CS434: Machine Learning & Data Mining (Spring 2016)
Implementation Assignment 2
Muratcan CICEK
(No Group)

NOTES:

- I used Python for the implementation of this assignment and I printed some of my experiments, I also plotted some results. They are very below.

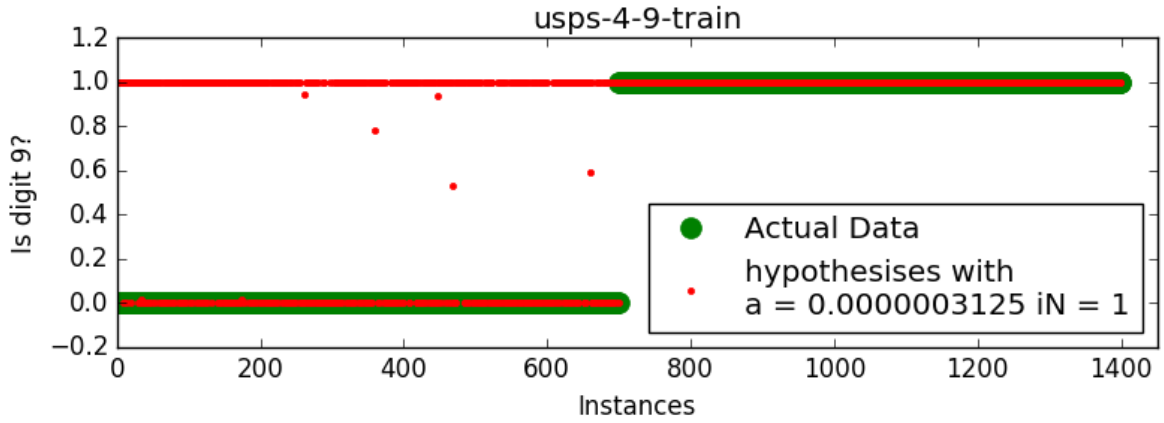
1. I have Implement the batch gradient descent algorithm in 'HW2' as gradientDescentWithLR(x, y, alpha, iterationNumber) where x is a Python Numpy Array of the training images and y is actual labels like described in the assignment document. I have also implemented functions to experience different learning rates and different iteration numbers I have run these experiences in a nested way and I have concluded the following results:

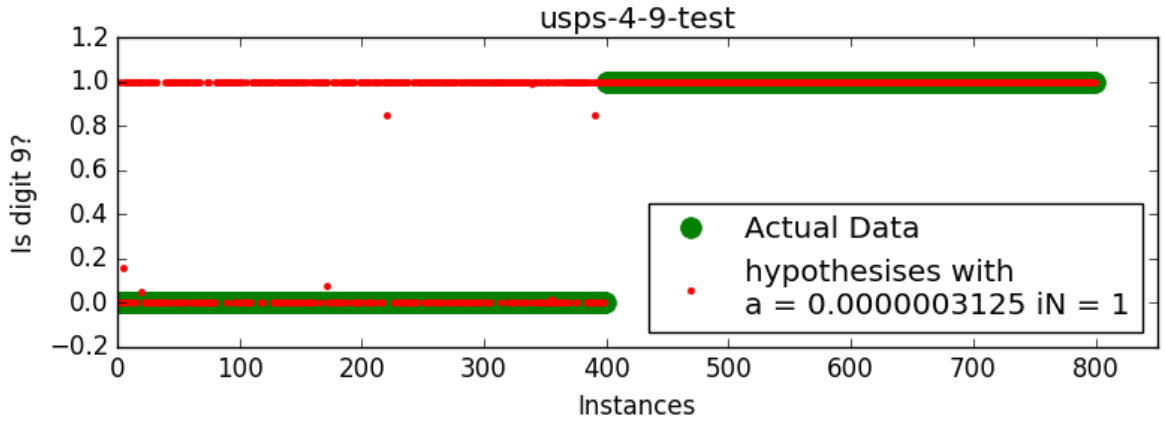
FIXED ITERATION NUMBERS	OPTIMAL LEARNING RATES	CORRESPONDING LOSS VALUES
1	0.0000003125	402.82
10	0.05	55.00
100	0.00000625	4.00
1000	0.05	0.00
10000	0.05	0.00

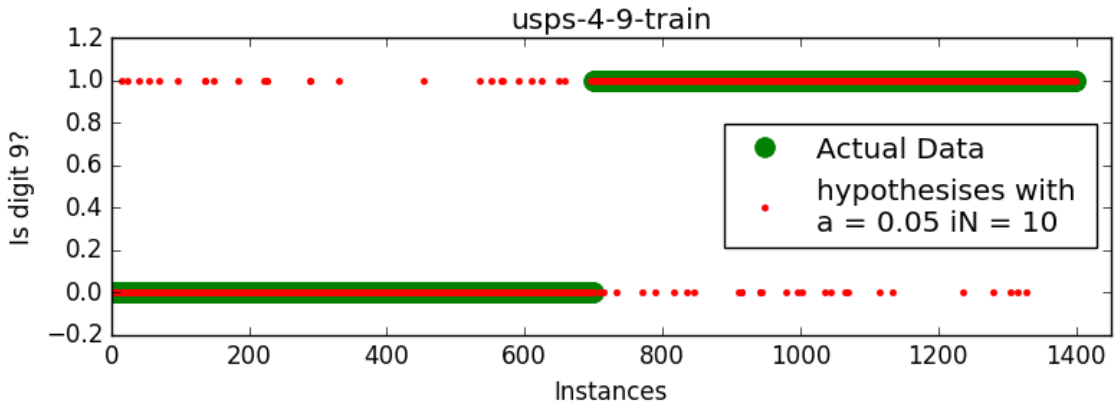
According to these results, 0.05 is an optimal Learning Rate when we use iterate the gradient descent 1000 timely.

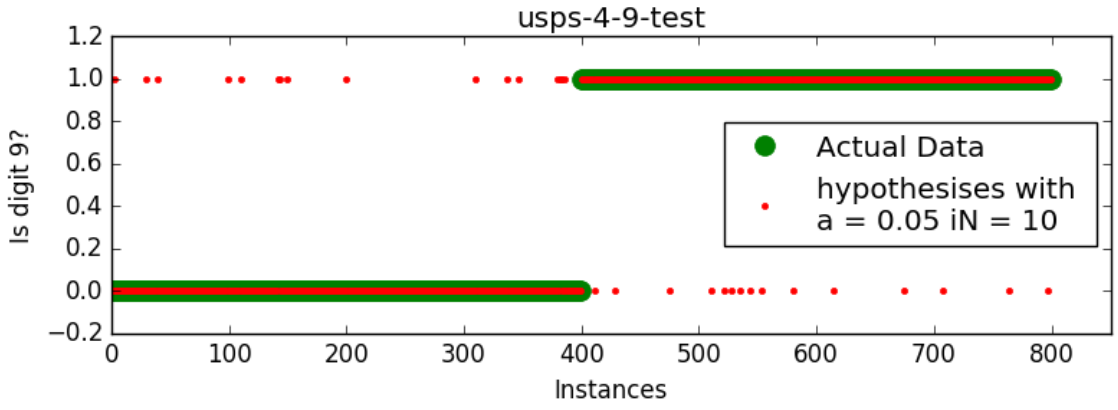
2. PLOTS

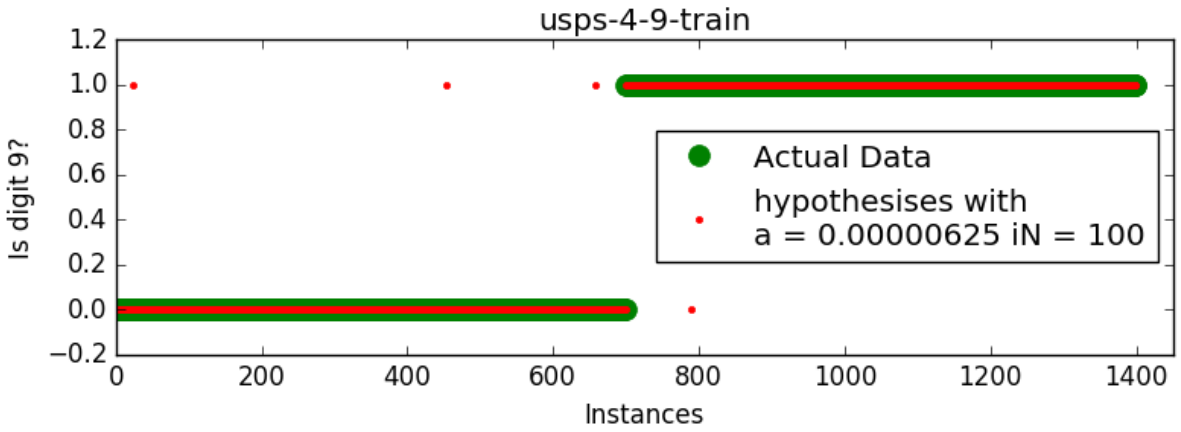
I have reran the gradient descent when I observed the optimal learning rating with corresponding iteration numbers from the experiences that I mentioned above. Then, I have done predictions for both, the training and testing data with the weight vector that each gradient descent returned and I have plotted them like below.

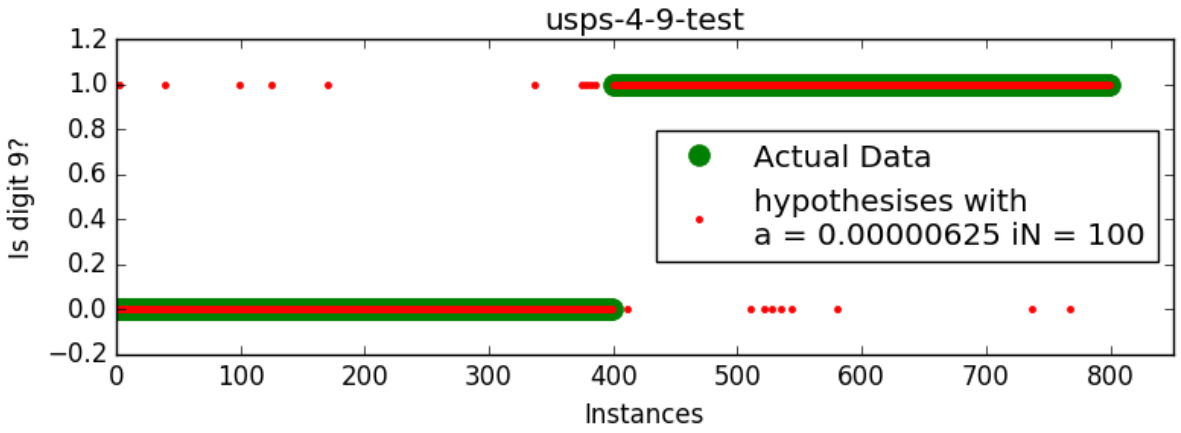

usps-4-9-train
Is digit 9?
Instances
Actual Data
hypothesises with
a = 0.0000003125 iN = 1

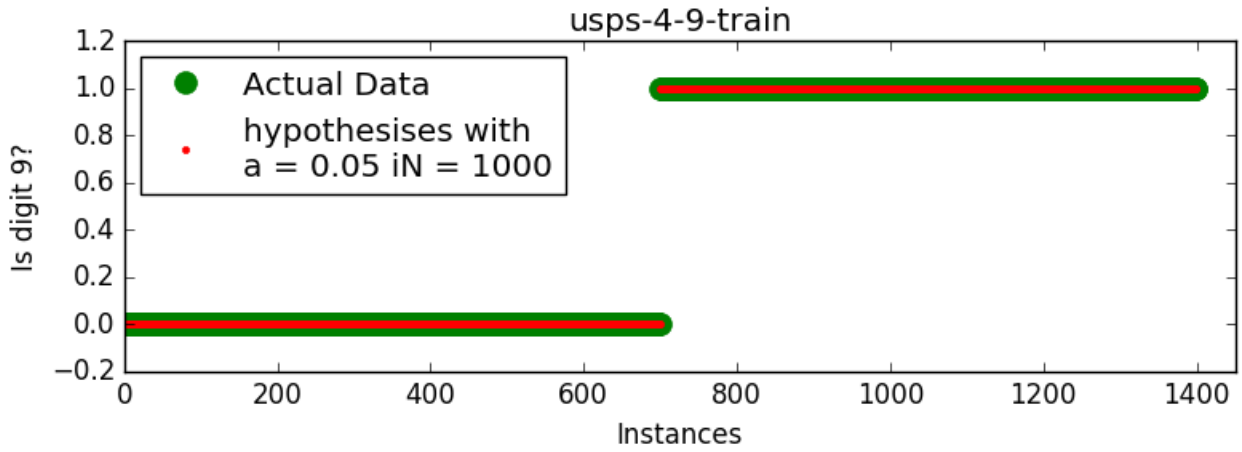

usps-4-9-test
Is digit 9?
Instances
Actual Data
hypothesises with
a = 0.0000003125 iN = 1

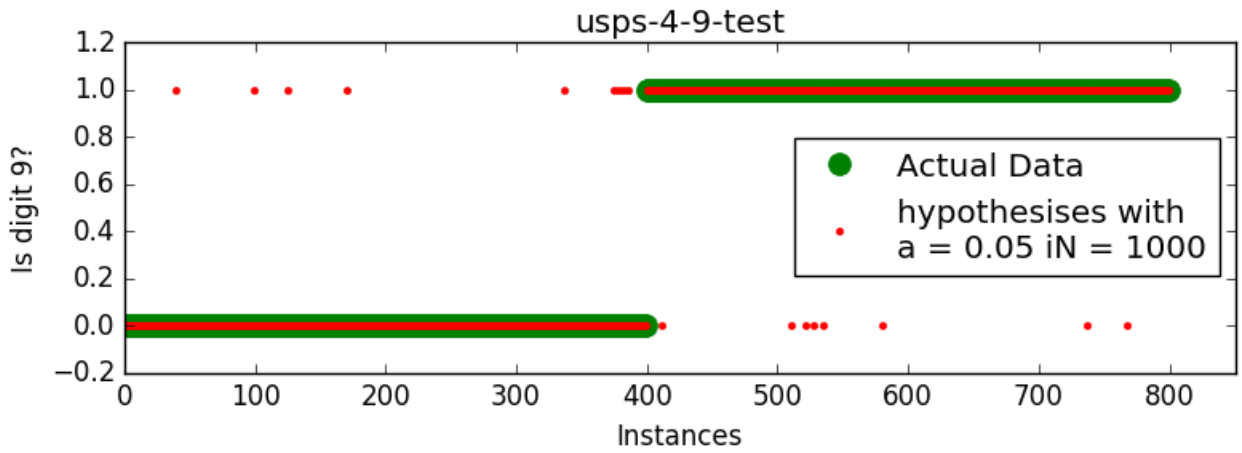

usps-4-9-train
Is digit 9?
Instances
Actual Data
hypothesises with
a = 0.05 iN = 10


usps-4-9-test
Is digit 9?
Instances
Actual Data
hypothesises with
a = 0.05 iN = 10


usps-4-9-train
Is digit 9?
Instances
Actual Data
hypothesises with
a = 0.00000625 iN = 100


usps-4-9-test
Is digit 9?
Instances
Actual Data
hypothesises with
a = 0.00000625 iN = 100


usps-4-9-train
Is digit 9?
Instances
Actual Data
hypothesises with
a = 0.05 iN = 1000

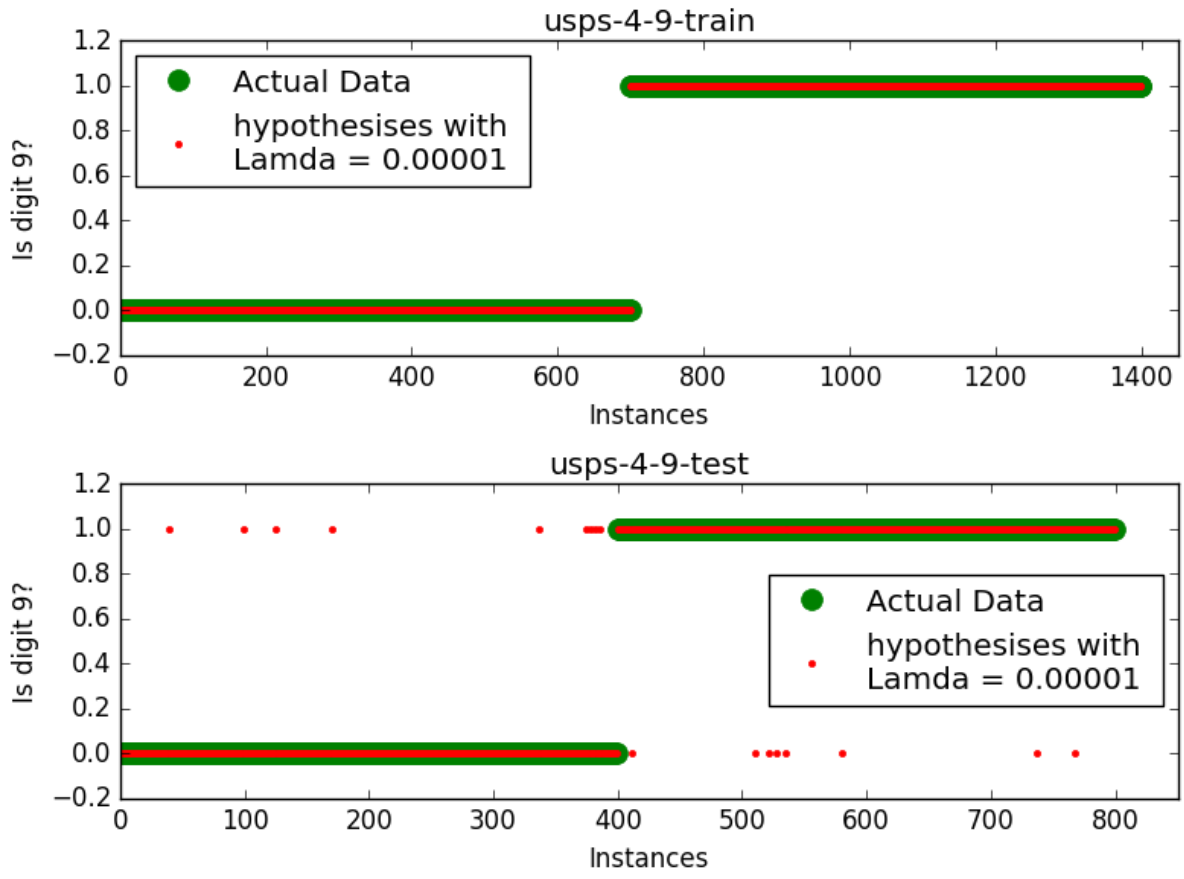

usps-4-9-test
Is digit 9?
Instances
Actual Data
hypothesises with
a = 0.05 iN = 1000

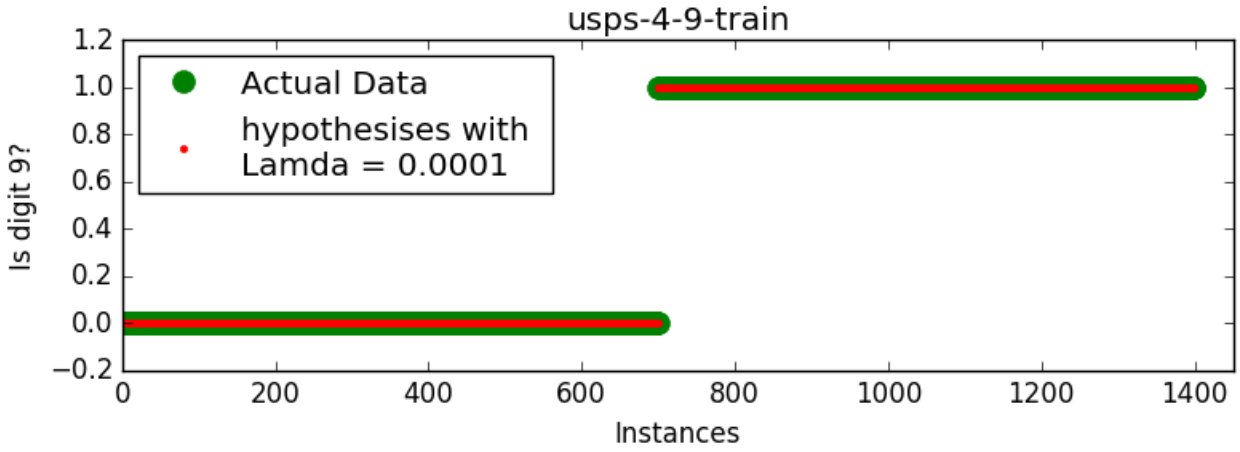
3. **Pseudocode**

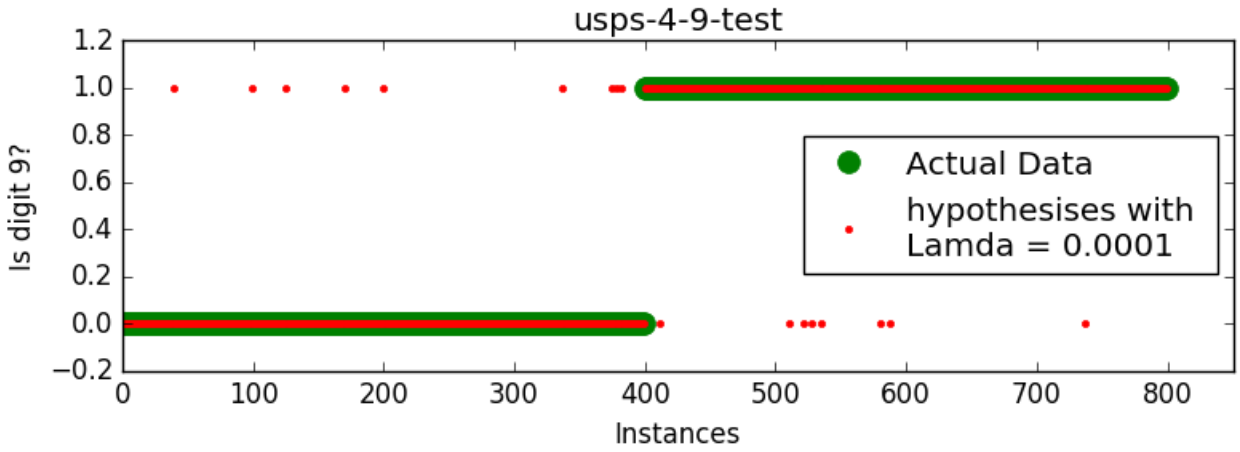
Given: training examples (x^i, y^i),i = 1,…,N
Let w = (0,0,0,…,0)
Repeat until convergence
    d = (0,0,0,…,0)
    for i = 1 to N do
        $y'^i = \frac{1}{1+e^{-w \cdot x^i}}$
        error = $y^i - y'^i$
        d = d + error*x + λ*w
    w = w + ηd

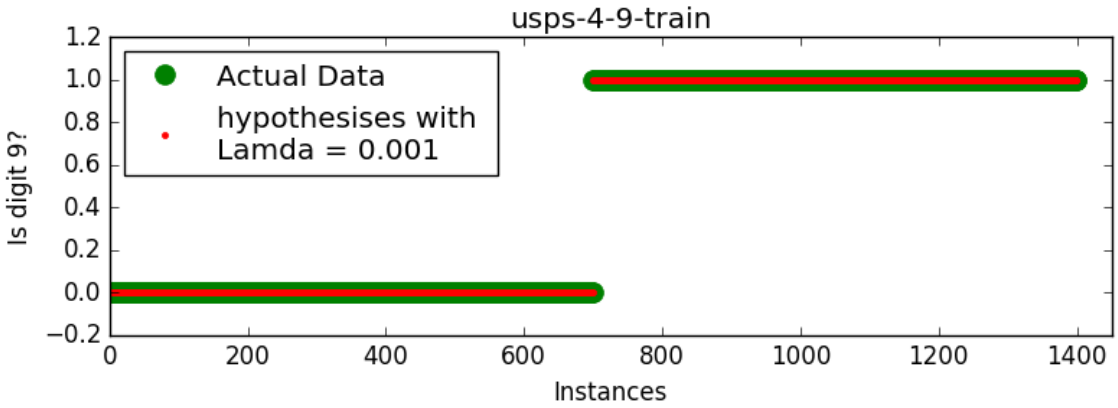
4. **Plots**

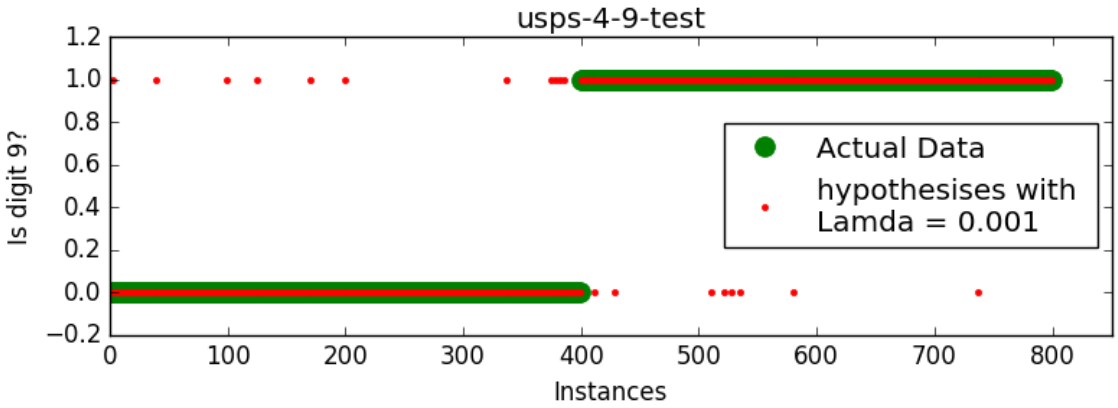
Honestly, I did a lot of implementation to try different λ values (10^{-5} to 10) and plotted the predictions as below. But I was not able to detect the difference, when λ < 1 there were a few different predictions and when λ >= 1, I got Python nan values on the weight vector and no predictions.

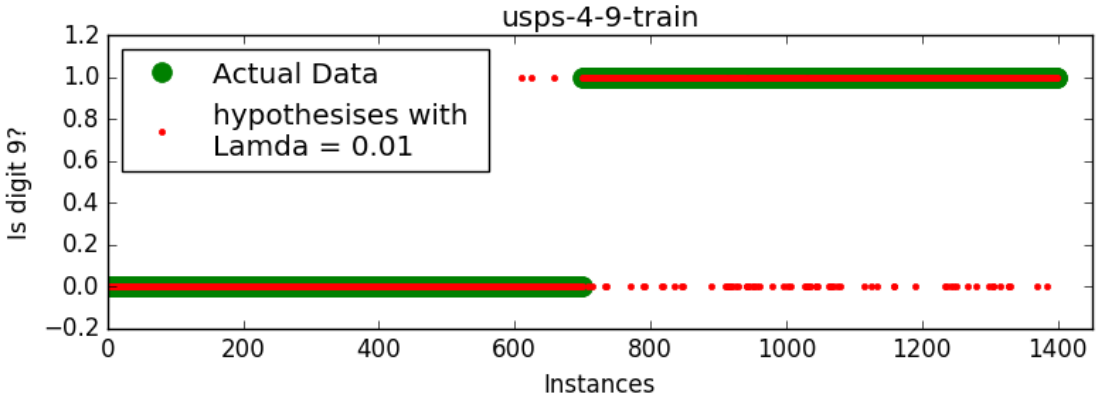


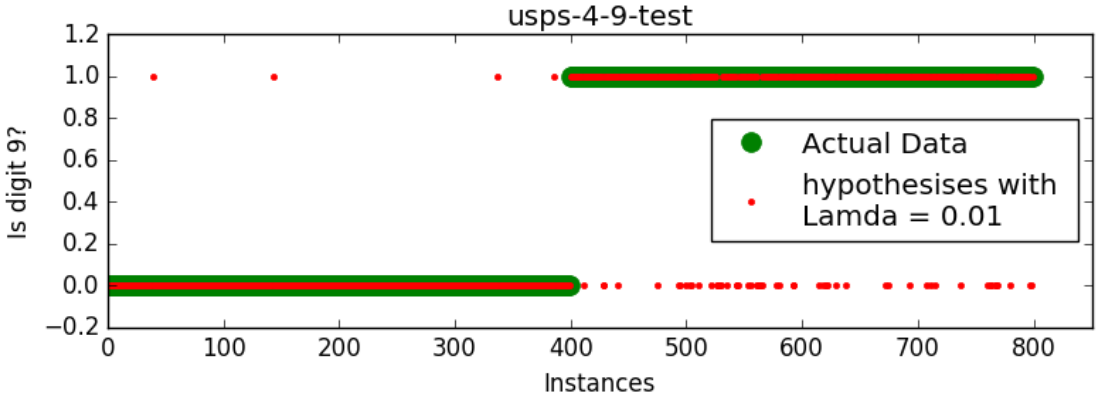

usps-4-9-train
Actual Data
hypothesises with
Lamda = 0.0001
Is digit 9?
Instances
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
200
400
600
800
1000
1200
1400

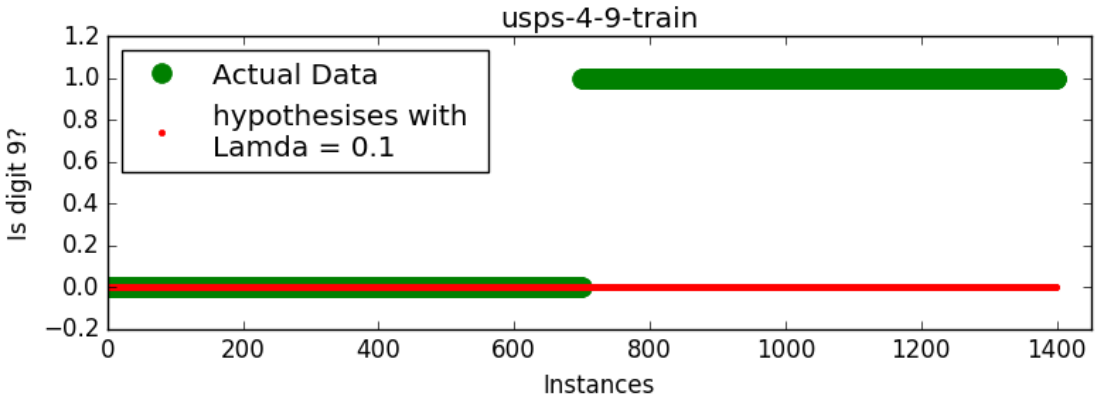

usps-4-9-test
Actual Data
hypothesises with
Lamda = 0.0001
Is digit 9?
Instances
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
100
200
300
400
500
600
700
800

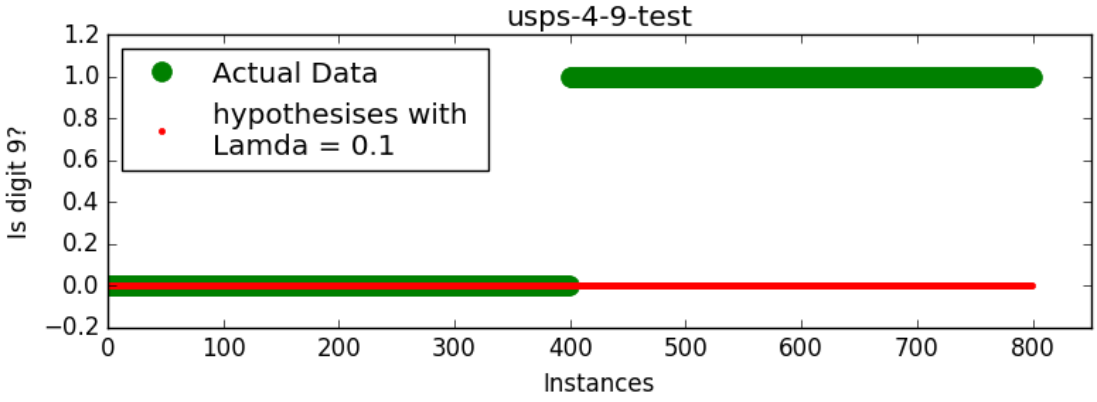

usps-4-9-train
Actual Data
hypothesises with
Lamda = 0.001
Is digit 9?
Instances

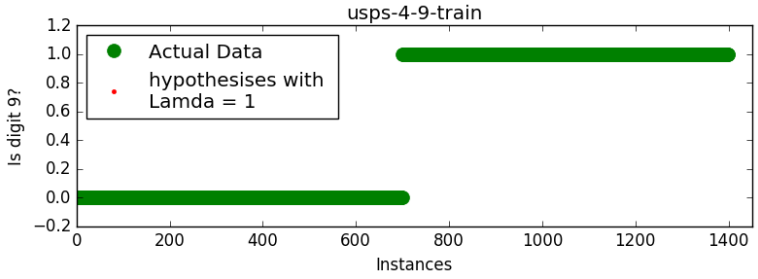

usps-4-9-test
Actual Data
hypothesises with
Lamda = 0.001
Is digit 9?
Instances

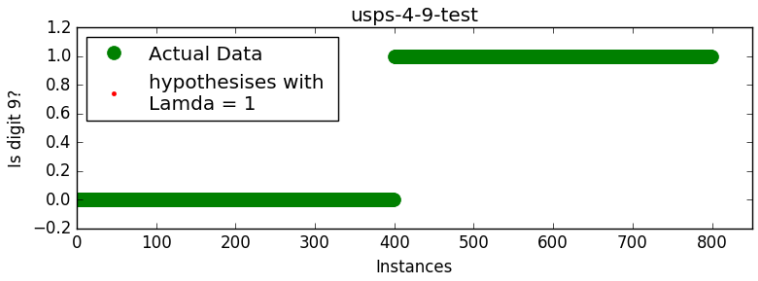

usps-4-9-train
Actual Data
hypothesises with
Lamda = 0.01
Is digit 9?
Instances

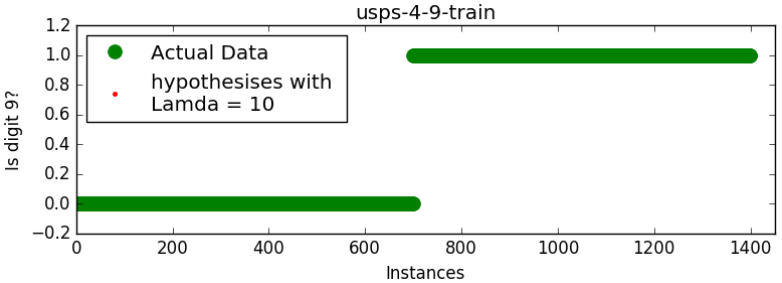

usps-4-9-test
Actual Data
hypothesises with
Lamda = 0.01
Is digit 9?
Instances

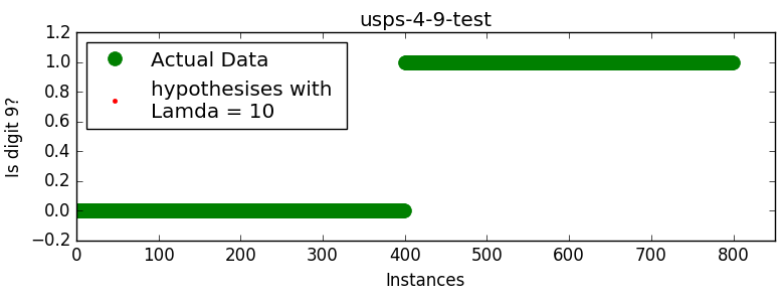

usps-4-9-train
Actual Data
hypothesises with
Lamda = 0.1
Is digit 9?
Instances


usps-4-9-test
Actual Data
hypothesises with
Lamda = 0.1
Is digit 9?
Instances


usps-4-9-train
Actual Data
hypothesises with
Lamda = 1
Is digit 9?
Instances


usps-4-9-test
Actual Data
hypothesises with
Lamda = 1
Is digit 9?
Instances


usps-4-9-train
Actual Data
hypothesises with
Lamda = 10
Is digit 9?
Instances


usps-4-9-test
Actual Data
hypothesises with
Lamda = 10
Is digit 9?
Instances