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
ANS:--A

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

ANS:--A

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

A) Positive B) Negative
C) Zero D) Undefined
ANS:--A

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
ANS:--B

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
ANS:--A

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

ANS:--B

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization
ANS:--D

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

A) Cross validation B) Regularization
C) Kernel D) SMOTE
ANS:--D

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
ANS:--A

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True B) False
ANS:--B

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 stops words
D) Forward selection
ANS:--D

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.
ANS:--A, B AND C

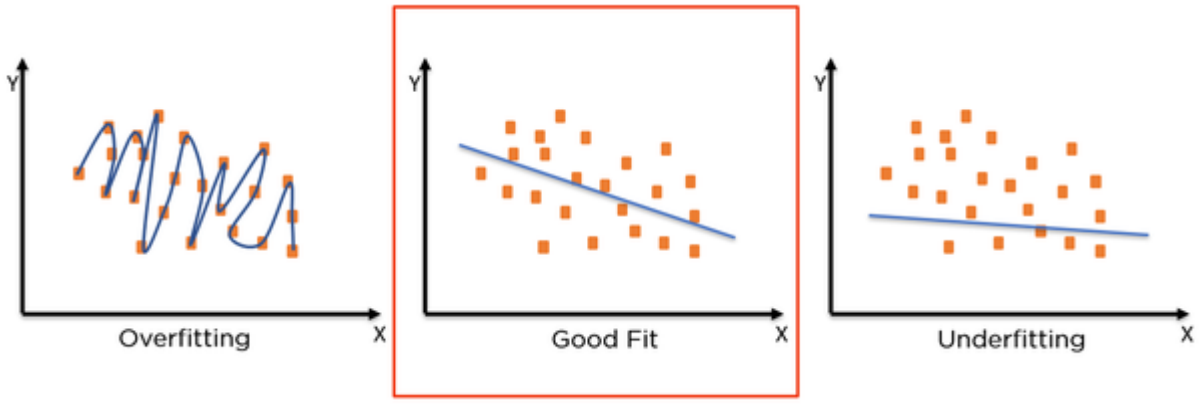
Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Sometimes while training a model, over fitting and under fitting of data occur. This leads to the inaccuracy of the trained model.

Regularization is the technique used to reduce the error by fitting a function appropriately on the given training set and avoid over fitting and under fitting.

Regularization techniques are used to calibrate the linear regression models in order to minimize the adjusted loss function and prevent over fitting or under fitting.



There are types of regularization methods:

1. **Ridge Regression**: It modifies the over fitted and under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

2. **Lasso Regression**: It modifies the over fitted and under fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.

14. Which particular algorithms are used for regularization?

Here are three main regularization techniques, namely:

- Ridge Regression (L2 Norm)
- Lasso (L1 Norm)
- Dropout

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.

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.

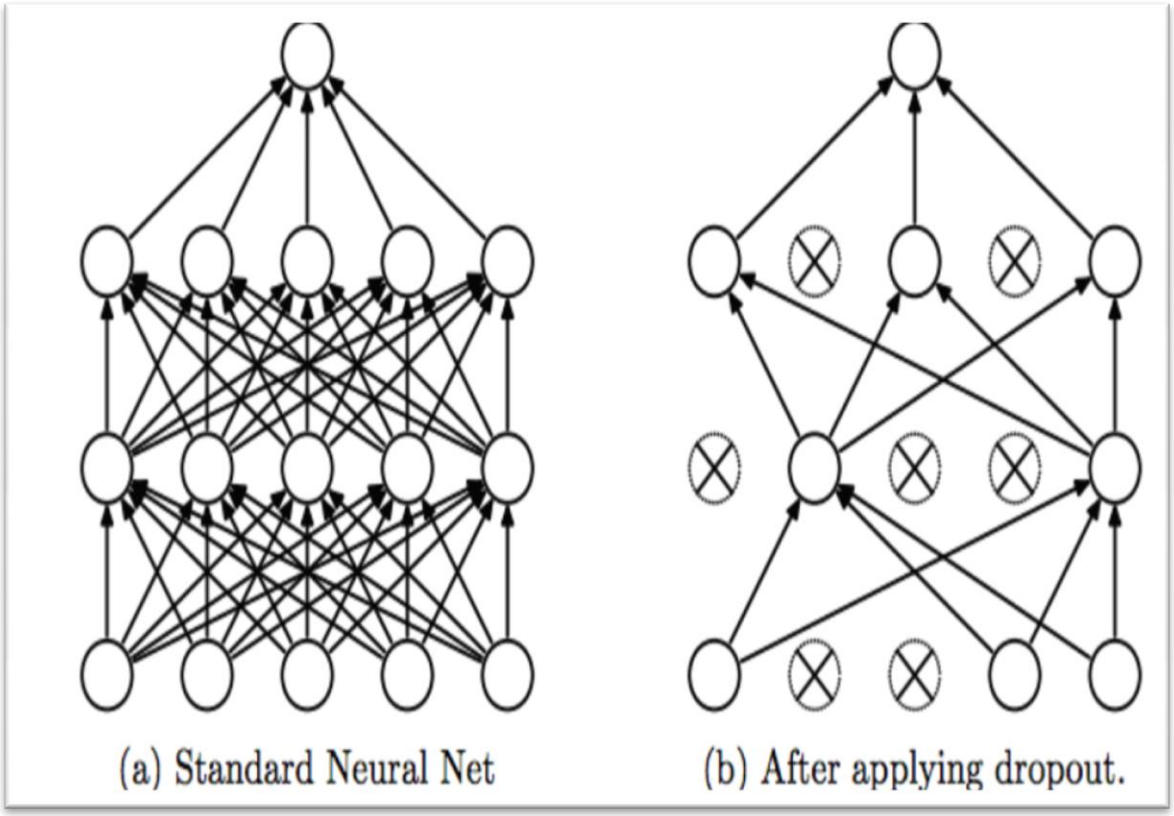
As loss function only considers absolute weights, optimization algorithms penalize higher weight values.

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 over fit 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.



With dropout, you are left with a reduced network as dropped out neurons are left out during that training iteration.

Dropout decreases over fitting 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 you make your model learn rather than just memorize.

Conclusion

Be it an over-fitting or under-fitting problem, it will lower down the overall performance of a machine learning model. To get the best out of machine learning models, you must optimize and tune them well. At info chips, we deliver machine learning services that help businesses optimize the utilization AI technology. We have machine learning capabilities across cloud, hardware, neural networks, and open source frameworks. Connect with our team to learn more about how machine learning can be useful to you.

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

This seems a sort of strange way to go about things. "Properties" are often grouped in different ways. But, for an ordinary least squares regression model, errors should be:

- Independent
- Identically and normally distributed
- Have mean 0
- Have constant variance

But this isn't so much about *utility* as assumptions. Sometimes, models with violated assumptions can be useful.

In addition, there are variations of regression that relax the assumptions.