Quiz 8 - Model Evaluation

1. A model that generalizes well means that

- The model is overfitting.
- The model does a good job of fitting to the noise in the data.
- **The model performs well on data not used in training.**
- The model performs well on data used to adjust its parameters.

2. What indicates that the model is overfitting?

- High training error and low generalization error
- **Low training error and high generalization error**
- High training error and high generalization error
- Low training error and low generalization error

3. Which method is used to avoid overfitting in decision trees?

- Post-pruning
- None of these
- Pre-pruning
- **Pre-pruning and post-pruning**

4. Which of the following best describes a way to create and use a validation set to avoid overfitting?

- leave-one-out cross-validation
- random sub-sampling
- k-fold cross-validation
- **All of these**

5. Which of the following statements is NOT correct?

- The test set is used to evaluate model performance on new data.
- The validation set is used to determine when to stop training the model.
- The training set is used to adjust the parameters of the model.
- **The test set is used for model selection to avoid overfitting.**

6. How is the accuracy rate calculated?

- Add the number of true positives and the number of false negatives.
- Divide the number of true positives by the number of true negatives.
- **Divide the number of correct predictions by the total number of predictions**
- Subtract the number of correct predictions from the total number of predictions.

7. Which evaluation metrics are commonly used for evaluating the performance of a classification model when there is a class imbalance problem?

- **precision and recall**
- precision and accuracy
- accuracy and error
- precision and error

8. How do you determine the classifier accuracy from the confusion matrix?

- Divide the sum of the diagonal values in the confusion matrix by the sum of the off-diagonal values.
- Divide the sum of all the values in the confusion matrix by the total number of samples.
- **Divide the sum of the diagonal values in the confusion matrix by the total number of samples.**
- Divide the sum of the off-diagonal values in the confusion matrix by the total number of samples.