Machine Learning Worksheet-1

Q.no	Solutions
1.	(B) O(n) (Simple Linear Regression where Y=MX+C)
2.	(B) Logistic Regression.
3.	(B) Gradient Descent.
4.	(C) Lasso
5.	(A)Stochastic Gradient Descent
6.	(B) False.
7.	(A) Scaling cost function half makes gradient descent converge faster.
8.	(D) None of these.
9.	(A,B) A:- We don't have to choose learning rates B:- It becomes slow when number of features are large.
10.	(A,C) A:-Linear Regression With high bias and low variance C:-Polynomial with degree 5 will have low bias and high variance
11.	(C,D) C:- It discovers casual relationship. D:-No inference can be made from regression line.

Q:12 Which algorithm will not suffer or might suffer, if the features in the training set have different scales?

Ans:- Feature scaling(Standard scaler) is a method through which all the independent variables/features on the same scale or to bring them on the same range usually the range of all the parameters are from -1 to 1 i.e normal distribution range. It is one the very important practice that needs to be followed as it helps the algorithm works faster and the optimal results are achieved.

Algorithms which will suffer if features are not scaled are the algos like K-Nearest Neighbors(KNN), Support vector Mechanism(SVM), Gradient Descent Algorithm.

KNN and SVM algos are the distance based algos.

 KNN usually works on the mechanism where it will check the nearest neighbor of the node and will categories as per the nearest neighbor. If the nodes are the not scaled it will tough be to locate the nearest node and will take large computational time to predict the result.

Algorithm which will not suffer if feature scaling is not done are Decision Tree Algorithms, Random Forest Algorithms.

As Decision tree algorithm works on root node concept it has nothing to do with the scaling of features similarly random forest algo is an ensemble technique which follows the concept of decision tree forest only and works on the same pattern as of decision tree only.

Q:13 Which Linear Regression training algorithm can we use if we have a training set with millions of features?

Ans:- For training large datasets where millions of features are present I will suggest to use Net Elastic Regression model as it works on the concept of LI(Lasso) [will vanish those features which contribute least in output variable] and L2(Ridge) [will give less importance to those features which contribute less in output variable] where net elastic will inherit the properties of both and will penalize those features with huge weightage to reduce its weight.

We can also use PCA techniques i.e Principal Component Analysis to reduce the number of variables when the variables are large it will still preserve the large and important information required to generate the output variable.

We can also use gradient descent algorithms also the main objective of the gradient descent algorithms is to find the minima of the function using the iterative process so as to reduce the cost function to its minimum value. Gradient Descent algos. Give the best optimal results as it minimize the cost function and the best results are achieved.