PLA learning algorithm (using Python) and analyze its performance using

different training sets.

The first program is used to generate the training data. It takes a three-dimensional vector

<w0,w1,w2> as input and generate data points x=<x1,x2> with sign (w1x1+w2x2+w0)>0 or

(w1x1+w2x2+w0)<0. If the sign is negative, the data point has label “-“. Otherwise it has label “+”.

Name this program “DataEmit”. It must be run using the following format:

DataEmit [w0,w1,w2] m n // [w0,w1,w2] specifies the line. m is the number of points with label

“+”. n is the number of points with label “-“. w0,w1,w2 are separated by just “,”. No extra space

is allowed. [] are used to surround the weight vector w and should be used as part of the

input.

The program must output a file named “train.txt”, which contains all the data points with

labels. Each point takes one line with format: x1 x2 label.

For example, if we test your program using the following command:

DataEmit <0,-1,1> 5 4

The output file “train.txt” may look like the following:

10 11.9 +

100.9 20 -

-1 0 +

7 29.8 +

0 99.8 +

0 -18.9 –

23.8 0 –

-45.6 -90.8 –

-6 29 +

Program 2 will take “train.txt” as input and then output the learned weight using PLA. It must

be named as “PLA”. To run it, we will use:

PLA train.txt

The output should be a weight vector <w0,w1,w2> and also a plot that contains all the training

data points and the line (represented by <w0,w1,w2>). The detailed format of the plot: blue

circle represents positive labeled data points and red cross refers to negative labels. The line

can be black (refer to the note about the PLA). No need to show the vector w. Just plot the line:

w1x1+w2x2+w0=0. Show the axis name (x1 or x2).

Each program will be tested using three input cases. Each case is 12 pts. In total: 72 points for

testing the two programs.

Once you finish the programs. Do the following experiments and record the results in the

report.

DataEmit [5,2,3] 10 10

DataEmit [5,2,3] 50 50

DataEmit [5,2,3] 100 100

DataEmit [5,2,3] 150 150

DataEmit [5,2,3] 200 200

For each training data, run PLA program and compared what you learned with the known

“line”. Analyze how the size of the training data affects the output of PLA?

Then, choose your own vector w and repeat the above experiment again. Test PLA’s

performance with 1) increase of the training data; 2) the ratio of the two labels in the training

data (balanced to unbalanced). Clearly describe your designed experiment. Use tables or figures

to summarize the results.

20 points for the report containing the above analysis.

8 pts for following all the instructions and a readme file containing any specific instructions for

running your programs. For example, what is the python version and what is the running

environment.