

Machine Learning in Economics

School of Economics, QAU, Islamabad

Answer all questions. Show all working for Section B questions.

Section A: Multiple Choice Questions (25 marks)

Select the best answer for each question.

Question 1

Which of the following best describes the purpose of k-fold cross-validation?

- a) To increase the size of the training dataset
- b) To estimate the test error of a statistical learning method
- c) To reduce the variance of predictions
- d) To eliminate the need for a validation set

Question 2

In the bootstrap method, approximately what fraction of observations are NOT included in a given bootstrap sample?

- a) $1/n$
- b) $1/e \approx 0.368$
- c) $1/2$
- d) $1/3$

Question 3

The bias-variance trade-off in k-fold cross-validation suggests that:

- a) Larger k leads to higher bias and lower variance
- b) Larger k leads to lower bias and higher variance
- c) Larger k leads to both lower bias and lower variance
- d) k does not affect bias or variance

Question 4

Leave-one-out cross-validation (LOOCV) is equivalent to:

- a) 2-fold cross-validation
- b) 5-fold cross-validation
- c) n-fold cross-validation
- d) Bootstrap validation

Question 5

Ridge regression addresses multicollinearity by:

- a) Removing highly correlated predictors
- b) Setting some coefficients exactly to zero
- c) Shrinking all coefficients toward zero
- d) Orthogonalizing the predictor matrix

Question 6

Which statement about Lasso regression is TRUE?

- a) It can only shrink coefficients, not set them to zero
- b) It performs variable selection by setting some coefficients to zero
- c) It requires all predictors to be on the same scale
- d) Both b and c are correct

Question 7

In ridge regression, as the tuning parameter λ increases:

- a) Model flexibility increases and training RSS decreases
- b) Model flexibility decreases and training RSS increases
- c) Coefficient estimates approach their least squares values
- d) The bias-variance trade-off does not change

Question 8

Principal Components Regression (PCR) differs from ordinary least squares by:

- a) Using only the first M principal components as predictors
- b) Performing variable selection automatically
- c) Requiring standardization of predictors
- d) Both a and c are correct

Question 9

Which regularization method is more likely to produce a sparse model (many coefficients exactly zero)?

- a) Ridge regression
- b) Lasso regression
- c) Principal Components Regression
- d) Partial Least Squares

Question 10

Elastic Net combines:

- a) Ridge and Lasso penalties
- b) PCR and PLS methods
- c) Forward and backward selection
- d) Bagging and boosting

Question 11

A natural cubic spline with K knots is represented by:

- a) K basis functions
- b) K + 1 basis functions
- c) K + 2 basis functions

d) $K + 4$ basis functions

Question 12

Which method would be most appropriate for modeling a non-linear relationship when you suspect different predictors have different functional forms?

- a) Polynomial regression
- b) Regression splines
- c) Generalized Additive Models (GAMs)
- d) Step functions

Question 13

A key advantage of regression splines over polynomial regression is:

- a) They are always more flexible
- b) They provide better behavior at the boundaries
- c) They require fewer basis functions
- d) They automatically select the degree of the polynomial

Question 14

Smoothing splines minimize:

- a) RSS only
- b) A combination of RSS and a roughness penalty
- c) The number of knots
- d) The prediction error on validation set

Question 15

Local regression (LOESS) assigns weights to observations based on:

- a) Their variance
- b) Their distance from the target point
- c) Their correlation with the response
- d) Their influence on the model

Question 16

In decision tree construction, which measure is NOT typically used for node splitting in classification trees?

- a) Gini index
- b) Cross-entropy
- c) Mean squared error
- d) Classification error rate

Question 17

Bagging (Bootstrap Aggregation) reduces:

- a) Bias only
- b) Variance only
- c) Both bias and variance equally
- d) Neither bias nor variance

Question 18

Random forests improve upon bagging by:

- a) Using a larger number of trees
- b) Randomly selecting a subset of predictors at each split
- c) Using deeper trees

d) Pruning each tree more aggressively

Question 19

In boosting, each new tree is fit to:

- a) The original response values
- b) The residuals from the current model
- c) A bootstrap sample of the data
- d) A random subset of predictors

Question 20

Which tree-based method typically uses shallow trees (stumps)?

- a) Bagging
- b) Random forests
- c) Gradient boosting
- d) Single decision trees

Question 21

The learning rate (shrinkage parameter) in boosting:

- a) Should always be set to 1 for best performance
- b) Controls the contribution of each tree to the ensemble
- c) Determines the depth of each tree
- d) Has no effect on model performance

Question 22

In a neural network, the activation function serves to:

- a) Initialize the weights
- b) Introduce non-linearity into the model
- c) Reduce overfitting
- d) Select important features

Question 23

Dropout is a regularization technique that:

- a) Removes less important input features
- b) Randomly sets a fraction of neurons to zero during training
- c) Reduces the number of hidden layers
- d) Decreases the learning rate over time

Question 24

Convolutional Neural Networks (CNNs) are particularly well-suited for:

- a) Time series forecasting
- b) Text classification
- c) Image recognition
- d) Structured tabular data

Question 25

Recurrent Neural Networks (RNNs) differ from standard neural networks by:

- a) Having connections that form cycles
- b) Using convolutional layers
- c) Requiring less training data
- d) Being limited to binary classification

Section B: Short Questions (25 marks)

Question 1 (8 marks)

Suppose we have a dataset with $n = 100$ observations and $p = 10$ predictors. We want to compare three approaches: least squares regression, ridge regression, and lasso regression.

(a) (3 marks) Explain conceptually why we might prefer ridge or lasso regression over least squares regression in this setting. Discuss the bias-variance trade-off.

(b) (3 marks) Describe a procedure to select the optimal tuning parameter λ for ridge regression using cross-validation. Include specific details about the steps involved.

(c) (2 marks) Suppose after fitting lasso regression with the optimal λ , you find that only 4 of the 10 predictors have non-zero coefficients. How would you interpret this result?

Question 2 (8 marks)

Consider fitting a generalized additive model (GAM) to predict a continuous response Y using three predictors X_1 , X_2 , and X_3 .

(a) (3 marks) Write down the general form of a GAM with these three predictors. Explain what each component represents and why GAMs are useful for economic modeling.

(b) (3 marks) Describe two advantages and one disadvantage of using a GAM compared to a standard linear regression model.

(c) (2 marks) Suppose diagnostic plots suggest that the relationship between Y and X_1 is approximately linear, while the relationships with X_2 and X_3 are highly non-linear. Write the modified model equation that accounts for this.

Question 3 (9 marks)

You are building a classification model to predict customer churn (binary outcome) using a dataset with 20 predictor variables. You are considering using random forests and boosting.

(a) (3 marks) Explain the key algorithmic difference between random forests and boosting. How does each method combine the predictions from multiple trees?

(b) (3 marks) Both methods have tuning parameters that need to be selected. For random forests, identify two key tuning parameters and explain their role. For boosting, identify two key tuning parameters and explain their role.

(c) (3 marks) After fitting both models, you find that the random forest has a test error rate of 12% and the boosting model has a test error rate of 10%, but the boosting model takes much longer to train. Discuss the trade-offs you would consider when choosing between these two models for deployment in a production environment, considering factors like interpretability, computational cost, and business requirements.