MACHINE LEARNING

In Q1 to Q5, only one option is correct, Choose the correct option:

1. In which of the following you can say that the model is overfitting?

Ans. B) Low R-squared value for train-set and High R-squared value for test-set.

2. Which among the following is a disadvantage of decision trees?

Ans. B) Decision trees are highly prone to overfitting.

3. Which of the following is an ensemble technique?

Ans. C) Random Forest

4. Suppose you are building a classification model for detection of a fatal disease where detection of the disease is most important. In this case which of the following metrics you would focus on?

Ans. A) Accuracy

5. The value of AUC (Area under Curve) value for ROC curve of model A is 0.70 and of model B is0.85. Which of these two models is doing better job in classification?

Ans. B) Model B

In Q6 to Q9, more than one options are correct, Choose all the correct options:

6. Which of the following are the regularization technique in Linear Regression??

Ans. Ridge and Lasso

7. Which of the following is not an example of boosting technique?

Ans. Decision Tree and Random Forest

8. Which of the techniques are used for regularization of Decision Trees?

Ans. Pruning technique

9. Which of the following statements is true regarding the Adaboost technique?

Ans. D) None of the above

Q10 to Q15 are subjective answer type questions, Answer them briefly.

10. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?

Ans. The adjusted R-squared is a modified version of R-squared that adjusts for predictors that are not significant in a regression model. Compared to a model with additional input variables, a lower adjusted R-squared indicates that the additional input variables are not adding value to the model.

11. Differentiate between Ridge and Lasso Regression.

Ans. Ridge and Lasso regression uses two different penalty functions for regularisation. Ridge regression uses L2 on the other hand lasso regression go uses L1 regularisation technique. In ridge regression, the penalty is equal to the sum of the squares of the coefficients and in the Lasso, penalty is considered to be the sum of the absolute values of the coefficients. In lasso regression, it is the shrinkage towards zero using an absolute value (L1 penalty or regularization technique) rather than a sum of squares (L2 penalty or regularization technique).

Since we know that in ridge regression the coefficients can't be zero. Here, we either consider all the coefficients or none of the coefficients, whereas Lasso regression algorithm technique, performs both parameter shrinkage and feature selection simultaneously and automatically because it nulls out the co-efficient of collinear features. This helps to select the variable(s) out of given n variables while performing lasso regression easier and more accurate.

12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?

Ans. A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis. Multicollinearity exists when there is a correlation between multiple independent variables in a multiple regression model. This can adversely affect the regression results.

The default VIF cutoff value is 5; only variables with a VIF less than 5 will be included in the model. However, note that many sources say that a VIF of less than 10 is acceptable.

13. Why do we need to scale the data before feeding it to the train the model?

Ans. Scaling of the data comes under the set of steps of data pre-processing when we are performing machine learning algorithms in the data set. As we know most of the supervised and unsupervised learning methods make decisions according to the data sets applied to them and often the algorithms calculate the distance between the data points to make better inferences out of the data.

In real life, if we take an example of purchasing apples from a bunch of apples, we go close to the shop, examine various apples and pick various apples of the same attributes. Because we have learned about the attributes of apples and we know which are better and which are not good also we know which attributes can be compromised and which can not. So if most of the apples consist of pretty similar attributes we will take less time in the selection of the apples which directly affect the time of purchasing taken by us. The moral of the example is if the apples every apple in the shop is good we will take less time to purchase or if the apples are not good enough we will take more time in the selection process which means that if the values of attributes are closer we will work faster and the chances of selecting good apples also strong.

Similarly in the machine learning algorithms if the values of the features are closer to each other there are chances for the algorithm to get trained well and faster instead of the data set where the data points or features values have high differences with each other will take more time to understand the data and the accuracy will be lower

So if the data in any conditions has data points far from each other, scaling is a technique to make them closer to each other or in simpler words, we can say that the scaling is used for making data points generalized so that the distance between them will be lower..