MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression? **A) Least Square Error**

2. Which of the following statement is true about outliers in linear regression? **A) Linear regression is sensitive to outliers**

3. A line falls from left to right if a slope is _____? **B) Negative**

4. Which of the following will have symmetric relation between dependent variable and independent variable? **A) Regression**

5. Which of the following is the reason for over fitting condition? **C) Low bias and high variance**

6. If output involves label then that model is called as: **D) All of the above**

7. Lasso and Ridge regression techniques belong to _____? **D) Regularization**

8. To overcome with imbalance dataset which technique can be used? **D) SMOTE**

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph? A) TPR and FPR B) Sensitivity and precision C) Sensitivity and Specificity D) Recall and precision

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less. **B) False**

11. Pick the feature extraction from below: **B) Apply PCA to project high dimensional data**

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression? **B) It becomes slow when number of features is very large.**

13. Explain the term regularization

This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting. Regularization, significantly reduces the variance of the model, without substantial increase in its bias

14. Which particular algorithms are used for regularization

Two types of regularization techniques, which are given below:

Ridge Regression, Lasso Regression.

Ridge regression helps us to reduce only the overfitting in the model while keeping all the features

present in the model. It reduces the complexity of the model by shrinking the coefficients whereas

Lasso regression helps in reducing the problem of overfitting in the model as well as automatic feature selection.

Lasso Regression tends to make coefficients to absolute zero whereas Ridge regression never sets the value of coefficient to absolute zero

15. Explain the term error present in linear regression equation

Error term in a regression equation represents the effect of the variables that were omitted from the equation.

the graph won’t actually go through all our data points. The distance between each point and the linear graph (shown as black arrows on the above graph) is our error term

not all real-world data won't follow the function we're trying to fit to it… because of observed predictors, poor choice of model, measurement error, and/or plain ol' randomness (I once made an exception for physics in front of a physicist and they assured me they see error terms, too). If you leave it out, then most likely you won't be able to find a function that fits your data… or the function that fits your data will be so horrible and data specific that you won't want to use it, and any new data you come upon is likely to not fit

add a new independent variable to your original regression, and when you would do a second regression on the residual from your first regression, to reduce residuals Error term is come into picture.