ECE/CS 559 - Fall 2021
Homework #2
Due: 09/15/2022, 11:00pm.

- You are allowed to discuss the homework problems with your classmates, but you are supposed to do your assignment individually.

- **READ VERY CAREFULLY:** You cannot use an existing machine learning / neural network library.

- **READ VERY CAREFULLY:** You need to submit a hard copy of your codes (in the same PDF file as your report) to Gradescope. The same code should be uploaded to Box (I shared you a separate link for code repository where you can upload) in the format XX-YYYYYYYYY-LastName.py, where XX is the homework number, and YYYYYYYYY is your UIC net ID. For example, for Homework 2, if your UIC net ID is 12345678, and your name is James Bond, you should name the file 02-12345678-Bond.py. Your single code file should be able to compile and execute without any errors. Submissions to Gradescope and Box should both be before the deadline. Failure to follow this procedure or submitting a non-functional code will be interpreted as failure of the programming assignment.

1. **(100 pts)** Write a computer program that runs the perceptron training algorithm with the step activation function $u(\cdot)$. Implement the following steps and report your results. Upload your code that provides these results as described above.

   (a) Pick (your code should pick it) w_0 uniformly at random on $[-\frac{1}{4}, \frac{1}{4}]$.

   (b) Pick w_1 uniformly at random on $[-1, 1]$.

   (c) Pick w_2 uniformly at random on $[-1, 1]$.

   (d) Pick $n = 100$ vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ independently and uniformly at random on $[-1, 1]^2$, call the collection of these vectors $\mathcal{S}$.

   (e) Let $\mathcal{S}_1 \subset \mathcal{S}$ denote the collection of all $\mathbf{x} = [x_1 \ x_2] \in \mathcal{S}$ satisfying $[1 \ x_1 \ x_2][w_0 \ w_1, \ w_2]^T \geq 0$.

   (f) Let $\mathcal{S}_0 \subset \mathcal{S}$ denote the collection of all $\mathbf{x} = [x_1 \ x_2] \in \mathcal{S}$ satisfying $[1 \ x_1 \ x_2][w_0 \ w_1, \ w_2]^T < 0$.

   (g) In one plot, show the line $w_0 + w_1 x_1 + w_2 x_2 = 0$, with x_1 being the "x-axis" and x_2 being the "y-axis." In the same plot, show all the points in $\mathcal{S}_1$ and all the points in $\mathcal{S}_0$. Use different symbols for $\mathcal{S}_0$ and $\mathcal{S}_1$. Indicate which points belong to which class. An example figure may be as shown in Fig. 1 (My labels look bad, I expect you to do a better job!).

   (h) Use the perceptron training algorithm to find the weights that can separate the two classes $\mathcal{S}_0$ and $\mathcal{S}_1$ (Obviously you already know such weights, they are w_0, w_1 and w_2 above, but we will find the weights from scratch, and the training sets $\mathcal{S}_0$ and $\mathcal{S}_1$). In detail,

      i. Use the training parameter $\eta = 1$.

      ii. Pick w'_0, w'_1, w'_2 independently and uniformly at random on $[-1, 1]$. Write them in your report.

      iii. Record the number of misclassifications if we use the weights $[w'_0, w'_1, w'_2]$.

      iv. After one epoch of the perceptron training algorithm, you will find a new set of weights $[w''_0, w''_1, w''_2]$.
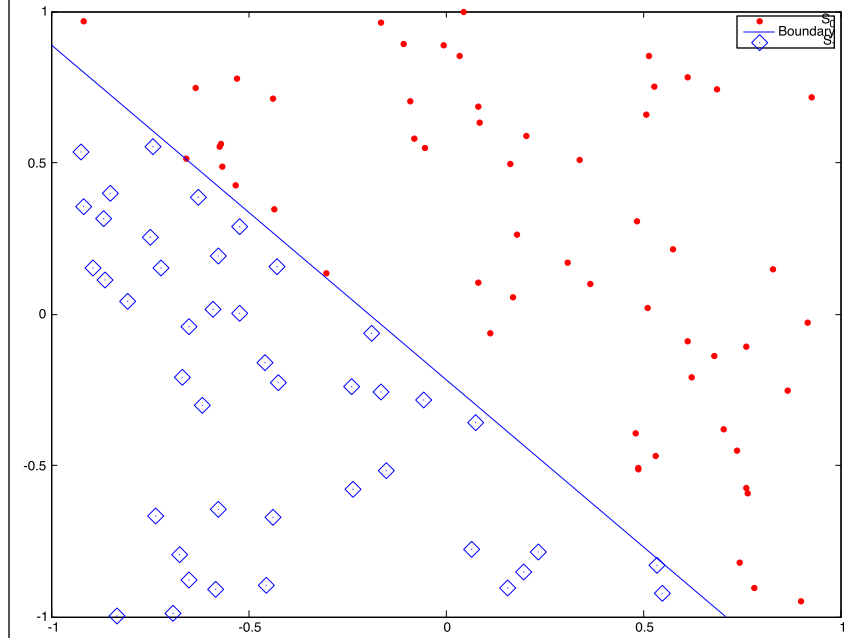
Figure 1: An example figure for Problem 1g.

  v. Record the number of misclassifications if we use the weights $[w_0'', w_1'', w_2'']$.

  vi. Do another epoch of the perceptron training algorithm, find a new set of weights, record the number of misclassifications, and so on, until convergence.

  vii. Write down the final weights you obtain in your report. How does these weights compare to the "optimal" weights $[w_0, w_1, w_2]$?

(i) Regarding the previous step, draw a graph that shows the epoch number vs the number of misclassifications.

(j) Repeat the same experiment with $\eta = 10$. Do not change $w_0, w_1, w_2, \mathcal{S}, w_0', w_1', w_2'$. As in the case $\eta = 1$, draw a graph that shows the epoch number vs the number of misclassifications.

(k) Repeat the same experiment with $\eta = 0.1$. Do not change $w_0, w_1, w_2, \mathcal{S}, w_0', w_1', w_2'$. As in the case $\eta = 1$, draw a graph that shows the epoch number vs the number of misclassifications.

(l) Comment on how the changes in η effect the number of epochs needed until convergence.

(m) Comment on whether we would get the exact same results (in terms of the effects of η on training performance) if we had started with different $w_0, w_1, w_2, \mathcal{S}, w_0', w_1', w_2'$.

(n) Do the same experiments with $n = 1000$ samples. Comment on the differences compared to $n = 100$.