QF632-2025-W1

Number of participants: 41

1. **Which of the following is an example of unstructured data?**

A CSV file containing tabular sales records

0%   0 votes

A JSON file with user demographic information

8%   3 votes

A collection of tweets without a predefined format

93%   37 votes

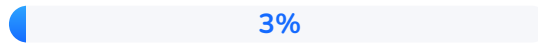A spreadsheet of financial numbers

0%   0 votes

2. **What is the primary distinguishing feature of a dataframe compared to a generic 2D matrix?**

Dataframes can be used only for storing Boolean values.

3% — 1 vote

Dataframes allow labeled rows and columns, enabling mixed data types in different columns.
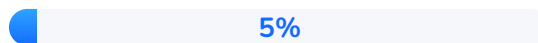
92% — 34 votes

Dataframes store data only in numeric form without any text labels.

0% — 0 votes

Dataframes must be 3D, making them unsuitable for flat data storage.

5% — 2 votes

3. **Which of the following best describes a scenario for using a while loop instead of a for loop?**

You have a finite list of numbers to process in sequence.

0% — 0 votes

You know exactly how many times you need to repeat the loop beforehand.

5% — 2 votes

You want to iterate until a certain condition is no longer satisfied, but you don't know how many iterations it will take in advance.

89% — 34 votes

You need to iterate over an array in a strictly defined order until the array ends.

5% — 2 votes

4. What is the defining characteristic of a recursive function?

It requires an iterative loop inside its body.

33% — 13 votes

It calls itself within its own definition to solve a problem. ✓

64% — 25 votes

It can only be used to solve mathematical problems like factorials.

3% — 1 vote

It is automatically optimized by the compiler to run in parallel.

0% — 0 votes

5. A researcher wants to run time-series analysis on repeated measurements for each subject across multiple time points. Converting the dataset into long format would be advantageous because:

It makes it easier to create pivot tables for summarizing data.

6%

2 votes

Each time point appears in a separate row, allowing easy grouping by time and subject.

84%

26 votes

It automatically encodes categorical variables as integers.

6%

2 votes

It reduces memory usage compared to wide format.

3%

1 vote

6. **What is the primary purpose of performing one-hot encoding on a categorical variable?**

To reduce the dimensionality of the dataset.

0%

0 votes

To represent each category with a separate binary feature (column), thus making it numeric.

97%

33 votes

To convert text data into continuous numeric ranges.

0%

0 votes

To compress continuous variables into a single categorical feature.

3%

1 vote

7. **Which statement best summarizes the difference between one-hot encoding and label encoding?**

25 correct answers out of 31 respondents

One-hot encoding is used for numeric data, while label encoding is used for text data only.
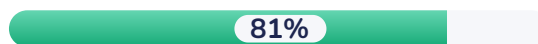
10% — 3 votes

One-hot encoding assigns each category a unique integer, while label encoding uses a binary column for each category.

10% — 3 votes

One-hot encoding creates multiple binary features for each category, whereas label encoding replaces each category with a single integer label.

81% — 25 votes

Both methods always reduce the overall number of

0% — 0 votes

columns needed in the dataset.

8. Why is it important to split the dataset into training, validation, and test sets?

To increase the overall size of the dataset for training

0%

0 votes

To ensure that performance metrics are estimated on unseen data and to reduce overfitting

0%

0 votes

To reduce the computational cost of the training process

0%

0 votes

To merge different datasets for a larger training set

0%

0 votes

9. What does hyperparameter tuning involve in the context of training a supervised model?

Adjusting the internal model weights during training by minimizing the loss function

0% — 0 votes

Searching for the best configuration of settings (like learning rate, regularization strength, or network architecture) that control the training process

0% — 0 votes

Evaluating the final model on the test dataset

0% — 0 votes

Preprocessing the dataset for better performance

0% — 0 votes

10. **Which property is most essential for a loss function to be effectively minimized using gradient descent?**

Boundedness between 0 and 1 — 0% — 0 votes

Differentiability — 0% — 0 votes

Invariance to feature scaling — 0% — 0 votes

Non-convexity — 0% — 0 votes

11. Which loss function is designed to combine the robustness of MAE with the smooth differentiability of MSE, making it less sensitive to outliers?

Log-Cosh Loss   0%   0 votes

Huber Loss   0%   0 votes

Cross-Entropy Loss   0%   0 votes

Quantile Loss   0%   0 votes