INTERNSHIP

MACHINE LEARNING

ANSWERS

1. Maximum likelihood
2. Linear regression is sensitive to outliers.
3. Negative
4. Correlation
5. Low bias and high variance
6. Predictive modal
7. Removing outliers
8.
9. TPR and FPR
10. False
11. Apply PCA to project high dimensional data
12. (a)We don't have to choose the learning rate   (b) It becomes slow when number of features is very large
13. Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using regularization, we can fit our machine learning model appropriately on a given test set and hence the errors in it.
REGULARIZATION TECHNIQUE
There are two main types of regularization technique: Ridge regularization and Lasso regularization

14. There are three main regularizations techniques, namely:
    (a) Ridge Regression (L2 Norm)
    (b) Lasso (L1 Norm)
    (c) Dropout

Ridge and Lasso can be used for any algorithm 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.

Ridge Regression (L2 Regularization)

Ridge regression is also called L2 norm or regularization

When using this technique, we add the sum of weights 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 weight. Here normalized weight are in the form of squares.

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.

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 $1\text{-}p$ probability for each of the specified layers. Where p is called **keep probability parameter** and which needs to be tuned.