Midterm Exam

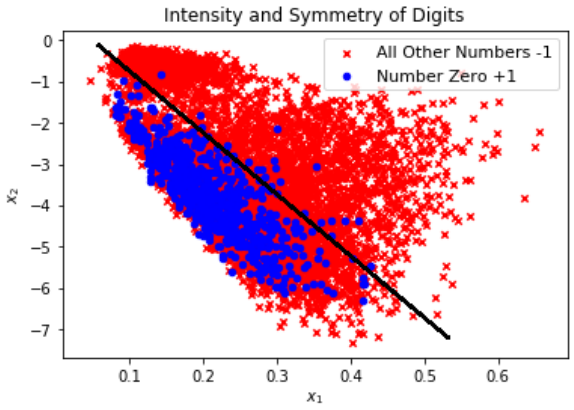
Max Schemitsch

3/13/2019

1 Problem 1

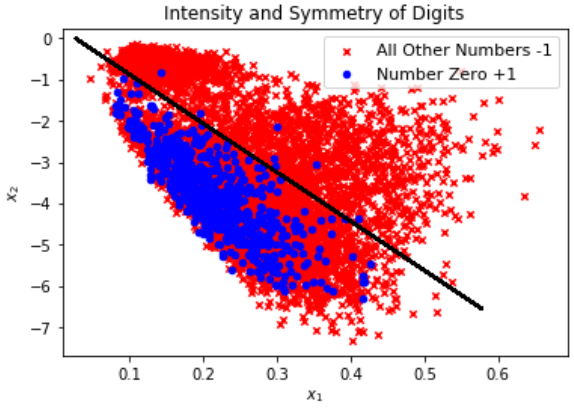
1.1 Pocket Algorithm, w = 0

Although I couldn't quite get it working, the code from *Midterm_Problem1a.py* is similar to what would be required for classification. This uses the PLA algorithm to separate $y = 4$ and $y \neq 4$. The resulting classification would look something like this:



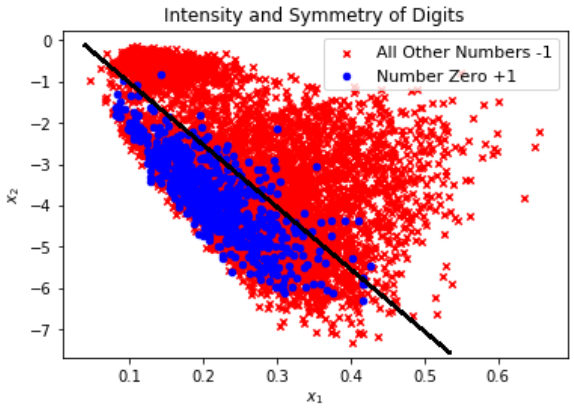
1.2 Linear Regression

Similar to Part A, I couldn't get the code working. *Midterm_Problem1b.py* replicates a program similar to what would classify this dataset. The result would look something like this:



1.3 Pocket Algorithm, Start with Linear Solution

Again, I couldn't figure out the code on this one, but I have something that is close to what we would want. *Midterm_Problem1c.py* is similar to what would be needed for classification. Our resulting plot would look something like this:



1.4 Comments

Even though I couldn't actually run the code, I think the technical best algorithm would be the Pocket Algorithm starting with Linear Regression. This is because it is two programs in one, and may provide a middle-ground solution. Alternatively, the fastest algorithm, as we've discussed in class, is the Linear Regression algorithm. The amount of code and processing time is significantly reduced.

2 Problem 2

We can use the code below (from HW2) to solve Problem 2.12 in an iterative manner.
I've also added a list, *storedN*, that stores values of N as it iterates.

```
import numpy as np
storedN = []
def get_N(dvc=10, delta=0.05, epsilon=0.05, initial_N=1000, tolerance = 1):

    new_N = 8 / epsilon**2 * np.log((4 * ((2 * initial_N)**dvc + 1)) / delta)
    storedN.append(new_N)

    if abs(new_N - initial_N) < tolerance: # Did it converge?
        return new_N

    else: # If so return N
        return get_N(dvc, delta, epsilon, new_N, tolerance) # Iterate

print("Our sample size must be at least " + format(int(get_N())) + ".")
print("Sample Size Estimates: " + str([int(x) for x in storedN]) + ".")
```

```
Our sample size must be at least 452956.
Sample Size Estimates: [257251, 434853, 451651, 452864, 452950, 452956, 452956].
```

3 Problem 3 (Extra Credit)

I like the idea that Marist contains both letters a and i, and in order.
I'm awful at drawing (especially on a computer), but maybe something like this: