MACHINE LEARNING WORKSHEET 6

Q1. 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.A) ) High R-squared value for train-set and High R-squared value for test-set.

Q2. 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.

Q3. Which of the following is an ensemble technique?

A) SVM B) Logistic Regression

C) Random Forest D) Decision tree

Ans.c)Random forest

Q4. 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.A)Accuracy

Q5. 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 because it has more AUC**

**In Q6 to Q9, more than one options are correct, Choose all the correct options:**

Q6. Which of the following are the regularization technique in Linear Regression??

A) Ridge B) R-squared

C) MSE D) Lasso

Ans A and C

Q7. Which of the following is not an example of boosting technique?

A) Adaboost B) Decision Tree

C) Random Forest D) Xgboost.

Ans.C)Random forest

Q8. 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

Q9. 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.B)** ) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well

**Q10 to Q15 are subjective answer type questions, Answer them briefly.**

Q10. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?

Ans. 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.

Q11. Differentiate between Ridge and Lasso Regression.

Ans. The cost function for both ridge and lasso regression are similar. However, ridge regression takes the square of the coefficients and lasso takes the magnitude.

Lasso regression can be used for automatic feature selection, as the geometry of its constrained region allows coefficient values to inert to zero.

An alpha value of zero in either ridge or lasso model will have results similar to the regression model.

The larger the alpha value, the more aggressive the penalization.

Q12. 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. Multicollinearity exists when there is a correlation between multiple independent variables in a multiple regression model. This can adversely affect the regression results. Thus, the variance inflation factor can estimate how much the variance of a regression coefficient is inflated due to multicollinearity.

As a rule of thumb, 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.

Q13. Why do we need to scale the data before feeding it to the train the model?

Ans. 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. Having features on a similar scale will help the gradient descent converge more quickly towards the minima.

Specifically, in the case of Neural Networks Algorithms, feature scaling benefits optimization by:

It makes the training faster

It prevents the optimization from getting stuck in local optima

It gives a better error surface shape

Weight decay and Bayes optimization can be done more conveniently

Q14. What are the different metrics which are used to check the goodness of fit in linear regression?

Ans. There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are: Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE)

Q15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.

|  |  |  |
| --- | --- | --- |
| Actual/Predicted | True | False |
| True | 1000 | 50 |
| False | 250 | 1200 |

Ans.Sensitivity=TP/TP+FN=1000/1250=0.8

Recall=0.8

Accuracy=(TP +TN)/(TP+TN+RP+FN)=2200/2500=0.88

Specificity=TN/(TN+FP)=1200/1250=0.96

Precision=TP/(TP+FP)=1000/1050=0.958