

MACHINE LEARNING

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

- 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
- 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
- A line falls from left to right if a slope is _____?
A) Positive
B) **Negative**
C) Zero
D) Undefined
- 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
- 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
- If output involves label then that model is called as:
A) **Descriptive model**
B) **Predictive modal**
C) Reinforcement learning
D) All of the above
- Lasso and Ridge regression techniques belong to _____?
A) Cross validation
B) Removing outliers
C) SMOTE
D) **Regularization**
- To overcome with imbalance dataset which technique can be used?
A) Cross validation
B) Regularization
C) Kernel
D) **SMOTE**
- 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
- In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True
B) **False**
- 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

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.

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?
14. Which particular algorithms are used for regularization?
15. Explain the term error present in linear regression equation?

13. Answer:-

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.

14. Answer:-

Ridge Regularization:-

Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present.

In the cost function, the penalty term is represented by $\text{Lambda } \lambda$. By changing the values of the penalty function, we are controlling the penalty term. The higher the penalty, it reduces the magnitude of coefficients. It shrinks the parameters. Therefore, it is used to prevent multicollinearity, and it reduces the model complexity by coefficient shrinkage.

Lasso Regularization:-

It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients. Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients. Lasso regression line fits the model more accurately than the linear regression line.

MACHINE LEARNING**15. Answer:-**

The error term of a regression equation represents all of the variation in the dependent variable *not* explained by the weighted independent variables.

A regression equation is the formula for a straight line — in this case, the best-fit line through a scatterplot of data. If there were no error, all the data points would be located *on* the regression line; to the extent they are not represents error; this is what the error term summarizes.
