

MACHINE LEARNING

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: B) Model B

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: B) Decision Tree and 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

Answer: A) Pruning and 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

Answer: 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?

Answer: The adjusted R^2 will penalize you for adding independent variables (K in the equation) that do not fit the model. Some of those variables will be significant, but you can't be sure that significance is just by chance. The adjusted R^2 will compensate for this by that penalizing you for those extra

variables. While values are usually positive, they can be negative as well. This could happen if your R^2 is zero; After the adjustment, the value can dip below zero. This usually indicates that your model is a poor fit for your data. Other problems with your model can also cause sub-zero values, such as not putting a constant term in your model.

11. Differentiate between Ridge and Lasso Regression.

Answer: Differences between Ridge and Lasso Regression:

Ridge Regression	Lasso Regression
Ridge regression performs 'L2 regularization' i.e. it adds a factor of sum of squares of coefficients in the optimization objective. Thus, ridge regression optimizes the following: Objective = $RSS + \alpha * (\text{sum of square of coefficients})$	LASSO stands for Least Absolute Shrinkage and Selection Operator. Lasso regression performs L1 regularization, i.e. it adds a factor of sum of absolute value of coefficients in the optimization objective. Thus, lasso regression optimizes the following: Objective = $RSS + \alpha * (\text{sum of absolute value of coefficients})$
Adds penalty term which is equal to the square of the coefficient.	Adds penalty term to the cost function.
Ridge regression decreases the complexity of a model but does not reduce the number of variables	If $p > n$, the lasso selects at most n variables.

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

Answer: VIF stands for Variance Inflation Factor. A variance inflation factor(VIF) is used to find multicollinearity in regression analysis. Multicollinearity occurs when two or more independent variables are highly correlated with one another in a regression model.

$$VIF = 1/(1-R^2)$$

VIF score of an independent variable represents how well the variable is explained by other independent variables. Suitable values of a VIF 1 = not correlated, Between 1 and 5 = moderately correlated and Greater than 5 = highly correlated.

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

Answer: Scaling puts the values in the same range or same scale so that no variable is dominated by the other.

If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values. In k-nearest neighbors with an Euclidean distance measure is sensitive to magnitudes and hence should be scaled for all features to weigh in equally. Scaling is critical, while performing Principal Component Analysis (PCA). PCA tries to get the features with maximum variance and the variance is high for high magnitude features. This skews the PCA towards high magnitude features.

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

Answer: There are three main metrics which are used to check the goodness of fit in linear regression. They are: R-squared, the overall F-test, and the Root Mean Square Error (RMSE).

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

Actual/Predicted	True	False
True	1000	50
False	250	1200

Answer:

$$\text{sensitivity} = \text{TP} / (\text{TP} + \text{FN}) = 1000 / (1000 + 50) = 0.95$$

$$\text{specificity} = \text{TN} / (\text{TN} + \text{FP}) = 1200 / (1200 + 250) = 0.827$$

$$\text{precision} = \text{TP} / (\text{TP} + \text{FP}) = 1000 / (1000 + 250) = 0.80$$

$$\text{recall} = \text{sensitivity} = \text{TP} / (\text{TP} + \text{FN}) = 1000 / (1000 + 50) = 0.95$$

$$\text{accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) = (1000 + 1200) / (1000 + 1200 + 250 + 50) = 0.88$$