MACHINE LEARNING

ASSIGNMENT – 6

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

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:

5. The value of AUC (Area under Curve) value for ROC curve of model A is 0.70 and of model B is 0.85. Which of these two models is doing better job in classification?

   Ans:

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

   Ans: A) Ridge

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

   Ans: C) Random Forest

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

   Ans: A) Pruning

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

   Ans:

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

    Ans: Adjusted R-squared is an modified version of R-squared that has been adjusted for the number of predictors I the model. The adjusted R-squared increases when the ne term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected. Typically the adjusted R-squared is positive not negative it is always lower than the R-squared.

    R-squared is a statistical measure that representas the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a model R-squared explains to what extent the variance of one variable explains the variance of the second variable. So if the r2 of a model is 0.50then approximately half of the observed variation can be explained by the model inputs.

11. Differentiate between Ridge and Lasso Regression.

Ans: In Ridge regression, we add a penalty term which is equal to the square of the coefficient. The *L2* term is equal to the square of the magnitude of the coefficients. We also add a coefficient to control that penalty term. In this case if is zero then the equation is the basic OLS else if then it will add a constraint to the coefficient. As we increase the value of this constraint causes the value of the coefficient to tend towards zero. This leads to tradeoff of higher bias (dependencies on certain coefficients tend to be 0 and on certain coefficients tend to be very large, making the model less flexible) for lower variance.

Lasso regression stands for Least Absolute Shrinkage and Selection Operator. It adds penalty term to the cost function. This term is the absolute sum of the coefficients. As the value of coefficients increases from *0* this term penalizes, cause model, to decrease the value of coefficients in order to reduce loss. The difference between ridge and lasso regression is that it tends to make coefficients to absolute zero as compared to Ridge which never sets the value of coefficient to absolute zero.

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

Ans: A VIF equal to one means variables are not correlated and multicollinearity does not exist in the regression model.

The overall model might show strong, statistically sufficient explanatory power, but be unable to identify if the effect is mostly due to the unemployment rate or to the new initial jobless claims. This is what the VIF would detect, and it would suggest possibly dropping one of the variables out of the model or finding some way to consolidate them to capture their joint effect depending on what specific hypothesis the researcher is interested in testing.

While a moderate amount of multicollinearity is acceptable in a regression model, a higher multicollinearity can be a cause for concern.

Two measures can be taken to correct high multicollinearity, First, one or more of the highly correlated variables can be removed, as the information provided by these variables is redundant. The second method is to use principal components analysis or partial least square regression instead of OLS regression, which can respectively reduce the variables to a smaller set with no correlation, or create new uncorrelated variables. This will improve the predictability of a model.

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

    Ans: After building an understanding of how to do data scaling and which data scaling techniques to use, we can now talk about where to use these data scaling techniques.

In this article, we shall discuss one of the ubiquitous steps in the machine learning pipeline Feature Scaling. This article's origin lies in one of the coffee discussions in my office on what all models actually are affected by feature scaling and then what is the best way to do it to normalize or to standardize or something else?

In this article, in addition to the above, we would also cover a gentle introduction to feature scaling, the various feature scaling techniques, how it might lead to data leakage, when to perform feature scaling, and when NOT to perform feature scaling. So tighten your reading glasses and read on

An increasing number of features on different scales make machine learning problems hard to handle, is feature scaling the solution?

Feature scaling is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step.

14. What are the different metrics which are used to check the goodness of fit in linear regression?

Ans: Regression refers to predictive modeling problems that involve predicting a numeric value.It is different from classification that involves predicting a class label. Unlike classification, you cannot use classification accuracy to evaluate the predictions made by a regression model. Instead, you must use error metrics specifically designed for evaluating predictions made on regression problems. In this tutorial, you will discover how to calculate error metrics for regression predictive modeling projects.