Assignment 7

Alberto Mejia
RIN: 661514960

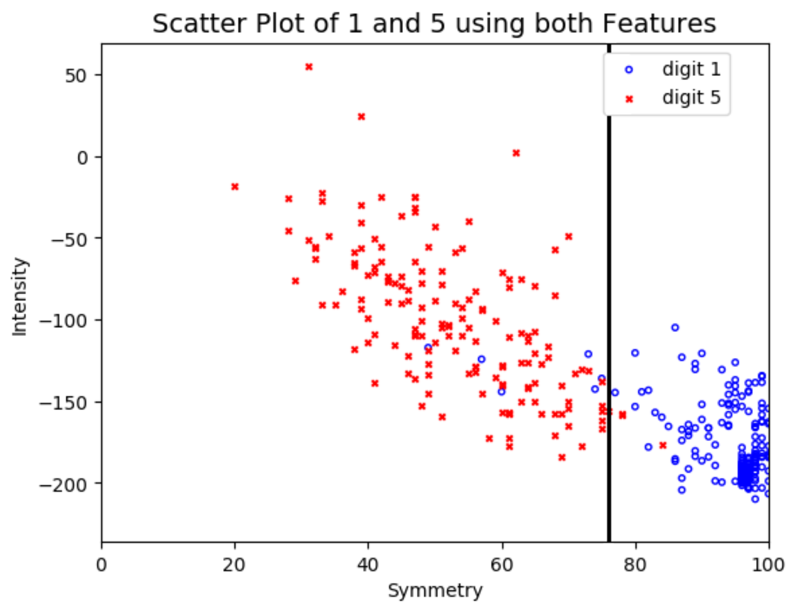
CSCI 4100 - Machine Learning from Data

October 20, 2019

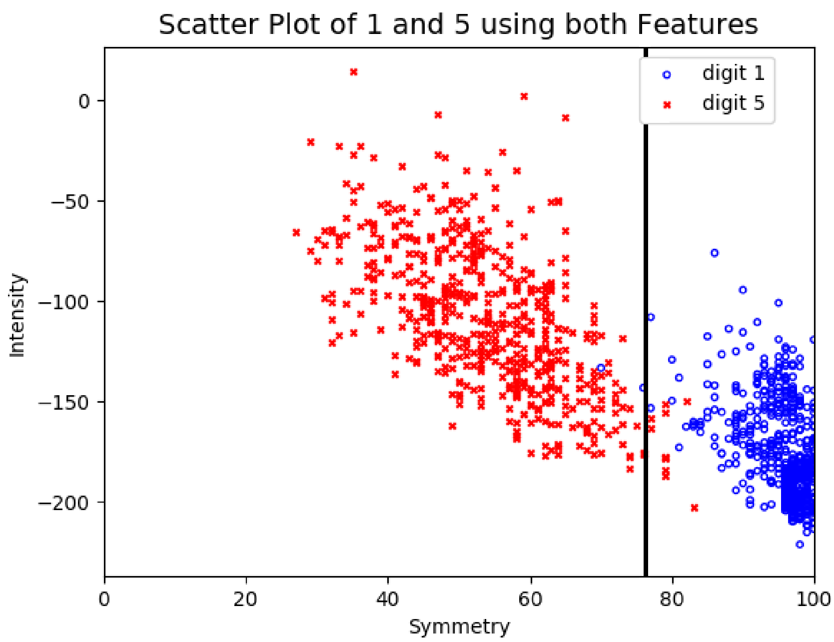
1. (500) Classifying Handwritten Digits: 1 vs. 5

(a)

Plot on Test Data



Plot on Training Data



(b) Calculate the error by dividing the number of incorrect classifications by the total sample:

$E_{in} = 0.00427$

$E_{test} = 0.01663$

(c) There are 1561 examples in the training set with 2 features.

$E_{out} \le E_{in} + \sqrt{\frac{8}{N} ln(\frac{4((2N)^{d_{vc}}+1)}{\delta})}$

N = 1561
δ = 0.05
d_{vc} = (2+1) = 3

$E_{out} \le 0.00427 + \sqrt{\frac{8}{\mathbf{1561}} ln(\frac{4((2(\mathbf{1561}))^3+1)}{\mathbf{0.05}})}$

= 0.38658

Bound of E_{in} = 0.38658

There are 424 examples in the test set.

$E_{out} \le E_{test} + \sqrt{\frac{1}{2N} ln(\frac{2M}{\delta})}$

N = 424
δ = 0.05
d_{vc} = (2+1) = 3

$E_{out} \le 0.01663 + \sqrt{\frac{1}{2(424)} ln(\frac{2}{0.05})}$

= 0.08258

Bound of E_{test} =0.08258

(d)

L[x] -> [x0, x_1, x_1^2, x_2^2, x_1x_2, x_1^3, x_2^3, $x_2^2x_1$, $x_1^2x_2$]
[1, x_1, x_1^2, x_2^2, x_1x_2, x_1^3, x_2^3, $x_2^2x_1$, $x_1^2x_2$]
Applying this transform gives us:

E_{in} = 0.00248025
E_{test} = 0.016925

$E_{out} \le E_{in} + \sqrt{\frac{8}{N} ln(\frac{4((2N)^{d_{vc}}+1)}{\delta})}$

N = 1561
δ = 0.05
d_{vc} = 10

$E_{out} \leq 0.00248025 + \sqrt{\frac{8}{1561} ln(\frac{4((2(1561))^{10}+1)}{0.05})}$

$= 0.66188$

Bound of $E_{in} = 0.66188$

$E_{out} \leq E_{test} + \sqrt{\frac{1}{2N} ln(\frac{2M}{\delta})}$

N = 424
$\delta = 0.05$
$d_{vc} = (2+1) = 3$

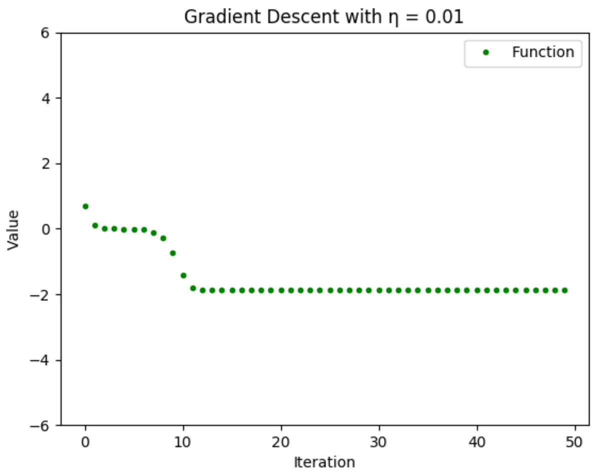
$E_{out} \leq \ 0.016925 + \sqrt{\frac{1}{2(424)} ln(\frac{2}{0.05})}$

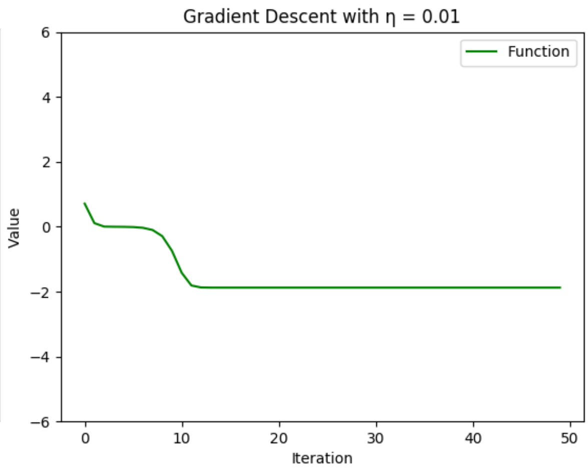
$= 0.08288$
Bound of $E_{test} = 0.08288$

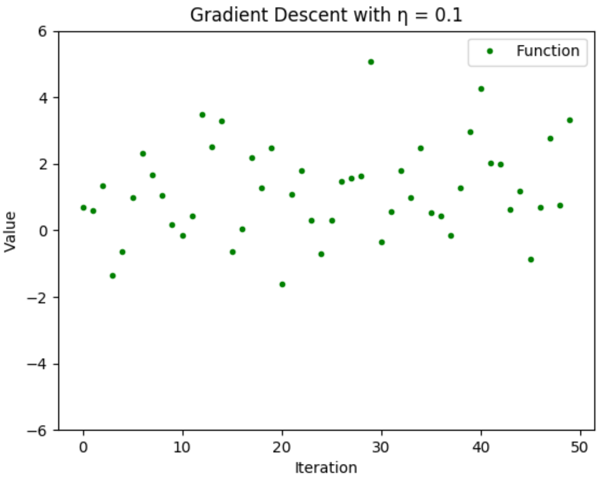
(e) A final deliverable I would give the customer is a linear model. As we can see based on the calculations, the transformation actually gives us worse error. This is because the 3^{rd}-polynomial transformation is very complicated compared to the simple linear model. This complexity gives the 3^{rd}-polynomial transformation a larger VC Dimension. This will cause it to capture more noise, which could potentially lead to overfitting.

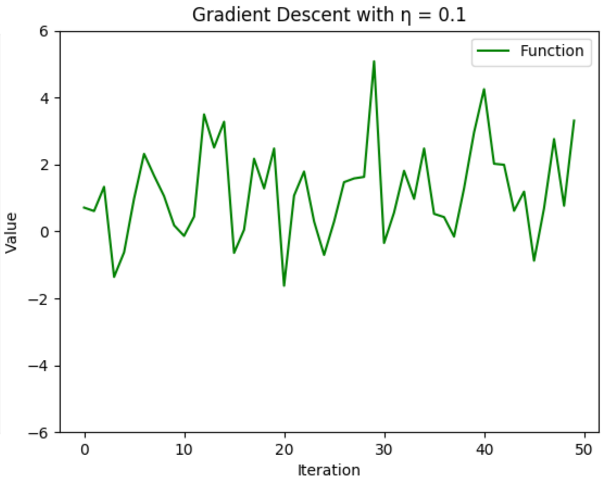
2. (200) Gradient Descent on a "Simple" Function

(a)









(b)

Initial Points	X	Y	Min f(x,y)
(x_0, y_0)			
(0.1, 0.1)	0.2438...	-0.2349...	-1.848...
(1, 1)	1.218...	0.7128...	0.535...
(-0.5, -0.5)	-0.73..	-0.23..	-1.35...
(-1, -1)	-1.12...	-0.71..	0.53..

3. (300) LFD Problem 3.16

$$P(Y|X) = \begin{cases} f(x) & y = +1 \\ (1 - f(x) & y = -1 \end{cases}$$

(a) g(x) = P[y = +1 | x]
P[y = -1 | x] - > 1 – g(x)
cost(accept) = (1-g(x))c_a
cost(reject) = g(x)c_a

(b)
(1-g(x))$c_a \leq$ g(x)c_a
g(x)(c_r+c_a) $\geq$ c_a
g(x) $\geq$ (c_a)/(c_a+ c_r)
$\therefore$ k = (c_a)/(c_a+ c_r)

(c)
Supermarket: c_r = 10, c_a = 1
k = 1/(1+10) = 1/11 = 0.0909090909
CIA: c_r = 1000, c_a = 1

k = 1000/(1000+1) = 1000/1001 = 0.999000999

CIA will only accept a fingerprint if they are 99% sure that the fingerprint is correct. Where as the supermarket will accept it if they are atleast 9% sure.