Machine Learning

Regression

DPP: 1

Q1 The parameters acquired through linear regression:

(A) can take any value in the real space

(B) are strictly integers

(C) always lie in the range [0,1]

(D) can take only non-zero values

Q2 Which of the statements is/are True ?

(A) Ridge has sparsity constraint, and it will drive coefficients with low values to 0.

(B) Lasso has a closed form solution for the optimization problem, but this is not the case for Ridge.

(C) Ridge regression does not reduce the number of variables since it never leads a coefficient to zero but only minimizes it.

(D) If there are two or more highly collinear variables, Lasso will select one of them randomly

Q3 The relation between studying time (in hours) and grade on the final examination (0-100) in a random sample of students in the Introduction to Machine Learning Class was found to be : Grade = 30.5 + 15.2 (h) How will a student's grade be affected if she studies for four hours ?

(A) It will go down by 30.4 points.

(B) It will go down by 30.4 points.

(C) It will go up by 60.8 points.

(D) The grade will remain unchanged.

(E) It cannot be determined from the information given

Q4 Which of the following statements about principal components in Principal Component Regression (PCR) is true?

(A) Principal components are calculated based on the correlation matrix of the original predictors.

(B) The first principal component explains the largest proportion of the variation in the dependent variable.

(C) Principal components are linear combinations of the original predictors that are uncorrelated with each other.

(D) PCR selects the principal components with the highest p-values for inclusion in the regression model.

(E) PCR always results in a lower model complexity compared to ordinary least squares regression.

Q5 Which statement is true about outliers in Linear regression ?

(A) Linear regression model is not sensitive to outliers

(B) Linear regression model is sensitive to outliers

(C) Can't say

(D) None of these

Q6 What does the slope coefficient in a linear regression model indicate?

(A) The point where the regression line intersects the y-axis

(B) The dependent variable changes for every one-unit change in the independent variable

(C) The average value of the dependent variable

(D) The dispersion of the dependent variable

Q7 Find the mean of squared error for the given predications :

Y	F(X)
1	2

2	3
4	5
8	9
16	15
32	31

Hind : Find the squared error for each predication and take the mean of that.

(A) 1 (B) 2

(C) 1.5 (D) 0

Q8 What is the primary assumption of linear regression regarding the relationship between the independent and dependent variables?

(A) Non-linearity

(B) Independence of errors

(C) Homoscedasticity

(D) Linearity

Q9 Which of the following statements is true regarding Partial Least Squares (PLS) regression?

(A) PLS is a dimensionality reduction technique that maximizes the covariance between the predictors and the dependent variable.

(B) PLS is only applicable when there is no multicollinearity among the independent variables.

(C) PLS can handle situations where the number of predictors is larger than the number of observations.

(D) PLS estimates the regression coefficients by minimizing the residual sum of squares.

(E) PLS is based on the assumption of normally distributed residuals.

(F) All of the above.

(G) None of the above.

Q10 The confidence interval is an interval which is an estimate of -

(A) The mean value of the dependent variable

(B) The standard deviation value of the dependent variable

(C) The mean value of the independent variable

(D) The standard deviation value of the independent variable

Q11 In the regression model ($y = a + bx$) where $x = 2.50$, $y = 5.50$ and $a = 1.50$ (x and y denote mean of variables x and y and a is a constant), which one of the following values of parameter 'b' of the model is correct?

(A) 1.75 (B) 1.60

(C) 2.00 (D) 2.50

Q12 There is no value of x that can simultaneously satisfy both the given equations. Therefore, find the 'least squares error' solution to the two equations, i.e ., find the value of x that minimize the sum of squares of the errors in the two equations. ____________.

$2x = 3$

$4x = 1$,

Q13 For a bivariate data set on (x, y), if the means, standard deviations and correlation coefficient are

$x = 1.0$, $y = 2.0$, $sx = 3.0$, $sy = 9.0$, $r = 0.8$

Then the regression line of y on x is:

(A) $y = 1 + 2.4(x-1)$

(B) $y = 2 + 0.27(x - 1)$

(C) $y = 2 + 2.4(x-1)$

(D) $y = 1 + 0.27(x-2)$

Q14 What is the purpose of regularization in linear regression?

(A) To make the model more complex

(B) To avoid underfitting

(C) To encourage overfitting

(D) To reduce the complexity of the model

Q15 A set of observations of independent variable (x) and the corresponding dependent variable (y) is given below :

X	5	2	4	3
Y	16	10	13	12

Based on the data, the coefficient a of the linear regression model.

y = a + bx is estimated as 6.1

The coefficient b is ___________ .(round off to one decimal place)

Q16 The purpose of using a dummy variable in the regression model is -

(A) Some of the independent variables are categorical data

(B) The dependent variable is categorical data

(C) Both independent and dependent variable may have a categorical data value

(D) The dependent and independent variable must be a numerical data value

Q17 The random error e in multiple linear regression model y = Xβ + e are assumed to be identically and independently distributed following the normal distribution with zero mean and constant variance. Here y is a n × 1 vectors of observations on response variable, X is a n × K matrix of n observations on each of the K explanatory variables, β is a K × 1 vectors of regression coefficients and e is a n × 1 vectors of random errors. The residuals $\hat{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}}$ based on the ordinary least squares estimator of b have , in general.

(A) Zero mean, constant variance and are independent

(B) Zero mean, constant variance and are not independent

(C) Zero mean, non constant variance and are not independent

(D) non Zero mean, non constant variance and are not independent

Q18 The linear regression model y=a0+a1x1+a2x2+...+apxp is to be fitted to a set of N training data points having p attributes each. Let X be Nx(p+1) vectors of input values (augmented by 1's), Y be Nx1 vector of target values, and theta (0) be (p+1)×1 vector of parameter values (a0, a1, a2,...,ap). If the sum squared error is minimized for obtaining the optimal regression model, which of the following equation holds ?

(A) XTX=XY (B) XO=XTY

(C) XTXO=Y (D) XTX0=XTY

Q19 Use the regression equation to predict the glucose level given the age. Consider the following is the data set for understanding the concept of Linear Regression Numerical Example with One Independent Variable.

SUBJECT	AGE X	GLUCOSE LEVEL Y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81
7	55	?

Q20 In linear regression, what is the primary difference between Lasso (L1 regularization) and Ridge (L2 regularization)?

(A) Lasso tends to produce sparse coefficient vectors, while Ridge does not.

(B) Ridge tends to produce sparse coefficient vectors, while Lasso does not.

(C) Both Lasso and Ridge produce sparse coefficient vectors.

(D) Both Lasso and Ridge tend to produce non-sparse coefficient vectors.

Q21 Consider the following set of points: {(-2 , -1) , (1 , 1) , (3 , 2)}

(a) Find the least square regression line for the given data points.

(b) Plot the given points and the regression line in the same rectangular system of axes.

Q22 In the table below, the xi column shows scores on the aptitude test. Similarly, the yi column shows statistics grades. The last two columns show deviations scores - the difference between the student's score and the average score on each measurement. The last two rows show sums and mean scores.

Find the regression equation

Student	xi	yi	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	95	85	289	64
2	85	95	49	324
3	80	70	4	49
4	70	65	64	144
5	60	70	324	49
Sum	390	385	730	630
Mean	78	77		

Q23 What does $(x^{(5)}, y^{(5)})$ represent or imply ?

(A) There are 5 training examples

(B) The values of x and y are 5

(C) The fourth training examples

(D) The fifth training example.

Q24 The values of y and their corresponding values of y are shown in the table below.

X	0	1	2	3	4
y	2	3	5	4	6

(a) Find the least square regression line y = a x + b.

(b) Estimate the value of y when x = 10.

Q25 In the context of linear regression, what is the purpose of the F-test?

(A) To determine the significance of individual coefficients.

(B) To test the overall significance of the regression model.

(C) To assess the presence of multicollinearity among independent variables.

(D) To evaluate the normality of residuals.

Q26 When performing linear regression, multicollinearity can be problematic. Which of the following statements about multicollinearity is true?

(A) Multicollinearity occurs when there is no correlation between independent variables.

(B) Multicollinearity makes it easier to interpret the individual coefficients in the regression model.

(C) Multicollinearity inflates the standard errors of the regression coefficients.

(D) Multicollinearity always improves the predictive performance of the model.

Q27 What is heteroscedasticity, and how does it affect the assumptions of linear regression?

(A) Heteroscedasticity refers to the presence of outliers in the dataset, violating the assumption of linearity.

(B) Heteroscedasticity refers to the non-constant variance of residuals, violating the assumption of homoscedasticity.

(C) Heteroscedasticity occurs when there is perfect multicollinearity among independent variables, violating the assumption of independence.

(D) Heteroscedasticity refers to the presence of correlated errors, violating the assumption of normality.

Q28 When should one prefer ridge regression over lasso regression?

(A) When the goal is to select a subset of important predictors.

(B) When the coefficients of irrelevant predictors should be exactly zero.

(C)

When there is multicollinearity among the independent variables.

(D) When the dataset has a large number of observations.

Answer Key

Q1	(C)
Q2	(A)
Q3	(C)
Q4	(C)
Q5	(B)
Q6	(B)
Q7	(A)
Q8	(D)
Q9	(C)
Q10	(A)
Q11	(B)
Q12	(5)
Q13	(C)
Q14	(B)
Q15	(1.9)
Q16	(A)
Q17	(C)
Q18	(D)
Q19	86.327
Q20	(A)
Q21	0.
Q22	0.644x.
Q23	(D)
Q24	11.2
Q25	(B)
Q26	(C)
Q27	(B)
Q28	(C)

Hints & Solutions

Q1 Text Solution:

1. "The parameters obtained in linear regression can take any value in the real space: This statement is True. In linear regression, the model's parameters (coefficients) can take any real value, including positive, negative, or zero.
2. "Are strictly integers": This statement is not true in the context of typical linear regression. In linear regression, the parameters are usually real numbers, not strictly integers. However, there are cases where specialized variants of linear regression, such as integer linear regression or integer programming, exist and can deal with parameters that are required to be integers.
3. "Always lie in the range [0,1]": This statement is not true. In standard linear regression, the parameters are not constrained to lie within the range [0, 1]. They can take values anywhere on the real number line.
4. "Can take only non-zero values": This statement is not true. In linear regression, the parameters can take any real value, including zero.

Q2 Text Solution:

True. Ridge regression adds an L2 regularization term to the linear regression cost function, which imposes a penalty on the magnitude of the coefficients. As the regularization strength increases, the coefficients with low values tend to be driven closer to zero, effectively introducing sparsity in the model.

2. Lasso has a closed form solution for the optimization problem, but this is not the case for Ridge. True. Lasso regression adds an L1 regularization term to the linear regression cost function. The L1 regularization introduces sparsity and often leads to some coefficients being exactly equal to zero. Due to the nature of the L1 regularization term, the optimization problem has a closed-form solution, which allows for an efficient and direct calculation of the coefficients.

3. Ridge regression does not reduce the number of variables since it never leads a coefficient to zero but only minimizes it.

True. Ridge regression tends to shrink the coefficients towards zero but does not lead them exactly to zero (except in cases where the predictors are perfectly collinear). Consequently, Ridge does not perform variable selection or reduce the number of variables, as all the predictors remain in the model, albeit with smaller weights.

4. If there are two or more highly collinear variables, Lasso will select one of them randomly.

True. In situations of high collinearity, Lasso regularization may randomly select one of the correlated variables to include in the model while driving the coefficients of others to exactly zero. The choice of which variable is kept and which ones are eliminated may vary depending on the algorithm or software implementation used.

In summary, all four statements are true.

Q3 Text Solution:

To calculate how a student's grade will be affected if she studies for four hours, we can use the given regression equation:

Grade 30.5+15.2(h)

where "h" is the studying time in hours.

To find the grade for studying four hours (h = 4):

Grade 30.5+15.2(4)

Grade = 30.5+60.8

Grade = 91.3

So, if the student studies for four hours, her grade will be 91.3 points.

The correct answer is: "It will go up by 60.8 points."

Q4 Text Solution:

The true statement about principal components in Principal Component Regression (PCR) is:

Principal components are linear combinations of the original predictors that are uncorrelated with each other.

1. Principal components are calculated based on the correlation matrix of the original predictors.

False. Principal components are calculated based on the covariance matrix (or equivalently, the correlation matrix after standardization) of the original predictors, not the correlation matrix directly.

2. The first principal component explains the largest proportion of the variation in the dependent variable. False. The first principal component explains the largest proportion of the variation in the predictors, not the dependent variable. It captures the direction of maximum variance in the predictor space.

3. Principal components are linear combinations of the original predictors that are uncorrelated with each other.

True. Principal components are linear combinations of the original predictors that are constructed in such a way that they are uncorrelated with each other. Each principal component represents a unique orthogonal direction in the predictor space.

4. PCR selects the principal components with the highest p-values for inclusion in the regression model.

False. PCR does not involve p-values or hypothesis testing. It is a dimensionality reduction technique that aims to reduce multicollinearity and model complexity by selecting a subset of the principal components that capture most of the variance in the predictors.

5. PCR always results in a lower model complexity compared to ordinary least squares regression.

False. PCR can result in a lower model complexity compared to ordinary least squares (OLS) regression when a small number of principal components are retained. However, if all principal components are used in PCR, the model complexity can be similar to the full OLS regression model.

Q5 Text Solution:

The slope of the regression line will change due to outliers in most of thecases.

Q6 Text Solution:

In a linear regression model, the slope coefficient represents the rate of change in the dependent variable (Y) for each one-unit change in the independent variable (X). Specifically, it indicates how much the predicted value of the dependent variable changes for every one-unit increase (or decrease) in the independent variable, holding all other variables constant.

Q7 Text Solution:

Calculate the squared error for each prediction, which is the square of the difference between each predicted value (F(x)) and the corresponding true value (y).

Given predictions : y = [1, 2, 4, 8, 16 , 32] F(x) = [2 , 3 , 5 , 9 ,15 , 31]

Squared error for each prediction :

Prediction 1 : (2 - 1) ∧2 = 1

Prediction 2 : $(3-2) \wedge 2=1$

Prediction 3 : $(5-4) \wedge 2=1$

Prediction 4 : $(9-8) \wedge 2=1$

Prediction 5 : $(15-16) \wedge 2=1$

Prediction 6 : $(31-32) \wedge 2=1$

2. Calculate the mean of squared error by taking the sum of squared errors and dividing by the number of predictions (samples).

Mean squared error (MSE) = (Squared error 1 + Squared error 2 +Squared error 3 + Squared error 4 + Squared error 5 +Squared error 6) / 6

Mean squared error (MSE) = (1 + 1 + 1 + 1 + 1+ 1) / 6 = 6 / 6 = 1.

So, the mean squared error for the given predicitions is 1.

Q8 Text Solution:

The primary assumption of linear regression is that there exists a linear relationship between the independent variables (predictors) and the dependent variable (outcome). This means that the change in the dependent variable is proportional to the change in the independent variables. The model assumes that the relationship can be described by a straight line.

Q9 Text Solution:

PLS is a dimensionality reduction technique that maximizes the covariance between the predictors and the dependent variable.

True. PLS aims to find a low-dimensional latent space that maximizes the covariance between the predictors (independent variables) and the dependent variable while considering their relationship.

(2). PLS is only applicable when there is no multicollinearity among the independent variables.

False. Unlike traditional multiple linear regression, PLS can handle multicollinearity among the independent variables. It deals with multicollinearity by creating latent variables (components) that are linear combinations of the original predictors.

(3). PLS can handle situations where the number of predictors is larger than the number of observations.

True. PLS is particularly useful when dealing with high-dimensional datasets, where the number of predictors (independent variables) is larger than the number of observations. It can effectively reduce the dimensionality and handle the "small n, large p" problem.

(4)PLS estimates the regression coefficients by minimizing the residual sum of squares.

False. PLS estimates the regression coefficients by maximizing the covariance between the predictors and the dependent variable, not by minimizing the residual sum of squares as in ordinary least squares (OLS) regression.

(5) PLS is based on the assumption of normally distributed residuals.

False. PLS does not assume normally distributed residuals. It is a non-parametric method and makes fewer assumptions about the underlying data distribution compared to linear regression.

Q10 Text Solution:

(A) The mean value of the dependent variable: This is correct. The confidence interval estimates the range within which the true mean of the dependent variable is likely to fall.

(B) The standard deviation value of the dependent variable: This is incorrect. The confidence interval is not typically used to estimate the standard deviation of the dependent variable. Instead, it estimates the mean of the dependent variable.

(C) The mean value of the independent variable: This is incorrect. The confidence interval is related to estimating the mean or other parameters of the dependent variable, not the independent variable.

(D) The standard deviation value of the independent variable: This is incorrect for the same reason as option B. The confidence interval is not typically used to estimate the standard deviation of the independent variable.

Q11 Text Solution:

(y = a + bx)

where,

x = 2.50

y = 5.50

a = 1.50

(x and y denote mean of variables x and y and a is a constant)

Putting values in the formula:

5.50 =1.50+ b*2.50

b*2.50 = 4

b= 4/2.5 =1.60

Therefore, 'B' is the correct answer.

Q12 Text Solution:

Given the functions are : $2x = 3 \Rightarrow 2x - 3 = 0$ and $4x = 1 \Rightarrow 4x - 1 = 0$

$\therefore \ R = (2x - 3)^2 + (4x - 1)^2$

Hence, to minimize the value of R, $\frac{dR}{dx} = 0$

$\therefore \ \frac{dR}{dx} = 2 \times 2\,(2x - 3) + 4 \times 2\,(4x - 1) = 0$

$\therefore \ x = \frac{1}{2}\&R_{min} = \left(2 \times \frac{1}{2} - 3\right)^2 + \left(4 \times \frac{1}{2} - 1\right)^2 = 5$

$\therefore$ The value of x that minimizes the sum of squares of the errors in the two equation is 1/2.

Q13 Text Solution:

According to the question

y - 2 = 0.8 x 9 (x-1)/3

⇒ y - 2 2.4(x - 1)

y = 2 + 2.4(x - 1)

Q14 Text Solution:

The purpose of regularization in linear regression is to prevent underfitting by allowing the model to capture more complex relationships in the data while still avoiding overfitting. Regularization achieves this by penalizing overly complex models and encouraging simpler models that generalize better to new data.

Q15 Text Solution:

Given Data and Calculation :

x	y	x^2	xy
5	16	25	80
2	10	4	20
4	13	16	52
3	12	9	36
Σ x = 14	Σ y = 51	Σx^2 = 54	Σ xy = 188

n = 4 So

51 = 4a + 14b

188 = 14a + 54 b

Solving the above two equations a = 6.1 and b = 1.9.

Q16 Text Solution:

This is the correct purpose of using a dummy variable in a regression model. When you have categorical variables (variables that represent categories or groups), such as gender, ethnicity, or geographic region, in your dataset, you need to convert them into a numerical format to include them in a regression model. Dummy variables are created to represent different categories of the categorical variable.

Q17 Text Solution:

$\hat{\varepsilon} = y - \widehat{y}$

= $y - Xb$ where $b = (X'X)^{-1}X'y$

= $(I - H)y$ where $H = X(X'X)^{-1}X'$

= $(I - H)\,\varepsilon$

$E\,(\hat{\varepsilon}) = 0$

$V\,(\hat{\varepsilon}) = \sigma^2\,(I - H)$

Since $E\,(\hat{\varepsilon}) = 0$, $\hat{\varepsilon}_i$'s have zero mean.

Since I − H is not generally a diagonal matrix , So $\hat{\varepsilon}_i$'s do not have necessarily the same variances.

The off - diagonal elements in (I − H) are not zero , in general. So $\hat{\varepsilon}_i$'s are not independent.

Q18 Text Solution:

0 minimizes the sum of squared errors and obtain the optimal linear regression model, we need to solve for the parameter vector 0. The equation that holds in this context is:

XTXO = XTY

Where

X is the Nx(p+1) matrix of input values (augmented by 1's) with N data points and p attributes each.

Y is the Nx1 vector of target values (the dependent variable).

O(thetha) is the (p+1)x1 vector of parameter values (a0, a1, a2, ..., ap).

To understand why this equation holds, let's briefly describe the steps of linear regression. The goal of linear regression is to find the parameter vector 0 that minimizes the sum of squared errors (SSE). The SSE is given by:

SSE(0) = (Y - XO)T (Y - XQ)

where (Y-XO) is the vector of residuals (the difference between the actual target values Y and the predicted values XO).

To find the optimal 0 that minimizes SSE, we take the derivative of SSE with respect to 0 and set it to zero. The solution for 8 that satisfies this condition is:

0 = (XTX)^(-1)XTY

Substituting this value of 0 back into the SSE equation, we get:

SSE(0) = (Y - XO)T (Y - X8) SSE(0) = (Y - X(XTX)^(-1)XTY)T (Y - X(XTX)^(-1)XTY) SSE(0) = (Y - Xe)T (Y- xe)

Q19 Text Solution:

$$b_0 = \frac{(\sum y)(\sum x^2)-(\sum x)(\sum xy)}{n(\sum x^2)-(\sum x)^2}$$

$$b_1 = \frac{n(\sum xy)-(\sum x)(\sum xy)}{n(\sum x^2)-(\sum x)^2}$$

SUBJECT	AGE X	GLUCOSE LEVEL Y	XY	X^2	Y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
E	247	486	20485	11409	40022

Find b_0 :

$$b_0 = \frac{(\sum y)(\sum x^2)-(\sum x)(\sum xy)}{n(\sum x^2)-(\sum x)^2}$$

$$b_0 = \frac{(486)(11409)-(247)(20485)}{6(11409)-(247)^2}$$

$$b_0 = \frac{4848979}{7445} = 65.14$$

Find b_1 :

$$b_1 = \frac{n(\sum xy)-(\sum x)(\sum xy)}{n(\sum x^2)-(\sum x)^2}$$

$$b_1 = \frac{6(20485)-(247)(486)}{6(11409)-(247)^2}$$

$$b_1 = \frac{2868}{7445} = 0.385335$$

Insert the values into the equation.

y' =bo +b1 * x

y' = 65.14 + (0.385225 * x)

Prediction – the value of y for the given value of x = 55

y' = 65.14 +(0.385225 ∗55)

y' = 86.327

Q20 Text Solution:

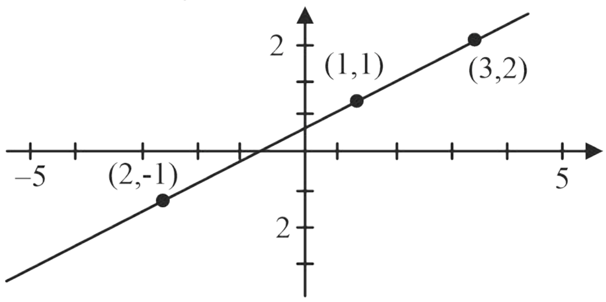
Lasso tends to produce sparse coefficient vectors, while Ridge does not. Lasso regularization includes an L1 penalty term that encourages some coefficients to be exactly zero, leading to sparsity in the coefficient vector, while Ridge regularization (L2 penalty) does not force coefficients to be exactly zero.

Q21 Text Solution:

x	y	xy	x^2
-2	-1	2	4
1	1	1	1
3	2	6	9
?x = 2	?y = 2	?xy = 2	$?x^2$ = 14

A = 23/38 b = 5/19

(b) now graph the regression line given by y = a x + b and the given points.



Q22 Text Solution:

Student	xi	yi	$(x_i - \bar{x})$ $(y_i - \bar{y})$

1	95	85	136
2	85	95	126
3	80	70	- 14
4	70	65	96
5	60	70	126
Sum	390	385	470
Mean	78	77	

The regression equation is a linear equation of the form ; $\widehat{y} = b_0 + b_1 x$ To contact a regression analysis, we need to solve for b_0 and b_1 , Computations

First , we solve for the regression coefficient (b1);

$$b_1 = \sum[(x_i - \bar{x})(y_i - \bar{y})] / \sum\left[(x_i - \bar{x})^2\right]$$

b_1 = 470 /730

b_1 = 0.644

Once we know the value of the regression coefficient (b1), we can solve for the regression slope (b0) :

$$b_0 = \bar{y} - b_1 * \bar{x}$$

b_0 = 77 - (0.644) (78)

b_0 = 77 - (0.644) (78)

b_0 = (26.768)

Therefore , the regression equations is $\widehat{y}$ = 26.768 + 0.644 x.

Q23 Text Solution:

In a linear regression model, the set $(x^{(i)}, y^{(i)})$ represents the i^{th} example in the training set. $x^{(i)}$ gives the value of i^{th} x, $y^{(i)}$ gives the i^{th} value of y.

Q24 Text Solution:

x	y	xy	x^2
0	2	0	0
1	3	3	1
2	5	10	4
3	4	12	9
4	6	24	16
?x = 10	?y = 20	?xy = 49	?x^2 = 30

We now calculate a and b using the least square regression formulas for a and b.

A = 0.9 b = 2.2

(b) Now that we have the least square regression line y = 0.9 x + 2.2, substitute x by 10 to find the value of the corresponding y.

y = 0.9 * 10 + 2.2 = 11.2

Q25 Text Solution:

To test the overall significance of the regression model. The F-test is used to determine whether the regression model as a whole is statistically significant. It compares the variance explained by the model to the variance not explained by the model.

Q26 Text Solution:

Multicollinearity inflates the standard errors of the regression coefficients. Multicollinearity refers to the situation where independent variables in a regression model are highly correlated with each other. It can cause instability in the estimation of coefficients and inflate their standard errors, making the interpretation of individual coefficients less reliable.

Q27 Text Solution:

Heteroscedasticity refers to the non-constant variance of residuals, violating the assumption of homoscedasticity. In linear regression, homoscedasticity assumes that the variance of the residuals is constant across all levels of the independent variables. Heteroscedasticity violates this assumption by causing the variance of residuals to vary systematically with the independent variables.

Q28 Text Solution:

When there is multicollinearity among the independent variables. Ridge regression is particularly useful when multicollinearity is present among the independent variables because it shrinks the coefficients towards zero without setting them exactly to zero. Lasso regression, on the other hand, can set coefficients exactly to zero, which may not be desirable when multicollinearity is present.

Therefore, ridge regression is preferred in such situations.