CMSC 409: Artificial Intelligence

Fall 2020, Instructor: Dr. Milos Manic, http://www.people.vcu.edu/~mmanic

Project 1

CMSC 409: Artificial Intelligence
Project No. 1
Due Thursday, September ~~10~~ 15, 2020, 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.1.1

A) Understand and explore a data set (10 pts)

Three data sets (set A, B, and C) have been created following normally distributed classes. These data sets provide examples of car models where:

- The first column represents the cost in USD.
- The second column represents the weight in pounds.
- The third (last) column corresponds to the type (0 for small, 1 for big car).

Each data set contains 2,000 samples for each type.
Update: an alternate dataset was recently posted (Pr1_Data_ALT.zip). This dataset demonstrates possible scenarios slightly better. You may consider using that one, or the one originally posted (Pr1_Data.zip). Your choice will have no effect on the grade.

For each data set, do the following:

1. Plot the data for two vehicle types.
2. Estimate a separation line and draw it manually (by hand) on that plot. This line will be a linear separator (or decision function), which separates small and big cars. At this time, no running algorithm is needed (we will do that in next assignment).
3. Determine the mathematical definition of this linear separator
    a. This really is a definition of a neuron. Note: Think of the inequality we covered in class.
    b. Determine the weights and threshold. Comment.
4. Calculate false positives and false negatives (refer to confusion matrix).
5. Calculate accuracy, error, true positive rate and true negative rate, false positive rate and false negative rate.

6. Compare results for each data set and explain the differences. How are these datasets different?

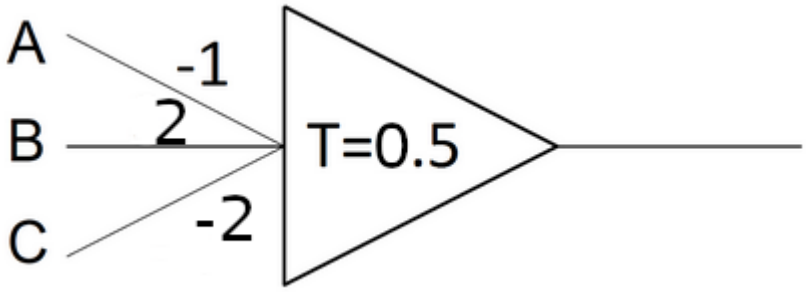
Important: Assume the example of true positive: the class is "it is a small car" and prediction is "small car"
Hint: Consider normalizing data.

B) **McCulloch-Pitts neurons (5 pts)**

1. Create a truth table for the artificial neuron below. What is the functionality of this neuron?
2. Given the same set of weights and the determined functionality of a neuron, what would be the range of possible values for threshold?

Note: Consider unipolar hard threshold activation function (possible inputs/outputs are obviously 0 & 1). Always start with the unit definition (net, output).

Hint: The truth table (similar to the one in class) should present inequalities that will evidence the functionality of a neuron (prove that it works as promised).



This assignment may be updated!

Note:

1. Compile all your deliverables into a single file (word or pdf).
2. Submit your file in Blackboard. Please name the zip file as GroupName_Project1.zip.