

# Machine Learning Worksheet 1

Name- Amit Matreja  
Internship 21  
Datrained batch -1834

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
- B) Maximum Likelihood
- C) Logarithmic Loss
- D) Both A and B

Ans 1. A)Least Square Error

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

- A) Linear regression is sensitive to outliers
- B) linear regression is not sensitive to outliers
- C) Can' t say
- D) none of these

Ans 2. A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is \_\_\_\_\_?

- A) Positive
- B) Negative
- C) Zero
- D) Undefined

Ans 3. B) Negative

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

- A) Regression
- B) Correlation
- C) Both of them
- D) None of these

Ans 4. B) Correlation

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

- A) High bias and high variance

# Machine Learning Worksheet 1

- B) Low bias and low variance
- C) Low bias and high variance
- D) none of these

Ans 5. C) Low bias and high variance

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

- A) Descriptive model
- B) Predictive modal
- C) Reinforcement learning
- D) All of the above

Ans 6. B) Predictive Model

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_?

- A) Cross validation
- B) Removing outliers
- C) SMOTE
- D) Regularization

Ans 7. A) Regularization

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

- A) Cross validation
- B) Regularization
- C) Kernel
- D) SMOTE

Ans 8. A) Cross Validation

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

Ans 9. A) TPR and FPR

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

- A) True

# Machine Learning Worksheet 1

B) False

Ans 10. B) False

11. Pick the feature extraction from below:

- A) Construction bag of words from a email
- B) Apply PCA to project high dimensional data
- C) Removing stop words
- D) Forward selection

Ans 11. C) Removing stop words

In Q12, more than one options are correct, choose all the correct options:

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

- A) We don't have to choose the learning rate.
- B) It becomes slow when number of features is very large.
- C) We need to iterate.
- D) It does not make use of dependent variable.

Ans 12. A) We don't have to choose the learning rate. & B) It become slow when number of features is very large.

Ans 13. Regularization is selecting preferred complexity level of model/ to reduce the complexity of the model, so that models will start predicting in better way. if we don't use regularization models might be too complex and overfit, and might also give poor predictions. For Regularization we need couple of things, we need a way of testing to check how good or accurate models are at predictions, and a selecting parameter which lets you change the complexity of the model.

We need to adjust the complexity/model to find the value which gives the best model predictions.

Ans 14. we use two types of algorithms for Regularization

1. Ridge Regression- Ridge regression is also known as L2 Regularization. In simple term, Ridge regression shrinks the co-efficient, which helps to reduce the complexity and multi-

# **Machine Learning Worksheet 1**

collinearity of the model. In this we introduce small amount of bias, we call it Ridge regression penalty to get better long-term predictions.

2. Lasso Regression – It is another type of regularization technique which helps in reducing the complexity of the model. Lasso stands for Least Absolute and Selection operator. Lasso regression is also known as L1 regularization. Lasso regression is different from ridge regression as it uses absolute weight values for normalization.

Ans 15. A regression equation always have an error term because independent variables never predicts perfectly. The error term of a regression equation is the variation in the dependent variable not explained by the independent variables. A regression equation is a formula for straight line. Suppose, the best fit line in a scatterplot of data. If there were no error, all the data points would be located on the regression line.