Mock Exam Questions - Module 1

Medium Answer

1. In a real estate regression model, the sale price (in thousands of dollars) is predicted by the equation: price = 50 + 8 × size − 5 × bedrooms, where size is measured in hundreds of square feet. If a house is 3 (i.e., 300 sq ft) with 4 bedrooms and actually sells for 75 (thousand dollars), what is the residual (error)?

2. Consider the following dataset of two predictors X1 and X2 and a response Y:

X1	X2	Y
1	4	3
3	2	5
2	5	7

   Using a regularization parameter $\lambda = 2$, compute the Ridge Regression coefficient estimates for this dataset.

3. You are analyzing factors influencing crop yield. The predictors are X1 = Sunlight hours per day, X2 = Soil type (1 for Loamy, 0 for Sandy), X3 = Fertilizer type (1 for Organic, 0 for Synthetic), X4 = Interaction between Sunlight and Soil type, and X5 = Interaction between Soil type and Fertilizer type. The response is the yield per plant in grams. After fitting the model, you obtain $\hat{\beta}_0 = 200, \hat{\beta}_1 = 15, \hat{\beta}_2 = -50, \hat{\beta}_3 = 30, \hat{\beta}_4 = 5, \hat{\beta}_5 = -10$. Predict the yield for a plant that receives 8 hours of sunlight per day, is grown in loamy soil, and is given organic fertilizer.

4. In a few sentences, contrast the process of training a predictive model with the conventional approach of writing explicit code step by step. How do these two methodologies differ?

5. When using multiple linear regression to forecast monthly retail sales from advertising spend, pricing, and seasonality factors, what common shortcomings might you encounter and what strategies can you use to address these limitations?

6. Suppose data are collected for a group of job applicants undergoing an online preparation course, with the following variables: W1 = number of practice interview hours completed, W2 = prior work experience score (on a scale from 1 to 5), and z = receiving a formal job offer (yes or no). A logistic regression model was fitted, yielding estimated coefficients: $\gamma_0 = -4.5, \gamma_1 = 0.06, \gamma_2 = 0.9$. Estimate the probability that an applicant who completes 50 practice hours and has a work experience score of 4 will receive a job offer.

7. Suppose you're developing a credit scoring model; explain how applying LASSO regression helps in selecting the most relevant financial indicators. Provide a concise overview of the method.

8. Consider a dataset D=(x_i, y_i) where x_i denotes the number of tutoring hours per week and y_i denotes the resulting improvement in test scores (in percentage points) after one month. The observed values are: X = $[2, 4, 6, 8, 10]^T$ (hours per week) y = $[5, 9, 14, 18, 25]^T$ (percentage points improvement) Apply simple linear regression to compute the coefficients β_0 (intercept) and β_1 (slope), then use the fitted model to predict the expected score improvement for a student who receives 7 hours of tutoring per week.

9. Consider a dataset $C = \{(h_i, d_i)\}$ for $i = 1, \ldots, M$, where $h_i \in \mathbb{R}^q$ represents borrower characteristics and $d_i \in \{0, 1\}$ indicates whether borrower i defaulted on a loan. Under a logistic model $P(d_i = 1|h_i; \theta) = 1/(1 + \exp(-\theta^T h_i))$, (a) write down the likelihood function of θ for the entire dataset C, and (b) state the corresponding negative log-likelihood (cross-entropy) objective that must be minimized to estimate θ.

10. Suppose one collects data on a group of sales representatives with X1 = number of client meetings, X2 = hours of product training completed, and y = makes at least one sale (1 = yes, 0 = no). A logistic regression fit produces estimated coefficients $\beta_0 = -4, \beta_1 = 0.10, \beta_2 = 0.50$. Estimate the probability that a representative who holds 20 client meetings and completes 10 hours of training makes at least one sale.

11. Imagine you are building a logistic regression model to predict customer churn using 12 candidate predictors. Outline the step-by-step algorithm for the best subset selection technique to identify the optimal subset of features.

12. A data scientist develops a multiple linear regression model to predict monthly apartment rental prices using ten predictor variables and reports an R^2 of 0.76 and an adjusted R^2 of 0.72. Define R^2 and adjusted R^2, include their mathematical formulas, and explain why the adjusted R^2 is typically lower than the R^2.

13. Explain the phenomenon of overfitting in predictive modeling. Describe how introducing a penalty on model parameters can help control it. Finally, write down the objective function for linear regression with an L2 penalty (ridge regression) and explain each term.

14. Consider a dataset $D = \{(u_i, s_i)\}_{i=1}^{M}$, where each $u_i \in \mathbb{R}^d$ represents biometric measurements of a patient and $s_i \in \{0, 1\}$ indicates the presence (1) or absence (0) of a disease. Under a logistic regression model: (a) Write down the likelihood function $L(\theta|D)$ for the observed labels given the feature vectors and parameter θ. (b) Derive the corresponding objective (loss) function that must be minimized to estimate θ.

15. Consider a logistic regression classifier trained to differentiate between spam and non-spam emails. For a new email represented by a feature vector $x \in \mathbb{R}^p$, what exactly does the model output, and how do you convert that output into a predicted class label?

16. In building a retail sales forecasting model with an L1 penalty, how are the regression coefficients estimated? And why might this L1-regularized approach be more advantageous than using an L2 (ridge) penalty?

17. Imagine a scenario where you must fit a regression model using a dataset featuring many highly correlated predictors and limited sample size. What are the benefits of using ridge regression instead of standard least squares regression in this situation?

Long Answer

1. Suppose you want to predict the selling price of used cars based on features such as mileage, engine displacement, and age. (a) Write down the general form of a multiple linear regression model for N cars and k predictors. (b) Derive the normal-equation solution for the coefficient vector $\hat{\beta}$ using least squares. (c) Explain how you would use $\hat{\beta}$ to estimate the price of a new car with feature vector x_0.

2. Suppose we collect data on a group of sales representatives with features X_1 = number of client meetings per month, X_2 = years of experience, and Y = whether they meet their quarterly sales goal (1 = yes, 0 = no). We fit a logistic regression model and obtain coefficient estimates $\beta_0 = -4, \beta_1 = 0.06, \beta_2 = 0.8$. (a) Estimate the probability that a representative with 25 client meetings in a month and 4 years of experience meets the sales goal. (b) For a representative with 4 years of experience, how many client meetings per month are needed to have a 50% chance of meeting the quarterly sales goal?

3. Consider a dataset with five predictors: X1 = hours studied per day, X2 = number of practice problems solved, X3 = tutoring status (1 if the student received tutoring, 0 otherwise), X4 = X1 $\times$ X2, and X5 = X1 $\times$ X3. The response Y is the student's final exam score (out of 100). A least squares regression yields the estimates $\beta0 = 20, \beta1 = 4, \beta2 = 1, \beta3 = -5, \beta4 = 0.05$, and $\beta5 = 2$. (i) For what value of X1 (hours studied per day) will students who received tutoring outperform those who did not on average? (ii) Predict the exam score for a student who received tutoring, studies 5 hours per day, and solves 30 practice problems.

4. In a two-class problem for distinguishing patients with and without a certain disease based on biomarker readings, what key assumption does Linear Discriminant Analysis make about the distribution of these biomarker values? Provide a clear derivation of the LDA decision rule for multiple biomarkers (multivariate case), explicitly showing each step that leads to the final linear discriminant functions.

5. Consider the following multiple linear regression model for predicting a student's exam score:

$$\text{score} = 20 + 5h + 10p + 3s$$

   where score is the predicted exam score, h is the number of hours studied, p is the number of practice tests taken, and s is the hours of sleep the night before the exam. a) Predict the exam score for a student who studied for h = 4 hours, took p = 2 practice tests, and slept s = 6 hours. b) If the student's actual exam score is 90, calculate the residual (prediction error) for this data point. c) Keeping h and s constant, how much would the predicted score change if the student takes 3 additional practice tests?

6. Discuss the key drawbacks of using ordinary multiple linear regression in practice, particularly when predictors are highly correlated or the sample size is small. Then derive the closed-form formula for estimating the coefficient vector in ridge regression, clearly showing how the regularization parameter enters the solution.

7. What is a regression task in machine learning? Using a dataset with feature matrix X and response vector y, derive the closed-form solution (normal equations) for calculating the parameter vector θ in a multiple linear regression model.

8. Given a dataset with feature matrix $Z \in \mathbb{R}^{m\times k}$ and response vector $y \in \mathbb{R}^m$, derive the closed-form expression for the parameter vector β that solves the regularized least squares problem

$$\text{minimize}_\beta \|Z\beta - y\|^2 + \alpha\|\beta\|^2,$$

   where $\alpha > 0$ is the penalty parameter. Then explain how you would use this solution to compute the predicted response at a new feature vector $z^* \in \mathbb{R}^k$.

9. A marketing analyst collects data on customers in an email campaign with predictors X_1 = number of promotional emails received, X_2 = customer loyalty score (on a scale of 1–10), and response y = made a purchase (1 = yes, 0 = no). A logistic regression fit yields estimated coefficients $\beta_0 = -2.5, \beta_1 = 0.12$, and $\beta_2 = 0.6$. (i) Estimate the probability that a customer who has received 20 emails and has a loyalty score of 7 will make a purchase. (ii) For a customer with loyalty score 7, how many promotional emails must they receive to have a 40% chance of making a purchase?

10. Suppose we collect data on recent hires at a technology firm with five predictors: X1 = years of programming experience, X2 = technical assessment score (0–100), X3 = participation in a leadership program (1 if yes, 0 if no), X4 = the interaction between experience and assessment score, and X5 = the interaction between assessment score and leadership participation. The response Y is the year-end bonus (in thousands of dollars). After fitting a least squares regression, we obtain estimates: $\beta_0 = 25, \beta_1 = 12, \beta_2 = 0.15, \beta_3 = 20, \beta_4 = 0.03$, and $\beta_5 = -6$.

11. You're tasked with selecting the most relevant predictors for a house-price regression model from an initial pool of variables (e.g. square footage, lot size, age, number of bedrooms, proximity to schools, etc.). Describe in detail how you would carry out a forward stepwise selection procedure: list the steps you would follow from start to finish, including how you choose which variable to add at each iteration and when you stop. Finally, explain whether this greedy algorithm is guaranteed to identify the absolute best subset of predictors and justify your reasoning.