WORKSHEET SET – I Assignment
MACHINE LEARNING

QUESTION 1. 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

ANSWER: A) Least Square Error

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

ANSWER: A) Linear regression is sensitive to outliers

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

A) Positive B) Negative

C) Zero D) Undefined

ANSWER: B) Negative

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

ANSWER: B) Correlation

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

ANSWER: C) Low bias and high variance

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

ANSWER: B) Predictive modal

7.] Lasso and Ridge regression techniques belong to _________?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization

ANSWER: D) Regularization

MACHINE LEARNING

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

A) Cross validation B) Regularization

C) Kernel D) SMOTE

ANSWER: D) SMOTE

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

ANSWER: A) TPR and FPR

10.] In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

A) True B) False

ANSWER: A) True

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 stop words

D) Forward selection

ANSWER: B) Apply PCA to project high dimensional data

QUESTION 2. 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.

ANSWER: A), B) & C)

QUESTION 3. are subjective answer type questions, Answer them briefly?

13.] Explain the term regularization?

ANSWER:
Regularization is a technique used for tuning the function by adding an additional penalty term in the error function.

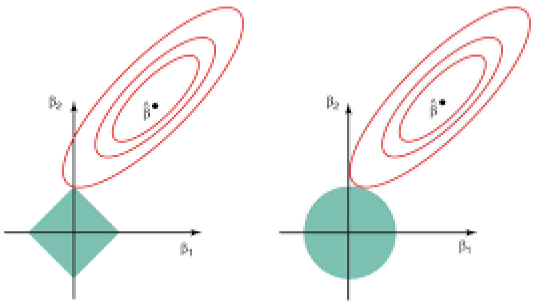
WORKSHEET SET – I Assignment

MACHINE LEARNING

Regularization consists of different techniques and methods used to address the issue of over-fitting by reducing the generalization error without affecting the training error much. Choosing overly complex models for the training data points can often lead to overfitting.

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. This technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting. A simple relation for linear regression.



There are two types of regularization as follows:

L1 Regularization or Lasso Regularization.

L2 Regularization or Ridge Regularization.

L1 Regularization or Lasso Regularization

L1 Regularization or Lasso Regularization adds a penalty to the error function. The penalty is the sum of the absolute values of weights.

$$Min(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + p\sum_{i=1}^{n}|w_i|)$$

p is the tuning parameter which decides how much we want to penalize the model.

L2 Regularization or Ridge Regularization

L2 Regularization or Ridge Regularization also adds a penalty to the error function. But the penalty here is the sum of the squared values of weights.

$$Min(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + p\sum_{i=1}^{n}(w_i)^2)$$

Similar to L1, in L2 also, p is the tuning parameter which decides how much we want to penalize the model.
This is known as regularization.

14.] Which particular algorithms are used for regularization?

ANSWER:

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.

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:

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

The original loss function is modified by adding normalized weights. Here normalize weights are in the form of squares.

The 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.

Thus, parameters are used to 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.

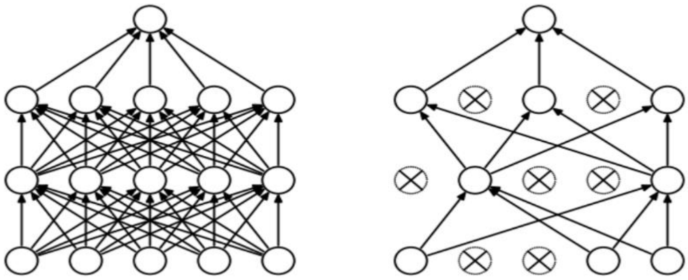
As loss function only considers absolute weights, optimization algorithms penalize higher weight values.

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, they only 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, can also be used as a regularization method.

For real-world applications, it is a must that a model performs well on unseen data.

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

ANSWER:

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables.

As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

WORKSHEET SET – I Assignment

MACHINE LEARNING

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters *e,* ε, or u.

An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model.

The error term is a residual variable that accounts for a lack of perfect goodness of fit.

Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely.

An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.

The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications.

For example, assume there is a multiple linear regression, function that takes the following form:

$$Y=\alpha X+\beta\rho+\epsilon$$

Were,
α, β=Constant parameters X,
ρ=Independent variables,
ϵ=Error term

Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed.

In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.

Points that do not fall directly on the trend line exhibit the fact that the dependent variable, in this case, the price, is influenced by more than just the independent variable, representing the passage of time. The error term stands for any influence being exerted on the price variable, such as changes in market sentiment.

The two data points with the greatest distance from the trend line should be an equal distance from the trend line, representing the largest margin of error.

If a model is heteroskedastic, a common problem in interpreting statistical models correctly, it refers to a condition in which the variance of the error term in a regression model varies widely.