Machine Learning worksheet1_Answerkey

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: 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 _____?

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: C) Sensitivity and Specificity

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: D) Forward selection

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.
D) It does not make use of dependent variable.

13. Explain the term regularization?

Ans: Def: Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting

There are mainly two types of regularization techniques

1. L1 regularization
2. L2 regularization
3. Dropout regularization

L1 regularization

A regression model which uses L1 Regularization technique is called LASSO(Least Absolute Shrinkage and Selection Operator) regression. Lasso Regression adds *"absolute value of magnitude"* of coefficient as penalty term to the loss function(L).

L2 regularization

Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions A regression model that uses L2 regularization technique is called Ridge regression.
Ridge regression adds "*squared magnitude*" of coefficient as penalty term to the loss function(L)

14. Which particular algorithms are used for regularization?

Ans: Lasso Regression:

- Lasso regression is regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.

- Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.
- It is also called as L1 regularization. Some of the features in this technique are completely neglected for model evaluation.
- Hence, the Lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

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

Ans: An **error term** in statistics is a value which represents how observed data differs from actual population data. It can also be a variable which represents how a given statistical model differs from reality. The error term is often written ε.

It is often said that the error term in a regression equation represents the effect of the variables that were omitted from the equation. This is unsatisfactory, even in simple contexts, as the following discussion should indicate. Suppose subjects are IID, and all variables are jointly normal with expectation 0. Suppose the explanatory variables have variance 1. The explanatory variables may be correlated amongst themselves, but anyp of them have a non-singularp-dimensional distribution. The parameters αj are real. Let

(1) Yi = ∞ j=1 αjXij For each p = 1, 2,..., consider the regression model

(2) Yi = p j=1 αjXij + i(p) where (

3) i(p) = ∞ j=p+1 αjXij

The αj are identifiable. If the Xij are independent for j = 1, 2,..., the standard assumptions hold, and i(p) does indeed represent the effect on Yi of the omitted variables{Xij : j = p+1,...},