HomeWork -2

Name : Prakash Wagle
Student ID: 800889950

Algorithm : Implemented PLA algorithm to calculate weights to classify a given data set. Compared performance of PLA with and without weights been initialized using pseudo inverse method. pseudo inverse method is used to minimize the error in linear regression which can be also used to calculate initial weights for PLA algorithm.

Assumption : Data created is linearly separable

a. main() - This is a driver function which will automatically call generateData() pla() and pseudoinverse() method 100 times for each of the six data size ([10,50,100,200,500,1000]). It will also calculate average iterations and plot a graph for those average iterations.

b. [X,Y] =generateData(N)
❖ Uses rand() function to create N size uniformly distributed data between rang [-1,+1].
❖ Again use rand() function to get two more distributed points to construct line.
❖ Substitute input data [X] in line equation if result > 0 mark it +1 or of result < 0 mart it -1

c. [W,iters] = pla(X,Y,W0)
❖ First check whether initial weights exists or not | if it does not exists then initialize to W0= [0;0;0]
❖ Add X0 to input [X]
❖ Multiply input [X] with initial weights [W0] and take sign of the result [sign(X*W0)]
❖ Compare the result of above operation to label [Y] to check whether an ypoint is misclassified.
❖ If there exists any misclassified point then choose a random misclassified point to update the weights. Also update iterator counter.
❖ Continue this process till all points in input[X] are classified correctly
❖ Return calculated weights and no. of Iterations

d. [W] = pseudoinverse(X, Y)

First calculates the pseudo inverse of input data[X] and multiplies it with it's given label [Y] to get which can be used as initial weights [W] for PLA

e. test(X,Y,W)

Calculate sign of product of input and it's calculated weights [sign(X*W)] and compare with the given labels [Y], if all labels are matched correctly the test is passed.

Result of Experiment :

	No. of Iteration	
N	**Without Pinv**	**With Pinv**
10	8.72	6.09
50	67.59	56.08
100	118.39	95.26
200	174.12	147.35
500	756.98	589.37
1000	940.82	880.99

Without Pinv and With Pinv
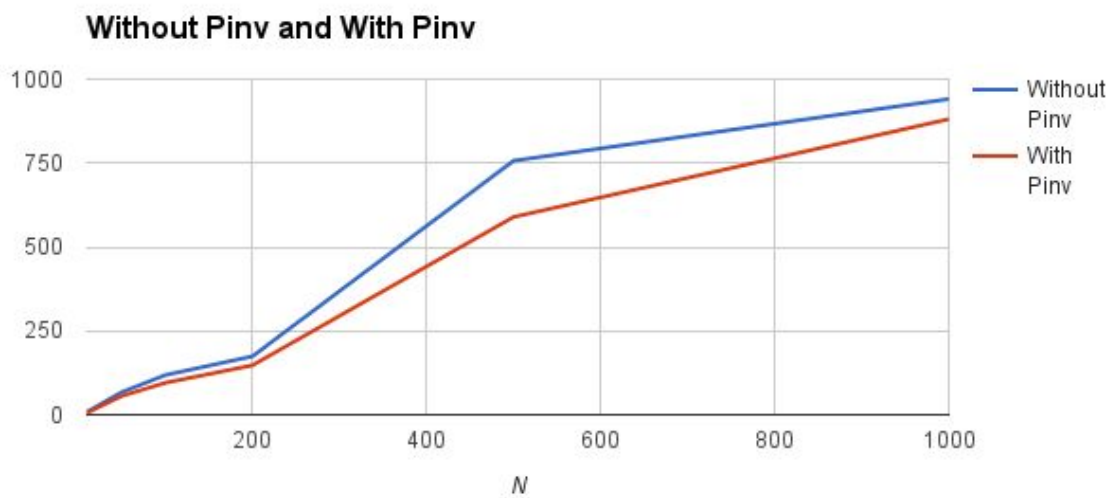
	Time	
N	Without Pinv	With Pinv
10	0.038223	0.037525
50	0.214103	0.157983
100	0.335507	0.33196
200	0.737472	0.664229
500	1.8741	1.572383
1000	3.267891	3.644986

Without Pinv and With Pinv
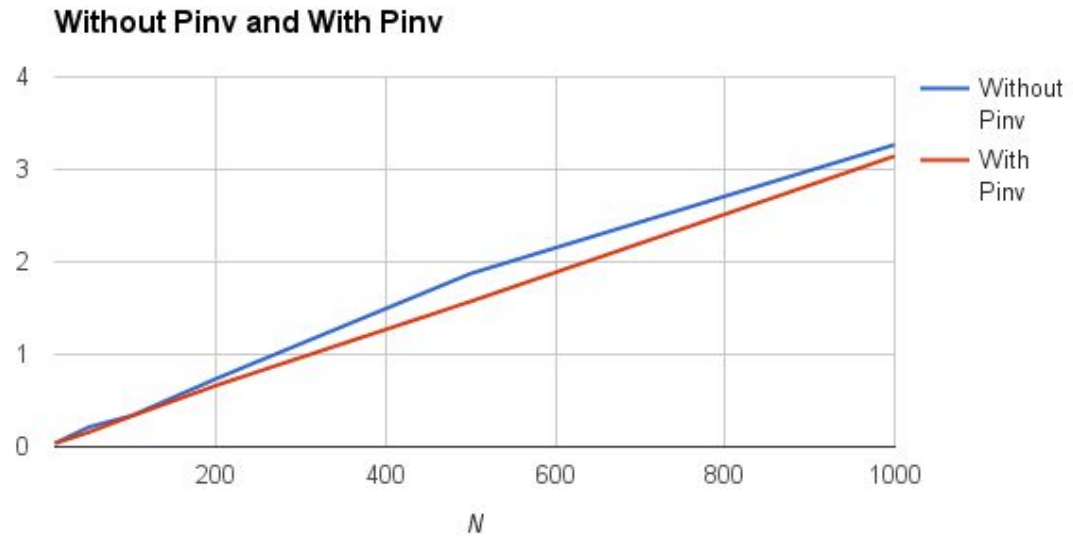
● **Observations :**

1) The above experiment was run 100 times on each of the following size data set Size=[10,50,100,200,500,1000] with weights initialized to zero [0;0;0] and again with weights initialized based on pseudo inverse calculations of input data set and it's label.

2) Number of Iteration and time increases as size of Data increases.

3) As shown above almost all time Iterations required by PLA(**Perceptron Learning Algorithm**) with Weights =0 were more than if they were initialized using pseudo inverse method.

4) Time taken by PLA with W=0 were more than PAL with pseudo inverse initialized weights

5) In some cases though number of Iterations for pseudo inverse initialized weights were more than the PLA weights initialized to zero.

6) Although initializing weights with pseudo inverse method may increase efficiency of PLA algorithm to calculate linearly separable weights it is not always true.