



Your grade: 100%

Your latest: 100% • Your highest: 100% • To pass you need at least 60%. We keep your highest score.

Next item →

1. Which of the following best describes the role of the test set in model validation?

1 / 1 point

- ☐ The test set is used to transform the target variable.
- ☐ The test set is used to optimize hyperparameters.
- ☒ The test set is used to evaluate the model after it has been trained and validated.
- ☐ The test set is used to train the model.

✓ **Correct**

The test set is essential for evaluating model performance after training and validation are complete, ensuring no data snooping or overfitting.

2. What is the benefit of transforming a skewed target variable, such as using a Box-Cox or logarithmic transformation?

1 / 1 point

- ☐ It increases the complexity of the model.
- ☐ It improves the interpretability of the model's predictions.
- ☐ It guarantees better model performance on training data.
- ☒ It makes the model fit the target variable more easily by reducing skewness.

✓ **Correct**

Transforming a skewed target variable makes it more suitable for models that assume a normal distribution, thus improving model fitting.

3. In the context of regularization, what does the lambda parameter control?

1 / 1 point

- ☒ The penalty term's influence on the cost function
- ☐ The number of features in the data set
- ☐ The number of iterations in the training process
- ☐ The learning rate of the model

✓ **Correct**

Lambda is the regularization parameter that controls how much influence the penalty term has in the regularized cost function. It helps balance model fit and complexity.

4. In which scenario is Lasso regression most beneficial?

1 / 1 point

- ☐ When all features are equally important
- ☒ When the data is sparse and has many irrelevant features
- ☐ When the model requires high computational efficiency
- ☐ When the data set is very small

✓ **Correct**

Lasso regression is beneficial in sparse data scenarios where many features are irrelevant. It performs well in feature selection by shrinking some coefficients to zero, eliminating less important features.

5. What is the primary cause of data leakage in machine learning?

1 / 1 point

- ☐ Using outdated information
- ☐ Not scaling the features correctly
- ☒ Using features that would not be available to a deployed model in the real world
- ☐ Using a larger data set for training

✓ **Correct**

Data leakage occurs when your model is trained on data that wouldn't be accessible in a real-world deployment scenario. This leads to misleading performance during training and validation.