ML4E Exam 2024

School of Economics, Quaid-i-Azam University, Islamabad

Encircle/Cross the correct answer.

1. **Which of the following is an example of a supervised learning problem?**
   A. Clustering stocks into groups based on historical returns
   B. Identifying whether an email is spam or not given a set of labeled emails
   C. Finding principal components of a set of gene expression measurements
   D. Determining the underlying structure in unlabeled text documents

2. **In a statistical learning context, which of the following best defines 'prediction'?**
   A. Estimating the exact underlying function that generated the data
   B. Inferring relationships among features without any labeled outcomes
   C. Using a model trained on labeled data to estimate an outcome for previously unseen inputs
   D. Testing how well a model's parameters approximate the true parameters

3. **Which of the following is a primary goal of statistical learning?**
   A. To eliminate randomness from data
   B. To identify and model relationships between predictors and response variables
   C. To ensure all predictors have equal importance
   D. To randomly split data without any purpose

4. **A 'qualitative' response variable refers to what kind of output?**
   A. A continuous numerical value
   B. A binary or categorical value
   C. A value that increases linearly with a predictor
   D. A value that is missing or incomplete

5. **Which statement correctly differentiates supervised from unsupervised learning?**
   A. Supervised learning deals only with categorical predictors, while unsupervised handles numeric predictors.
   B. Supervised learning uses known response variables for training, whereas unsupervised learning works with unlabeled data.
   C. Unsupervised learning always performs better than supervised learning.
   D. Unsupervised learning is another name for supervised dimensionality reduction.

6. **What is meant by the "bias-variance tradeoff"?**
   A. It describes how increasing sample size increases both bias and variance.
   B. It is the relationship where reducing model bias often increases model variance, and vice versa.
   C. It states that bias and variance are independent of each other.
   D. It shows that bias equals variance at the optimal model complexity.

7. **Which of the following characterizes a parametric approach to modeling?**
   A. It makes fewer assumptions and tries to get as close to the data as possible.
   B. It involves assuming a functional form for the relationship between predictors and response.
   C. It is always more flexible and less prone to overfitting than non-parametric methods.
   D. It requires no estimation of parameters from the data.

8. **What is the primary drawback of choosing a model that is too flexible?**
   A. It will always have high bias.
   B. It may overfit the training data and perform poorly on new data.
   C. It cannot achieve a low training error.
   D. It leads to a decrease in the variance of the estimates.

9. **In general, as model complexity increases, which of the following tends to decrease?**
   A. Training error
   B. Variance of the model's estimates
   C. Test error
   D. Overfitting likelihood

10. **What is the main purpose of splitting data into training and test sets?**
    A. To ensure the model fits the training data perfectly
    B. To allow tuning parameters until the test set error is minimized
    C. To obtain an unbiased estimate of the model's generalization error
    D. To reduce computational time required for training

11. **Which of the following is the primary goal of linear regression?**
    A. To minimize the sum of squared residuals between observed and predicted values
    B. To maximize the correlation between predictors
    C. To ensure that all predictors have the same coefficient value
    D. To produce classifications rather than predictions

12. **In simple linear regression, which parameter is typically chosen to minimize the residual sum of**

squares (RSS)?
A. The intercept only
B. The slope parameter only
C. Both the intercept and slope parameters
D. None of the parameters; RSS is not relevant

13. **The coefficient of determination (R^2) in a linear regression model:**
A. Ranges from -∞ to +∞
B. Ranges from 0 to 1 and measures the proportion of variance explained by the model
C. Must always be close to 1 for a good model
D. Is unaffected by the scale of the predictors

14. **In multiple linear regression, multicollinearity refers to:**
A. The presence of interaction terms among predictors
B. The presence of highly correlated predictors, which can destabilize coefficient estimates
C. The lack of any correlation among predictors
D. The need for polynomial expansions

15. **Including interaction terms in a linear regression model allows us to:**
A. Reduce the number of predictors needed
B. Test non-linear relationships directly
C. Model situations where the effect of one predictor depends on the value of another predictor
D. Increase multicollinearity intentionally

16. **When adding polynomial terms (e.g., (X^2)) to a linear regression model, we are:**
A. Performing a non-linear regression
B. Converting the model into a classification model
C. Restricting the model to linear relationships only
D. Using linear regression on transformed predictors, thus still considered a linear model

17. **The Residual Standard Error (RSE) measures:**
A. The average magnitude of the residuals
B. The proportion of variance explained by the model
C. The correlation between predictors
D. The slope of the regression line

18. **A high p-value for a predictor's coefficient in a linear regression model typically suggests:**
A. The predictor is definitely important
B. The predictor is likely not statistically significant
C. The model fits the data perfectly
D. The predictor always improves the R^2

19. **In the context of classification, the Bayes classifier assigns an observation to the class for which:**
A. The class name is alphabetically first
B. The posterior probability is highest given the predictor values
C. The prior probability of the class is the lowest
D. The predictor values are closest to the class centroid

20. **Logistic regression models the probability of a binary outcome using:**
A. A linear function directly on the probability
B. A logarithmic function of the probability (log-odds)
C. A polynomial function of the probability
D. A decision tree structure

21. **Linear Discriminant Analysis (LDA) assumes that within each class:**
A. Predictors are distributed according to a multivariate normal distribution with a class-specific mean but the same covariance matrix across classes
B. Predictors have a uniform distribution with identical means
C. Classes have completely different covariance matrices
D. No assumptions are made about the distribution of predictors

22. **One key difference between LDA and QDA is that QDA:**
A. Uses linear boundaries between classes
B. Assumes all classes share the same covariance matrix
C. Allows each class to have its own covariance matrix, resulting in quadratic decision boundaries
D. Is never used for more than two classes

23. **The k-Nearest Neighbors (kNN) classifier:**
A. Uses a linear model to make predictions
B. Determines class membership by looking at the majority class among the k closest observations in the feature space
C. Requires estimating parameters for a parametric model
D. Always outperforms LDA and logistic regression

24. **Which of the following metrics is used to evaluate the performance of a classification model on a binary outcome?**
A. R-squared
B. Residual Standard Error
C. Confusion Matrix
D. Sum of Squared Residuals

25. **An ROC curve is used to:**
A. Show the tradeoff between the true positive rate and false positive rate as a classification threshold varies
B. Evaluate whether the predictors are linearly related to the response
C. Display the residuals of a linear regression model
D. Show the correlation structure between predictors

Short answers

1.

Describe the main difference between supervised and unsupervised learning.

OR

Give an example of a real-world application where supervised learning is appropriate and explain why.

2.
What is the bias-variance tradeoff, and why is it important in model selection?

OR

Briefly explain the concept of model flexibility and how it relates to overfitting.

3.
What does the coefficient of determination (R^2) represent in a linear regression model?
OR
Explain why adding interaction or polynomial terms can improve a linear regression model's fit.

4.
How does Logistic Regression model the probability of a binary outcome, and why is it preferred over linear regression for classification?

OR

Describe how the Linear Discriminant Analysis (LDA) approach differs conceptually from a simple logistic regression model.