MACHINE LEARNING WORKSHEET

- **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 modal
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 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?
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?
Answer :-
This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In other words, *this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.*

14. Which particular algorithms are used for regularization?

Answer :-

There are three main regularization techniques, namely:

1. Ridge Regression (L2 Norm)
2. Lasso (L1 Norm)
3. Dropout

1.Ridge Regression (L2 Regularization)

Ridge regression is also called L2 norm or regularization.

When using this technique, we add the sum of weight's square to a loss function and thus create a new loss function which is denoted thus:

$$\text{Loss} = \sum_{j=1}^{m} \left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji \right)^2 + \lambda \sum_{i=1}^{n} Wi^2$$

As seen above, the original loss function is modified by adding normalized weights. Here normalized weights are in the form of squares.

2.Lasso Regression (L1 Regularization)

Also called lasso regression and denoted as below:

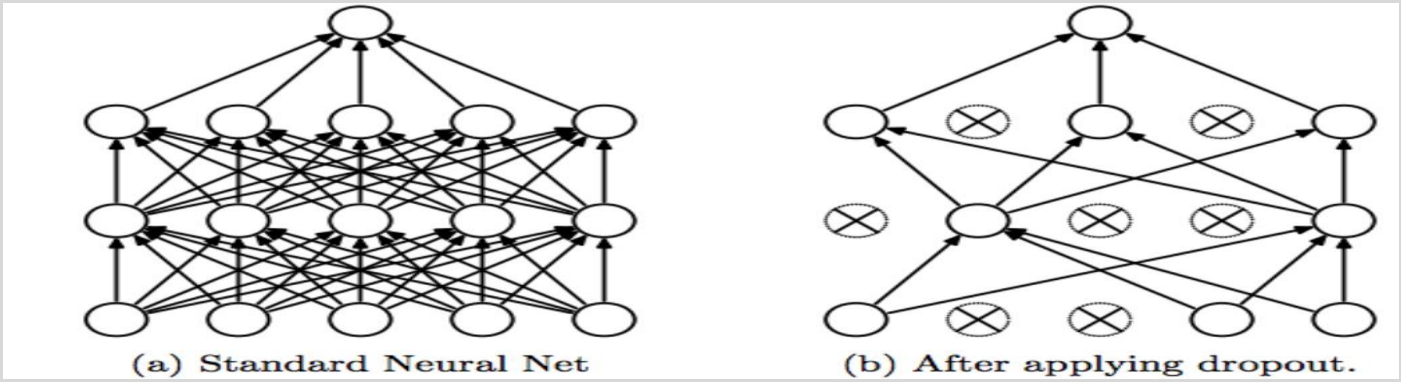
$$\text{Loss} = \sum_{j=1}^{m} \left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji \right)^2 + \lambda \sum_{i=1}^{n} |Wi|$$

This technique is different from ridge regression as it uses absolute weight values for normalization. λ is again a tuning parameter and behaves in the same as it does when using ridge regression.

3.Dropout

Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.

In neural nets, fully connected layers are more prone to overfit on training data. Using dropout, you can drop connections with *1-p* probability for each of the specified layers. Where *p* is called **keep probability parameter** and which needs to be tuned.



(a) Standard Neural Net (b) After applying dropout.

With dropout, you are left with a reduced network as dropped out neurons are left out during that training iteration.
Dropout decreases overfitting by avoiding training all the neurons on the complete training data in one go. It also improves training speed and learns more robust internal functions that generalize better on unseen data.

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

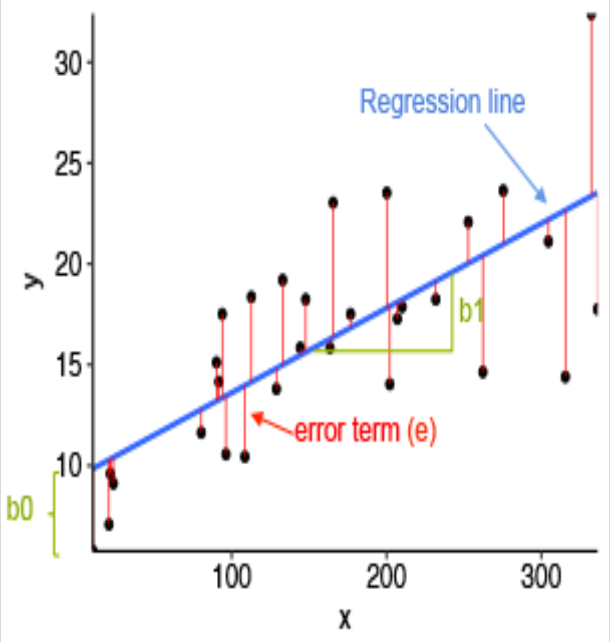
Answer :-

A Linear Regression model' focus to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized.
Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.

Let's understand this with the help of a diagram.

On the basis of above diagram, we can say that:-

- x is our dependent variable which is plotted on the x-axis and y is the dependent variable which is plotted on the y-axis.
- Black dots are the data points i.e the actual values.
- b_o is the intercept which is 10 and b_1 is the slope of the x variable.
- The blue line is the best fit line predicted by the model i.e the predicted values lie on the blue line.



The vertical distance between the data point and the regression line is known as error or residual. Each data point has one residual and the sum of all the differences is known as **the Sum of Residuals/Errors.**