CMSC 436: Artificial Intelligence

Fall 2025, Instructor: Dr. Milos Manic, http://www.people.vcu.edu/~mmanic

Project 2

CMSC 436: Artificial Intelligence
Project No. 2
Due Monday, Oct. 6, 2024 at noon

Student certification:
Team member 1:
Print Name: _____ *Date:* _____
I have contributed by doing the following: _____
Signed: _____ *(you can sign/scan or use e-signature)*

Team member 2:
Print Name: _____ *Date:* _____
I have contributed by doing the following: _____
Signed: _____ *(you can sign/scan or use e-signature)*

Team member 3:
Print Name: _____ *Date:* _____
I have contributed by doing the following: _____
Signed: _____ *(you can sign/scan or use e-signature)*

Pr.2.1 Perceptron-based classifier (10 pts)
In this assignment please use the datasets from Project 1. In the language of your preference (Python, Java, Matlab, C++), implement a perceptron-based classifier that will iterate until the **total error** is:

Dataset	Total Error
A	Epsilon $\varepsilon < 10^{-5}$
B	Epsilon $\varepsilon < 40$
C	Epsilon $\varepsilon < 700$

In addition, you may want to introduce a limit on the maximum number of iterations (let that be *ni*=5,000).
Please normalize the datasets first. Initialize your neuron using random values between (-0.5, 0.5).

Please use:
- **a)** Hard unipolar activation function
- **b)** Soft unipolar activation function

Note: For either activation function, you may need to experiment with different learning rates (alpha). Smaller alpha (especially with hard activation function) may yield smaller errors. When using a soft activation function, in addition to alpha, you can adjust the gain (smaller gain may result in more

iterations, but more "stable" convergence).

For activation function in a) perform the following steps for each of the datasets.

1. Choose 75% of the data for training and the rest for testing. Train and test your neuron. Plot the data and decision line for training and testing data (separately). Note and report the final Total Error (TE) of training. Create a confusion matrix for the testing dataset, as well as rates (true positive, false positive, etc), and compare those to the one from Project 1. (2 pts)
2. Choose 25% of the data for training and the rest for testing. Train and test your neuron. Plot the data and decision line for training and testing data (separately). Note and report the final Total Error (TE) of training. Create a confusion matrix for the testing dataset, as well as rates (true positive, false positive, etc), and compare it to the one from Project 1. (2 pts)
3. Compare **training** TE between steps 1 & 2 above. For **testing** datasets, compare accuracy, confusion matrices, and rates between steps 1 & 2 above. Answer the following questions: (1 pt)
    a. Are error rates different, and if so, why?
    b. What is the effect of different data sets and the effect of different training/testing distributions of TEs on the accuracy, confusion matrices, and rates (true positive, false positive, etc).?
    c. When would you go with step 1 and when with step 2 from above?
    d. Comment and discuss.

Repeat steps 1. through 3. **for activation function in b).**

Important: The data sets list the data points for both types of patterns ("small" car and "big" car).

Extra credit question: When given a data set, how would you approach selection of which data points to use for training, and which data points to use for testing? Note that the algorithm may fail if trained on one type of pattern, and tested on another, different type of pattern. (1 pt)

Pr. 2.2 Soft vs. hard activation function (5 pts)
Compare and discuss results obtained with hard unipolar activation, vs. the results obtained with soft unipolar activation function. You should include the plots and provide quantitative comparisons. (3 pts)
Comment on each training/testing data split (75% vs 25%, etc), for each data set (A, B, and C). (2 pt)

Note:

1. The code must be user-friendly. The TA must be able to test your code by simply executing it.
2. Project deliverable should be two files:
    a. Written report with all the plots and answers to all of the questions above, in **pdf** format.
    b. A zip file containing:
        i. Training/testing data sets as specified in Pr. 2.1 steps 1 & 2
        ii. Source code. For example, Python code can be in a Python notebook file (.ipynb) or a Python file (.py).
3. Submit your deliverables (Canvas). Please name files as:
    a. GroupName_Project2.pdf and
    b. GroupName_Project2.zip