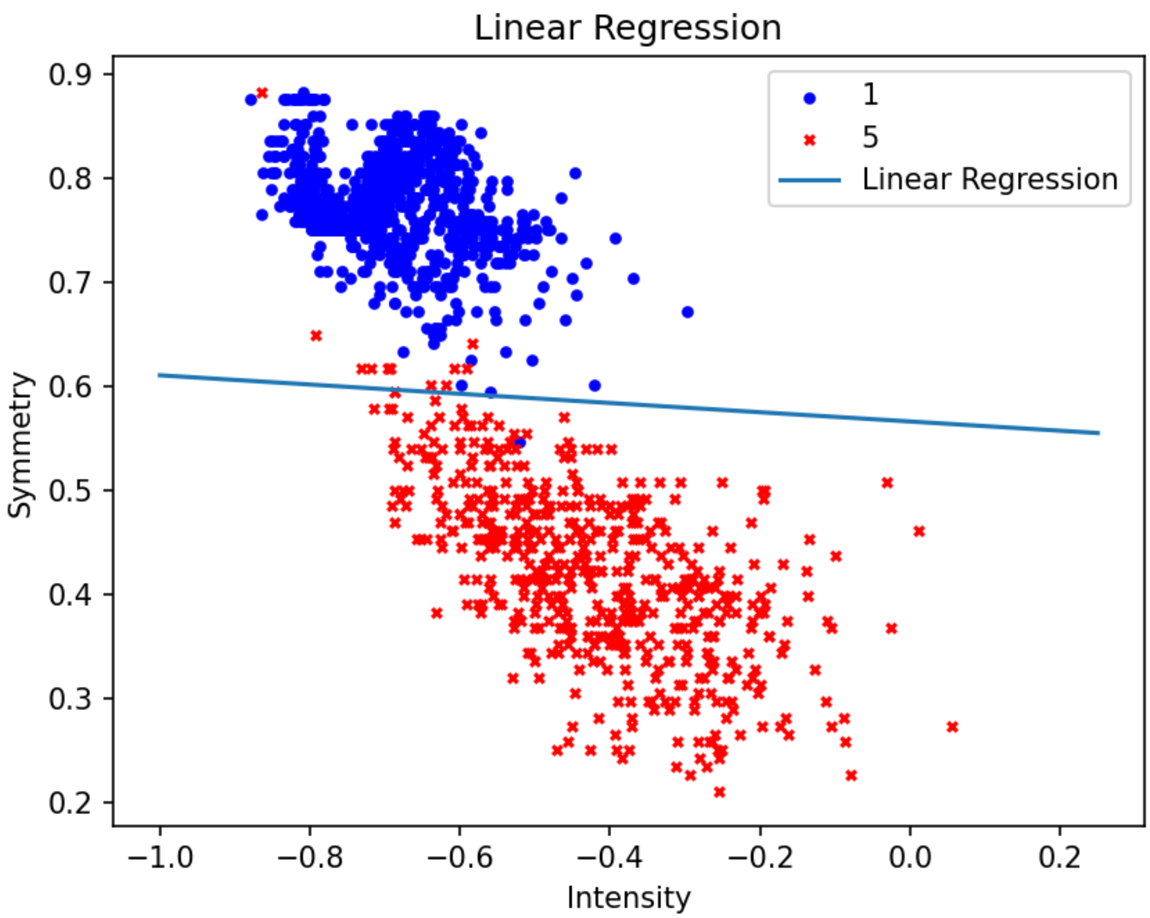
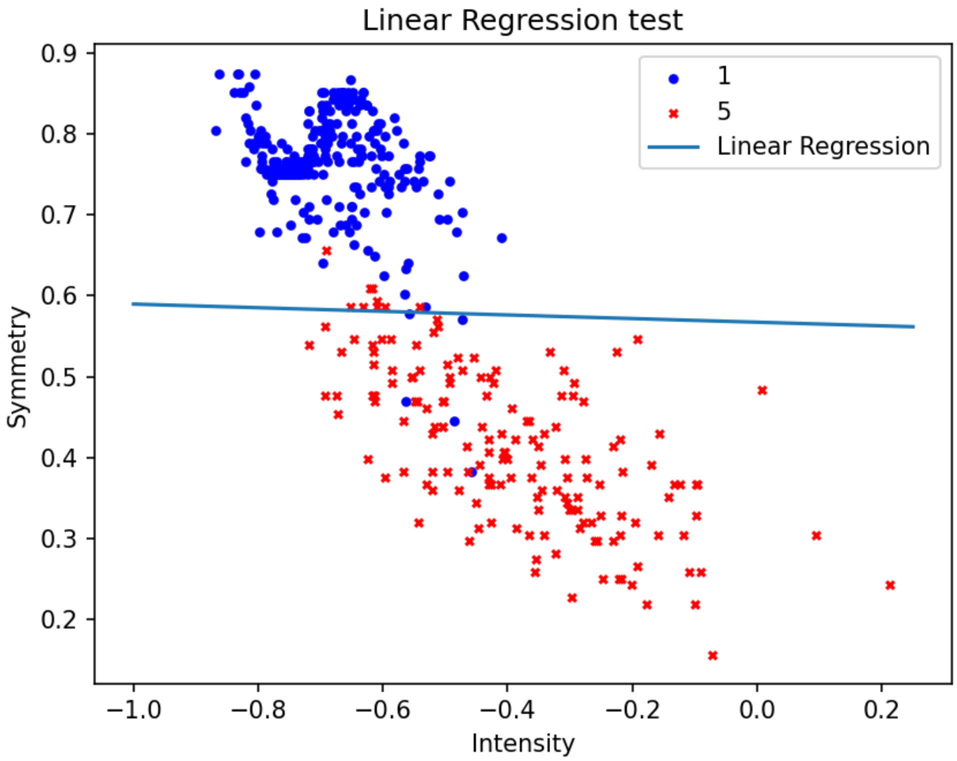
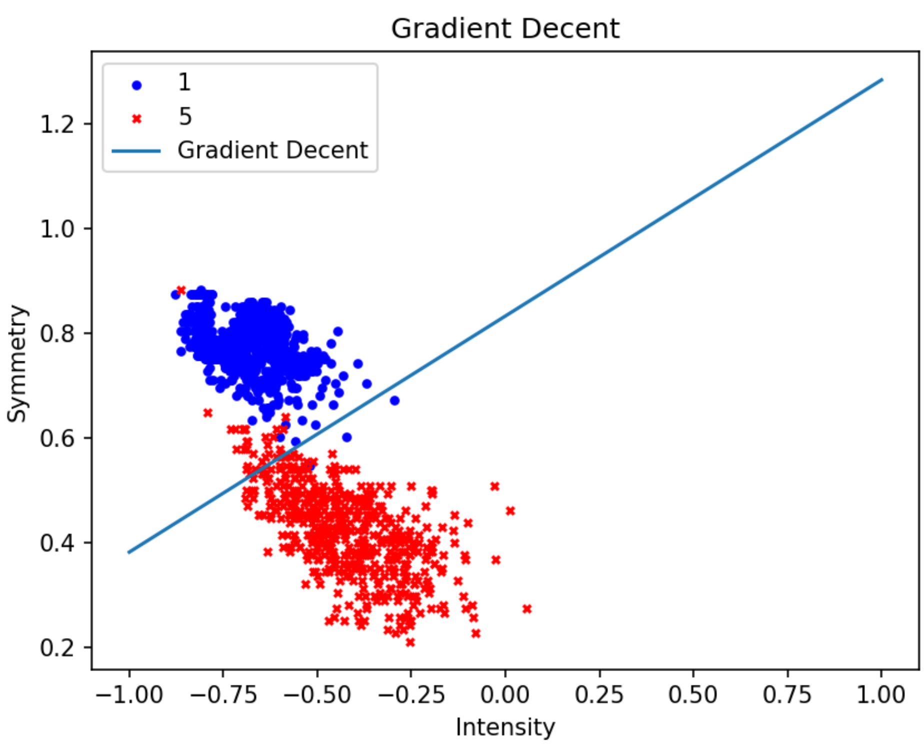
Problem 1:
a:
Method: Linear Regression:
Training data:



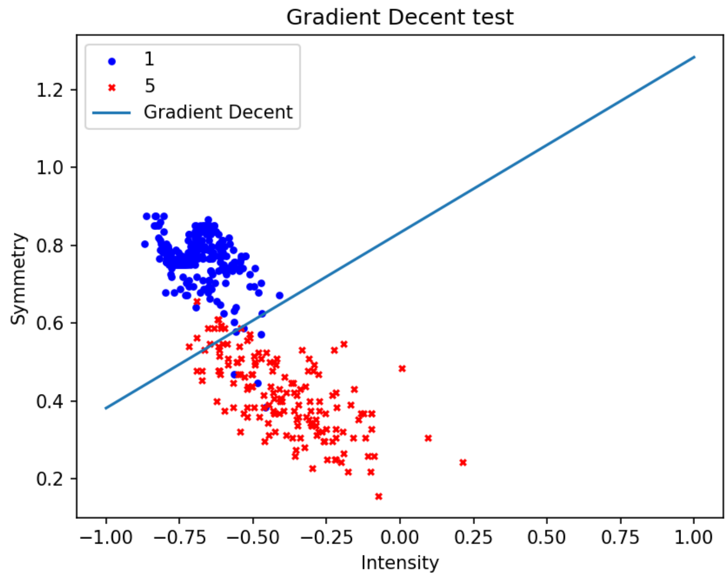
Testing data:



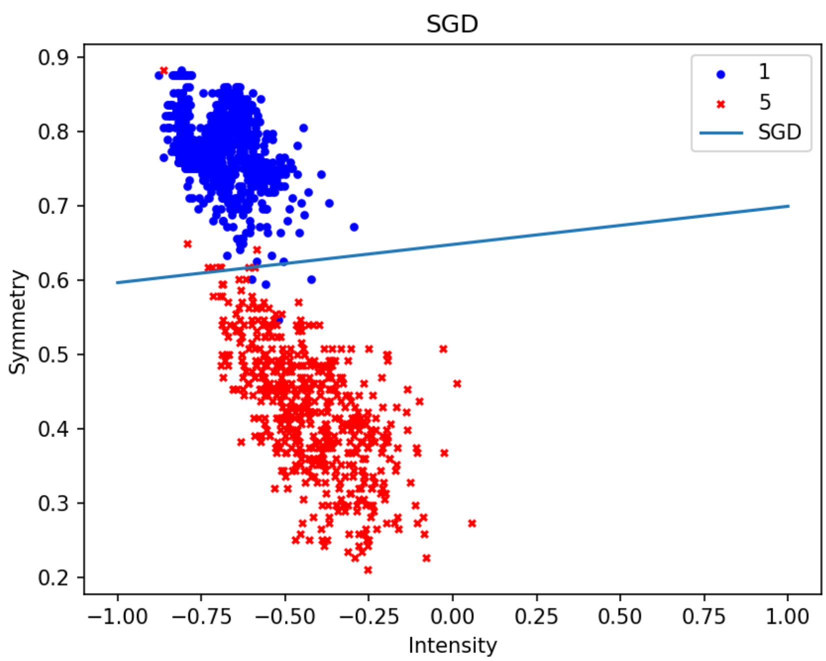
Gradient descent:
Training data



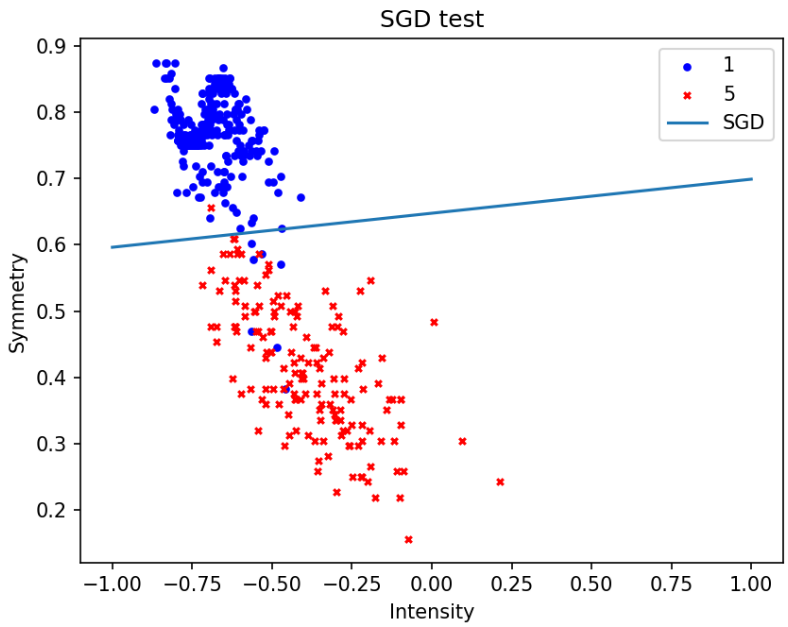
Testing data:



SGD: trainning



Test:



b:

Linear Regression:

Ein: 0.008327994875080076

Etest: 0.02358490566037736nbbbbbb

Gradient descent with eta = 0.01, iteration = 2000

Ein: 0.021780909673286355

Etest: 0.04009433962264151

SGD with eta = 1, iteration = 1000000:

Ein: 0.008327994875080076

Etest: 0.018867924528301886

Part c:

Eout <= Ein + $\sqrt{\frac{8}{N} ln(\frac{4((2*N)^{dvc}+1)}{\delta})}$

Min Ein = 0.008327994875080076

N = 1561

δ = 0.05

dvc = 3

Eout <= 0.008327994875080076 + $\sqrt{\frac{8}{1561} ln(\frac{4((2*1561)^{3}+1)}{0.05})}$

Eout <= 0.39064510

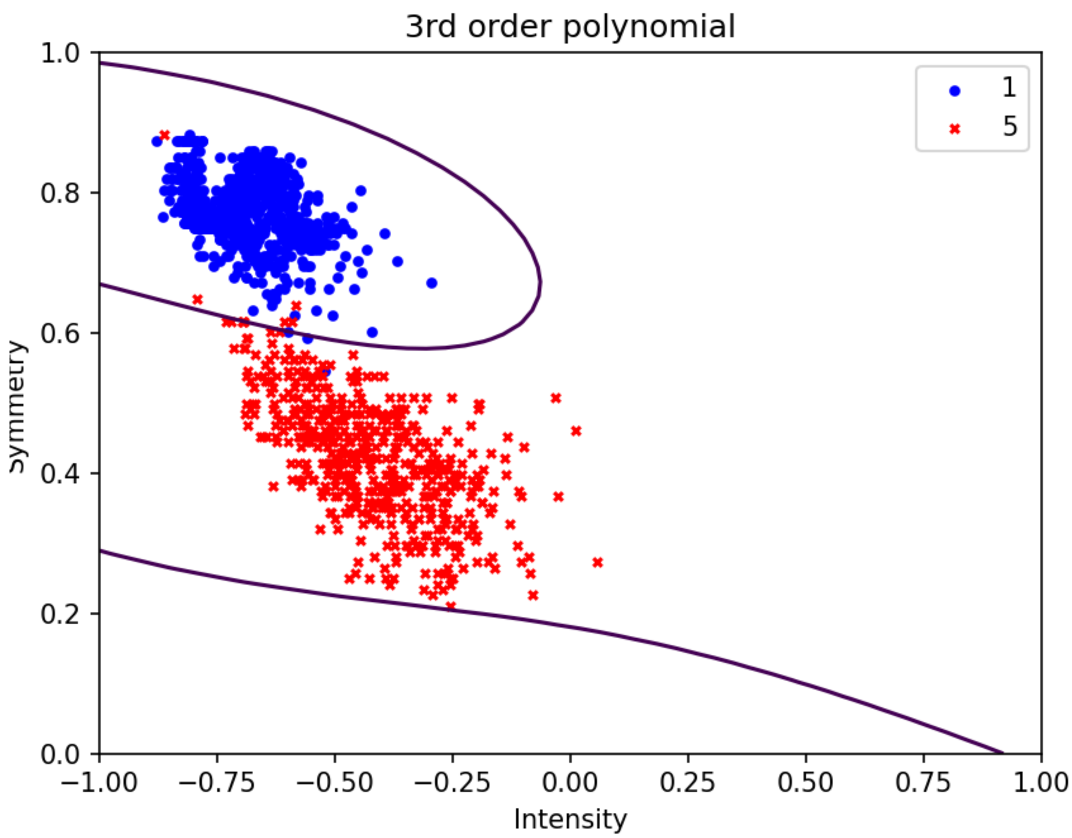
Part d:

Eout <= Ein + $\sqrt{\frac{1}{2N} ln(\frac{2M}{\delta})}$

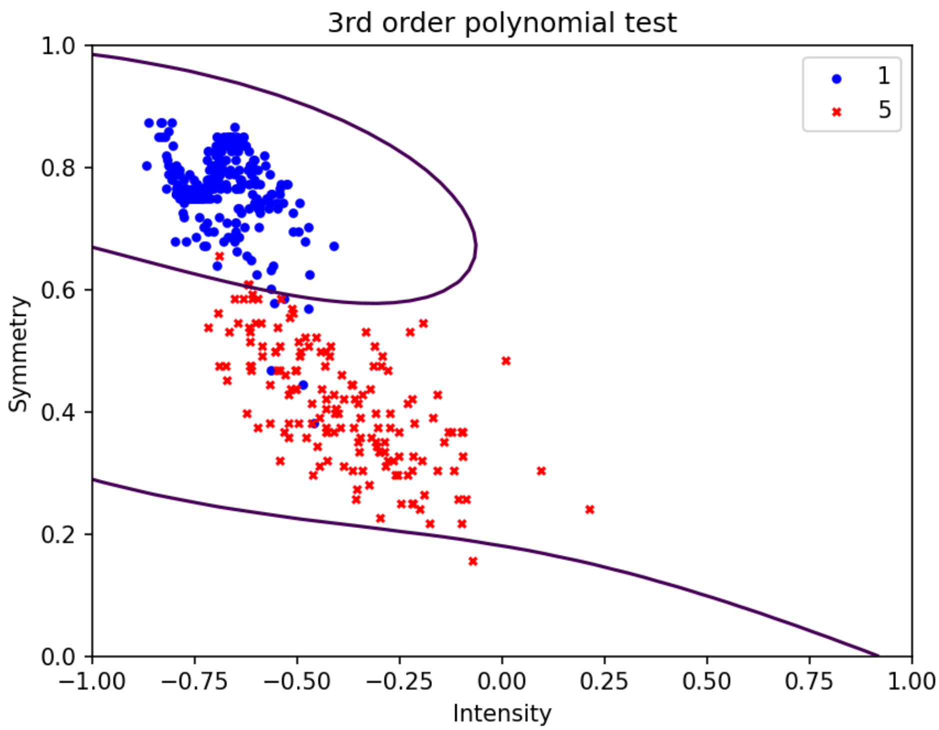
Min Etest = 0.018867924528301886
N = 424
δ = 0.05
M = 1

Eout <= 0.018867924528301886 + $\sqrt{\frac{1}{2*424} ln(\frac{2*1}{0.05})}$

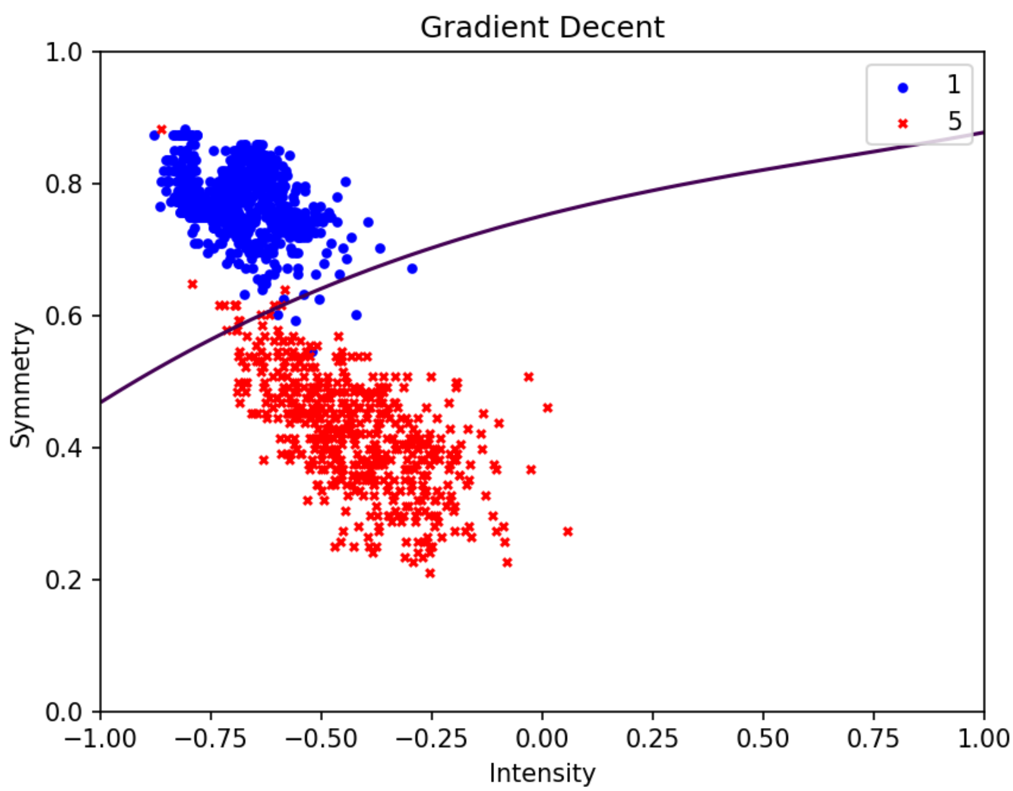
Eout <= 0.08482316
Part: e:
e.a
Linear regression:
Training data:



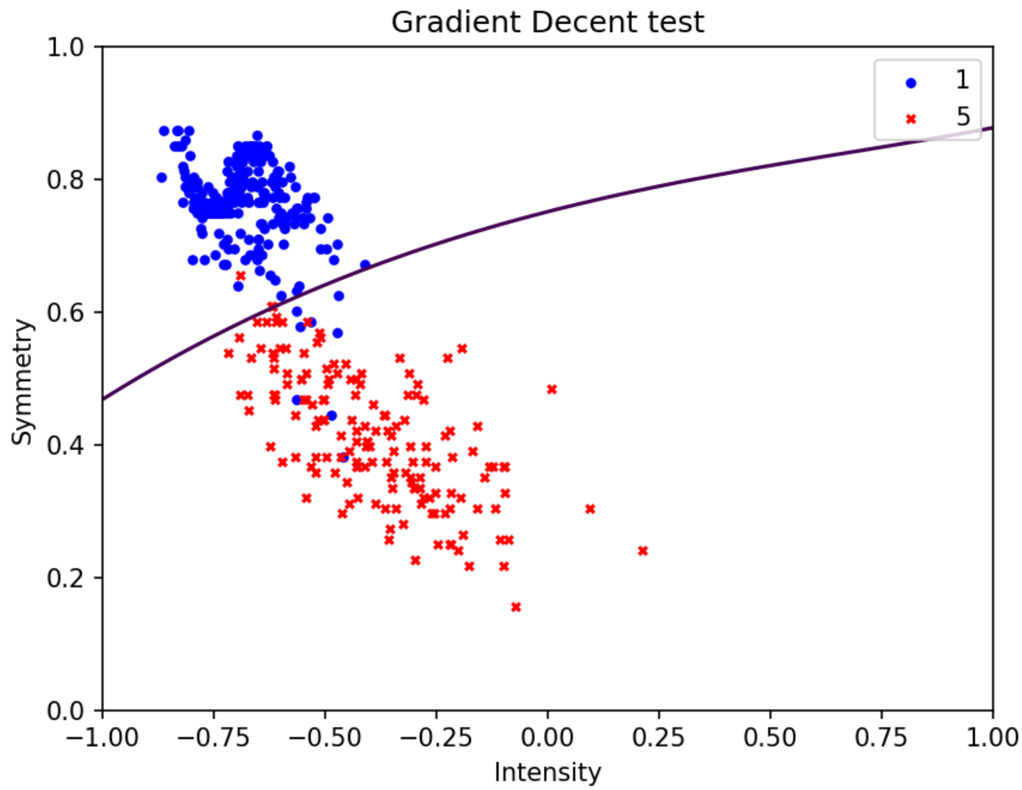
Testing data:



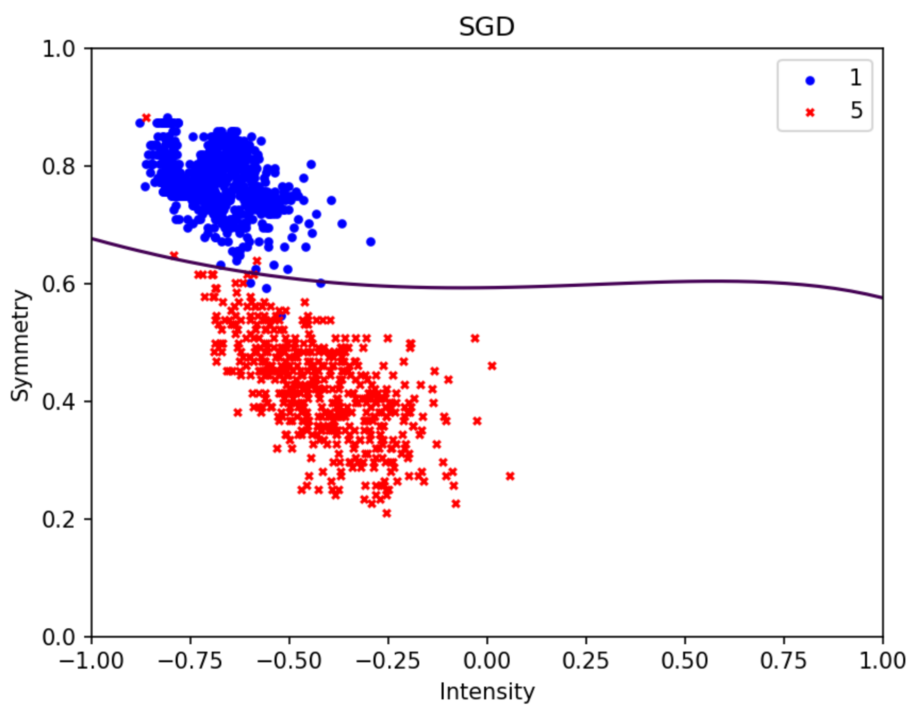
Gradient Descent:
Training data:



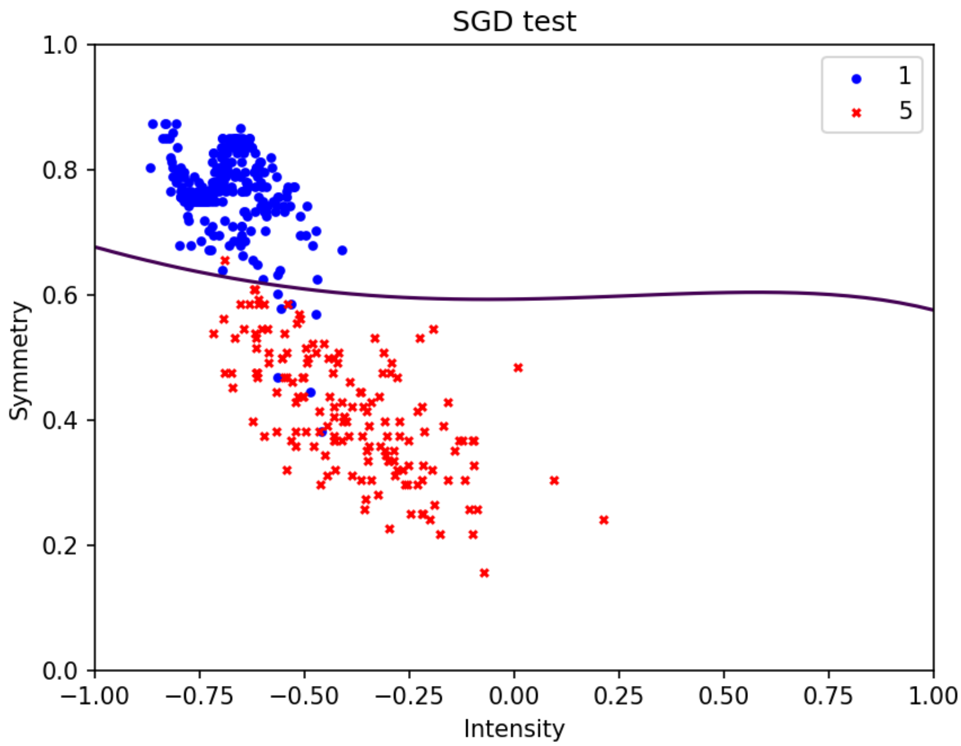
Testing data:



SGD:
Training:



Testing:



e.b:
Linear regression:
Ein = 0.0064061499039077515
Eout = 0.02358490566037736
Gradient Descent:
Ein = 0.012812299807815503
Eout = 0.025943396226415096
SGD:
Ein = 0.004484304932735426
Eout = 0.018867924528301886
e.c:

Eout <= Ein + $\sqrt{\frac{8}{N} ln(\frac{4((2*N)^{dvc}+1)}{\delta})}$

Min Ein = 0.004484304932735426
N = 1561
δ = 0.05
dvc = 3

Eout <= 0.004484304932735426 + $\sqrt{\frac{8}{1561} ln(\frac{4((2*1561)^{3}+1)}{0.05})}$

Eout <= 0.38680141
e.d:

Eout <= Ein + $\sqrt{\frac{1}{2N} ln(\frac{2M}{\delta})}$

Min Etest = 0.018867924528301886
N = 424
δ = 0.05
M = 1

Eout <= 0.018867924528301886 + $\sqrt{\frac{1}{2*424}ln(\frac{2*1}{0.05})}$

Eout <= 0.084326

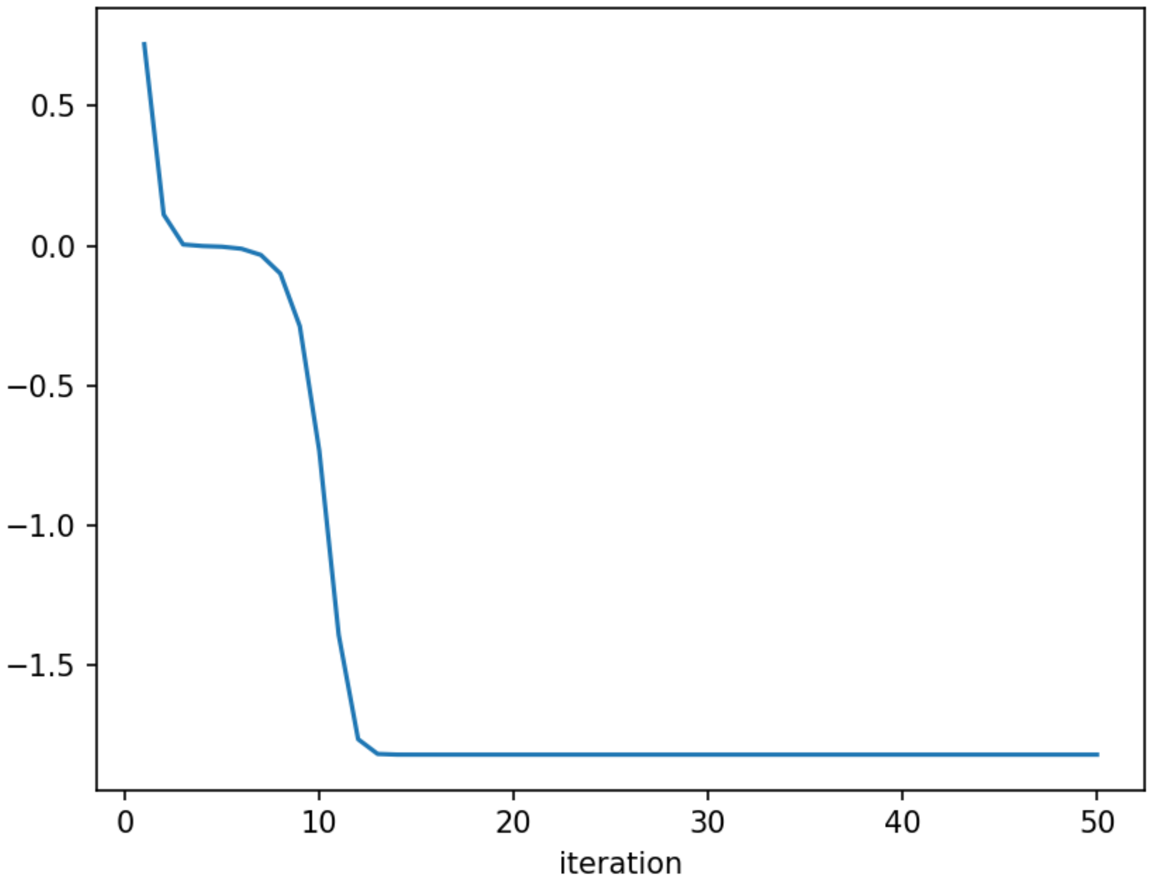
Part: f:
Although there are some improvements by using 3rd order polynomial transform, I will still choose to use a linear model. Because the improvement is very small, and using a linear model can avoid overfitting. (can save time also)

Problem 2:
f(x,y) = $x^2 + y^2 + 2sin(2x\pi)sin(2y\pi)$

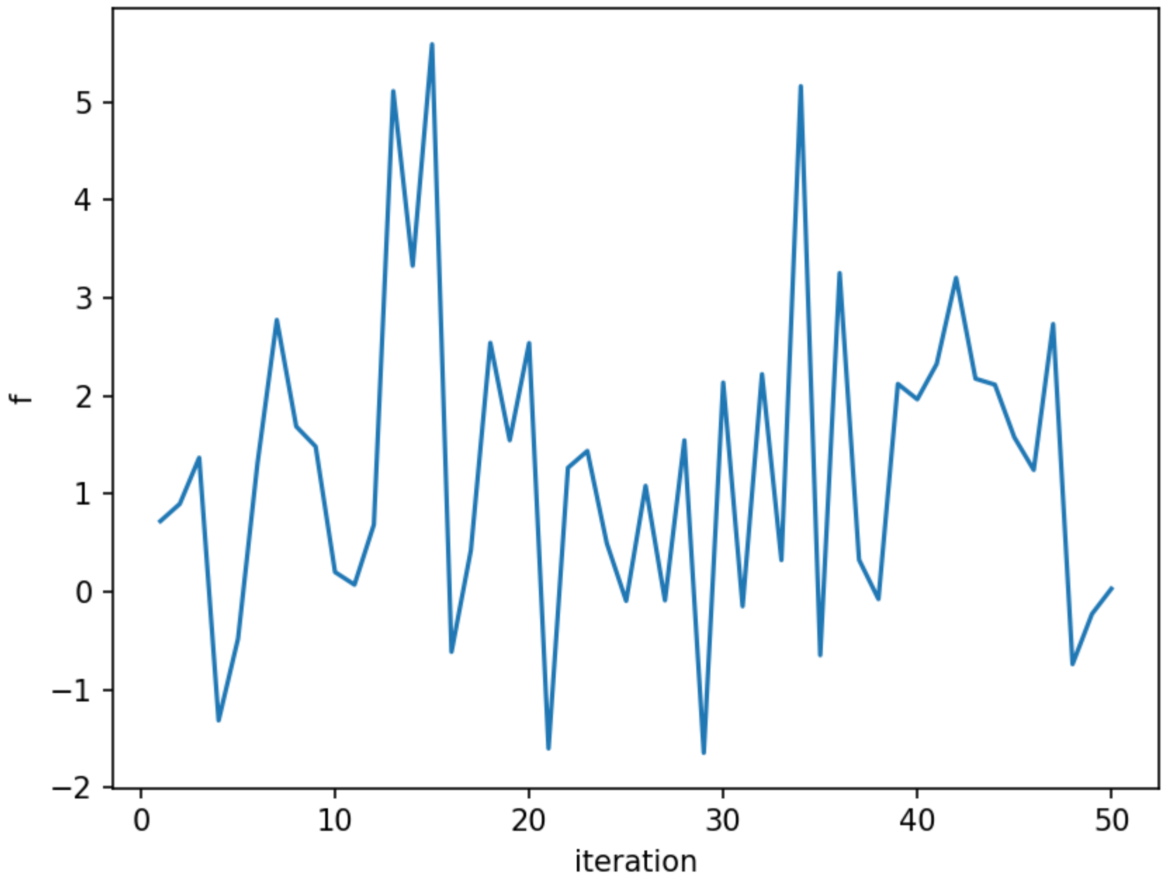
$\frac{d}{dx} = 2x + 4cos(2\pi x)sin(2y\pi)$

$\frac{d}{dy} = 2x + 4cos(2\pi y)sin(2x\pi)$

Graph:
η = 00.1



$\eta = 0.1$:



After change η to 0.1, we can see f is increasing and then decreasing and not able to find min.

Part b:

Initial point	η	x	y	min_value
(0.1,0.1)	0.1	0.236240854	0.229220365	-1.645221618
(0.1,0.1)	0.01	0.243804969	-0.237925821	-1.820078542
(1,1)	0.1	0.563844817	-0.035035466	-1.699744167
(1,1)	0.01	1.218070301	0.71281195	0.593269374
(-0.5,-0.5)	0.1	-0.01859951	0.400986378	-1.396466643
(-0.5,-0.5)	0.01	-0.73137746	0.035035466	-1.332481062
(-1,-1)	0.1	-0.563844817	0.035035466	-1.699744167
(-1,-1)	0.01	-1.218070301	-0.71281195	0.593269374

Problem 3:

Part a:

From the question we know:

$$g(\mathbf{x}) = \mathbb{P}[y = +1 \mid \mathbf{x}],$$

P[y = -1|x] = 1 - g(x)
cost(accept) = P[y = -1|x] * 0 + P[y = -1|x]*c_a = P[y = -1|x]*c_a = (1-g(x))c_a
cost(reject) = P[y = -1|x] * c_r + P[y = -1|x]*0 = g(x) c_r

Part b:
We accept person when cost(reject) >= cost(accept)
g(x) c_r >= (1-g(x))c_a
g(x) c_r >= c_a - g(x)c_a
g(x)(c_r + c_a) >= c_a
g(x) >= $\frac{c_a}{c_r + c_a}$
From the question we know g(x) >= k
K = $\frac{c_a}{c_r + c_a}$

Part c:
Example 1.1

Supermarket

		f: +1	f: −1
h	+1	0	1
	−1	10	0

CIA

		f: +1	f: −1
h	+1	0	1000
	−1	1	0

For Supermarket: c_a = 1, c_r = 10
K = 1/1+10 = 0.091

For CIA: c_a = 1000, c_r = 1
K = 1000/1+1000 = 0.99
High K means the cost of false rejection is much less than false agreement.
For the supermarket, a false rejection is costly because if a customer gets wrongly rejected, she may be discouraged from patronizing the supermarket in the future. All future revenue from this annoyed customer is lost. On the other hand, the cost of a false acceptance is minor. You just gave away a discount to someone who didn't deserve it, and that person left their fingerprint in your system - they must be bold indeed.
For the CIA, a false acceptance is a disaster. False rejects, on the other hand, can be tolerated since authorized persons are employees