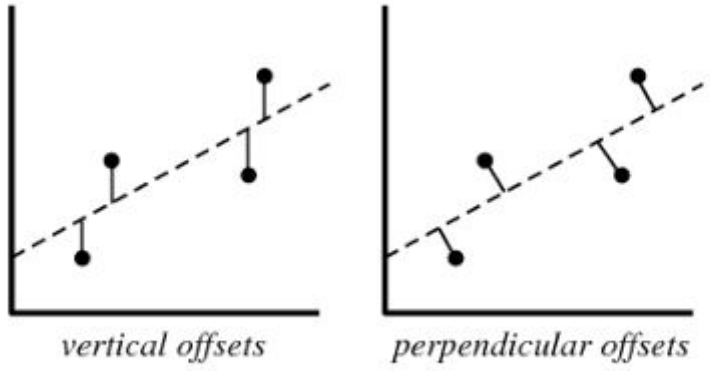
Q1 **Which of the following methods do we use to find the best fit line for data in Linear Regression?**

a. Least Square Error
b. Maximum Likelihood
c. Logarithmic Loss
d. Both A and B

Solution: (A)

In linear regression, we try to minimize the least square errors of the model to identify the line of best fit.

Q2 **Which of the following offsets, do we use in linear regression's least square line fit? Suppose horizontal axis is independent variable and vertical axis is dependent variable.**



a. Vertical offset
b. Perpendicular offset
c. Both, depending on the situation
d. None of above

Solution: (A)

We always consider residuals as vertical offsets. We calculate the direct differences between actual value and the Y labels. Perpendicular offset are useful in case of PCA.

Q3 **True- False: Overfitting is more likely when you have huge amount of data to train?**

a. TRUE
b. FALSE

Solution: (B)

With a small training dataset, it's easier to find a hypothesis to fit the training data exactly i.e. overfitting.

Q4 **Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?**

a. AUC-ROC
b. Accuracy
c. Logloss

d. Mean-Squared-Error

Solution: (D)

Since linear regression gives output as continuous values, so in such case we use mean squared error metric to evaluate the model performance. Remaining options are use in case of a classification problem.

Question Context 5-6:

Suppose that you have a dataset D1 and you design a linear regression model of degree 3 polynomial and you found that the training and testing error is “0” or in another terms it perfectly fits the data.

Q5 **What will happen when you fit degree 4 polynomial in linear regression?**

a. There are high chances that degree 4 polynomial will over fit the data
b. There are high chances that degree 4 polynomial will under fit the data
c. Can’t say
d. None of these

Solution: (A)

Since is more degree 4 will be more complex(overfit the data) than the degree 3 model so it will again perfectly fit the data. In such case training error will be zero but test error may not be zero.

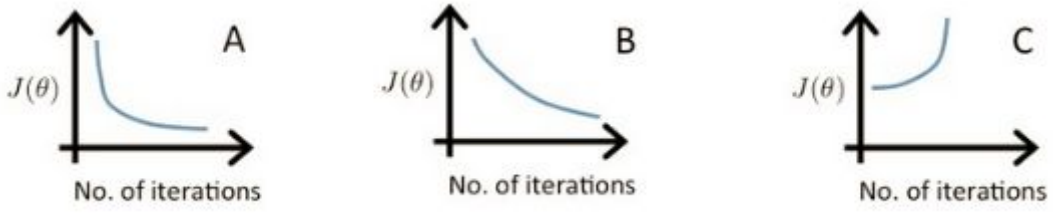
Q6 **What will happen when you fit degree 2 polynomial in linear regression?**

a. It is high chances that degree 2 polynomial will over fit the data
b. It is high chances that degree 2 polynomial will under fit the data
c. Can’t say
d. None of these

Solution: (B)

If a degree 3 polynomial fits the data perfectly, it’s highly likely that a simpler model(degree 2 polynomial) might under fit the data.

Q7 **Which of the following is true about below graphs(A,B, C left to right) between the cost function and Number of iterations?**

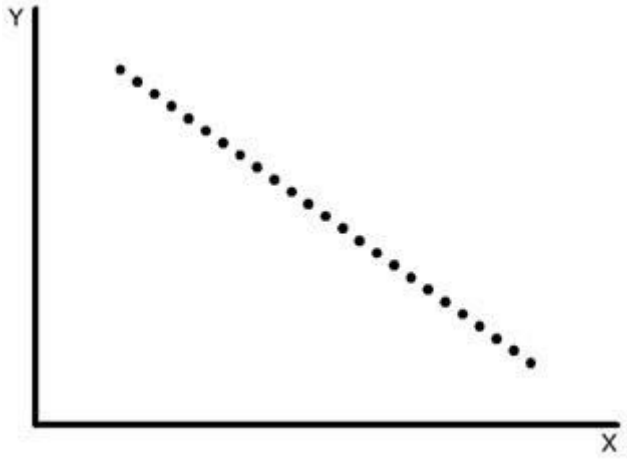


Suppose l1, l2 and l3 are the three learning rates for A,B,C respectively.

a. l2 < l1 < l3
b. l1 > l2 > l3
c. l1 = l2 = l3
d. None of these

Solution: (A)
In case of high learning rate, step will be high, the objective function will decrease quickly initially, but it will not find the global minima and objective function starts increasing after a few iterations.
In case of low learning rate, the step will be small. So the objective function will decrease slowly

Q8 **Consider the following data where one input(X) and one output(Y) is given.**



What would be the root mean square training error for this data if you run a Linear Regression model of the form (Y = A0+A1X)?

a. Less than 0
b. Greater than zero
c. Equal to 0
d. None of these

Solution: (C)
We can perfectly fit the line on the following data so mean error will be zero.

Q9 **Suppose, you got a situation where you find that your linear regression model is under fitting the data. 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

Solution: (A)
In case of under fitting, you need to induce more variables in variable space or you can add some polynomial degree variables to make the model more complex to be able to fir the data better.

Q10 **What does a simple linear regression analysis examine?**

a. The relationship between only two variables
b. The relationship between one dependent and one independent variable
c. The relationship between many variables
d. The relationship between two dependent and one independent variable

Solution: (A) (C)

Q11 **Which of the following are correct for simple regression?**

a. The intercept/constant ($\boldsymbol{\beta}_0$) is the Y when X=0
b. The intercept is the amount of change in Y when X=0
c. The coefficient is the Y at a certain value of X
d. The coefficient (***w***) is the amount of change in Y for every unit increase in X

Answer:

a. The intercept/constant ($\boldsymbol{\beta}_0$) is the Y when X=0
d. The coefficient (***w***) is the amount of change in Y for every unit increase in X

Q12 **What does the MSE method do exactly in simple regression?**

a. Minimizes the distance between the data points
b. Finds the least problematic regression line
c. Finds those (best) values of the intercept and slope that provide us with the smallest value of the sum of squared errors
d. Finds those (best) values of the intercept and slope that provide us with the smallest value of the sum of errors

Answer:

c. Finds those (best) values of the intercept and slope that provide us with the smallest value of the residual sum of squares

Q13 **What does having w = 0 in a simple regression model mean?**

a. Y changes as a result of a change in *X*
b. Y does not change as a result of a change in *X*
c. Y value becomes 0 as a result of a change in *X*
d. Y value is equal to 0 when X=0

Answer:

b. Y does not change as a result of a change in *X*

Q14 **What does a multiple linear regression analysis examine?**

a. The relationship between more than one dependent and only one independent variable
b. The relationship between one or more than one dependent and only one independent variable
c. The relationship between one dependent and more than one independent variables
d. The relationship between more than one independent variables

Answer:

c. The relationship between one dependent and more than one independent variables

Q15 **For a linear regression model with two independent variables we have, $w_1 = w_2 = 0$, what does it mean?**

a. One of the independent variables is useful in predicting the dependent variable
b. Both of the independent variables are useful in predicting the dependent variable
c. None of the independent variables is useful in predicting the dependent variable
d. There is a third independent variable predicting the dependent variable

Answer:

c. None of the independent variables is useful in predicting the dependent variable