

Status	Finished
Started	Sunday, 4 January 2026, 13:41
Completed	Sunday, 4 January 2026, 13:44
Duration	3 mins 2 secs
Marks	6.00/6.00
Grade	100.00 out of 100.00

Question 1

Correct

Mark 1.00 out of 1.00

When splitting a dataset into train and test sets, it's important to:

Select one or more:

- ☐ a. a test dataset is not always necessary
- ☐ b. always pick 80% of the data for training and 20% for testing
- ☐ c. shuffle the dataset and pick samples at random
- ☒ d. depending on the data: either shuffle datapoints or maintain order for timeseries data with temporal dependencies ✓

Your answer is correct.

The correct answer is: depending on the data: either shuffle datapoints or maintain order for timeseries data with temporal dependencies

Question 2

Correct

Mark 1.00 out of 1.00

Data leakage can be avoided by:

Select one or more:

- ☒ a. chaining transformations as part of a pipeline ✓
- ☐ b. scaling the dataset before splitting into train and test sets
- ☐ c. cleaning the data
- ☐ d. outlier detection

Your answer is correct.

The correct answer is: chaining transformations as part of a pipeline

Question 3

Correct

Mark 1.00 out of 1.00

Which of the following is true about sklearn pipelines?

Select one or more:

- ☒ a. Pipelines ensure that the same data transformations are applied consistently during both training and prediction, reducing the risk of data leakage ✓
- ☒ b. Transformers are a type of Estimators. ✓
- ☒ c. Some estimators can also do predictions via predict(). ✓
- ☒ d. All steps but the last one in a pipeline are Transformers. ✓
- ☐ e. You need to call fit and predict for every step in the pipeline.

Your answer is correct.

The correct answers are: All steps but the last one in a pipeline are Transformers., Transformers are a type of Estimators., Some estimators can also do predictions via predict()., Pipelines ensure that the same data transformations are applied consistently during both training and prediction, reducing the risk of data leakage

Question 4

Correct

Mark 1.00 out of 1.00

Which of the following is a key advantage of XGBoost over traditional gradient boosting methods?

Select one or more:

- ☒ a. XGBoost uses approximate histogram computation to speed up threshold finding. ✓
- ☐ b. XGBoost can only work with numerical features and requires all categorical variables to be dropped before training.
- ☐ c. It uses a linear regression model instead of decision trees.
- ☒ d. It employs optimization techniques like parallel processing. ✓

Your answer is correct.

The correct answers are: It employs optimization techniques like parallel processing., XGBoost uses approximate histogram computation to speed up threshold finding.

Question 5

Correct

Mark 1.00 out of 1.00

Which of the following statements about neural networks are true?

Select one or more:

- ☒ a. Backpropagation computes gradients by propagating errors backward from the output layer to the input layer. ✓
- ☐ b. Activation functions must be linear to allow the network to learn complex patterns.
- ☐ c. There is an input neuron for each sample in the data.
- ☒ d. The forward pass involves computing predictions by passing input data through the network layers, applying weights, biases, and activation functions at each layer. ✓

Your answer is correct.

The correct answers are: Backpropagation computes gradients by propagating errors backward from the output layer to the input layer., The forward pass involves computing predictions by passing input data through the network layers, applying weights, biases, and activation functions at each layer.

Question 6

Correct

Mark 1.00 out of 1.00

What are some advantages of fully sharded data parallelism (FSDP) over data parallelism (DP) in distributed learning?

Select one or more:

- ☒ a. FSDP improves memory efficiency by avoiding the redundancy of storing identical copies of model parameters on every GPU. ✓
- ☐ b. FSDP allows each GPU to train on different model architectures simultaneously, enabling ensemble learning within a single training run.
- ☐ c. FSDP eliminates the need for gradient synchronization between GPUs, making it faster than traditional data parallelism for all model sizes.
- ☒ d. FSDP reduces memory consumption per GPU by sharding model parameters, gradients, and optimizer states across devices, enabling training of larger models. ✓

Your answer is correct.

The correct answers are: FSDP reduces memory consumption per GPU by sharding model parameters, gradients, and optimizer states across devices, enabling training of larger models., FSDP improves memory efficiency by avoiding the redundancy of storing identical copies of model parameters on every GPU.