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] B) Maximum Likelihood

C) Logarithmic Loss D) Both A and B

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

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

A) Positive [✓ B) Negative]

C) Zero D) Undefined

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

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

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:

A) Descriptive model [✓ B) Predictive model]

C) Reinforcement learning D) All of the above

7. Lasso and Ridge regression techniques belong to __________?

A) Cross validation B) Removing outliers

C) SMOTE [✓ D) Regularization]

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

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?

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.

A) True B) False

11. Pick the feature extraction from below:

A) Construction bag of words from an 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?

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.

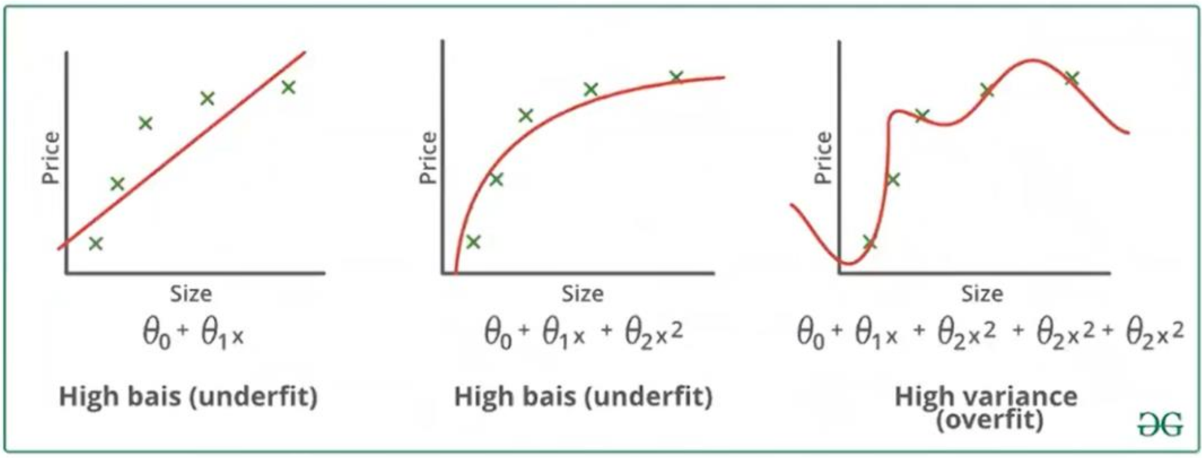
D) It does not make use of dependent variable.

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

13. Explain the term regularization?

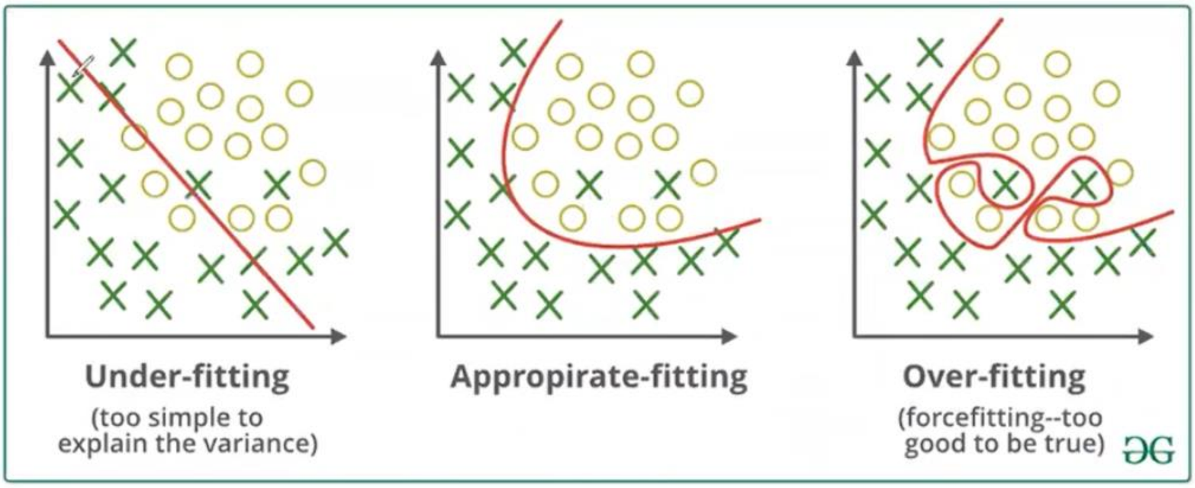
Ans: Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

The image shows High bias (under fit) and High variance (over fit) models.



- A model with High bias pay very little attention to the training data and it leads to high error on training and test data.
- A model with High variance pay a lot of attention to the training data and such model perform very well on training data but have high error rates on test data.

Figure represents Under-fitting, Appropriate-fitting and Over-fitting models.


Under-fitting
(too simple to explain the variance)
Appropirate-fitting
Over-fitting
(forcefitting--too good to be true)

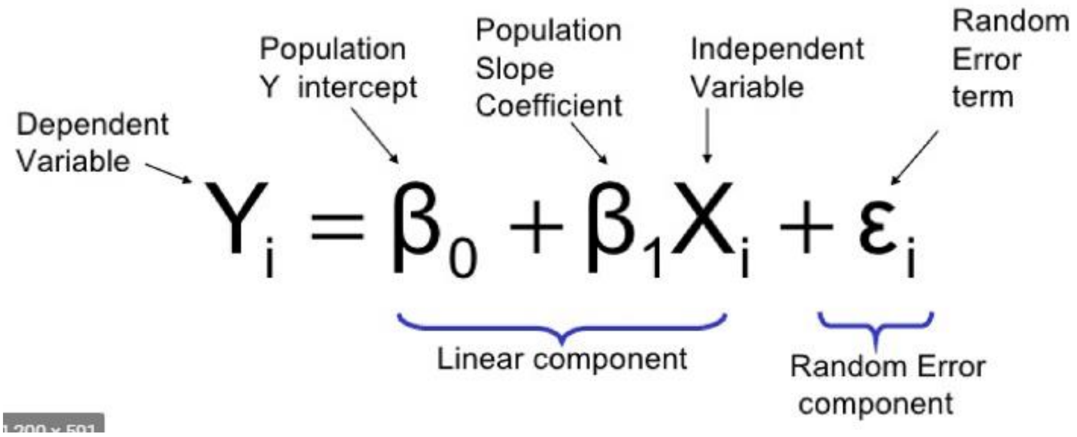
With the help of Regularization in Machine learning, we can reduce the errors of the models after that Under-fit and Over-fit models converted into Appropriate-fit models.

14. Which particular algorithms are used for regularization?

Ans: There are three main regularization techniques, namely:
1. Ridge Regression.(L1)
2. LASSO (Least Absolute Shrinkage and Selection Operator) Regression.(L2)
3. Elastic-Net Regression (combination of both lasso and Ridge)

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

Ans: An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.



Error term ε_i is not an error in the sense of a mistake. The error term was put into the estimating equation to capture missing variables and errors in measurement that may have occurred in the dependent variables.