Dept of Mathematics and Statistics
King Fahd University of Petroleum & Minerals
AS491: Topics in AS & Fin. Math. 2
Dr. Ridwan A. Sanusi
Mid-Term Exam, Term 252
Thursday, April 2, 2026
3:00 PM – 4:30 PM

Name: _____ ID#: _____

Instructions:

1. Any student caught with mobile phones on during the exam will be considered under the cheating rules of the University.
2. If you need to leave the room, please do so before the exam starts. Nobody will be allowed to leave the room once the exam starts.
3. Only materials provided by the instructor can be present on the table during the exam.
4. Do not spend too much time on any one question. If a question seems too difficult, leave it and go on.
5. Use the blank portions of each page for your work. Extra blank pages can be provided if necessary.
6. While every attempt is made to avoid defective questions, sometimes they do occur. In the rare event that you believe a question is defective, the instructor cannot give you any guidance beyond these instructions.
7. Mobile calculators, I-pad, or communicable devices are disallowed. Use regular scientific calculators, financial calculators, or SOA approved calculators only.

GOOD LUCK, and you may begin now!

Question 1: Which of the following best describes the bias-variance trade-off?

(A) As model complexity increases, bias increases and variance decreases

(B) As model complexity increases, bias decreases and variance increases

(C) As model complexity increases, both bias and variance increase

(D) As model complexity increases, both bias and variance decrease

(E) Bias and variance are unrelated to model complexity

Question 2: A data scientist fits a model with 5 predictors and obtains an R^2 of 0.95 on training data but 0.45 on test data. This is an example of:

(A) Underfitting

(B) Overfitting

(C) The curse of dimensionality

(D) Multicollinearity

(E) Heteroscedasticity

Question 3: Which of the following is NOT a characteristic of non-parametric methods?

(A) They make no explicit assumptions about the form of f

(B) They typically require more observations than parametric methods

(C) They are generally more flexible than parametric methods

(D) They have a fixed number of parameters regardless of sample size

(E) They can fit a wider range of possible shapes for f

Question 4: The irreducible error in a prediction problem comes from:

(A) Using too flexible a model

(B) Using too inflexible a model

(C) The variance of the error term ϵ

(D) The bias of the model

(E) Not having enough training data

Question 5: Which statement about supervised vs unsupervised learning is TRUE?

(A) Supervised learning has no response variable

(B) Unsupervised learning predicts a response variable

(C) Clustering is an example of supervised learning

(D) Linear regression is an example of supervised learning

(E) PCA is an example of supervised learning

Question 6: As the sample size n increases, what happens to the test MSE of a flexible method relative to an inflexible method?

(A) The flexible method always performs worse

(B) The flexible method always performs better

(C) The flexible method may perform better as n increases

(D) Both methods perform identically

(E) The comparison is impossible to make

Question 7: The Bayes error rate is analogous to:

(A) The training error rate

(B) The test error rate

(C) The irreducible error

(D) The reducible error

(E) The classification error rate

Question 8: In KNN classification with $K = 1$, the decision boundary will be:

(A) Very smooth with low variance

(B) Very wiggly with high variance

(C) Linear regardless of data

(D) Always perfectly correct on test data

(E) Independent of the training data

Question 9: You are given a training data set with six observations. The Euclidean distances from a test point $(0, 0, 0)$ to the six observations are: 3.00, 2.00, 3.16, 2.24, 1.41, and 1.73. The corresponding class labels are: Red, Red, Red, Green, Green, Red. Using $K = 3$, what is the predicted class?

(A) Red

(B) Green

(C) Cannot be determined

(D) Both classes equally likely

(E) Neither class

Question 10: Determine which of the following statements is/are true.
I. The number of clusters must be pre-specified for both K-means and hierarchical clustering. II. The K-means clustering algorithm is less sensitive to the presence of outliers than the hierarchical clustering algorithm. III. The K-means clustering algorithm requires random assignments while the hierarchical clustering algorithm does not.

(A) I only

(B) II only

(C) III only

(D) I, II, and III

(E) The correct answer is not given by (A), (B), (C), or (D)

Question 11: In a simple linear regression, the 95% confidence interval for β_1 is approximately:

(A) $\hat{\beta}_1 \pm \text{SE}(\hat{\beta}_1)$

(B) $\hat{\beta}_1 \pm 1.96 \times \text{SE}(\hat{\beta}_1)$

(C) $\hat{\beta}_1 \pm t_{n-2,0.975} \times \text{SE}(\hat{\beta}_1)$

(D) $\hat{\beta}_1 \pm \text{RSE}$

(E) $\hat{\beta}_1 \pm t_{n-2,0.95} \times \text{RSE}$

Question 12: The R^2 statistic measures:

(A) The standard deviation of the residuals

(B) The proportion of variance in Y explained by X

(C) The correlation between X and Y

(D) The p-value for the regression

(E) The number of observations

Question 13: For a simple linear regression with $n = 25$, the residual sum of squares is 230 and $R^2 = 0.64$. The total sum of squares (TSS) is:

(A) 605.94

(B) 638.89

(C) 690.77

(D) 701.59

(E) 750.87

Question 14: The residual standard error (RSE) estimates:

(A) The standard deviation of Y

(B) The standard deviation of the residuals

(C) The standard deviation of X

(D) The standard deviation of $\hat{\beta}_1$

(E) The mean of the residuals

Question 15: Which of the following indicates a potential collinearity problem?

(A) Low R^2 value

(B) Large p-values for all predictors

(C) Large VIF (Variance Inflation Factor) values

(D) Small standard errors

(E) High F-statistic

Question 16: Suppose we have a model: $\hat{Y} = 50 + 20X_1 + 0.07X_2 + 35X_3 + 0.01X_1X_2 - 10X_1X_3$, where $X_3 = 1$ for College, $X_3 = 0$ for High School. For fixed IQ, high school graduates earn more than college graduates provided that:

(A) GPA is less than 3.5

(B) GPA is greater than 3.5

(C) GPA is equal to 3.5

(D) IQ is greater than 100

(E) IQ is less than 100

Question 17: You are given the following results from a regression model:

Observation (i)	y_i	$\hat{f}(x_i)$
1	2	4
2	5	3
3	6	9
4	8	3
5	4	6

Calculate the sum of squared errors (SSE).

(A) -35

(B) -5

(C) 5

(D) 35

(E) 46

Question 18: Determine which of the following statements is/are true for a simple linear relationship, $y = \beta_0 + \beta_1 x + \epsilon$.
I. If the estimated variance of $\epsilon = 0$, the 95% confidence interval is equal to the 95% prediction interval. II. The prediction interval is always at least as wide as the confidence interval. III. The prediction interval quantifies the possible range for $E(y|x)$.

(A) I only

(B) II only

(C) III only

(D) I, II, and III

(E) The correct answer is not given by (A), (B), (C), or (D)

Question 19: Which of the following statements about LOOCV is TRUE?

(A) LOOCV has higher bias than 5-fold CV

(B) LOOCV has lower variance than 5-fold CV

(C) LOOCV requires fitting the model n times

(D) LOOCV always gives the same results as 10-fold CV

(E) LOOCV cannot be used for classification problems

Question 20: In k-fold cross-validation, as k increases from 2 to n:

(A) Bias decreases and variance decreases

(B) Bias decreases and variance increases

(C) Bias increases and variance decreases

(D) Bias increases and variance increases

(E) Both bias and variance remain unchanged

Question 21: A bootstrap sample of size $n = 100$ is drawn with replacement. The probability that a specific observation appears at least once is approximately:

(A) 0.01

(B) 0.37

(C) 0.50

(D) 0.63

(E) 0.99

Question 22: For linear regression, the LOOCV shortcut formula uses:

(A) The residual standard error

(B) The leverage h_i

(C) The F-statistic

(D) The p-value

(E) The R^2 value

Question 23: Which resampling method has the lowest bias?

(A) Validation set approach

(B) 5-fold CV

(C) 10-fold CV

(D) LOOCV

(E) All have equal bias

Question 24: The main disadvantage of LOOCV compared to k-fold CV is:

(A) Higher bias

(B) Higher variance

(C) Cannot be used for classification

(D) Requires more data

(E) Not applicable to linear models

Question 25: The bootstrap is most useful for:

(A) Estimating test error

(B) Estimating standard errors when no formula exists

(C) Reducing model bias

(D) Increasing model flexibility

(E) Feature selection

Question 26: For a bootstrap sample of size $n = 5$, the probability that a specific observation appears in the sample is:

(A) 0.32768

(B) 0.50000

(C) 0.63210

(D) 0.67232

(E) 0.75000

Question 27: Which of the following statements about best subset selection is FALSE?

(A) It requires fitting 2^p models

(B) It is computationally feasible for $p = 50$

(C) The model with all predictors always has smallest training RSS

(D) For logistic regression, deviance is used instead of RSS

(E) Cross-validation can be used to select the final model

Question 28: In ridge regression, as $\lambda \to \infty$:

(A) All coefficient estimates go to zero

(B) All coefficient estimates go to OLS values

(C) The model becomes the full least squares model

(D) Bias decreases to zero

(E) Variance increases

Question 29: Which is TRUE about the lasso compared to ridge regression?

(A) Lasso uses an ℓ_2 penalty

(B) Lasso always includes all predictors

(C) Lasso can produce coefficient estimates exactly zero

(D) Lasso has a circular constraint region

(E) Lasso is equivalent to best subset selection

Question 30: The regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon$ is being investigated. The following maximized log-likelihoods are obtained:

- Using only intercept: -1126.91
- Using intercept, X_1, X_2: -1122.41
- Using all four terms: -1121.91

The null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ is being tested at the 5% significance level using the likelihood ratio test. Which of the following is true?

(A) The test statistic is 1 and the hypothesis cannot be rejected

(B) The test statistic is 9 and the hypothesis cannot be rejected

(C) The test statistic is 10 and the hypothesis cannot be rejected

(D) The test statistic is 9 and the hypothesis should be rejected

(E) The test statistic is 10 and the hypothesis should be rejected

Answer Key

Chapter 2 Answers

Q#	Answer	Brief Explanation
1	B	As complexity increases, bias decreases, variance increases
2	B	High training R^2, low test R^2 = overfitting
3	D	Non-parametric methods have parameters that grow with sample size
4	C	Irreducible error = Var(ϵ), cannot be reduced
5	D	Linear regression predicts a response $\rightarrow$ supervised
6	C	More data helps flexible methods reduce variance
7	C	Bayes error rate = irreducible error for classification
8	B	$K = 1$ gives wiggly boundary, high variance
9	A	Three nearest neighbors: Obs5(Green), Obs6(Red), Obs2(Red) $\rightarrow$ 2 Red, 1 Green
10	C	I and II are false; III is true

Chapter 3 Answers

Q#	Answer	Brief Explanation
11	C	Rule of thumb: $\hat{\beta}_1 \pm 2 \times \text{SE}(\hat{\beta}_1)$
12	B	$R^2 = 1 - \text{RSS/TSS}$ = proportion of variance explained
13	B	TSS = RSS/(1-R^2) = 230/0.36 = 638.89
14	B	RSE estimates standard deviation of ϵ
15	C	VIF >5 or 10 indicates problematic collinearity
16	B	Difference: College - High School = $35 - 10X_1$. Positive when $X_1 > 3.5$
17	E	SSE = $(2-4)^2 + (5-3)^2 + (6-9)^2 + (8-3)^2 + (4-6)^2 = 4+4+9+25+4 = 46$
18	E	I and II are true; III is false (confidence interval quantifies $E(y\|x)$)

Chapter 5 Answers

Q#	Answer	Brief Explanation
19	C	LOOCV fits model n times, once per observation
20	B	As k increases, bias decreases, variance increases
21	D	$1 - (0.99)^{100} \approx 1 - e^{-1} \approx 0.632$
22	B	h_i is leverage; formula: $\frac{y_i - \hat{y}_i}{1 - h_i}$
23	D	LOOCV uses $n - 1$ observations for training
24	B	LOOCV models are highly correlated $\rightarrow$ higher variance
25	B	Bootstrap useful when no formula for SE exists
26	D	$1 - (4/5)^5 = 1 - 0.32768 = 0.67232$

Chapter 6 Answers

Q#	Answer	Brief Explanation
27	B	$2^{50} \approx 1.1 \times 10^{15}$ models - infeasible
28	A	Large λ = large penalty $\rightarrow$ coefficients $\rightarrow 0$
29	C	ℓ_1 penalty allows coefficients exactly zero
30	E	Test statistic $= 2 \times [(-1121.91) - (-1126.91)] = 10$, df=3, critical value=7.81 $\rightarrow$ rejec

Formula Reference Sheet

Concept	Formula
MSE	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$
R^2	$1 - \frac{\text{RSS}}{\text{TSS}}$
t-statistic	$\frac{\hat{\beta}_1}{\text{SE}(\hat{\beta}_1)}$
F-statistic	$\frac{(\text{TSS}-\text{RSS})/p}{\text{RSS}/(n-p-1)}$
LOOCV	$\frac{1}{n}\sum_{i=1}^{n}\text{MSE}_i$
k-fold CV	$\frac{1}{k}\sum_{j=1}^{k}\text{MSE}_j$
Bootstrap Probability	$1-(1-\frac{1}{n})^n \rightarrow 1-e^{-1} \approx 0.632$
C_p	$\frac{1}{n}(\text{RSS} + 2d\hat{\sigma}^2)$
BIC	$\frac{1}{n}(\text{RSS} + \log(n)d\hat{\sigma}^2)$
Adjusted R^2	$1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}/(n-1)}$
Ridge	$\text{RSS} + \lambda\sum\beta_j^2$
Lasso	$\text{RSS} + \lambda\sum\lvert\beta_j\rvert$