

In Q1 to Q11, only one option is correct, choose the correct option:

1. 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

Answer is A)Least Square Error

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

2. Which of the following statement is true about outliers in linear regression?

- A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers
- C) Can't say D) none of these

Answer: (A) Linear regression is sensitive to outliers

Explanation: The slope of the regression line will change due to outliers in most of the cases. So Linear Regression is sensitive to outliers.

3. A line falls from left to right if a slope is _____?

- A) Positive B) Negative
- C) Zero D) Undefined

Answer B)Negative

Explanation :

If the graph of a line rises from left to right, the slope is positive. If the graph of the line falls from left to right the slope is negative.

4. Which of the following will have symmetric relation between dependent variable and independent variable?

- A) Regression B) Correlation
- C) Both of them D) None of these

Answer B) Correlation

Explanation:

The relationship is symmetric between x(dependent) and y(Independent) in case of correlation but in case of regression it is not symmetric.

5. Which of the following is the reason for over fitting condition?

- A) High bias and high variance B) Low bias and low variance
- C) Low bias and high variance D) none of these

Answer is C) Low bias and high variance

Overfitting: It is a Low Bias and High Variance model. Generally, Decision trees are prone to Overfitting.

6. If output involves label then that model is called as:

- A) Descriptive model B) Predictive model
- C) Reinforcement learning D) All of the above

Answer B) Predictive model

7. Lasso and Ridge regression techniques belong to _____?

- A) Cross validation B) Removing outliers
- C) SMOTE D) Regularization

Answer D) Regularization.

8. To overcome with imbalance dataset which technique can be used?

- A) Cross validation B) Regularization
- C) Kernel D) SMOTE

Answer A) Cross Validation

Explanation :It is noteworthy that cross-validation should be applied properly while using over-sampling method to address imbalance problems.

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

- A) TPR and FPR B) Sensitivity and precision
- C) Sensitivity and Specificity D) Recall and precision

Answer A) TPR and FPR

The ROC curve is plotted with TPR(True Positive Rate) against the FPR(False Positive Rate) where TPR is on the y-axis and FPR is on the x-axis

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

- A) True B) False

Answer B) False

Explanation: The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes.

11. Pick the feature extraction from below:
- A) Construction bag of words from a email
 - B) Apply PCA to project high dimensional data
 - C) Removing stop words
 - D) Forward selection

Answer D) Feature Selection.

In Q12, more than one options are correct, choose all the correct options:

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
- A) We don't have to choose the learning rate.
 - B) It becomes slow when number of features is very large.
 - C) We need to iterate.
 - D) It does not make use of dependent variable.

Answer A and B

Important : Q13 and Q 14 are explained with an example in Lasso And Ridge Python file.

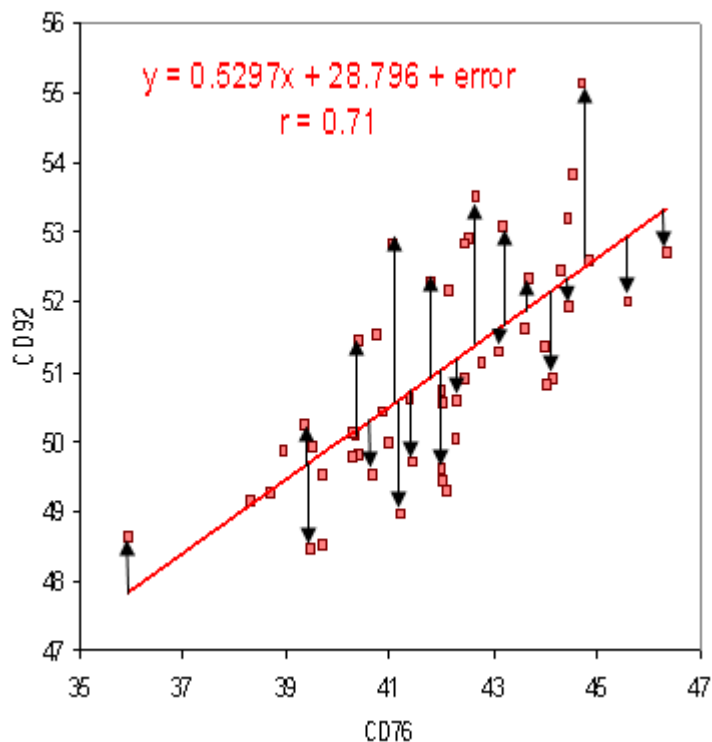
Q15 : Explain the term error present in linear regression equation?

Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.

Examples of the Error Term in Statistics

Basically, the classical normal linear regression model involves finding the best fitting linear model for observed data that shows the relationship between two variables.

For example, let's say you were running a study on the way the number of exams in a certain college affect the amount of red bull purchased from college vending machines. You could collect data which told you how many exams were given and how much red bull was purchased on a dozen or more days during the semester. This data can be plotted as a scatter plot, with exams (Ex) per given day on the x axis and red bull purchased (RB) per given day on the y axis. Then you would look for the line $y = \beta_0 + \beta_1 x$ that best fit the data.



“**Best fit**” here means that the **error term**, the distance from each point to the line, is minimized. Since the relationship between variables is probably not completely linear and because there are other factors outside the scope of our study (sales on red bull, sales on other caffeine drinks, difficult physics homework sets, etc.) the graph won’t actually go through all our data points. The distance between each point and the linear graph (shown as black arrows on the above graph) is our **error term**. So we can write our function as $RB = \beta_0 + \beta_1 Ex + \epsilon$ where β_0 and β_1 are constants and ϵ is an (non constant) error term.

Properties of the Error Term

The error term includes everything that separates your model from actual reality. This means that it will reflect nonlinearities, unpredictable effects, measurement errors, and omitted variables.