ASSIGNMENT

MACHINE LEARNING

Worksheet -4

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

Answer:- 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.

Answer:- 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

Answer:- 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

Answer:- A) Accuracy

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

Answer:- C) Both are performing equal

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

Answer:- a) Ridge and d) Lasso

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

Answer:- D) Decision Tree

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

Answer:- A) Pruning

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

Answer:- 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?

Answer:- Adjusted R^2 is a corrected goodness-of-fit (model accuracy) measure for linear models. It identifies the percentage of variance in the target field that is explained by the input or inputs. R^2 tends to optimistically estimate the fit of the linear regression. It always increases as the number of effects are included in the model. Adjusted R^2 attempts to correct for this overestimation. Adjusted R^2 might decrease if a specific effect does not improve the model. Adjusted R squared is calculated by dividing the residual mean square error by the total mean square error (which is the sample variance of the target field). The result is then subtracted from 1. Adjusted R^2 is always less than or equal to R^2. A value of 1 indicates a model that perfectly predicts values in the target field. A value that is less than or equal to 0 indicates a model that has no predictive value. In the real world, adjusted R^2 lies between these values.

11. Differentiate between Ridge and Lasso Regression.

Answer:-

Lasso Regression	**Ridge Regression**
Lasso is a modification of linear regression, where the model is penalized for the sum of absolute values of the weights. Thus, the absolute values of weight will be (in general) reduced, and many will tend to be zeros.	**Ridge** takes a step further and penalizes the model for the sum of squared value of the weights. Thus, the weights not only tend to have smaller absolute values, but also really tend to penalize the extremes of the weights, resulting in a group of weights that are more evenly distributed.
During training, the objective function become: $\frac{1}{2m}\sum_{i=1}^{m}(y-Xw)^2+alpha\sum_{j=1}^{p}\lvert w_j\rvert$	The objective function becomes: $\sum_{i=1}^{n}(y-Xw)^2+alpha\sum_{j=1}^{p}w_j^{\,2}$

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

Answer:- 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. When VIF = 1, it is considered to be a suitable value of a VIF for a feature to be included in a regression modelling.

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

Answer:- Scaling of the data makes it easy for a model to learn and understand the problem. In the case of neural networks, an independent variable with a spread of values may result in a large loss in training and testing and cause the learning process to be unstable. Normalization and Standardization are the two main methods for the scaling of the data. Which are widely used in the algorithms where scaling is required. Both of them can be implemented by the scikit-learn libraries pre-process package.

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

Answer:- There are 3 main metrics for model evaluation in regression:

- **R Square/Adjusted R Square: -**
-

  R Square measures how much variability in dependent variable can be explained by the model. It is the square of the Correlation Coefficient(R) and that is why it is called R Square.

  $$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

  R Square is calculated by the sum of squared of prediction error divided by the total sum of the square which replaces the calculated prediction with mean. R Square value is between 0 to 1 and a bigger value indicates a better fit between prediction and actual value.
  It does not take into consideration of overfitting problem. If your regression model has many independent variables, because the model is too complicated, it may fit very well to the training data but performs badly for testing data. That is why Adjusted R Square is introduced because it will penalize additional independent variables added to the model and adjust the metric to prevent overfitting issues.

- **Mean Square Error (MSE)/Root Mean Square Error (RMSE): -**

  Mean Square Error is an absolute measure of the goodness for the fit.

  $$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

  MSE is calculated by the sum of square of prediction error which is real output minus predicted output and then divide by the number of data points. It gives you an absolute number on how much your predicted results deviate from the actual number. You cannot interpret many insights from one single result but it gives you a real number to compare against other model results and help you select the best regression model.Root Mean Square Error(RMSE) is the square root of MSE. It is used more commonly than MSE because firstly sometimes MSE value can be too big to compare easily. Secondly, MSE is calculated by the square of error, and thus square root brings it back to the same level of prediction error and makes it easier for interpretation.

Mean Absolute Error (MAE): -

Mean Absolute Error (MAE) is similar to Mean Square Error(MSE). However, instead of the sum of square of error in MSE, MAE is taking the sum of the absolute value of error.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

Compare to MSE or RMSE, MAE is a more direct representation of sum of error terms. MSE gives larger penalization to big prediction error by square it while MAE treats all errors the same.

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)

Accuracy: - TP+TN/TP+FP+TN+FN
1000+1200/1000+50+1200+250 = 0.88

Precision: - TP/TP+FP
1000/1000+50 = 0.952

Recall/Sensitivity: - TP/TP+FN
1000/1000+250 = 0.8

Specificity: - TN/TN+FP
1200/1200+50 = 0.96