes are _____ type of attributes.

 Continuous

 Asymmetric

 Discrete

 Normal**Question 9****0.25 / 0.25 pts**

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. Can association analysis be used to analyze the results?

No

Yes

Question 10**0.25 / 0.25 pts**

Visualization techniques like box plots or scatter plots applied on given data is a good way to identify:

Data Entry errors

Outliers

Missing values

Errors against Codebook

Question 11**0.25 / 0.25 pts**

Genomic Sequence data is an example of Ordered Data

True

False

Question 12**0.25 / 0.25 pts**

What happens when a statistical method tries to match datapoints more closely or when a more flexible method is used?

-
- the bias reduces, but variance increases
 - the bias increases, and variance increases
 - the bias increases, but variance reduces
 - the bias reduces, and variance reduces
-

Question 13

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg.
Brightness as measured by a light meter. Continuous, quantitative, ratio.

'Laptop Model Number.'

-
- Discrete, qualitative, ordinal
 - Continuous, qualitative, nominal
 - Discrete, qualitative, nominal
-

Question 14

0.25 / 0.25 pts

When we build models for the training sets in terms of functions say $f(X)$, the estimate for $f(X)$ should not vary too much between training sets. If a method has high variance then small changes in the training data can result in large changes in $f(X)$

True False**Question 15****0.25 / 0.25 pts**

Choose the correct statements about association analysis.



For association problems, one binary attribute for each categorical value is introduced.



Association analysis requires asymmetric binary attributes.



Association analysis requires symmetric binary attributes.



In association analysis, only the attribute '1' is important.

Question 16**0.25 / 0.25 pts**

Match the following set of job descriptions to their roles:

warehousing the data and defining database architecture

Data Architect



implements, test and maintains infrastructure components

Data Engineer



solves problems using ML and DM

Data Scientist



data collection and interpretation

Data Analyst

**Question 17****0.25 / 0.25 pts**

Suppose you need to record the temperature of furnace in a database.
This type is an example for _____ attribute

 Discrete Continuous Asymmetric Normal**Question 18****0.25 / 0.25 pts**

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes
- C) It preserves all details even after aggregation at all times
- D) It does not complement well with statistical analysis

 B, D A , B C only B only

Question 19**0.25 / 0.25 pts**

Noise is a distortion of a value or addition of spurious values.

True

False

Question 20**0.25 / 0.25 pts**

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'The opinion in terms: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree'

Continuous, qualitative, ordinal

Discrete, qualitative, ordinal

Discrete, qualitative, nominal

Quiz Score: 5 out of 5

Quiz 2

Due Aug 1 at 7pm

Points 5

Questions 20

Available Jul 31 at 7pm - Aug 1 at 7pm 24 hours

Time Limit 60 Minutes

Instructions

- This Quiz is for **60 minutes** duration.
- Multiple attempts are **not** allowed.
- Choose the correct/most apt options from the given options.
- Correct Answers will be visible later. Please do not discuss answers.

All the Best!

Attempt History

| | Attempt | Time | Score |
|---------------|---------------------------|-------------|--------------|
| LATEST | Attempt 1 | 32 minutes | 4.5 out of 5 |

! Correct answers will be available on Aug 1 at 7:02pm.

Score for this quiz: **4.5** out of 5

Submitted Jul 31 at 10:04pm

This attempt took 32 minutes.

Question 1

0.25 / 0.25 pts

Linked Webpages is an example of which type of data

Graph Data

Internet Data

Record Data

Ordered Data

Question 2**0.25 / 0.25 pts**

Overfitting happens when:

- the model has low bias and low variance
- the model has high bias and high variance
- the model has low bias and high variance
- the model has high bias and low variance

Question 3**0.25 / 0.25 pts**

Visualization techniques like box plots or scatter plots applied on given data is a good way to identify:

- Data Entry errors
- Outliers
- Errors against Codebook
- Missing values

Question 4**0.25 / 0.25 pts**

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes

C) It preserves all details even after aggregation at all times

D) It does not complement well with statistical analysis

B, D

A , B

B only

C only

Question 5

0.25 / 0.25 pts

Network event logs are the example of _____.

Graph based

Natural language

Machine generated data

Structured data

Question 6

0.25 / 0.25 pts

When we build models for the training sets in terms of functions say $f(X)$, the estimate for $f(X)$ should not vary too much between training sets. If a method has high variance then small changes in the training data can result in large changes in $f(X)$

True

False**Question 7****0.25 / 0.25 pts**

The Normal distribution is fully specified by the following two parameters

-
- the mean and the standard deviation
 - the mean and the lower bound
 - the mean and the upper bound
 - the mean and the median
-

Question 8**0.25 / 0.25 pts**

What happens when a statistical method tries to match datapoints more closely or when a more flexible method is used?

-
- the bias reduces, and variance reduces
 - the bias increases, and variance increases
 - the bias reduces, but variance increases
 - the bias increases, but variance reduces
-

Incorrect

Question 9**0 / 0.25 pts**

These days purchasing used books through online store is becoming more popular. Assume that a user logs in to AbeBooks to do the same. Predicting whether he / she makes a purchase of a book is a _____ task.

Association

Reinforcement

Classification

Regression

Incorrect

Question 10

0 / 0.25 pts

The objective of Cross-Validation techniques is to improve _____ :

Precision and Recall

ROC Curve

Quality of Fit

Model evaluation

Question 11

0.25 / 0.25 pts

Noise is a distortion of a value or addition of spurious values.

True

False**Question 12****0.25 / 0.25 pts**

A statistical model serves which key purposes in a data analysis:

- A) Quantitative summary of data.
- B) Impose a specific structure on the population from which the data were sampled.
- C) Both (a) and (b)
- D) None of the above.

 A D B C**Question 13****0.25 / 0.25 pts**

Match the following type of data

Eye color

Nominal

**Education Level**

Ordinal



**Number of planets around
the Sun**

Discrete



The height of a student

Continuous



Question 14

0.25 / 0.25 pts

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. When data is converted to apply association analysis, what type of attributes would you have and how many?

- 100 symmetric binary attributes
- 100 asymmetric binary attributes
- 20 asymmetric binary attributes
- 20 symmetric binary attributes

Question 15

0.25 / 0.25 pts

Match the following set of job descriptions to their roles:

**warehousing the data and
defining database
architecture**

Data Architect



**implements, test and
maintains infrastructure
components**

Data Engineer



solves problems using
ML and DM

Data Scientist



data collection and
interpretation

Data Analyst



Question 16

0.25 / 0.25 pts

Which of the following are signs of a High Variance Machine Learning model?

- P) Noise in dataset
- Q) Overfitting
- R) Underfitting
- S) Failure to capture data trends

P, Q and S

P and Q

P only

R only

Question 17

0.25 / 0.25 pts

Binary attributes are _____ type of attributes.

Continuous

Asymmetric

Discrete Normal**Question 18**

0.25 / 0.25 pts

A survey is carried out to find the aftereffects of Covid. The test contains 20 questions with five possible answers each. Can association analysis be used to analyze the results?

 No Yes**Question 19**

0.25 / 0.25 pts

Genomic Sequence data is an example of Ordered Data

 True False**Question 20**

0.25 / 0.25 pts

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'Laptop Model Number.'

- Continuous, qualitative, nominal
- Discrete, qualitative, ordinal
- Discrete, qualitative, nominal

Quiz Score: **4.5** out of 5

Quiz 2

Due Aug 1 at 19:00

Points 5

Questions 20

Available Jul 31 at 19:00 - Aug 1 at 19:00 24 hours

Time Limit 60 Minutes

Instructions

- This Quiz is for **60 minutes** duration.
- Multiple attempts are **not** allowed.
- Choose the correct/most apt options from the given options.
- Correct Answers will be visible later. Please do not discuss answers.

All the Best!

Attempt History

| | Attempt | Time | Score |
|---------------|---------------------------|-------------|---------------|
| LATEST | Attempt 1 | 53 minutes | 4.92 out of 5 |

! Correct answers will be available on Aug 1 at 19:02.

Score for this quiz: **4.92** out of 5

Submitted Jul 31 at 22:00

This attempt took 53 minutes.

Question 1

0.25 / 0.25 pts

Match the following set of job descriptions to their roles:

warehousing the data and defining database architecture

Data Architect 

implements, test and maintains infrastructure components

Data Engineer 

solves problems using
ML and DM

Data Scientist

data collection and
interpretation

Data Analyst

Question 2

0.25 / 0.25 pts

Overfitting happens when:

- the model has high bias and high variance
- the model has low bias and high variance
- the model has high bias and low variance
- the model has low bias and low variance

Question 3

0.25 / 0.25 pts

Which of the following are signs of a High Variance Machine Learning model?

- P) Noise in dataset
- Q) Overfitting
- R) Underfitting
- S) Failure to capture data trends

- R only
- P, Q and S

P only P and Q**Question 4****0.25 / 0.25 pts**

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'The opinion in terms: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree'

 Discrete, qualitative, ordinal Continuous, qualitative, ordinal Discrete, qualitative, nominal**Question 5****0.25 / 0.25 pts**

Which of the following is/are true about data aggregation?

- A) It provides a high-level view of the data
- B) It can work with both quantitative and qualitative attributes
- C) It preserves all details even after aggregation at all times
- D) It does not complement well with statistical analysis

 B, D

C only A , B B only**Question 6****0.25 / 0.25 pts**

Match the following type of data

Eye color

Nominal

**Education Level**

Ordinal

**Number of planets around the Sun**

Discrete

**The height of a student**

Continuous

**Question 7****0.25 / 0.25 pts**

Quasi Structured data is considered to be more Structured than Semi-Structured data.

 True False

Question 8**0.25 / 0.25 pts**

Which of the following is not a Classification model metric?

- Precision
- R Squared or Adjusted R Square
- Recall
- Confusion Matrix

Question 9**0.25 / 0.25 pts**

When the model and the data don't match very well.

- A) Get a different model.
- B) Get different data.
- C) Do both or one of them.
- D) Only (A) or (B)

- D
- C
- B
- A

Question 10**0.25 / 0.25 pts**

Network event logs are the example of _____.

- Machine generated data
- Graph based
- Structured data
- Natural language

Partial

Question 11

0.17 / 0.25 pts

Choose the correct statements about association analysis.

- In association analysis, only the attribute '1' is important.
- For association problems, one binary attribute for each categorical value is introduced.
- Association analysis requires asymmetric binary attributes.
- Association analysis requires symmetric binary attributes.

Question 12

0.25 / 0.25 pts

Linked Webpages is an example of which type of data

- Record Data
- Ordered Data

Internet Data Graph Data**Question 13****0.25 / 0.25 pts**

Classify the following attribute as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation. Eg. Brightness as measured by a light meter. Continuous, quantitative, ratio.

'Laptop Model Number.'

 Continuous, qualitative, nominal Discrete, qualitative, ordinal Discrete, qualitative, nominal**Question 14****0.25 / 0.25 pts**

Noise is a distortion of a value or addition of spurious values.

 True False**Question 15****0.25 / 0.25 pts**

What happens when a statistical method tries to match datapoints more closely or when a more flexible method is used?

-
- the bias reduces, but variance increases
-
- the bias increases, and variance increases
-
- the bias increases, but variance reduces
-
- the bias reduces, and variance reduces

Question 16**0.25 / 0.25 pts**

The Normal distribution is fully specified by the following two parameters

-
- the mean and the standard deviation
-
- the mean and the lower bound
-
- the mean and the median
-
- the mean and the upper bound

Question 17**0.25 / 0.25 pts**

Genomic Sequence data is an example of Ordered Data

-
- True
-
- False

Question 18**0.25 / 0.25 pts**

Suppose you need to record the temperature of furnace in a database.
This type is an example for _____ attribute

- Asymmetric
- Discrete
- Normal
- Continuous

Question 19**0.25 / 0.25 pts**

Inconsistencies between data sources or against your company's standardised values is called Interpretation Error.

- True
- False

Question 20**0.25 / 0.25 pts**

A statistical model serves which key purposes in a data analysis:

- A) Quantitative summary of data.
- B) Impose a specific structure on the population from which the data were sampled.
- C) Both (a) and (b)

D) None of the above.

B

D

C

A

Quiz Score: **4.92** out of 5