

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.
2. Which among the following is a disadvantage of decision trees?
D) None of the above.
3. Which of the following is an ensemble technique?
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
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?
B) Model B
6. Which of the following are the regularization technique in Linear Regression??
A) Ridge D) Lasso
7. Which of the following is not an example of boosting technique?
B) Decision Tree
C) Random Forest
8. Which of the techniques are used for regularization of Decision Trees?
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

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

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 increase if the new predictor enhances the model above what would be obtained by probability.
11. Differentiate between Ridge and Lasso Regression.
Ans- Lasso is a short for least absolute shrinkage and selection operator, which is used both for regularization and model selection. If a model uses the L1 regularisation technique then it is called lasso regression while Ridge regression is similar to lasso regression, ridge regression puts a similar constraint on the coefficient by introducing a penalty factor. However, while lasso regression takes the magnitude of the coefficients, ridge regression takes the square.
12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?
Ans- VIF is stands for variance inflation factor is a measure of the amount of multicollinearity in regression analysis.
13. 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.
14. What are the different metrics which are used to check the goodness of fit in linear regression?
Ans- R-squared
The overall F-test
The root mean square error (RMSE).

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

Predicted		True	False
Ans-	True	1000	50
	False	250	1200