Machine Learning Worksheet 1 Solutions

Question 1: Which of the following methods do we use to find the best fit line for data in Linear Regression?
Solution: A) Least Square Error

Question 2: Which of the following statement is true about outliers in linear regression?
Solution: A) Linear regression is sensitive to outliers

Question 3: A line falls from left to right if a slope is _____?
Solution: B) Negative

Question 4: Which of the following will have symmetric relation between dependent variable and independent variable?
Solution: B) Correlation

Question 5: Which of the following is the reason for over fitting condition?
Solution: C) Low bias and high variance

Question 6: If output involves label, then that model is called as:
Solution: B) Predictive modal

Question 7: Lasso and Ridge regression techniques belong to _____?
Solution: D) Regularization

Question 8: To overcome with imbalance dataset which technique can be used?
Solution: D) SMOTE

Question 9: The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?
Solution: A) TPR and FPR

Question 10: In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
Solution: B) False

Question 11: Pick the feature extraction from below:
Solution: B) Apply PCA to project high dimensional data

Question 12: Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
Solution:
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.

Question 13: Explain the term regularization?
Solution:

Regularization is a technique that is usually used to reduce the adjusted loss function and thus prevent Overfitting and underfitting.
Here we reduce the error by fitting a function appropriately on the given training dataset and avoid overfitting and underfitting.
There are 2 main types of Regularization techniques.

a) **Ridge Regression**
b) **Lasoo Regression**

Question 14: Which particular algorithms are used for regularization?
Solution:

a) **Ridge Regression – Here, the magnitude of coefficients is squared and added**
   It is given as:
   Cost function = Loss + λ* Σ ||w||^2
   Using this we can get a model where the regression line fits more accurately than the normal linear model.

b) **Lasso Regression: This uses shrinkage.**
   It adds a penalty that is equal to the absolute value of the magnitude of the coefficient

Lasso Regression is much better than Ridge Regression because it selects only some of the features and decreases the coefficients of others to zero(shrinkage)

Question 15: Explain the term error present in linear regression equation?
Solution:
The standard error of the model is the calculation which represents the distance that the observations are located from the regression line.
The lower the error the higher the accuracy of the model.
It gives the difference between the predicted value of our model and the actual values.