

MACHINE LEARNING WORKSHEET 6

Q-1:- (C) High R-squared value for train-set and Low R-squared value for test-set.

Q-2:- (B) Decision trees are highly prone to overfitting.

Q-3:- (C) Random Forest.

Q-4:- (A) Accuracy.

Q-5:- (B) Model B.

Q-6:- (A) Ridge and (D) Lasso.

Q-7:- (B) Decision Tree and (C) Random Forest.

Q-8:- (A) Pruning and (C) Restricting the max depth of the tree.

Q-9:- (A) We initialize the probabilities of the distribution as $1/n$, where n is the number of data-points. **and**

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

Q-10:- Basically, The adjusted R-squared compensates for the addition of variables and only increases if the new predictor enhances the model above that what would be obtained by probability and conversely, it will decrease when a predictor improves the model less than what is predicted by chance.

Q-11:- Lasso:- It is a regularization technique used in feature selection using a Shrinkage method also called as the penalized regression method.

Lasso means Least Absolute Shrinkage and Selection Operator which is used both for regularization and model selection. If a model uses the L1 regularization technique, it is called lasso regression.

Ridge:- Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization.

Q-12:- Variance inflation factor (VIF) is a measure of the amount of multicollinearity present 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 also.

Usually, a VIF above 4 or tolerance below 0.25 indicates that multicollinearity might exist, and further investigation is required. If VIF is above 10, then the multicollinearity issue needs to be corrected.

Q-13:- Basically, it is done to make it easy for a model to learn and understand the problem clearly and To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same speed for all the features.

Q-14:- It is used to judge that how well a model fits a given set of data, and how well it will predict a future set of observations.

There are 3 metrics used to check goodness :-

1. MAE (Mean Absolute Error).
2. MSE (Mean Squared Error).
3. RMSE (Root Mean Squared Error).

Q-15:- SENSITIVITY:- 0.8 .

SPECIFICITY:- 0.96 .

PRECISION:- 0.8 .

ACCURACY:- 0.88 .