

# Beginning Machine Learning Quiz

Name: \_\_\_\_\_

Date: \_\_\_\_\_

## Multiple Choice

1. What is the main purpose of splitting data into training and testing sets?

- A. To make the dataset larger
- B. To evaluate how well the model performs on new, unseen data
- C. To speed up training
- D. To remove outliers

2. Which dataset is used to train the model?

- A. Test set
- B. Validation set
- C. Training set
- D. Feature set

3. What is overfitting?

- A. When the model performs well on training data but poorly on test data
- B. When the model performs poorly on both training and test data
- C. When the model trains too quickly
- D. When the dataset is too small

4. What is the main purpose of cross-validation?

- A. To increase the size of the dataset
- B. To evaluate model performance more reliably
- C. To remove features
- D. To normalize the data

5. In 5-fold cross-validation, how many times is the model trained?

- A. 1
- B. 5
- C. 10
- D. Depends on dataset size

6. Why is feature scaling important?

- A. It increases the number of features
- B. It makes all features have similar ranges
- C. It removes missing values
- D. It makes training slower

7. Which is a common scaling method?

- A. Decision Tree Scaling
- B. Standardization
- C. Random Scaling
- D. Feature Removal

8. What is PCA used for?

- A. Increase features
- B. Reduce features while keeping information
- C. Split data
- D. Scale features

Short Answer

9. Advantage of cross-validation:

10. Why scale before PCA: