MACHINE LEARNING ASSIGNMENT-6 ANSWER

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

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

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

Ans. Decision trees are highly prone to overfitting.

3. Which of the following is an ensemble technique?

Ans. 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. None of the above.

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. Model B

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

Ans. Ridge, Lasso

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

Ans. B) Decision Tree, C) Random Forest

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

Ans. A) Pruning C) Restricting the max depth of the tree

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

Ans. None of the above

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 the R-squared which takes into account the number of predictors used in a linear regression model. The R-squared measures how well the model fits the data, but it doesn't penalize the use of unnecessary predictors. The adjusted R-squared is calculated using the following formula:

Adjusted R-squared = 1 - [(1 - R-squared) * (n - 1) / (n - k - 1)]

where n is the sample size and k is the number of predictors in the model.

The adjusted R-squared penalizes the presence of unnecessary predictors in the model by reducing the value of the adjusted R-squared as the number of predictors increases. This is because the adjusted R-squared formula includes a penalty term that increases as the number of predictors increases. This penalty term adjusts for the fact that adding more predictors to the model will almost always increase the R-squared, even if those predictors are not really helping to explain the variation in the dependent variable.

Therefore, the adjusted R-squared is a better measure of model fit than R-squared when comparing models with different numbers of predictors. It provides a more accurate measure of how well the model is fitting the data, while taking into account the number of predictors used in the model. A higher adjusted R-squared indicates a better model fit, while a lower adjusted R-squared indicates that the model may be overfitting the data.

11. Differentiate between Ridge and Lasso Regression.

Ans. Ridge and Lasso Regression are two types of regularized regression techniques that are used to prevent overfitting in linear regression models. The main difference between Ridge and Lasso Regression is the way they add a penalty to the regression coefficients.

Ridge Regression adds a penalty to the sum of the squared values of the regression coefficients (L2 penalty) while Lasso Regression adds a penalty to the sum of the absolute values of the regression coefficients (L1 penalty). This difference leads to a difference in the way they shrink the coefficients of less important variables towards zero.

Ridge Regression tends to keep all the variables in the model and shrinks the coefficients of less important variables towards zero, while Lasso Regression tends to force the coefficients of less important variables to be exactly zero, which can lead to variable selection and result in a simpler model.

In summary, Ridge Regression and Lasso Regression differ in the type of penalty they add to the regression coefficients, which leads to a difference in the way they shrink the coefficients of less important variables. Ridge Regression keeps all variables in the model while Lasso Regression can result in variable selection and a simpler model.

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

Ans. VIF (Variance Inflation Factor) is a measure of collinearity among predictor variables in a regression model. It estimates how much the variance of an estimated regression coefficient increases if that predictor variable is added to the model.

The VIF value ranges from 1 upwards, and a value of 1 indicates no multicollinearity between the predictor variable and the other predictor variables. Generally, a VIF value greater than 1.5 or 2 indicates high multicollinearity and should be investigated further. A commonly used rule of thumb is that a VIF value of 5 or above indicates that there is serious multicollinearity.

In regression modeling, it is recommended to remove the predictors with high VIF values to obtain a better model with lower variance. However, this decision depends on the research question, and the researcher's discretion, as the choice of variable selection methods may have an impact on the final model.

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

Ans. Scaling the data is an important step in preparing data for training a machine learning model. Here are some reasons why we need to scale the data before feeding it to the model:

- Scaling can help to normalize the data and remove the effects of different scales: Many machine learning algorithms work better when the input data is normalized or standardized. If the data has different scales, some features may dominate the others, leading to bias in the model. Scaling can help to remove the effects of different scales in the data.

- Scaling can help to speed up the training process: Many machine learning algorithms use iterative optimization techniques, and scaling can help to speed up the convergence of the algorithm.
- Scaling can improve the performance of some machine learning algorithms: Some algorithms, such as k-Nearest Neighbors, are sensitive to the scale of the input data. Scaling can help to improve the performance of such algorithms.

There are different methods of scaling, including standard scaling, min-max scaling, and robust scaling. The suitable method of scaling depends on the type of data and the requirements of the model.

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

Ans. There are several metrics used to check the goodness of fit in linear regression. Some of the commonly used metrics are:

- R-squared (R2) - R-squared measures the proportion of variance in the dependent variable that is explained by the independent variables in the model.
- Mean Squared Error (MSE) - MSE is the average of the squared differences between the predicted and actual values.
- Root Mean Squared Error (RMSE) - RMSE is the square root of the MSE and provides a measure of the average magnitude of the errors in the predicted values.
- Mean Absolute Error (MAE) - MAE is the average of the absolute differences between the predicted and actual values.
- Residual Standard Error (RSE) - RSE measures the standard deviation of the residuals (the difference between the predicted and actual values).

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

Actual/Predicted	**True**	**False**
True	1000 (TP)	50 (FP)
False	250 (FN)	1200 (TN)

Ans. The following confusion metrics can be calculated:

- **Sensitivity** (also known as **recall** or true positive rate): the proportion of actual positive instances that were correctly predicted by the model. It is calculated as TP / (TP + FN).
  1000 / (1000+250) = **0.8**
- **Specificity** (also known as true negative rate): the proportion of actual negative instances that were correctly predicted by the model. It is calculated as TN / (TN + FP).
  1200 / (1200+50) = **0.96**
- **Precision**: the proportion of predicted positive instances that were actually positive. It is calculated as TP / (TP + FP).
  1000 / (1000+50) = **0.95**
- **Accuracy**: the proportion of all instances that were correctly predicted by the model. It is calculated as (TP + TN) / (TP + TN + FP + FN).
  (1000+1200) / (1000+1200+50+250) = **0.88**