**MACHINE LEARNING WORKSHEET-1**

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

To evaluate the best fit line, the most common method is the Least Square Method.

In this method, the regression line is calculated by minimizing the least squared error between the regression line and the data points.

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

LINEAR REGRESSION is sensitive to outliers and poor-quality data—in the real world, data is often contaminated with outliers and poor-quality data.

If the number of outliers relative to non-outlier data points is more than a few,

then the linear regression model will be skewed away from the true underlying relationship.

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

A) Positive B) Negative

C) Zero D) Undefined

Ans: B

A line that is trending downward from left to right has a negative slope. In other words, the line's rise to run ratio is a negative value.

this represents an inverse relationship between these x and y. A negative slope moves in the downward direction or is downward sloping.

When a line slopes down from left to right, it has a negative slope. This means that a negative change in y is associated with a positive change in x.

When you are dealing with data points plotted on a coordinate plane, a negative slope indicates a negative correlation and the steeper the slope, the stronger the negative correlation.

On the other hand, when a line slopes up from left to right, it has a positive slope. This means that a positive change in y is associated with a positive change in x. The steeper the slope, the greater the rate of change in y in relation to the change in x.

When you are dealing with data points plotted on a coordinate plane, a positive slope indicates a positive correlation and the steeper the slope, the stronger the positive correlation.

slope= rise/run, i.e., slope = (change in y)/(change in x)

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

In regression, the independent variable X is considered to have some effect or influence on the dependent variable Y.

Correlation methods are symmetric with respect to the two variables,

with no indication of causation or direction of influence being part of the statistical consideration.

Correlation is a statistic metric that measures the linear association between two variables. It treats y and x symmetrically.

Regression is setup to predict y from x. The relationship is not symmetric.

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:C

Overfitting occurs when the model fits more data than required, and it tries to capture each and every datapoint fed to it.

Hence it starts capturing noise and inaccurate data from the dataset, which degrades the performance of the model.

An overfitted model doesn't perform accurately with the test/unseen dataset and can’t generalize well.

An overfitted model is said to have 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

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

SMOTE is an oversampling technique. In layman terms, SMOTE will create synthetic data points for the minority class.

It creates new instances between the points of the minority class.

We can apply SMOTE on the training dataset using the imblearn library.

The imbalanced dataset is extremely common when handling real-world scenarios.

A machine learning model is not robust if it uses an imbalanced dataset for training purposes.

Therefore, a balanced dataset is preferred for training machine learning models.

Techniques such as under sampling, oversampling, and SMOTE can be used to create balanced data.

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

The Receiver Operator Characteristic (ROC) curve is an evaluation metric for binary classification problems.

It is a probability curve that plots the TPR against FPR at various threshold values and

essentially separates the ‘signal’ from the ‘noise’.

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

The higher the AUC(area under curve), the better the performance of the model at distinguishing between the positive and negative classes.

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

Ans: A

In this process they extract the words or the features from a sentence, document, website, email etc.

and then they classify them into the frequency of use. So, in this whole process feature extraction is one of the most important parts.

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?

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

13. Explain the term regularization?

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

and avoid overfitting (overfitting results in poor accuracy, prediction and generalization power: it sticks too much to the data and the

model has probably learned the background noise while being fit, which isn’t acceptable. By noise we mean the data points that don’t

really represent the true properties of your data, but random chance.)

It is here where the regularization technique comes in handy.

=> You penalize your loss function by adding a multiple of an L1(LASSO) or an L2(Ridge) norm of your weights vector w(it is the vector

of the learned parameters in your linear regression).

You get the following equation:

L(X,Y)+ λN(w)

λ is the tuning parameter that decides how much we want to penalize the flexibility of our model, lower the constraint (low lambda) on

the features, the model will resemble linear regression cost function.

(N is either the L1, L2 or any other norm)

The commonly used regularization techniques are:

L1 regularization

L2 regularization

-> A regression model which uses L1 Regularization technique is called LASSO(Least Absolute Shrinkage & Selection Operator)regression.

Lasso Regression adds “absolute value of magnitude” of coefficient as penalty term(λ) to the loss function(L). This type of

regularization(L) can lead to zero coefficients i.e., some of the features are completely neglected for the evaluation of output.

So lasso regression not only helps in reducing over-fitting but it can help us in feature selection.

-> A regression model that uses L2 regularization technique is called Ridge regression.

Ridge regression adds “squared magnitude” of coefficient as penalty term(λ) to the loss function(L). Ridge regression puts

constraint on the coefficients (w). The penalty term(λ) regularizes the coefficients such that if the coefficients take large values

the optimization is penalized. So, ridge regression shrinks the coefficients close to zero, and it helps to reduce the model

complexity and multi-collinearity. When lambda tends to zero, the cost function becomes similar to linear regression cost function.

14. Which algorithms are used for regularization?

Ridge Regression

LASSO (Least Absolute Shrinkage and Selection Operator) Regression

Elastic-Net Regression

Working of Ridge, LASSO, and Elastic-Net Regression

The working of all these algorithms is quite similar to that of Linear Regression, it’s just the loss function that keeps on changing!

Ridge Regression:

Ridge regression is a method for analyzing data that suffer from multi-collinearity.

Ridge regression adds a penalty (L2 penalty) to the loss function that is equivalent to the square of the magnitude of the coefficients.

The regularization parameter (λ) regularizes the coefficients such that if the coefficients take large values, the loss function is

penalized.

λ → 0, the penalty term has no eﬀect, and the estimates produced by ridge regression will be equal to least-squares

i.e., the loss function resembles the loss function of the Linear Regression algorithm. Hence, a lower value of λ will resemble a model

close to the Linear regression model.

λ → ∞, the impact of the shrinkage penalty grows, and the ridge regression coeﬃcient estimates will approach zero (coefficients are

close to zero, but not zero).

Note: Ridge regression is also known as the L2 Regularization.

To sum up, Ridge regression shrinks the coefficients as it helps to reduce the model complexity and multi-collinearity.

LASSO Regression:

LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction

accuracy of the model.

LASSO regression adds a penalty (L1 penalty) to the loss function that is equivalent to the magnitude of the coefficients.

In LASSO regression, the penalty has the eﬀect of forcing some of the coeﬃcient estimates to be exactly equal to zero when the

regularization parameter λ is suﬃciently large.

Note: LASSO regression is also known as the L1 Regularization (L1 penalty).

To sum up, LASSO regression converts coefficients of less important features to zero, which indeed helps in feature selection,

and it shrinks the coefficients of remaining features to reduce the model complexity, hence avoiding overfitting.

Elastic-Net Regression:

Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods

respectively.

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

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.