MACHINE LEARNING WORKSHEET 6

ANSWER

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) (D)

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

ANS-- (B) (C)

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) (C)

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) (B)

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

ANS:- Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model.

In linear regression model, R-squared on training data increases as we increase the number of predictors because as we increase the number of predictors ,we are adding more information to the model .so the r-squared increases on the training dataset. But this can lead to overfitting of the model.

To prevent the overfitting in spite of R-squared we use adjusted r-squared. Its formula is given as follows:

$$R^2 \text{ adjusted} = 1 - (1 - R^2) (N - 1) / N - p - 1$$

where

R^2 = sample R-square

p = number of predictors

N = total sample size

As we can see from the above formula ,if R-squared remains constant and number of predictors increase then adjusted R-Squared decreases. So, if R-squared is not increased significantly on adding predictors, adjusted R-Squared will decrease. So, in this way adjusted R-squared will penalize the presence of unnecessary predictors.

11. Differentiate between Ridge and Lasso Regression.?

ANS— Ridge Regression is a type of regression algorithm and is usually considered when there is a high correlation between the independent variables or model parameters .

The word “LASSO” denotes Least Absolute Shrinkage and Selection Operator. Lasso regression follows the regularization technique to create prediction. It is given more priority over the other regression methods because it gives an accurate prediction. Lasso regression model uses shrinkage technique.

THE DIFFERENCE BETWEEN LASSO AND RIDGE REGRESSION:-

- Ridge and Lasso regression uses two different penalty functions for regularization. Ridge regression uses L2 ,on the other hand, lasso regression go uses L1 regularization 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 in ridge regression the coefficients can’t be zero. So, we either consider all the coefficients or none of the coefficients, whereas in 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**. It is a statistical concept that

indicates the increase in the variance of a regression coefficient as a result of collinearity.

The VIF estimates how much the variance of a regression coefficient is inflated due to multicollinearity in the model.

VIF=1/(1-Ri^2)=1/TOLERANCE

Where,

Ri^2 represents the unadjusted coefficient of determination for regressing the ith independent variable and the reciprocal of VIF is known as **Tolerance.**

Generally, a VIF above 4 or tolerance below 0.25 indicates that multicollinearity might exist, and further investigation is required. When VIF is higher than 10 or tolerance is lower than 0.1, there is significant multicollinearity that needs to be corrected.

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

ANS:-Scaling is one of the important pre-processing that is required for standardizing/normalization of input data. When the range of values are very distinct in each column , we need to scale them to the common level. The values are then brought to the common level and then we apply further machine learning algorithms to the input data.

And the Other is the Gradient Descent .if an algorithm uses gradient descent, it 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.

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

ANS: There are different metrics that we can use to check the goodness of fit in linear regression are as follow:-

- MAE -Mean absolute error
- MSE- mean square error
- RMSE- root mean square error
- R^2 SCORE

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

ANS:-SENSITIVITY OR RECALL—0.8

SPECIFICITY—0.96

ACCURACY—0.88

PRECISION—0.95

F1--0.86