Introduction to Machine Learning

HW 2

-Kaushal Vinay Nerkar

Collab Notebook link : - https://colab.research.google.com/drive/1icy-VheLeebTezEmiEx4Y3cjM1yKiAi6?authuser=1#scrollTo=vnd3ezO-qHXw&uniqifier=1

Question 2

Exercise 2.6

a) Error bar on train_data for N=400 ,tolerance = 0.05, Hypothesis size = 1000 is 0.11509037065006825
   Error bar on test_data for N=200 ,tolerance = 0.05 is 0.09603227913199208
b) If we reserve more samples for testing, we will get less accurate target function, thus E_test will much off compared to E_in.

Question 3

M T W T F S S
Page No.:
Date:
YOUVA

Problem 2.12

given $d_{vc} = 10$

confidence = 95% so $\delta = 0.05$

generalization error ; $\epsilon = 0.05$

So sample size (N) will be

$$N \geq \frac{8}{\varepsilon^2} \ln\left(\frac{4((2N)^{d_{vc}} + 1)}{\delta}\right)$$

So lets take $N = 10^{10}$ as intial in RHS

then we get

$$N \geq \frac{8}{(0.05)^2} \ln\left(\frac{4(2\times10^{10})^{10} + 4}{0.05}\right)$$

$$N \geq \frac{8}{(0.05)^2} (241.57)$$

$$N \approx 773{,}030$$

Now using this $N = 773{,}030$ we get

$$N \geq \frac{8}{(0.05)^2} \ln\left(\frac{4(2\times773030)^{10} + 4}{0.05}\right)$$

$$\geq \frac{8}{(0.05)^2} (146.89)$$

$$N \approx 470{,}062$$

Similarly in next iteration using $N = 470,062$

we get $N \geq \frac{8}{(0.05)^2} \times (141)$

$N \approx 454,143$

Now using $N \simeq 454,143$

$N \geq \frac{8}{(0.05)^2} (141.5)$

$N \approx 452,962$

Now using $N = 452,962$

$N \geq \frac{8}{(0.05)^2} (141.55)$

$N \approx 4,52,962$

So here N converges to 452,962

so sample size = 452,962

Question 4

a,b)

M T W T F S S
Page No.:
Date:
YOUVA

$$f(x) = x^2 \qquad h(x) = ax + b$$

data set has 2 points $\{x_1, x_2\}$

thus full data set is $\{(x_1, x_1^2), (x_2, x_2^2)\}$

$$E_{in}(h) = \sum_{i=1}^{2} (h(x_i) - f(x_i))^2$$

$$E_{in}(h) = \sum_{i=1}^{2} (ax_i + b - x_i^2)^2$$

$$\frac{\partial E_{in}(h)}{\partial a} = -2 \sum_{i=1}^{2} x_i (\cancel{x_i^2}\, ax_i + b - x_i^2) \rightarrow (1)$$

$$\frac{\partial E_{in}(h)}{\partial b} = -2 \sum_{i=1}^{2} (x_i^2 + ax_i + b) \rightarrow (2)$$

so for these 2 equation $\nabla = 0$

thus

from 1

$$x_1^2 - ax_1 - b = 0$$

from 2

$$x_2^2 - ax_2 - b = 0$$

$$a(x_2 - x_1) = x_2^2 - x_1^2$$

$$a = (x_2 + x_1) \quad ; \quad b = -x_1 x_2$$

So $h(x) = ax + b$

$\therefore$ So $h(x) = (x_1 + x_2)\, x - x_1 x_2$

So now average function is $\bar{g}(x) = E_D[h(x)]$

$$\bar{g}(x) = E_D[x_1 x + x_2 x - x_1 x_2]$$

$$\bar{g}(x) = E_D[x_1]\, x + E_D[x_2]\, x - E_D[x_1] E_D[x_2]$$

(b)

To calculate $\bar{g}(x)$

- Compute $h(x)$ for 2 points from $[-1, 1]$
- then take average using the points choosen above

for variance

we know variance $= E_D\left[\left(h(x) - \bar{g}(x)\right)^2\right]$

So for the 2 points choosen from $[-1, 1]$ find $(h(x) - \bar{g}(x))^2$ and then take average of this for 2 points.

$$\text{ie} \quad \frac{\sum_{i=1}^{2} (h(x) - \bar{g}(x))^2}{(N=2)}$$

bias is $E_D\left[\left(\bar{g}(x) - f(x)\right)^2\right]$

so similarly as above find average. of these value at 2 points

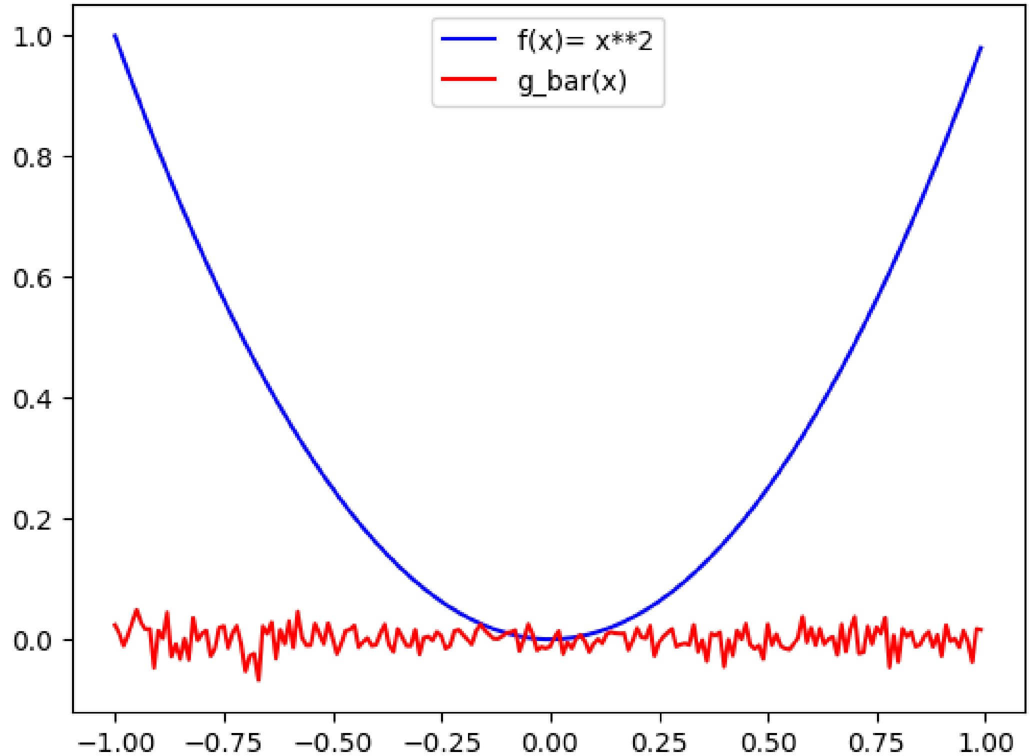
$$\text{ie} \quad \frac{\sum (\bar{g}(x) - f(x))^2}{(N=2)}$$

out of sample error is $E_D\left[\left(h(x) - f(x)\right)^2\right]$

so iteratively find average of 2 points

$$\text{ie} \quad \frac{\sum (h(x) - f(x))^2}{(N=2)}$$

c)



Variance = 0.33481
Bias = 0.20162
Eout = 0.5382
sum(variance, bias) = 0.5364

d)

M	T	W	T	F	S	S

Page No.:
Date:
YOUVA

d)

Variance $= E_D[(h(x) - \bar{g}(x))^2]$

$= E_D[((x_1 + x_2)x - x_1 x_2)^2]$

$= E_D(x_1^2 + x_2^2 + 2x_1x_2)\, x^2$
$+ E_D(x_1^2 x_2^2) - 2x\, E_D(x_1^2 x_2 + x_2^2 x_1)$

So $E_D[x^2] = \frac{1}{3}$

$E_D[x] = 0$

So $= x^2\left(\frac{1}{3} + \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{1}{3}\right) - 2(0+0)$

$= x^2\left(\frac{2}{3}\right) + \frac{1}{9}$

$= \frac{6x^2 + 1}{9}$

So $= \left(6 \times \frac{1}{3} + 1\right) / 9$ $\quad E_D[x^2] = \frac{1}{3}$

Variance $= \frac{1}{3}$

Page No.: Date:

$$\text{bias} = E_D\left[(\bar{g}(x) - f(x))^2\right]$$

$$= E_D\left[(0 - x^2)^2\right]$$

$$= E_D\left[x^4\right] \qquad \because E_D[x^4] = \frac{1}{5}$$

$$= \frac{1}{5}$$

$$E_{out} = E_D\left[(h(x) - f(x))^2\right]$$

$$= E_D\left[((x_1 + x_2)x - x^2)^2\right]$$

$$= (x_1 + x_2)^2 x^2 + x^4 - 2(x_1 + x_2)x^3$$

$$= x_1^2 x^2 + x_2^2 x^2 + 2x_1 x_2 x^2 + x^4 - 2x_1 x^3 - 2x_2 x^3$$

$$= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{5} - 0 - 0$$

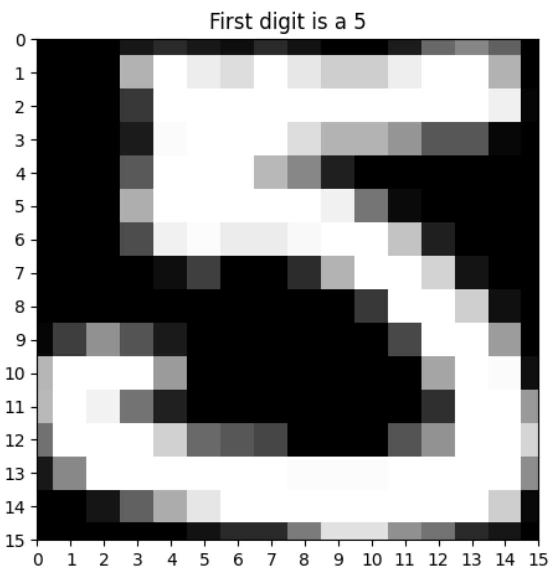
$$= \frac{1}{3} + \frac{1}{5}$$

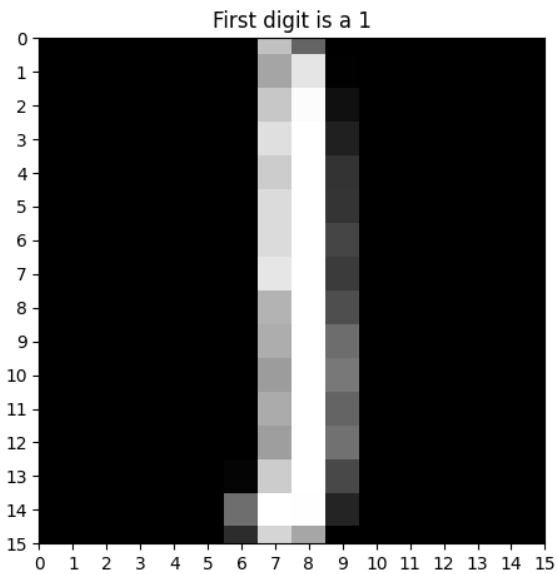
$$= \frac{8}{15}$$

also E_{out} = variance + bias.

Question 5

a)





b) Average intensity : For a 16 * 16 image with each element between [-1,1] showing intensity. Average intensity is the sum of all pixel intensity over total pixels .

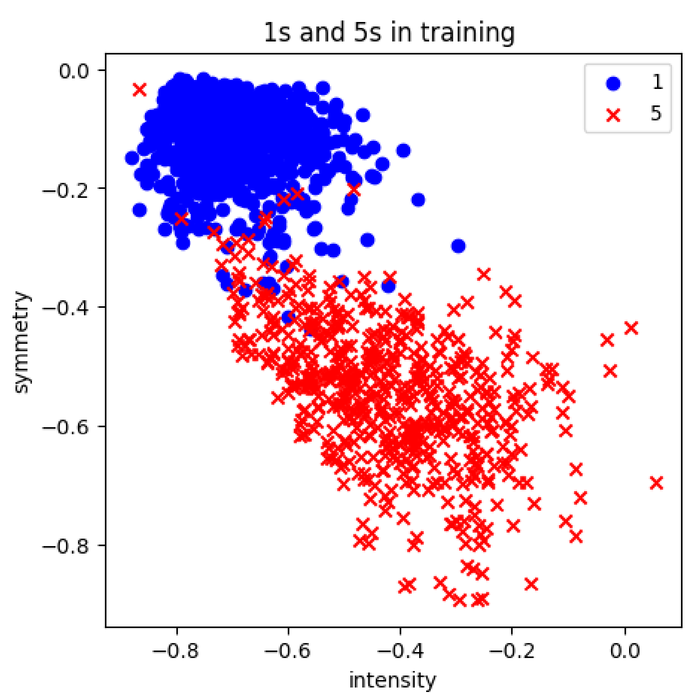
$$Average\ Intensity = \frac{\sum_0^{256} pixelIntensity(i)}{totalpixels}$$

Symmetry:

Asymmetry is average absolute difference between an image and its flipped versions, and symmetry is then negation of asymmetry. So let X be image and X_flipped be flipped image.

$$symmetryMeasure = -\frac{\sum_0^{256} \left|X(i) - X_{flipped(i)}\right|}{totalpixels}$$

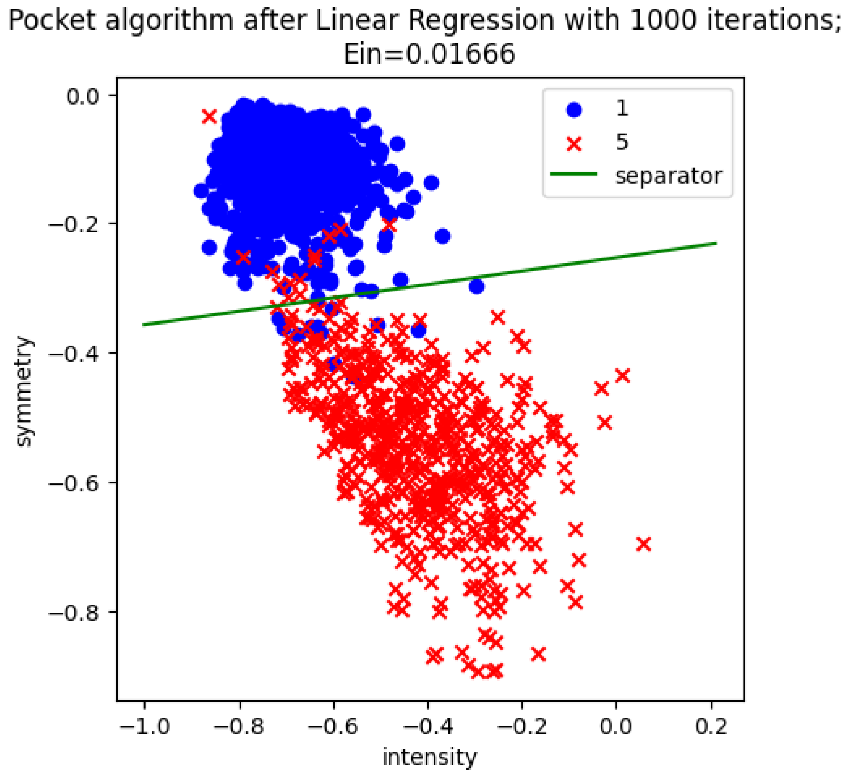
c)



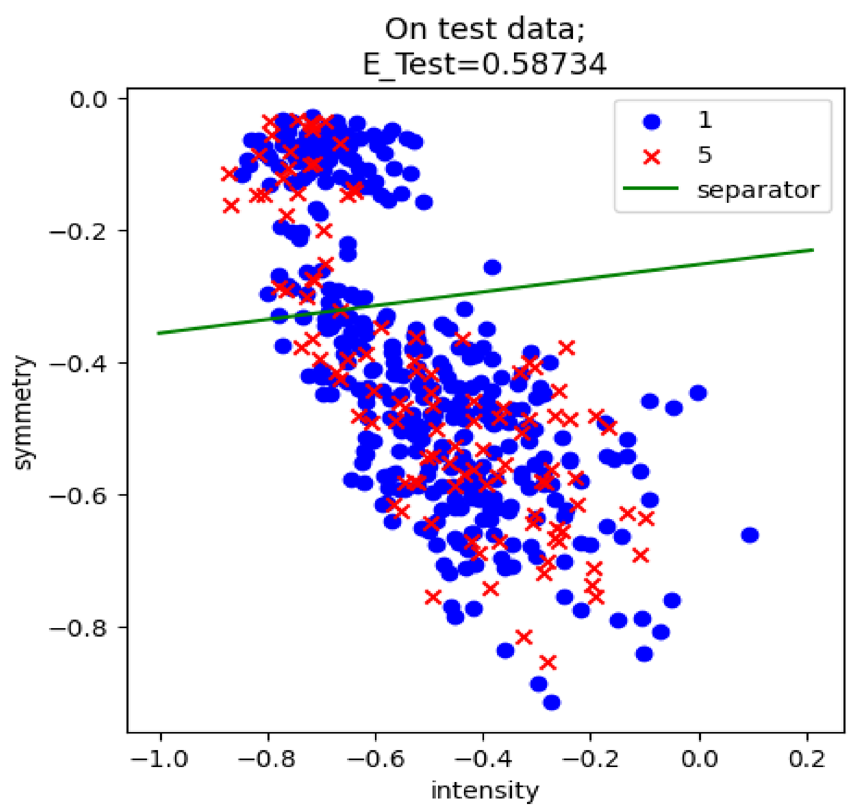
Question 6

Using Linear Regression

a) From training Data



Below is from Testing Data

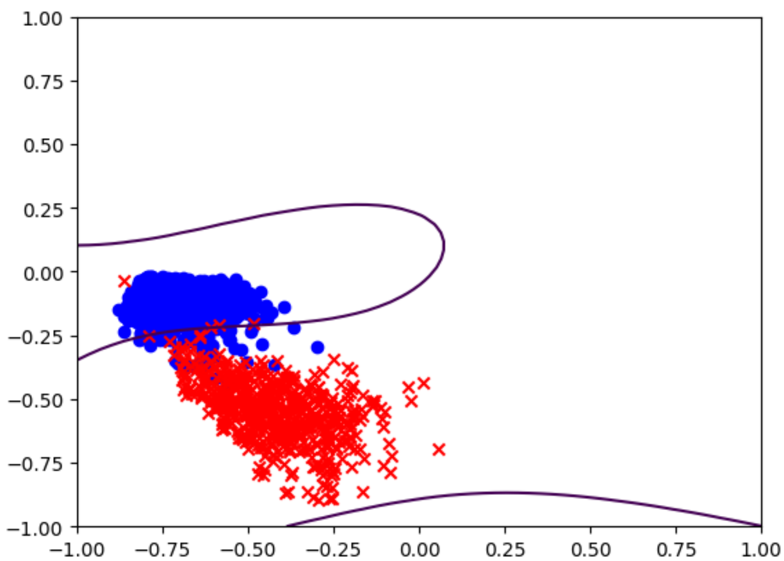


b) Ein on train data
Ein_best = 0.0166
w_best = [[3.0012006], [-1.23252726], [11.8526552]

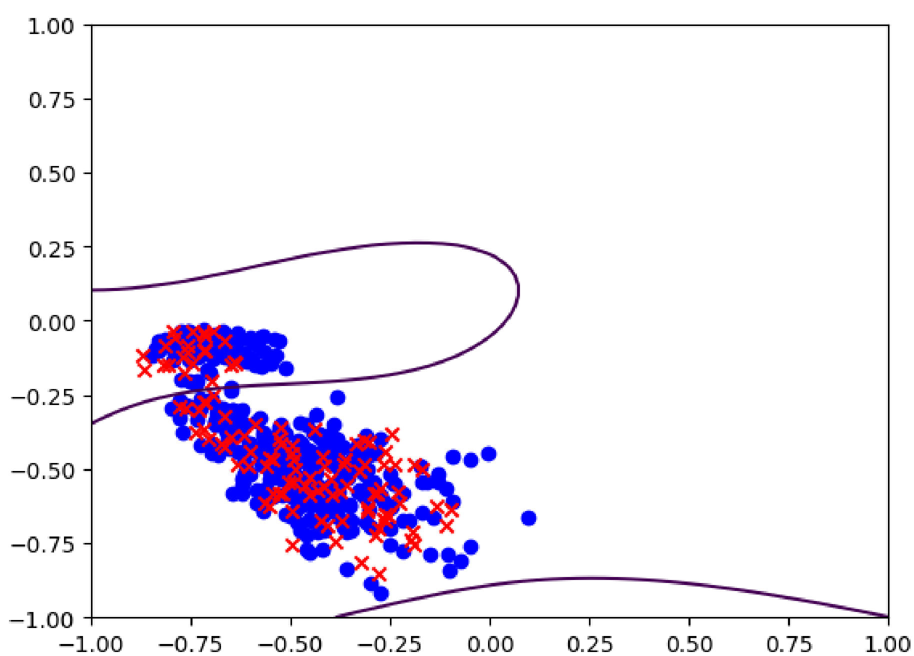
E_test on test data
E_test_best = 0.58734

c) Bounds
Bound on Eout in train : 0.07490960460609433
Bound on Eout in test : 0.6948963969719182
Eout on train data is better than Eout on test as hypothesis(target function) can classify train data much better than test data.

d) After 3rd degree Polynomial transform
On train data



On test data



On train data
Ein_best = 0.016015374759769378,
w_best= [[2.65603268],[-4.14485003],[6.26405377],[-6.78598422],[-0.92928893], [-13.14130524],[-3.09777974],[-2.89000424],[10.72310204],[-12.93772741]]

On test data
E_test = 0.58515

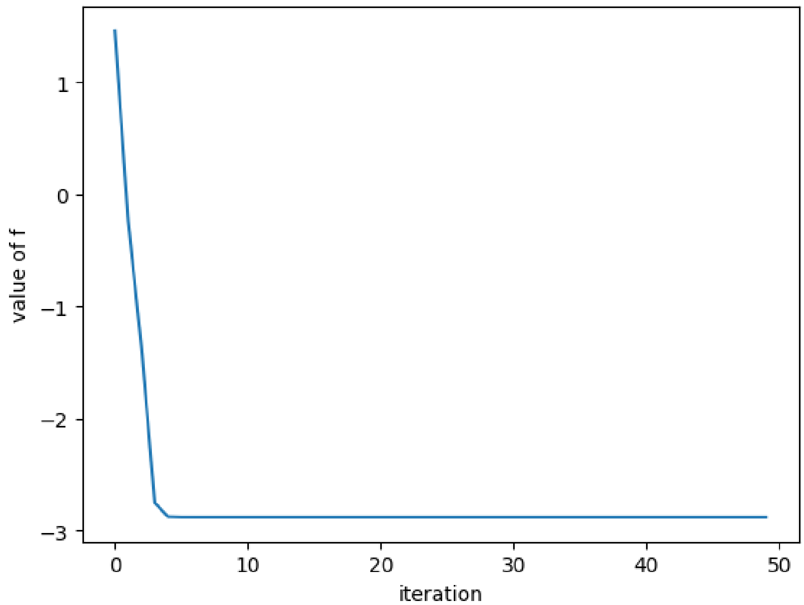
Bound on Eout in 3rd degree Poly train : 0.07619660460609433
Bound on Eout in 3rd degree Poly test : 0.6927063969719182

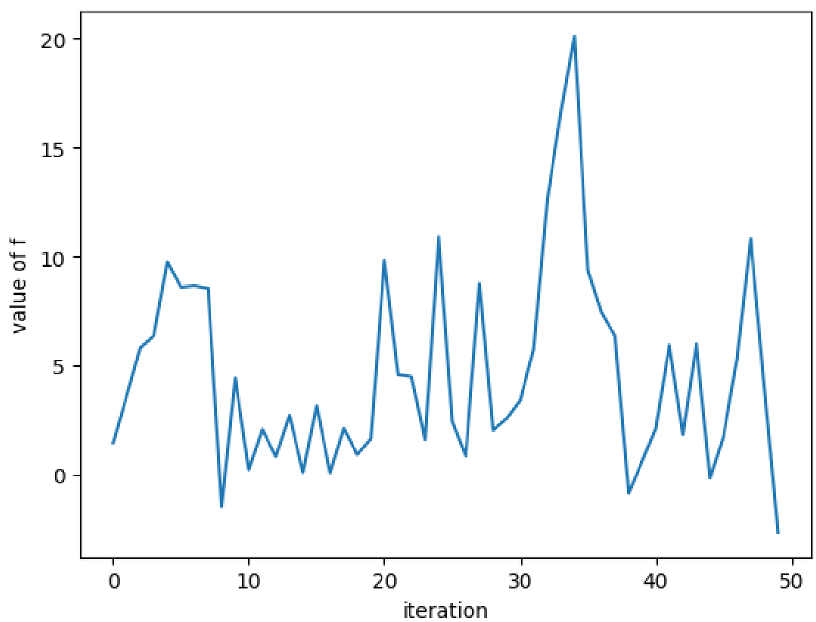
e) So Standard Error on Linear Regression = 2.34934498
and Standard Error on Linear Regression with polynomial transform = 2.34061135
So, both are comparatively similar thus we will use one without polynomial transform as it will give a simpler target function.

Question 7

f(x, y) = $2x^2 + y^2 + 3 \sin(2 \pi x) \cos(2 \pi y)$

a)

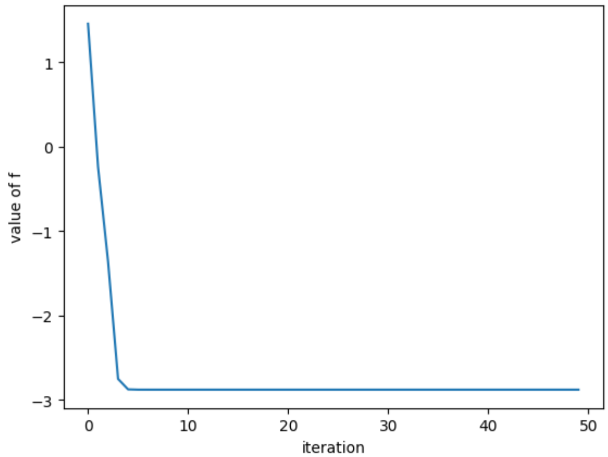




With the increase in Learning Rate, we observe abrupt increase in f values in iterations thus finding global minimum of function becomes difficult.

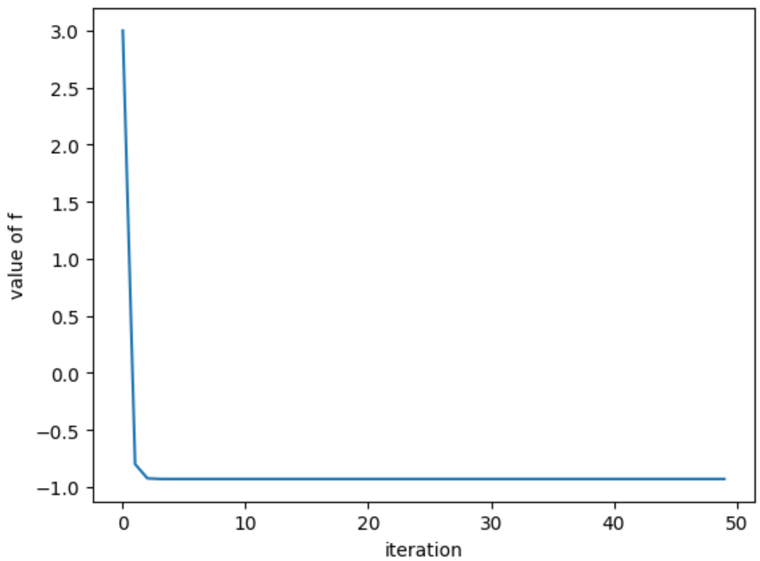
b) Starting points

a. (0.1, 0.1)



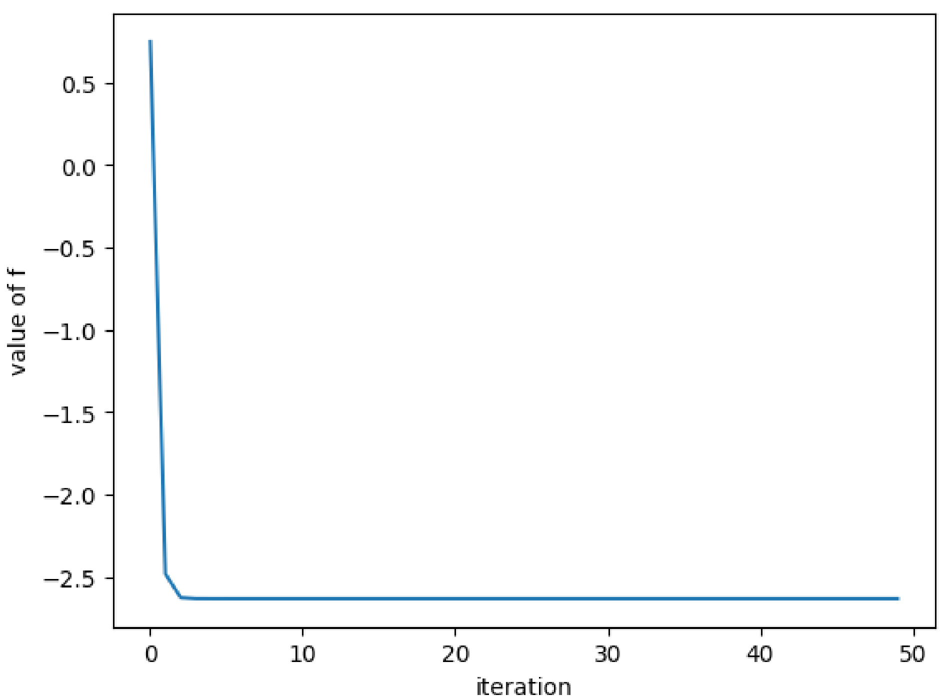
Min Value of f -2.8790846587644263 minX = -0.241828945494765 minY= 1.2269903410060357e-10

b. (1,1)



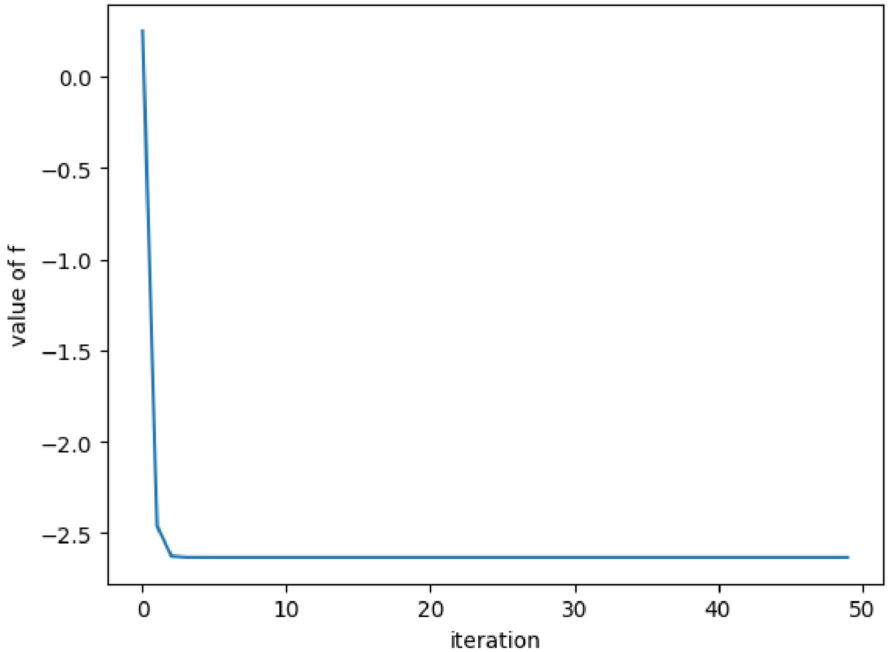
Min Value of f -0.9286447086312597 minX = 0.7252678803578847 minY= 0.9831635693235358

c. (0.5,0.5)



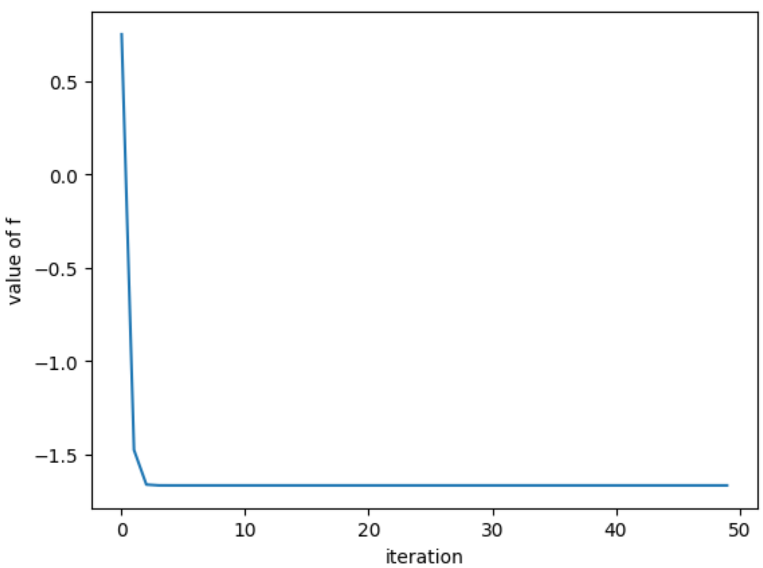
Min Value of f -2.6332425909756374 minX = 0.24181813075115313 minY= 0.4916822590684799

d. (0,0.5)



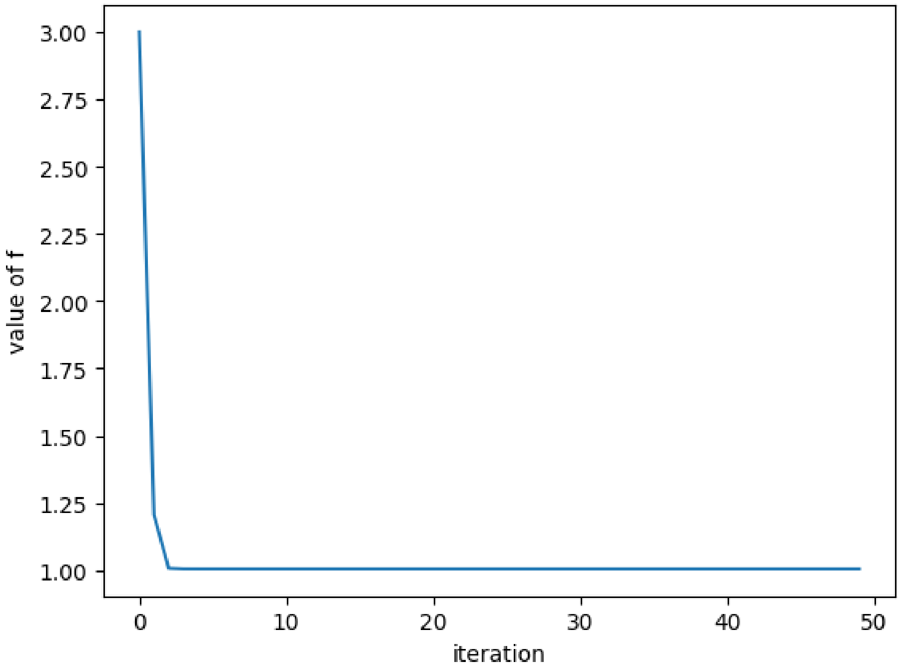
Min Value of f -2.6332425909756374 minX = 0.2418181308911715 minY= 0.49168225905153434

e. (-0.5,-0.5)



Min Value of f -1.6660267055389744 minX = -0.7253691676027778 minY= -0.49159419340047744

f. (-1,1)



Min Value of f 1.0051615759793306 minX = -1.2084759495263497 minY= 0.9827889176426049