Machine Learning Assignment 6

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

Answer. C) High R-squared value for train-set and Low R-squared value for test-set.

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

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

QUES 3. Which of the following is an ensemble technique?

Answer. C) Random Forest.

QUES 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?

Answer. A) Accuracy.

QUES 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?

Answer. B) Model B.

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

Answer. A) Ridge D) Lasso.

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

Answer. B) Decision Tree   C) Random Forest.

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

Answer. A) Pruning.

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

Answer. A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points.

C) It is example of bagging technique.

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

Answer. The adjusted R-squared compensates for the addition of variables and only increases if the new predictor enhances the model above what would be obtained by probability. Conversely, it will decrease when a predictor improves the model less than what is predicted by chance.

QUES 11. Differentiate between Ridge and Lasso Regression.

Answer. Ridge Regression: Ridge regression is a technique used to analyze multi-linear regression (multicollinear), also known as L2 regularization. It is Applied when predicted values are greater than the observed values.

$\sum i=1(yi-p\sum j=1xij\beta j)2+\lambda p\sum j=1\beta 2j$

Above equation represents the formula for Ridge Regression! where,

Lambda (λ) in the equation is tuning parameter which is selected using cross-validation technique which makes the fit small by making squares small ($\beta 2$) by adding shrinkage factor.

The shrinkage factor is lambda times the sum of squares of regression coefficients (The last element in the above equation).

LASSO REGRESSION: Lasso stands for – Least Absolute Shrinkage and Selection Operator. It is a technique where data points are shrunk towards a central point, like the mean. Lasso is also known as L1 regularization.

It is applied when the model is overfitted or facing computational challenges. It is a statistical formula for the regularisation of data models and feature selection.

Minimization objective = LS Obj + α * (sum of the absolute value of coefficients)

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

The above equation represents the formula for Lasso Regression! where, Lambda (λ) is a tuning parameter selected using the before Cross-validation technique.

Unlike Ridge Regression, Lasso uses |β| to penalize the high coefficients.

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

Answer. Full Form of VIF is Variance Inflation Factor. The Variance Inflation Factor (VIF) measures the severity of multicollinearity in regression analysis. It is a statistical concept that indicates the increase in the variance of a regression coefficient as a result of collinearity. Multicollinearity inflates the variance and type II error. It makes the coefficient of a variable consistent but unreliable. A VIF of three or below is not a cause for concern. As VIF increases, the less reliable your regression results are going to be.

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

Answer. We need to scale the data before feeding it to the train the model because to ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we scale the data before feeding it to the model. It makes it easy for a model to learn and understand the problem. 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.

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

Answer. There are many metrics to measure the performance of Linear Regression. Mostly used in Econometrics and Academics is R-squared and Restricted R-squared. R-squared or R2 explains the degree to which your input variables explain the variation of your output / predicted variable. So, if R-square is 0.8, it means 80% of the variation in the output variable is explained by the input variables. So, in simple terms, higher the R squared, the more variation is explained by your input variables and hence better is your model.

However, the problem with R-squared is that it will either stay the same or increase with addition of more variables, even if they do not have any relationship with the output variables. This is where "Adjusted R square" comes to help. Adjusted R-square penalizes you for adding variables which do not improve your existing model.

Hence, if we are building Linear regression on multiple variable, it is always suggested that we use Adjusted R-squared to judge goodness of model. In case you only have one input variable, R-square and Adjusted R squared would be exactly same.

Other commonly used metrics that are commonly used for evaluating and reporting the performance of a Linear Regression Model Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE).

QUES 15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.

Answer. Accuracy= TP+TN/TP+FP+FN+TN

Sensitivity= TP/TP+FN

Specificity= TN/TN+FP

Precision = TP / TP + FP

Recall = TP / TP + FN

Accuracy: 1000+1200/1000+1200+250+50=0.81

Sensitivity: 1000/1250=0.8

Specificity: 1200/1250=0.96.

Precision: 1000/1050=0.95

Recall: 1000/1250=0.8