Midterm Exam
DATA 440

Pablo Rivas

Assigned: Mar/7/18; Due: Mar/13/19; Points: 100

1 Instructions

This test could be written in LaTeX, just as all homework assignments. Write in understandable, easy to follow English. Make sure you provide good illustrations and figures. Remember to include your Python programs in your assignment.

Your assignment should be submitted in two ways: through GitHub, and in hardcopy (in class). Use the **same** repository you have been using and submit your work in a folder named "`lastname-midterm`", where lastname is your last name.

2 Problem Set

The following is a list of problems you will work on. When providing your solutions (hopefully using LaTeX), do not simply give the final answer, show how you arrived to the solution, justify your assumptions, and explain your results clearly.

1. Compare two algorithms on a classification task: the Pocket algorithm (designed for classification), and linear regression (not designed for classification). For linear regression, after learning the weights $\mathbf{w}$; we use $h(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x})$ to classify $\mathbf{x}$. For the dataset, start by using the following Python code:

```
import numpy as np
import matplotlib.pyplot as plt
from numpy import genfromtxt
# read digits data & split it into X (training input) and y (target output)
dataset = genfromtxt('features.csv', delimiter=' ')
y = dataset[:, 0]
X = dataset[:, 1:]
y[y<>0] = -1    #rest of numbers are negative class
y[y==0] = +1    #number zero is the positive class
# plots data
c0 = plt.scatter(X[y==-1,0], X[y==-1,1], s=20, color='r', marker='x')
c1 = plt.scatter(X[y==1,0], X[y==1,1], s=20, color='b', marker='o')
# displays legend
plt.legend((c0, c1), ('All Other Numbers -1', 'Number Zero +1'),
        loc='upper right', scatterpoints=1, fontsize=11)
# displays axis legends and title
plt.xlabel(r'$x_1$')
plt.ylabel(r'$x_2$')
plt.title(r'Intensity and Symmetry of Digits')
# saves the figure into a .pdf file (desired!)
plt.savefig('midterm.plot.pdf', bbox_inches='tight')
plt.show()
```

Create another dataset using the same method as above but now you will classify between 4 and the rest of the numbers. To do this you should change lines 8 and 9 as follows:

```
y[y<>4] = -1
y[y==4] = +1
```

Try the following three approaches using the dataset with the new changes, plot the final (best) hypothesis on each and then explain which works best in terms both E_{out} and the amount of computation required. E.g., what is the final classification error? after how many iterations did you stop the Pocket algorithm?

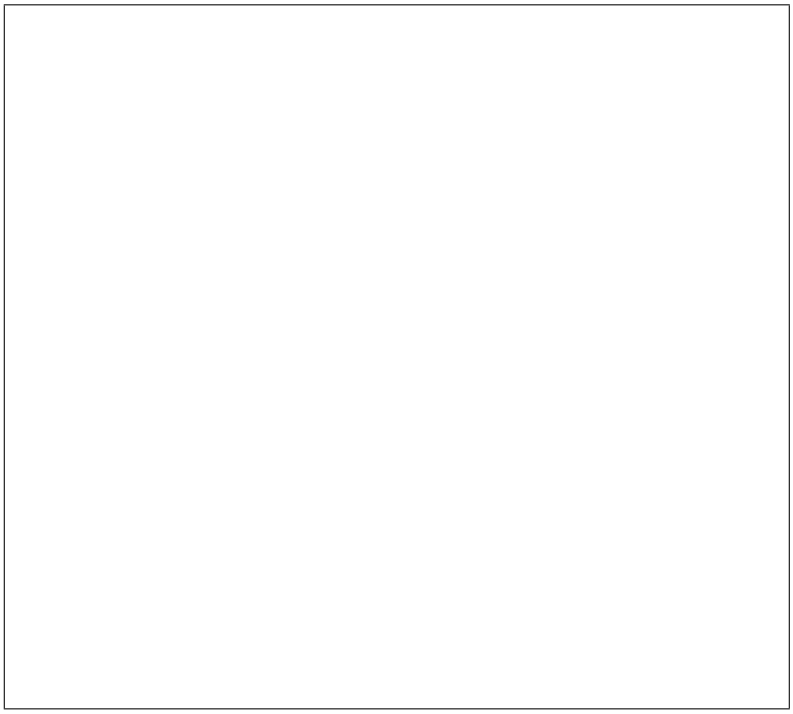
(a) The Pocket algorithm, starting from $\mathbf{w} = 0$.

(b) Linear regression (applied as a classification method).

(c) The Pocket algorithm, starting from the solution given by linear regression.

2. Write a Python program that solves **Problem 2.12** in an iterative manner. If you are feeling adventurous, plot the values of N as the program converges to a steady value of N.

3. (Extra Credit) Albert Einstein said "Creativity is intelligence having fun". You are very smart and talented. That is why you are in this class. With that in mind... I recently acquired the domain name `marist.ai` and I am looking to create site for students and faculty to share their AI-related projects; so, for a chance to be displayed our future website, draw some type of logo for the new site here[1]:



[1]For inspiration, check out the Mila logo here: `https://mila.quebec/` or the Montreal declaration logo here: `https://www.montrealdeclaration-responsibleai.com`