Score: **62%**

No. of questions: **11**

Correct answer: **7**

Incorrect answer: **4**

Show incorrect attempt only ⬤○

Question 1 ✓

1 Mark

Why do we need to combine sparse classes?

A	To reduce the number of categories	✓

B	It is necessary to combine sparse classes before using one-hot encoding

C	One-hot encoding does not work on categories with small frequency

D	None of the above

Correct Answer: A. To reduce the number of categories

It is not necessary to combine sparse classes before using one-hot encoding. Also, one-hot encoding can work on categories with small frequency. Hence, option b and c are not correct. Combining sparse classes helps us to reduce the number of categories.

Question 2 ✓

2 Marks

The scale of features does not affect the performance of distance based models like KNN.

| B | False | ✓ |

Correct Answer: B. False

Distance based models like KNN are affected by the scale of the features.

Question 3 ✓   3 marks

What will be the output of frequency encoding the following list?

["Shanghai", "Mumbai", "Osaka", "Mumbai", "Shanghai", "Cairo", "Mumbai"] ?

| A | [2, 2, 1, 3, 2, 1, 2] |

| B | [2, 3, 1, 3, 2, 1, 3] | ✓ |

| C | [2, 2, 1, 1, 2, 1, 1] |

| D | [2, 2, 1, 2, 1, 1, 1] |

Correct Answer: B. [2, 3, 1, 3, 2, 1, 3]

The frequency of different categories in the above list is:
Shanghai - 2
Mumbai - 3
Osaka - 1
Cairo - 1
Hence, the output will be: [2, 3, 1, 3, 2, 1, 3]

Question 4 ✓   1 Mark

Which of the following are examples of an ordinal variable?

1. Income - High, Average, Low
2. Room Size - Small, Medium, Big
3. Gender - Male, Female
4. City - Bangalore, Mumbai, Jaipur

A	1

B	3 and 4

C	1 and 2	✓

D	1, 2, 3, and 4

Correct Answer: C. 1 and 2
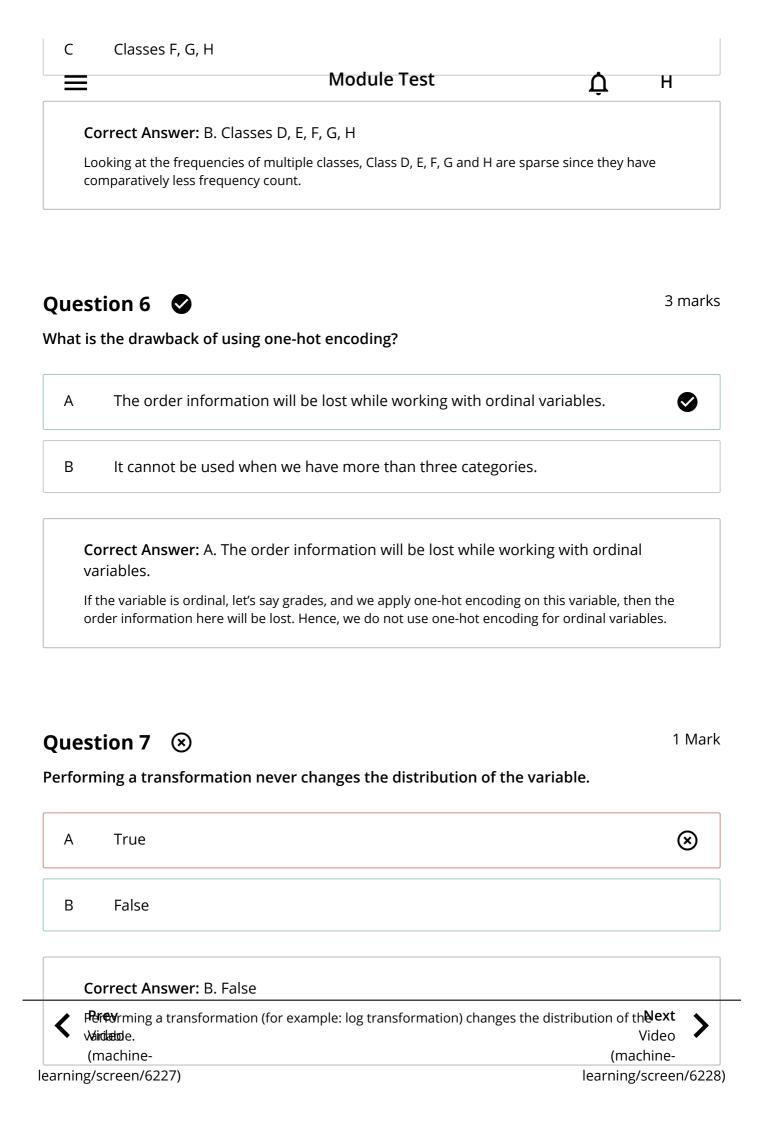
Here, Income and Room size variables have ordered categories and hence,are the only ordinal variables among the given options.

Question 5 ⊗   2 Marks

Identify the sparse classes based on the given frequency against each category?

1. Class A - 98
2. Class B - 72
3. Class C - 67
4. Class D - 12
5. Class E - 9
6. Class F - 2
7. Class G - 2
8. Class H - 1

A	Classes C, D, E, F, G, H	⊗

B	Classes D, E, F, G, H

C   Classes F, G, H

Correct Answer: B. Classes D, E, F, G, H

Looking at the frequencies of multiple classes, Class D, E, F, G and H are sparse since they have comparatively less frequency count.

Question 6 ✓   3 marks

What is the drawback of using one-hot encoding?

| A | The order information will be lost while working with ordinal variables. | ✓ |

| B | It cannot be used when we have more than three categories. | |

Correct Answer: A. The order information will be lost while working with ordinal variables.

If the variable is ordinal, let's say grades, and we apply one-hot encoding on this variable, then the order information here will be lost. Hence, we do not use one-hot encoding for ordinal variables.

Question 7 ⊗   1 Mark

Performing a transformation never changes the distribution of the variable.

| A | True | ⊗ |

| B | False | |

Correct Answer: B. False

Performing a transformation (for example: log transformation) changes the distribution of the variable.

Question 8 ✓

2 Marks

We perform log transformations to remove left skewness from the data.

A	True

B	False	✓

Correct Answer: B. False

Log transformations are used to remove the right skewness from the data.

Question 9 ⊗

3 marks

Multiplying all the values in a column by a constant number will change the distribution of the variable.
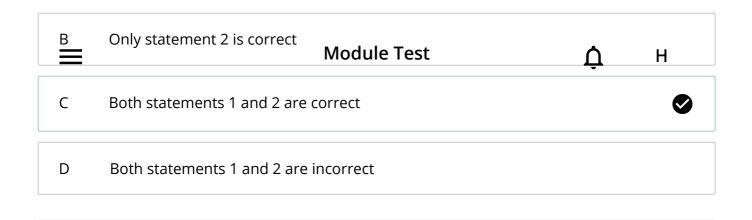
A	True	⊗

B	False

Correct Answer: B. False

The distribution remains the same when each observation is multiplied by a constant number in a column.

Question 10 ✓

1 Mark

Which of the below statements is/are correct regarding the use of feature transformations?

Statement 1: It is used to convert the nonlinear data to linear.
Statement 2: It is used to reduce the skewness of the variable.

A	Only statement 1 is correct

Prev
Video
(machine-learning/screen/6227)

Next
Video
(machine-learning/screen/6228)

C   Both statements 1 and 2 are correct   ✅

D   Both statements 1 and 2 are incorrect

Correct Answer: C. Both statements 1 and 2 are correct

Feature transformations are helpful to convert the non linear data to linear as well as to reduce the skewness of the variable.

Question 11 ⊗   2 Marks

While generating new features using the feature interaction method, we always use two categorical variables.

A   True   ⊗

B   False

Correct Answer: B. False

In feature interaction, we can use categorical as well as continuous variables to create new features.

Prev
Video
(machine-learning/screen/6227)

Next
Video
(machine-learning/screen/6228)