Machine Learning

1. A (Least Square Error)

2. A (Linear regression is sensitive to outliers)

3. A (Positive)

4. B (Correlation)

5. C (Low bias and high variance)

6. B (Predictive modal)

7. B (Removing outliers)

8. D (SMOTE)

9. A (TPR and FPR)

10. A(TRUE)

11. B (Apply PCA to project high dimensional data)

12. A) We don't have to choose the learning rate.

B) It becomes slow when number of features is very large

13. In Machine Learning term regularization refers to set of techniques that help to learn more than memorize data set. When you train a machine learning model and it is able to deliver correct result on training data set however provide relatively less result on test data set.

We can take an example for online marketing company, they want to build a model to predict if the user would buy product or not, based on customer history for last 10 days to show customer for digital advertisement. Customer usage history may include page visited, spent time, number of search done and etc..

Company build the model and it gives accurate result on existing data however when try to predict with current data it does not give a good result. So, it means model does more of memorization than learning. One possibility is that the model has overfitting problem, so it less performance on unseen data.

14. Regularization make machine learning algorithms learn not to memorize. When we use regression models to train some data there is a good chance that the model will overfit the given training data set. Regularization helps sort this overfiting problem by restricting the degrees of freedom of a given equation.

To Regularize the model, a shrinkage penalty is added to the cost function. There

Are three main regularization :-

1. Ridge Regression
2. LASSO (L1 Norm)
3. ELASTICNE (Less Popular)

15.   The error term of a regression equation represents all of the variation in the dependent variable not explained by the weighted independent variables. A regression equation is the formula for a straight line in this case, the best line through the scatterplot of data. If there were no error, all the data points would be located on the regression line to the extent they are not represents error.