1) Linear Regression is a supervised machine learning algorithm.

A. True
B. False

2) Regression is used in:

A. Predicting independent variables using dependent variables
B. Predicting dependent variables using independent variables
C. Predicting independence of independent variables
D. Predicting dependent coefficients

3) Suppose that we have N independent variables (X1, X2… Xn) and the dependent variable is Y. Now Imagine that you are applying linear regression by fitting the best fit line using least square error on this data.
You found that the correlation coefficient for one of its variables (Say X1) with Y is -0.95.

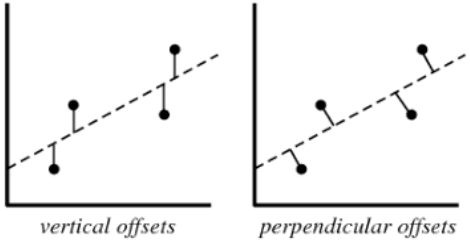
Which of the following is true for X1?

A. Relation between the X1 and Y is weak
B. Relation between the X1 and Y is strong
C. Relation between the X1 and Y is neutral
D. Correlation can’t judge the relationship

4)The linear regression model $y = a_0 + a_1x_1 + a_2x_2 + ... + a_px_p$ is to be fitted to a set of N training data points having p attributes each. Let X be N x (p+1) vectors of input values (augmented by 1's), Y be N x 1 vector of target values, and θ be (p+1) x 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. $X^TX = XY$
B. $X\theta = X^TY$
C. $X^TX\theta = Y$
D. $X^TX\theta = X^TY$

5) Which of the following offsets, do we use in linear regression's least-square line fit? Suppose the horizontal axis is the independent variable and the vertical axis is the dependent variable.



A. Vertical offset
B. Perpendicular offset
C. Both, depending on the situation
D. None of above

6) Suppose, you got a situation where you find that your linear regression model is under-fitting the data.

In such a situation which of the following options would you consider?

1. Add more variables
2. Start introducing polynomial degree variables
3. Remove some variables

A. 1 and 2
B. 2 and 3
C. 1 and 3
D. 1, 2 and 3

7) Now the situation is the same as written in the previous question(underfitting). Which of the following regularization algorithms would you prefer?

A. L1
B. L2
C. Any
D. None of these

8) Data was collected on two variables x and y and a least-squares regression line was fitted to the data. The resulting equation is y-hat = -2.29 + 1.70x. What is the residual for point (5,6)?

A. -2.91
B. -0.21
C. .21
D. 6.21
E. 7.91

9) What type of penalty is used on regression weights in Ridge regression?

A. L_0
B. L_1
C. L_2
D. None of the above

10) Which of the following coefficients is added as the penalty term to the loss function in Lasso regression?

A. Squared magnitude
B. The absolute value of magnitude
C. Number of non-zero entries
D. None of the above

11) What do you understand by L1 and L2 regularization?

12) When should ridge regression be preferred over lasso?

13) Prove that the Hat matrix(H) is symmetric($H^T = H$) and Idempotent(HH = H)

14) Why orthogonalization is applied on the dimensions in linear regression?

15) What happens in multiple regression, when the dimensions are nearly correlated?

16) What is the difference between Forward stepwise selection and Forward stagewise selection?