Project 1: 2021 (Adaline, Back Propagation and Some Transfer Learning)

Deadlines: Part A and B are due by April 29 , but may be submitted by May 13 with 10% penalty and not accepted after May 13. Part C and Part D are due by May 13 but may be submitted until May 20 with a 10% penalty. It will not be accepted afterwards.

Remember there will be a second project.

I prefer you submit in pairs, but I will accept a group of up to 4. (Project 2 will be ONLY groups of 2.)

Collaboration: You may *NOT* share code between groups, nor should you use code from other sources, but you may consult between groups freely on difficulties. I suggest you use the Forum to share questions and you may even schedule zoom meetings for discussions.

Presentations of Results:

Enter a pdf file in the submission box (one for the group).

It should have on pageAB 1, your names and T.Z. numbers.

Page A 2 should have all the tables of part A. You should place your discussions and illustrations on Page A3. (you can add additional pages if needed.)

Page A4 should have your code and any additional information

Page B 2 should have all tables of part B followed by the discussions and then code, etc.

You should add a summary and discussion page.

Page C D1 should have your names and T.Z.

Page C2 has the results and tables for part C , followed by

Page C3 with all illustrations, discussions etc.

Page C4 will have appropriate codes, etc.

Page D2, etc will do the same for Part D.

If you do not succeed on all the parts, you should hand in what you have done before the final deadlines.

The details of the Homework:

x, y <= 100. The data is all data points <x, y> where x is of the form m/100 where m is an integer between -10000 and +10000 and y is of the form n/100 with n an integer between -10000 and +10000. Suppose that all data points with y >1 have the value 1; all other points have the value -1.

Now suppose you do not know this; but you are given a random sample of data of size 1000 together with its value (e.g. the point <601/100, 802/100> has value 1; while the point <8000/100, 70/100> has the value -1.

(Part A and Part B are due by **April 29)**

Part A. Implement the Adaline learning algorithm and show how it generalizes to develop a decision that works on all the set. That is, you randomly select a training data set which is a random sample of 1000 data points of all the data points. You use the training data set as input to the Adaline algorithm.

Now randomly select a "test" set of 1000 data points and see how well your trained Adaline performs. Pick a second random test set and see if the results change.

What can you conclude about your results? How

Does the accuracy of the result depend on the training set? Present tables and possibly a picture indicating your results. Discuss the impact on the size of training set and choice of test set.

Part B: Now change the problem so that points such that <x.y> has value 1 only if

4 <= x^2 + y^2 <= 9

What are the best results you obtain using an Adaline? Does the quality of the results change if you use more data? Present tables and perhaps a figure.

Part C and Part D are due by **May 13**

Part C: Now try the same with a back-propagation algorithm instead of the adaline.

You will have to define the architecture (i.e number of neurons and number of levels) You may either implement the algorithm or use a package. **BUT YOU WILL NEED TO LOOK INSIDE** the results of each neuron separately for Part D

Show a geometric diagram in terms of the **inputs** of the training set for the *output* of *each neuron separately* in the neural network as well as for the output neuron. Present tables of results both for training and testing.

Part D: Now use the trained neurons from the next to last level of Part C as input and only an Adaline for the output. (That is, you will give the adaline the output of the neurons from Part 3 in the level below the output, and train only the Adaline.)

Describe how accurate the Adaline can be. Give diagrams.

Draw whatever conclusions you think are appropriate from your results and report them.