Amy Pitts
Assignment Number Midterm
Instructor: Dr. Pablo Rivas

Math 440L: Machine Learning
Due: March 13, 2019

Exercise List

Midterm

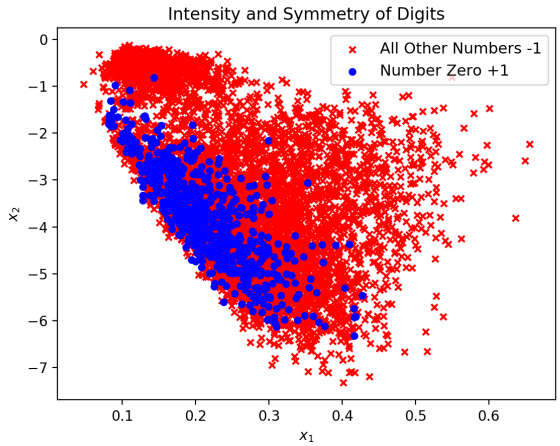
Exercise 1. *Compare two algorithms on a classification task: the Pocket algorithm (designed for classification), and linear regression (not designed for classification). For linear regression, after learning the weights w; ; we use $h(x) = sign(w^T x)$ to classify x. For the dataset, start by using the following Python code: In file labeled $midterm_generate.py$*
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 Eout and the amount of computation required. E.g., what is the final classification error? after how many iterations did you stop the Pocket algorithm?

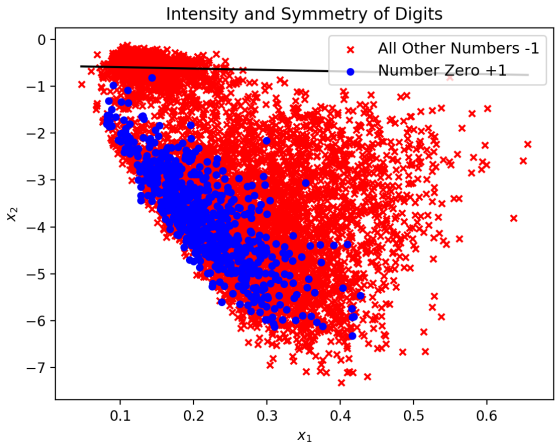
a) The Pocket algorithm, starting from $w = 0$

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

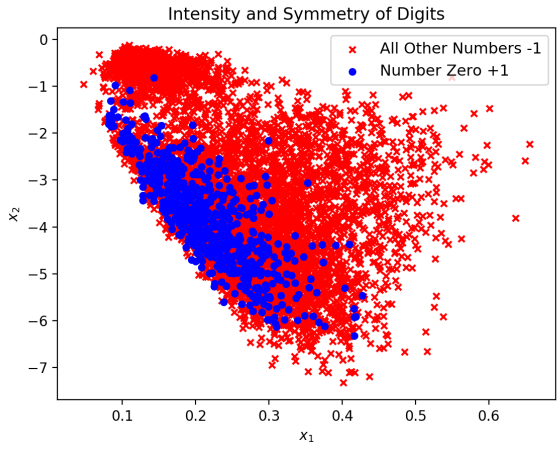
c) The Pocket algorithm, starting from the solution given by linear regression

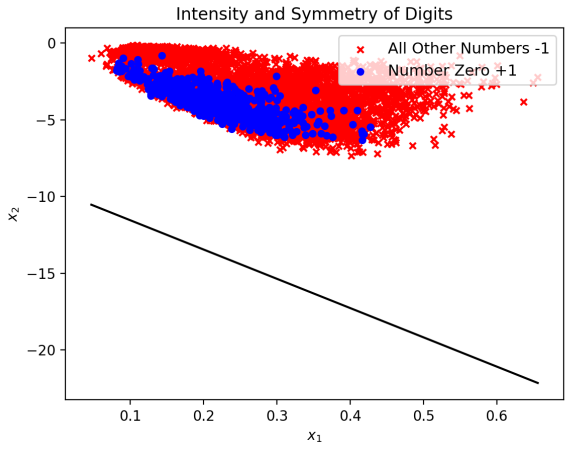
Solution. **a)** The number of iterations is 29 with a final classification error of 6182.



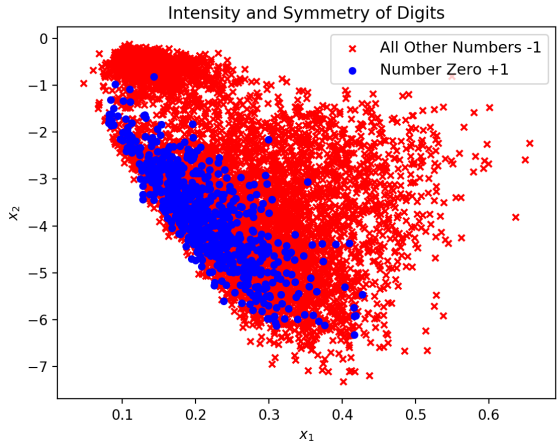


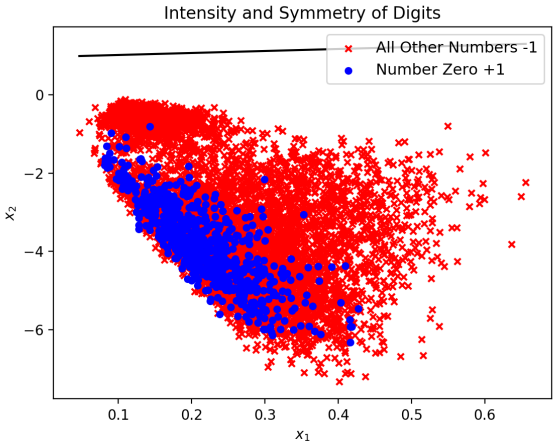
b) This was just plotting the linear regression on the data so there was no iterations.





c) The number of iterations is 29 with a final classification error of 652.





Summary: The linear regression produced a lower classification error but the graph itself does not make a whole lot of sense. This might be because in another dimensionality or with the line as a plane it would split the data in a way to classify 4. The code for this question is in q1.py.

Exercise 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.*

Problem 2.12: *For an H with* $d_{VC} = 10$*, what sample size do you need (as prescribed by the generalization bound) to have a 95% confidence that your generalization error is at most 0.05?*

Solution. Using the equation $N \geq \frac{8}{\epsilon^2} \ln\left(\frac{4((2N)^{d_{vc}}+1)}{\delta}\right)$ and plugging in we have

$$\begin{aligned} N &\geq \frac{8}{\epsilon^2} \ln\left(\frac{4\left((2N)^{d_{vc}}+1\right)}{\delta}\right) \\ &= \frac{8}{0.05^2} \ln\left(\frac{4(2N)^{10}+4}{0.05}\right) \\ &= 3200 \ln\left(81920N^{10}+80\right) \end{aligned}$$

We want to find the N which makes the equation true. We start with an initial seed of 20000. It does not mater what seed we start with, it will converge to the fixed point.

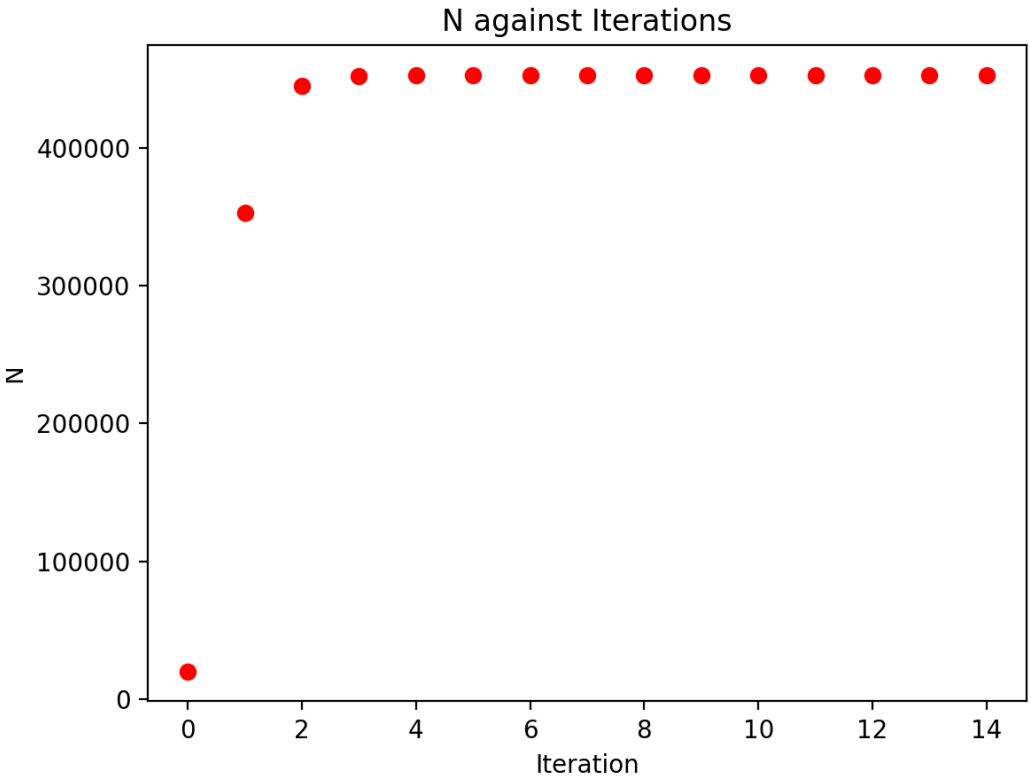
```
import numpy as np
import matplotlib.pyplot as plt


all_n = [] #empty list to save values of N
n = 20000 #initial starting place
all_n.append(n)
Sn = 3200 * np.log(81920*n**float(10)+80) #second N
all_n.append(Sn)


t=True
while(t):
    if(round(Sn,8) == round(n,8)): #checking if equal
        t=False
        print(Sn) #if equal then print
```

```
    else: #if not equal then recalculate
        n = Sn
        Sn = 3200 * np.log(81920*n**float(10)+80)
        all_n.append(Sn)
#plot the Ns
iterations = list(range(len(all_n))) #creating an x-axis
plt.plot(iterations,all_n, 'ro') #plotting the n against iteration
plt.title("N against Iterations")
plt.xlabel("Iteration")
plt.ylabel("N")
plt.show()
```

Using the code above the optimal number for N is 452957 rounding up when comparing 8 decimal places. Below is the graph of the values of N against the iteration. The graph starts at a low value of 20000 and then it levels of as it approaches 452957.



Exercise 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:*

Solution.


MARIST
MARIST
MARIST
MARIST