MACHINE LEARNING

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 - Correct
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B

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

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

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

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

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

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

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- Correct
D) none of these

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

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

7. Lasso and Ridge regression techniques belong to _____?

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

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

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

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 -Correct
D) Recall and precision

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

A) True
B) False - Correct

11**. Pick the feature extraction from below:**

A) Construction bag of words from a email
B) Apply PCA to project high dimensional data - Correct
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. - Correct
B) It becomes slow when number of features is very large. - Correct
C) We need to iterate.- Correct
D) It does not make use of dependent variable.

ASSIGNMENT – 39

MACHINE LEARNING

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

13. Explain the term regularization?

Regularization

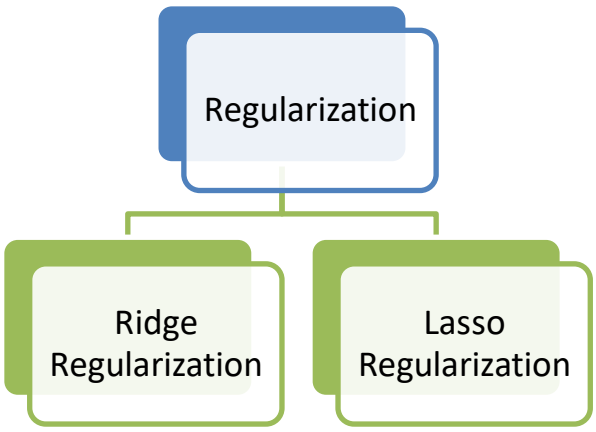
When we use regression models to train some data, there is a good chance that the model will over fit the given training data set. Regularization helps sort this over fitting problem by restricting the degrees of freedom of a given equation i.e. simply reducing the number of degrees of a polynomial function by reducing their corresponding weights. In a linear equation, we do not want huge weights/coefficients as a small change in weight can make a large difference for the dependent variable(Y). So, regularization constraints the weights of such features to avoid over fitting.

Why use Regularization?

Regularization helps to reduce the variance of the model, without a substantial increase in the bias. If there is variance in the model that means that the model won't fit well for dataset different than training data. The tuning parameter λ controls this bias and variance tradeoff. When the value of λ is increased up to a certain limit, it reduces the variance without losing any important properties in the data. But after a certain limit, the model will start losing some important properties which will increase the bias in the data. Thus, the selection of good value of λ is the key. The value of λ is selected using cross-validation methods. A set of λ is selected and cross-validation error is calculated for each value of λ and that value of λ is selected for which the cross-validation error is minimum.

14. Which particular algorithms are used for regularization?

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



LASSO (Least Absolute Shrinkage and Selection Operator) Regression (L1 Form)
LASSO regression penalizes the model based on the sum of magnitude of the coefficients. The regularization term is given by
Regularization=
Where, λ is the shrinkage factor.

Ridge Regression (L2 Form)
Ridge regression penalizes the model based on the sum of squares of magnitude of the coefficients. The regularization term is given by
Regularization=
Where, λ is the shrinkage factor.

Difference between Ridge and Lasso$\lambda * \Sigma|| \beta j \ \lambda * \Sigma|| \beta 2j$

Ridge regression shrinks the coefficients for those predictors which contribute very less in the model but have huge weights, very close to zero. But it never makes them exactly zero. Thus, the final model will still contain

all those predictors, though with less weights. This doesn't help in interpreting the model very well. This is where Lasso regression differs with Ridge regression. In Lasso, the L1 penalty does reduce some coefficients exactly to zero when we use a sufficiently large tuning parameter λ. So, in addition to regularizing, lasso also performs feature selection.

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

The error term is the difference between what the model is predicting and the actual value. This can range from being relatively small to huge, even within one model, across the observed data points.

This error term helps in the calculation of the R-squared value, that is, it tells us how good the model is overall. If the R-squared value of the model is 0.8, then your model explains 80% of the variation in your target variable.

Two other aspects of the error terms give us some help in improving our model.

If the error terms follow certain patterns, it's a warning that we might be using the wrong modelling technique.

First, if for example the error terms are mostly positive at the beginning and end of the data range, but negative in the middle, then you almost certainly need to try a model that will give you a curved regression line, so squaring one or more of your variables may give a better result, or perhaps a log transform may work better.

Second, and closely related to the first, is the distribution of the error terms as the data goes from 'small' to 'large'. For example, if the data points at the 'small' end of the chart have small error values, and at the other end large error values, then you probably need a log transform on your data.