

MACHINE LEARNING

Charu Saxena (Internship – 31)

Highlighted in Yellow Colour is correct Answer

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?

MACHINE LEARNING

- 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.

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

13. Explain the term regularization?

Regularization is a way to avoid overfitting by penalizing high-valued regression coefficients. It reduces parameters and shrinks (simplifies) the model. This more streamlined, more parsimonious model will likely perform better at predictions.

14. Which particular algorithms are used for regularization?

There are mainly two types of regularization techniques, which are given below:

- Ridge Regression
- Lasso Regression

Ridge Regression

- Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as **L2 regularization**.
- In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called Ridge Regression penalty.

Lasso Regression:

- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for **Least Absolute and Selection Operator**.
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
- Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0. It is also called as **L1 regularization**.

15. Explain the term error present in linear regression equation?

The distance between each point and the linear graph (shown as black arrows on the above graph) is our error term. The error term is a random variable with a mean of zero and a constant variance. The meaning of this is that the variances of the independent variables are independent of the value of the variable.

Example:- 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.
