MACHINE LEARNING WORKSHEET - 6

1. In which of the following you can say that the model is overfitting?
A) High R-squared value for train-set and High R-squared value for test-set.
B) Low R-squared value for train-set and High R-squared value for test-set.
C) High R-squared value for train-set and Low R-squared value for test-set.
D) None of the above

Ans: C) 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?
A) Decision trees are prone to outliers.
B) Decision trees are highly prone to overfitting.
C) Decision trees are not easy to interpret
D) None of the above.

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

3. Which of the following is an ensemble technique?
A) SVM B) Logistic Regression
C) Random Forest D) Decision tree

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?
A) Accuracy B) Sensitivity
C) Precision D) None of the above.

Ans: B) Sensitivity

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?
A) Model A B) Model B
C) both are performing equal D) Data Insufficient

Ans: B) Model B

6. Which of the following are the regularization technique in Linear Regression??
A) Ridge B) R-squared
C) MSE D) Lasso

Ans: A) Ridge, D) Lasso

7. Which of the following is not an example of boosting technique?
A) Adaboost B) Decision Tree
C) Random Forest D) Xgboost.

Ans: B) Decision Tree, C) Random Forest

8. Which of the techniques are used for regularization of Decision Trees?
A) Pruning B) L2 regularization
C) Restricting the max depth of the tree D) All of the above

Ans: A) Pruning, C) Restricting the max depth of the tree

9. Which of the following statements is true regarding the Adaboost technique?
A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points
B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well
C) It is example of bagging technique
D) None of the above

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

B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well

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

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.

11. Differentiate between Ridge and Lasso Regression.

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.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2=\mathrm{RSS}+\lambda\sum_{j=1}^{p}\beta_j^2$$

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 (β2) 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.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS}+\lambda\sum_{j=1}^{p}|\beta_j|.$$

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

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

The Variance Inflation Factor (VIF) is a measure of colinearity among predictor variables within a multiple regression. It is calculated by taking an independent variable and regressing it against every other predictor in the model.

$$VIF = \frac{1}{1-R_i^2}$$

Including highly correlated variables in your model can lead to overfitting. If we overfit, then the model performs extraordinarily well on the training data but doesn’t generalize well when we try to use it on new data. Small VIF values, VIF < 3, indicate low correlation among variables under ideal conditions. 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?
1)Feature scaling would help them all to be in the same range.
2)To prevent the data from data leakage that‘s why we need to scale the data.

14. What are the different metrics which are used to check the goodness of fit in linear regression?
1)R-squared
2)Adjusted R-squared
3)F-Test
4)RMSE

15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.
Accuracy = TP+TN/TP+TN+FP+FN =
1000+1200/1000+1200+250+50 =2200/2500 = 0.88

Recall = TP/TP+FN = 1000/1000+50 = 0.95

Precision = TP/TP+FP = 1000/1000+250 = 0.8

Sensitivity = TP/TP+FN = 1000/1000+50 = 0.95

Specificity = TN/FP+TN = 1200/50+1200 = 1200/1250 = 0.96