

Machin learning worksheet 1

1. Answer is option A)Least square Error.
2. Answer is option A)Linear regression is sensitive to Outliers.
3. Answer is option B)Negative.
4. Answer is option B)Correlation.
5. Answer is option C)Low bias and high variance.
6. Answer is option B)Predictive model.
7. Answer is option B)Removing Outliers.
8. Answer is option D)SMOTE
9. Answer is option A)TPR and FPR.
10. Answer is option B)False
11. Answer is option B)Apply PCA to project high dimensional data.
12. Answer is option A)We don't have to choose the learning rate and option B)It becomes slow when number of features is very large.
13. Regularization parameter in python's Scikit-learn C parameter used to maintain regularization. Here C is the penalty parameter, which represents misclassification or error term. The misclassification or error term tells the SVM optimization how much error is bearable. This is how you can control the trade-off between decision boundary and misclassification term. A smaller value of C creates a small-margin hyperplane and a larger value of C creates a larger-margin hyperplane.
14. There are two main algorithms used for regularization.
 - Ridge Regularization - Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients. This means that the mathematical function representing our machine learning model is minimized and coefficients are

calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present.

- Lasso Regularization - It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients. Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the

coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients.

15. The term error present in linear regression equation 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. An error term essentially means that the model is not completely accurate and results in differing results during real-world applications.