1) **Which of the following methods do we use to find the best fit line for data in linear regression?**

Ans:- Least Square Error

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

Ans:-Linear Regression is sensitive to outliers

3) **A line falls from left to right if a slope is?**

Ans:- 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 the model is called as**

Ans:- B) Predictive Model

7) **Lasso and Ridge regression techniques belongs to**

Ans:- D) Regularization

8) **To overcome with imbalance dataset which technique is used?**

Ans:- D) SMOTE

9) **The AUC Receiver Operator Characteristics(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 better model area under the curve should be less**

Ans:- B) False

11) **Pick the feature extraction from below**

Ans:- B) Pick 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) and B)

13) Explain the term regularization?

Ans:- The term regularize means to make the things regular or acceptable. This is exactly why we use it for. Regularization are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting

14) Which particular algorithms are used for regularizations?

Ans:- 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:

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

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

You may have noticed parameters λ along with normalized weights. λ is the parameter that needs to be tuned using a cross-validation dataset. When you use λ=0, it returns the residual sum of square as loss function which you chose initially. For a very high value of λ, loss will ignore core loss function and minimize weight's square and will end up taking the parameters' value as zero.

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 using ridge regression.

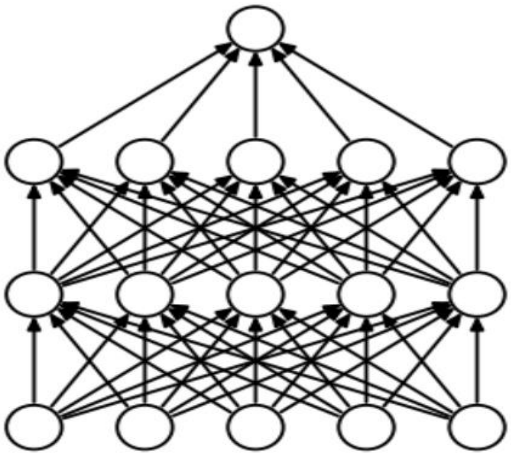
As loss function only considers absolute weights, optimization algorithms penalize higher weight values.

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.

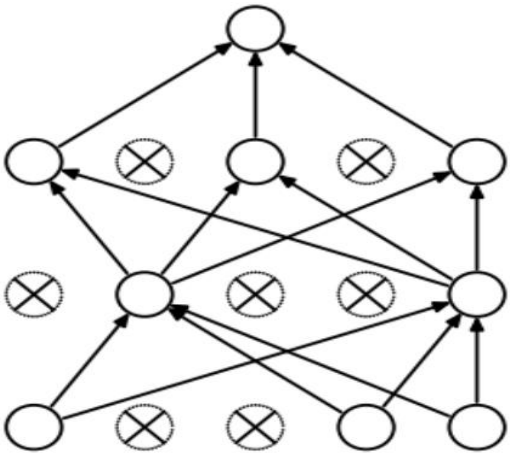
Dropout

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.

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. However, it is important to note that Dropout takes more epochs to train compared to training without Dropout (If you have 10000 observations in your training data, then using 10000 examples for training is considered as 1 epoch).

Along with Dropout, neural networks can be regularized also using L1 and L2 norms. Apart from that, if you are working on an image dataset, image augmentation can also be used as a regularization method.

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 model does not fully represent the actual relationship between the independent variables and dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis. The error term is also know as the residual disturbance or reminder term and is variously represented in models by the letters € or u. The error term is a residual variable the accounts for a lack of perfect goodness of fit.