

## MACHINE LEARNING

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

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

1. **Answer: - A**

A) Least Square Error

B) Maximum Likelihood

C) Logarithmic Loss

D) Both A and B

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

2. **Answer: - A**

A) Linear regression is sensitive to outliers

B) linear regression is not sensitive to outliers

C) Can't say

D) none of these

3. A line falls from left to right if a slope is \_\_\_\_\_? **Answer: - B**

A) Positive

B) Negative

C) Zero

D) Undefined

4. Which of the following will have symmetric relation between dependent variable and independent variable? **Answer: - B**

A) Regression

B) Correlation

C) Both of them

D) None of these

5. Which of the following is the reason for over fitting condition? **Answer: - C**

A) High bias and high variance

B) Low bias and low variance

C) Low bias and high variance

D) none of these

6. If output involves label then that model is called as: **Answer: - B**

A) Descriptive model

B) Predictive modal

C) Reinforcement learning

D) All of the above

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

7. **Answer: - B**

A) Cross validation

B) Removing outliers

C) SMOTE

D) Regularization

To overcome with imbalance dataset which technique can be used?

8. **Answer: - A**

A) Cross validation

B) Regularization

C) Kernel

D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses \_\_\_\_\_ to make graph? **Answer: - C**

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. **Answer: - A**

A) True

B) False

11. Pick the feature extraction from below: **Answer: - B**

A) Construction bag of words from a email

B) Apply PCA to project high dimensional data

C) Removing stop words

D) Forward selection

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? **Answer: - A, B, C**

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.

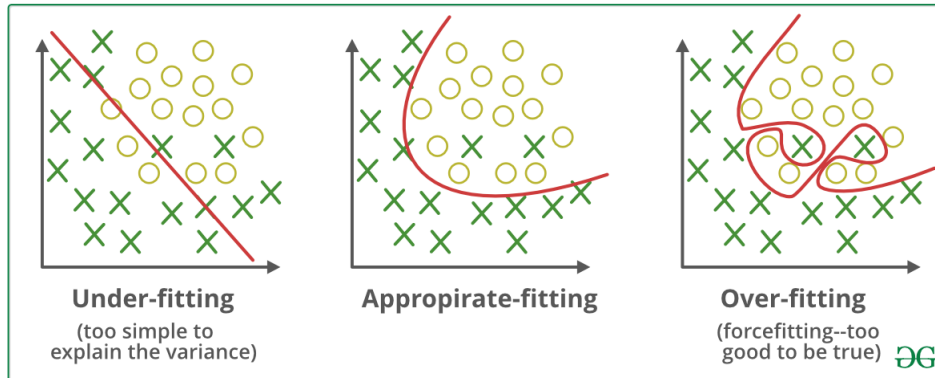
## ACHINE LEARNING

**Q13 and Q15 are subjective answer type questions, Answer them briefly.**

### **13. Explain the term regularization?**

**ANSWER:** Regularization is a technique which is used to reduce the errors by fitting the model or curve on the given training data set and avoid overfitting/ underfitting.

Following are graphs shows model is overfitting or underfitting.



Based on which regularization technique is further classified in to below categories.

1) Lasso Regression / L1 form

Lasso regression penalize the model based on sum of magnitude of the coefficients.

2) Ridge Regression/ L2 form

Ridge regression penalize the model based on sum of squared magnitude of the coefficients.

3) Elastic Net

Which is less popular in regularization. It is combination of Lasso and Ridge regressions.

### **14. Which particular algorithms are used for regularization?**

**ANSWER:** Widely used Regularization techniques are Lasso regression/ L1 form and Ridge regression/ L2 form to regularize modem from over fitting or under fitting.

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

**ANSWER:** The term error in linear regression means difference of actual target value and predicted target value. which is created when the model does not perfectly represent the actual relationship between the independent variables and the dependent variables. Which is also known as Residual (R).