MACHINE LEARNING ASSESMENT ANSWERS

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: **(B) Correlation**

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

ANS: **(C) Low bias and high variance**

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

ANS: **(B) Predictive model**

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: **(A) TPR and FPR**

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: **(B) Apply PCA to project high dimensional data**

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

ANS**: A) We don't have to choose the learning rate.**

B) It becomes slow when number of features is very large.

13. **Explain the term regularization?**

ANS: Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. The commonly used regularization techniques are:

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

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

ANS: The term 'regularization' refers to a set of techniques that regularizes learning from particular features for traditional algorithms or neurons in the case of neural network algorithms. It normalizes and moderate's weights attached to a feature or a neuron so that algorithms do not rely on just a few features or neurons to predict the result. This technique helps to avoid the problem of overfitting.

To do so, we need to use a loss function and find optimized parameters using gradient descent algorithms and its variants.

There are three main regularization techniques, namely:

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

Ridge and Lasso can be used for any algorithms involving weight parameters, including neural nets. Dropout is primarily used in any kind of neural networks e.g., ANN, DNN, CNN or RNN to moderate the learning. Let's take a closer look at each of the techniques.

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. Now the parameters are learned using a modified loss function. To minimize the above function, parameters need to be as small as possible. Thus, L2 norm prevents weights from rising too high.

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 use ridge regression.

As loss function only considers absolute weights, optimization algorithms penalize higher weight values.

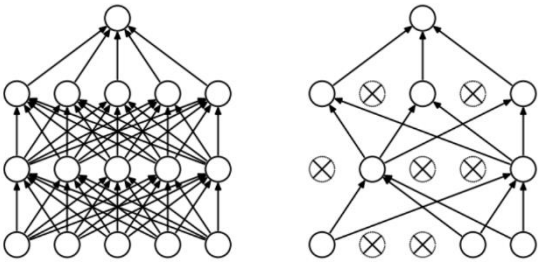
In ridge regression, loss function along with the optimization algorithm brings parameters near to zero but not actually zero, while lasso eliminates less important features and sets respective weight values

to zero. Thus, lasso also performs feature selection along with regularization.

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.

Along with Dropout, neural networks can be regularized also using L1 and L2 norms.

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

ANS: Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by: 1. measuring the distance of the observed y-values from the predicted y-values at each value of x; 2. squaring each of these distances; 3. calculating the mean of each of the squared distances. Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.

mean-square error (MSE) The Mean Squared Error (MSE) is a measure of how close a fitted line is to data points. For every data point, you take the distance vertically from the point to the corresponding y value on the curve fit (the error), and square the value. Then you add up all those values for all data points, and, in the case of a fit with two parameters such as a linear fit, divide by the number of points minus two.

The squaring is done so negative values do not cancel positive values. The smaller the Mean Squared Error, the closer the fit is to the data. The MSE has the units squared of whatever is plotted on the vertical axis. Another quantity that we calculate is the Root Mean Squared Error (RMSE). It is just the square root of the mean square error. That is probably the most easily interpreted statistic, since it has the same units as the quantity plotted on the vertical axis.