Statistical Learning and Fuzzy Logic Algorithms

CMSC 678

Project No. 1

Due Monday, September 24th, 2018, 2pm

1) **10 points for part 1)**

Create (in MATLAB) 20, 2-dimensional, normally distributed data with standard deviation 2, centered at [0; 0] for positive class, and 10, 2-dimensional, normally distributed data with standard deviation 2, centered at [5; 5] for negative class. Data should be created with a seed = 1. Train Perceptron with learning rate $\eta = 0.1$. Implement Perceptron update as given by Method 1 in the textbook. Initial weight vector $\mathbf{w} = [0\ 0\ 0]'$.

a) Show data and separation boundary in the **first graph**. How many epochs are needed?

b) What is the final weight vector **w**?

c) Run experiments with various learning rates η, say [1e-4 1e-3 1e-2 1e-1 1e0 1e 1e1 1e2 1e3 1e4]. Show in the **second graph** how number of epochs depends upon η.

d) Train linear neuron in a batch mode. Show its separation boundary in the first graph too.

Comment all the results in 1).

2) **5 points for part 2)**

Create an *outlier* at [20, 20] which belongs to negative class.

a) Train perceptron with $\eta = 1$. Show on the first graph from 1a) both the outlier and perceptron's separation boundary.

b) Train linear neuron in a batch mode again without regularization. Show its separation boundary too.

c) Train linear neuron in a batch mode again with penalty parameter $\lambda = 1$. Now the weight vector must be calculated as follows $\mathbf{w} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$. Show the new separation boundary too.

3) **10 points for part 3)**

Run 10-fold crossvalidation and find the best penalty parameter λ_{best}.

Using the λ_{best} design linear neuron and show its (best) separation boundary. What is the **w** now?

Comment all the results in 2) & 3).

Some hints:

SOFTWARE MUST BE USER FRIENDLY, so that I can run it easily too. At the top of your routine have the commands: `close all, format compact` *None of the calculations should be longer than 5 seconds on my laptop.*

Your report should STRICTLY be (in terms of everything; starting with 2 columns format up to fonts type and font size) in the form of IEEE journal (conference) paper. Use the template attached but **don't send to me your Word file, send to me PDF file**.

Submit both your written report and program to me by Email.

ZIP your report and programs into **a single zip file** (which will contain **max 2 files** (the report in PDF, and the MATLAB's m routine) name it with your family name (say, lee.zip) and send it to me. A **Subject** field in your Email MUST be **CMSC 678, Family name, Project 1**. Don't hesitate to contact me in the case of need. Use my **office hours –Tuesday 11am-12pm** (but, you can always drop by for up to 7 minutes questions and discussion. **Just knock on my door!**)

FINALLY: Any copying, "copying" or use in any form of somebody else's code or report will be treated as cheating and treated according to the VCU Honors Code.

My Email is: vkecman@vcu.edu.