MACHIN LEARNING _SHEET

1 > Which of the following methods do we use to find the best fit line for data in Linear Regression?

ANS : - (A) Least Square Error

2 > Which of the following statement is true about outliers in linear regression?

ANS : - (A) Linear regression is sensitive to outliers

3 > A line falls from left to right if a slope is _____?

ANS : - (B) Negative

4 > Which of the following will have symmetric relation between dependent variable and independent variable?

ANS : - (C) Both of them

5 > Which of the following is the reason for over fitting condition?

ANS : - (C) Low bias and high variance

6 > If output involves label then that model is called as:

ANS : - (B) Predictive

7 > Lasso and Ridge regression techniques belong to _____?

ANS : - (D) Regularization

8 > To overcome with imbalance dataset which technique can be used?

ANS : - (D)SMOTE

9 > The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

ANS : - (A) TPR and FPR

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

ANS : - (B) False

11 > Pick the feature extraction from below:

ANS : - (A) Construction bag of words from a email

12 > . Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

ANS : - (B) It becomes slow when number of features is very large.

13 > Explain the term regularization?

ANS : - **Regularization** is a technique to used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. This

technique discourages **learning** a more complex or flexible model, so as to avoid the risk of overfitting.
Overfitting is a phenomenon where a machine learning model models the training data too well but fails to perform well on the testing data. ***Regularization*** is one technique which help to prevent overfitting.
Overfitting happens when a model learns the details and noise in the training data to the extent that it negatively impacts the performance of the model on unseen data.
Regularization technique discourages learning a more complex or flexible model, avoiding the risk of Overfitting.

14 > Which particular algorithms are used for regularization?

ANS : - Regularization (commonly described as adding Ridge or Lasso Penalty) is difficult to enforce for Trees in the narrow sense. Tree is a heuristic algorithm.
Regularization for Trees is accomplished in a wider sense by limiting the maximum depth of tree ensembles / bagging more than one tree, and setting stricter stopping criteria for when to split a node further.

15 > Explain the term error present in linear regression equation?

ANS : - The error term in a regression equation represents the effect of the variable that were omitted from the equation. Basically, the error term in the regression equation help to explain the dependent variable the larger the error term the smaller the constant term none of the above is true.
The error term basically represent the combined effect of the omitted variables, assuming that

- Combined effect of omitted variables is independent of each variable included in the equation.
- Independent across various variables.
- Independent across subjects.
- Variables has expectation 0.