# Machine learning worksheet1

Answer 1. **A) Least Square Error** (By use of this, we make sure, smallest sum of squares of errors should be there, which leads to make best fit line)

Answer 2**. A) Linear regression is sensitive to outliers** (any presence of outlier in data, makes change in slop of regression therefore, Linear Regression is sensitive to outliers)

Answer 3. **B) Negative**
(# A line goes up from left to right if a slop is positive, means line is increasing
# A line falls from left to right if a slop is negative, means line is decreasing
# When a line is horizontal the slop is zero, (constant function)
# when a line is vertical the slop is undefined.

Answer 4. **C) Correlation** (Relationship between x and y is systematic in case of Correlation but in case of regression it becomes non-systematic)

Answer 5. **C) Low bias and high variance** (a model which have High bias and low variance, will under-fit the target while a model with low bias and high variance will over-fit the target)

Answer 6. **B) Predictive Model**

Answer 7. **D) Regularization** (Lasso: Remove those columns which are seems not useful, Ridge, minimize the coef_ of features so that gap coming between coef_ may decrease and ElasticNet : combination of LASSO and Ridge).

Answer 8. **D) SMOTE** (Synthetic Minority Oversampling Technique)

Duplicate random records from the minority class.

Answer 9. **A) TPR and FPR** (It is a probability curve that plots the TPR against FPR at various threshold values)

Answer 10. **B) False** (The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes.)

Answer 11. **B) Apply PCA to project high dimensional data**

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

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

Answer 13. Explain the term regularization

**********Regularization**********

Regularization is a technique to prevent the model from over fitting by adding extra information to it.

This is a form of regression that regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of over-fitting.

After the training of the model, sometimes, model don't able to predict target value or doesn't perform well when deal with unseen data by introducing noise in the output, hence model is called over fitting. This problem can be solved by the use of regularization.

It mainly regularizes or reduces the coefficient of features toward zero. In simple words, in regularization technique, we reduce the magnitude of the features by keeping the same number of features.

The commonly used regularization techniques are:

1. L1 LASSO (Least Absolute Shrinkage and selection Operator)
2. L2 Ridge (Ridge Regression )
3. Elastic (Combination of L1 and L2)

L1 (LASSO): Remove those columns which are seems not useful and goes to eliminate the feature.

L2 (Ridge): Minimize the coef_ of features so that gap coming between coef may decrease.

Elastic: It is a combination of both LASSO and Ridge

Answer 14. Which particular algorithms are used for regularization

*******Algorithms are used for regularization**.*****

The commonly used regularization techniques are:

1. L1 LASSO (Least Absolute Shrinkage and selection Operator)
2. L2 Ridge (Ridge Regression )
3. Elastic (Combination of L1 and L2)

**L1 (LASSO)**: Remove those columns which are seems not useful and goes to eliminate the feature.

Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

**L2 (Ridge)**: Minimize the coef_ of features so that gap coming between coef may decrease.

Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values to be far away from the actual values.

**ElasticNet:** It is a combination of both LASSO and Ridge

Elastic Net is a popular type of regularized linear regression that combines two popular penalties, specifically the L1 and L2 penalty functions.

Answer 15:

*****Error present in linear regression equation*****

An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model. The error term is a residual variable that accounts for a lack of perfect goodness of fit.

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

1. measuring the distance of the observed y-values from the predicted y-values at each value of x;
2. squaring each of these distances;
3. Calculating the mean of each of the squared distances.

**Python code** :

```
From sklearn.metries import import mean_squred_error
print(mean_squred_error (y_test, y_predict))
```