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?

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: C) Both of them

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 modal

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: C) Sensitivity and Specificity

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: A) Construction bag of words from an email, B) Apply PCA to project high dimensional data & C) Removing stop words

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?

<u>Ans:</u> A) We don't have to choose the learning rate, B) It becomes slow when number of features is very large

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

13. Explain the term regularization?

<u>Ans:</u> Regularization is a form of regression, that constrains or 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. It is a useful technique that can help in improving the accuracy of regression models.

14. Which particular algorithms are used for regularization?

<u>Ans:</u> There are majorly three regularization techniques are used as per my research:

Ridge Regression (L2 Norm), Lasso (L1 Norm) & 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.

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. 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.

2) Lasso Regression (L1 Regularization):

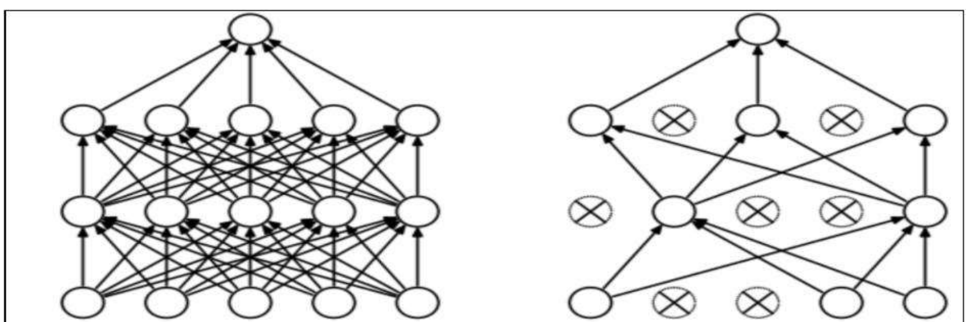
Lasso regression denoted as below:

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

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 sam e 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.

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.

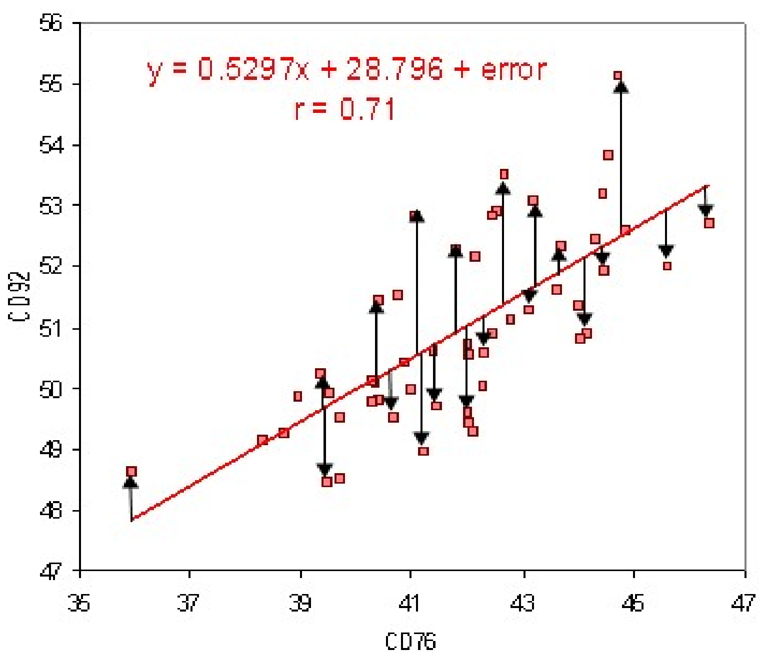
With dropout, we left with a reduced network as dropped out neurons are left out during that training iteration. Dropout decreases overfitting by avoiding training all the neurons on the complete training data in one go. It also improves training speed and learns more robust internal functions that generalize 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. For real-world applications, it is a must that a model performs well on unseen data. The techniques we discussed can help us make our model learn rather than just memorize.

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

<u>Ans:</u> An error term in statistics is a value which represents how observed data differs from actual population data. It can also be a variable which represents how a given statistical model differs from reality. The error term is often written ε.
The error term includes everything that separates your model from actual reality. This means that it will reflect nonlinearities, unpredictable effects, measurement errors, and omitted variables.

Examples of the Error Term in Statistics:
In econometric theory, the classical normal linear regression model (CNLRM) involves finding the best fitting linear model for observed data that shows the relationship between two variables.

For example, let's say you were running a study on the way the number of exams in a certain college affect the amount of red bull purchased from college vending machines. You could collect data which told you how many exams were given and how much red bull was purchased on a dozen or more days during the semester. This data can be plotted as a scatter plot, with exams (Ex) per given day on the x axis and red bull purchased (RB) per given day on the y axis. Then you would look for the line y = β0 + β1 x that best fit the data.



Errors on a scatter plot. "Best fit" here means that the error term, the distance from each point to the line, is minimized. Since the relationship between variables is probably not completely linear and because there are other factors outside the scope of our study (sales on red bull, sales on other caffeine drinks, difficult physics homework sets, etc.) the graph won't actually go through all our data points. The distance between each point and the linear graph (shown as black arrows on the above graph) is our error term. So, we can write our function as $RB = \beta 0 + \beta 1\ Ex + \varepsilon$ where $\beta 0$ and $\beta 1$ are constants and ε is an (non-constant) error term.