Solved 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

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) Ridge
   D) Lasso

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

Ans- 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?

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. The adjusted R-square is calculated as follows-

$$\text{Adjusted R-square} = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$$

Where, R^2 - sample R square

p - number of predictors/features

N - total sample size

In the above equation, when p is equal to 0, we can see that adjusted R-square equals to R^2. Thus adjusted R-square will always be less than or equal to R^2 and it penalises the excess of independent variables which do not affect the dependent variable

---

11. Differentiate between Ridge and Lasso Regression.

Ans.

Lasso Regression	Ridge Regression
LASSO (Least Absolute Shrinkage and Selection Operator) penalizes the model based on the sum of magnitude of the coefficients	Ridge regression penalizes the model based on the sum of squares of magnitude of the coefficients.
The regularization term is given by- Regularization = $\lambda * \sum\|B_j\|$	The regularization term is given by- Regularization = $\lambda * \sum\|B_j^2\|$
Lasso regression does reduce some coefficients exactly to zero when we use a sufficiently large tuning parameter.	Ridge regression shrinks the coefficients for those predictors which contribute very less in the model but have huge weights, very close to zero, but it never makes them exactly zero
Lasso also performs feature selection	Ridge does not perform feature selection

---

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. 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.

A VIF below 5 suggests 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?

Ans.

Scaling of the data makes it easy for a model to learn and understand the problem. If the data in any conditions has data points far from each other, scaling is a technique to make them closer to each other or in simpler words, we can say that the scaling is used for making data points generalized so that the distance between them will be lower. The machine learning models provide weights to the input variables according to their data points and inferences for output. In that case, if the difference between the data points is so high, the model will need to provide the larger weight to the points and in final results, the model with a large weight value is often unstable. This means the model can produce poor results or can perform poorly during learning.

---

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

Ans.

The different metrics which are used to check the goodness of fit in linear regression –

- Mean Squared Error: MSE or Mean Squared Error is one of the most preferred metrics for regression tasks. It is simply the average of the squared difference between the target value and the value predicted by the regression model. As it squares the differences, it penalizes even a small error which leads to over-estimation of how bad the model is. It is preferred more than other metrics because it is differentiable and hence can be optimized better.
- Root Mean Squared Error: RMSE is the most widely used metric for regression tasks and is the square root of the averaged squared difference between the target value and the value predicted by the model. It is preferred more in some cases because the errors are first squared before averaging which poses a high penalty on large errors. This implies that RMSE is useful when large errors are undesired.
- Mean Absolute Error: MAE is the absolute difference between the target value and the value predicted by the model. The MAE is more robust to outliers and does not penalize the errors as extremely as mse. MAE is a linear score which means all the individual differences are weighted equally. It is not suitable for applications where you want to pay more attention to the outliers.
- R^2 Error: Coefficient of Determination or R^2 is another metric used for evaluating the performance of a regression model. The metric helps us to compare our current model with a constant baseline and tells us how much our model is better. The constant baseline is chosen by taking the mean of the data and drawing a line at the mean. R^2 is a scale-free score that implies it doesn't matter whether the values are too large or too small, the R^2 will always be less than or equal to 1.

- Adjusted R^2: Adjusted R^2 depicts the same meaning as R^2 but is an improvement of it. R^2 suffers from the problem that the scores improve on increasing terms even though the model is not improving which may misguide the researcher. Adjusted R^2 is always lower than R^2 as it adjusts for the increasing predictors and only shows improvement if there is a real improvement.

---

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

Actual/Predicted	True	False
True	1000	50
False	250	1200

Ans.

Given, TP = 1000
FP = 50
FN = 250
TN = 1200

Sensitivity = TP/ (TP+FN) = 1000/(1000+250) = 0.8
Specificity = TN/ (TN+FP) = 1200/(1200+50) = 0.96
Precision = TP/ (TP+FP) = 1000/(1000+50) = 0.95
Recall = TP/ (TP+FN) = 1000/(1000+250) = 0.8
Accuracy = (TP+TN)/(TP+TN+FP+FN)=
(1000+1200)/(1000+1200+50+250) = 0.88
