MACHINE LEARNING WORKSHEET-1

Q1. Ans- Both A & B

Q2. Ans-Linear regression is sensitive to outliers

Q3. Ans- Negative

Q4. Ans- Correlation

Q5. Ans- High bias and high variance

Q6. Ans- Descriptive model

Q7.Ans – Regularization

Q8. Ans- SMOTE

Q9. Ans- Sensitivity and Specificity

Q10. Ans- False

Q11. Ans- Apply PCA to project high dimensional data

Q12. Ans- A & B

Q13.Regulization: -

It is one of the most important concepts of machine learning. This technique prevents the model from overfitting by adding extra information to it.

It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique forces us not to learn a more complex or flexible model, to avoid the problem of overfitting.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

How does Regularization work?

Regularization works by adding a penalty or complexity term to the complex model. Let's consider the simple linear regression equation:

y= β0+β1x1+β2x2+β3x3+⋯+βnxn +b

In the above equation, Y represents the value to be predicted

X1, X2 …Xn are the features for Y.

β0, β1…βn are the weights or magnitude attached to the features, respectively. Here represents the bias of the model, and b represents the intercept.

Linear regression models try to optimize the β0 and b to minimize the cost function. The equation for the cost function for the linear model is given below:

$$\sum_{i=1}^{M}\left(y_i - y'_i\right)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n} \beta_j * Xij\right)^2$$

Now, we will add a loss function and optimize parameter to make the model that can predict the accurate value of Y. The loss function for the linear regression is called as RSS or Residual sum of squares.

Q14. Regularization Methods

In order to create less complex (parsimonious) model when you have a large number of features in your dataset, some of the Regularization techniques used to address over-fitting and feature selection are:

- L1 Regularization
- L2 Regularization

A regression model that uses L1 regularization technique is called Lasso Regression and model which uses L2 is called Ridge Regression.

The key difference between these two is the penalty term.

Ridge regression adds "squared magnitude" of coefficient as penalty term to the loss function. Here the highlighted part represents L2 regularization element.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Cost function

Here, if lambda is zero then one can imagine to get back OLS. However, if lambda is very large then it will add too much weight and it will lead to under-fitting. Having said that it's important how lambda is chosen. This technique works very well to avoid over-fitting issue.

Lasso Regression (Least Absolute Shrinkage and Selection Operator) adds "absolute value of magnitude" of coefficient as penalty term to the loss function.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Cost function

Again, if lambda is zero then one will get back OLS whereas very large value will make coefficients zero hence it will under-fit.

The key difference between these techniques is that Lasso shrinks the less important feature's coefficient to zero thus, removing some feature altogether. So, this works well for feature selection in case we have a huge number of features.

Traditional methods like cross-validation, stepwise regression to handle overfitting and perform feature selection work well with a small set of features but these techniques are a great alternative when we are dealing with a large set of features.

Q15. Error Present in Linear Equation:-

It is often said that the error term in a regression equation represents the effect of the variables that were omitted from the equation. 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.