1a) Let w ◦ b be the element-wise product of the vector w and b. the results of that vector would be [w1b1, w2b2,w3b3, …, wnbn]. a^T (w ◦ b)= a1w1b1+a2w2b2+a3w3b3+…+anwnbn

To calculate the gradient with respect to w([d/w1,d/w2,d/w3,…,d/wn])_, the gradient would be [a1b1,a2b2,a3b3,…,anbn] which is just a ◦ b.

b) f(w)=tr(Bw w^T A) which we will use ‘ instead t to represent transpose

∇w f(w)= ∇w tr(B*w*w’*A)

=tr(∇w(B*w*w’*A)

=tr((B*w*w’*A)’ * ∇w(w*w’))

=tr((A’*B’*w*w’)’ * ∇w(w*w’))

=tr(w*w’*A’*B’* ∇w w)

=tr(w*(w’*A’*B’) ∇w w)

=w’ * (A*B)*w+w*(A’*B’)*w’

=(A*B*w)+(B’*A’*w)

The answer will be ∇w f(w)= A*B*w+B’*A’*w

c) f(w)=tr(Bw w^T A)
To find the Hessian matrix, we need to take the second partial derivative based on the gradient A*B*w+B’*A’*w. In this case it would be

H=∇w(A*B*w)+(B’*A’*w)

H=A*B+B’*A’

This will be the Hessian matrix for f(w)

D) In order to see if f(w) is a convex function, the eigenvalue of the Hessian matrix needs to be positive.

```python
import numpy as np

A = np.array([[2, -2], [-2, 3]])
B = np.array([[2, -1], [-1, 2]])
H = np.matmul(A,B)+np.matmul(np.transpose(B),np.transpose(A))
print(H)
eigenvalues= np.linalg.eigvals(H)
print(eigenvalues)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```
PS C:\Users\andre\OneDrive\Desktop\Uni\Year4\AI 2\Assignment1> python .\part_d.py
[[ 12 -13]
 [-13  16]]
[ 0.84705356 27.15294644]
PS C:\Users\andre\OneDrive\Desktop\Uni\Year4\AI 2\Assignment1>
```

The eigne value return [0.84,27.15] and since they are both positive, it means the f(w) is a convex function.

e) f(w)= log(σ(w^tx))= -log(1+e^-e(w'x))

= -d/dw log(1+e^-e(w'x))

= - 1/(1+e^-(w'x) d/dw(e^-(w'x)

= - 1/(1+e^-(w'x) * (e^-(w'x)) d/dw (-w'x)

= 1/(1+e^-(w'x) * (e^-(w'x))*x

= xe^(-w'x)/(1+e^-(w'x)

Which would be the gradient for f(w)

Question 2:

a)

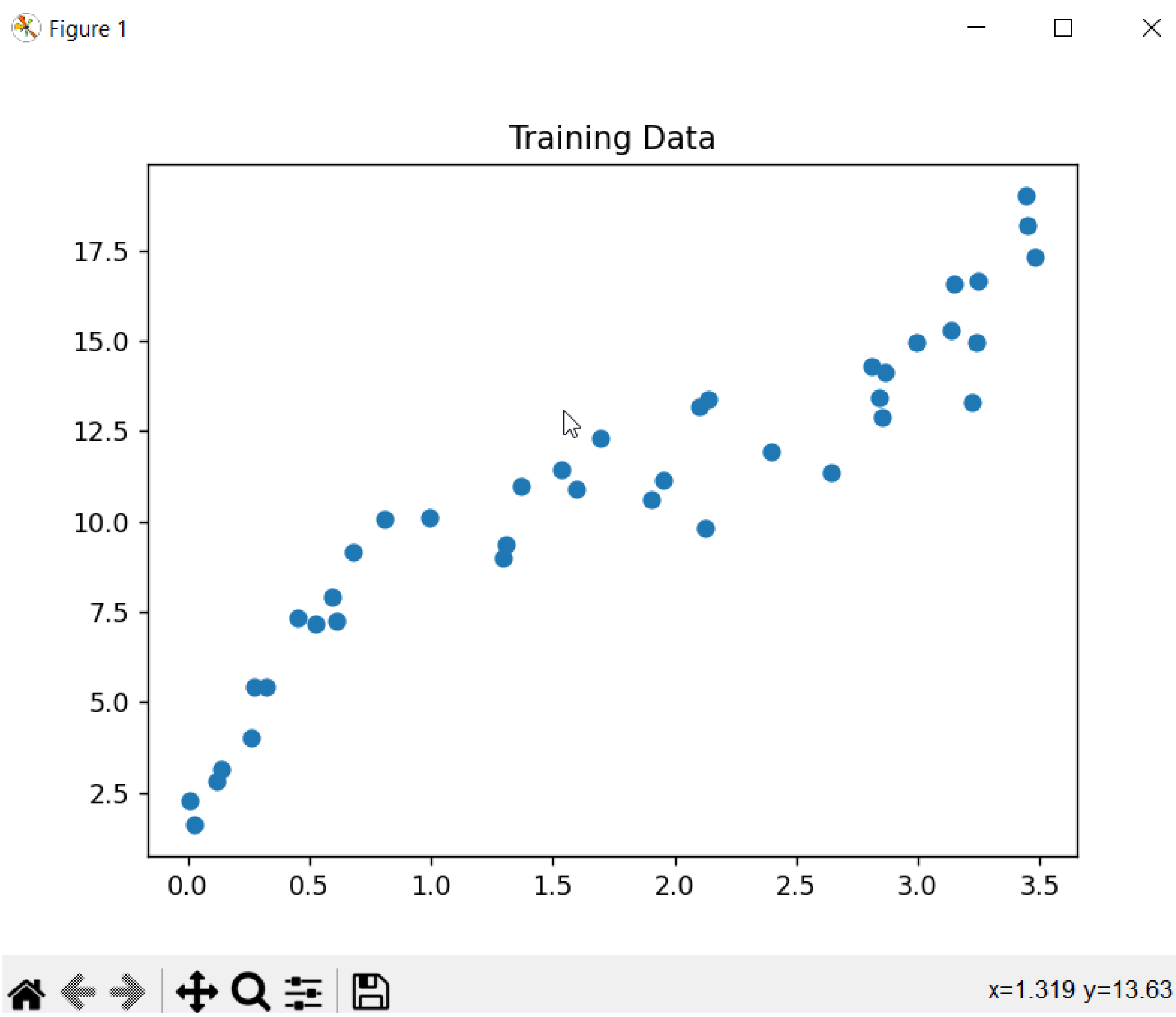


Figure 1
Training Data
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
x=1.319 y=13.63

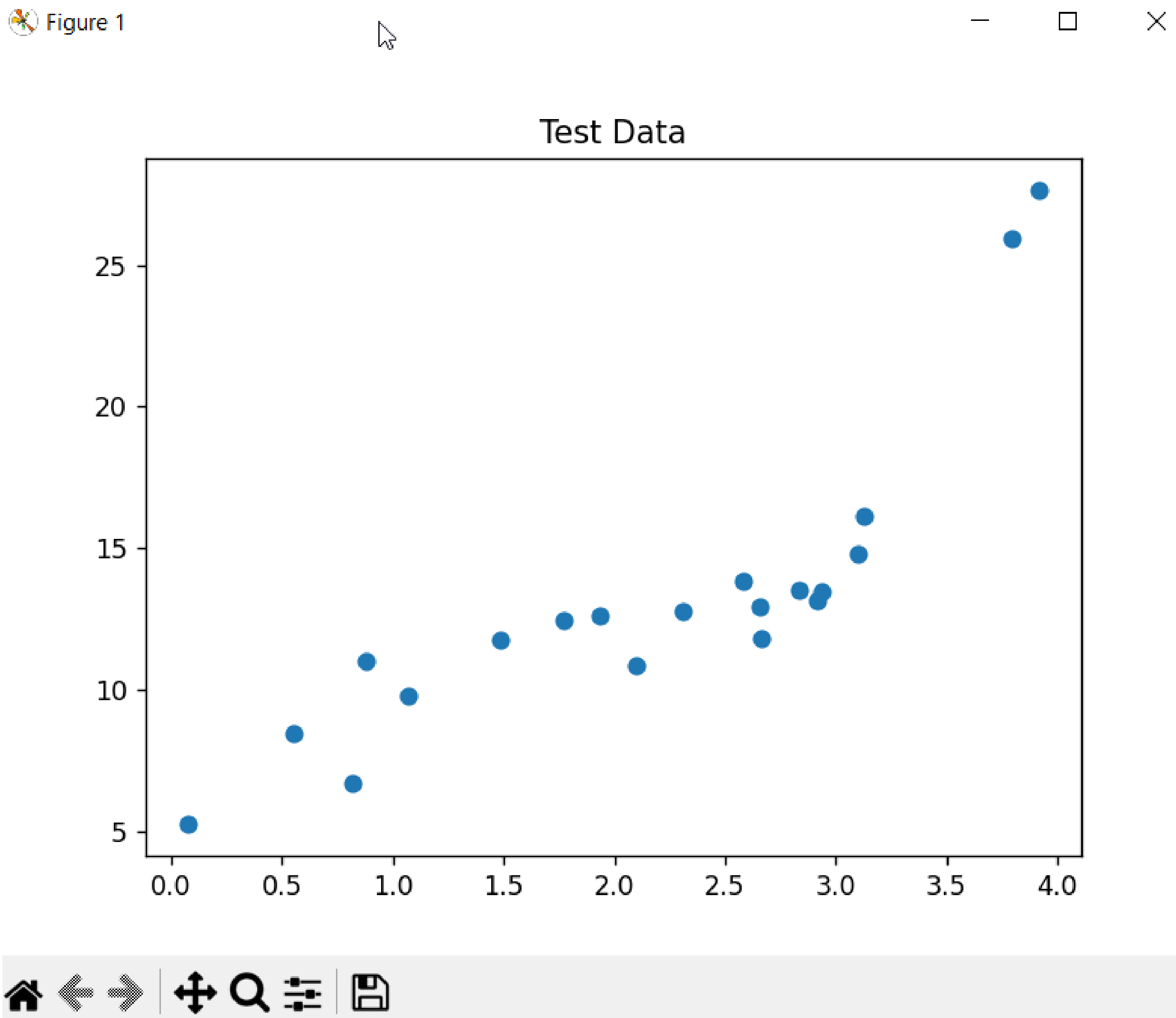
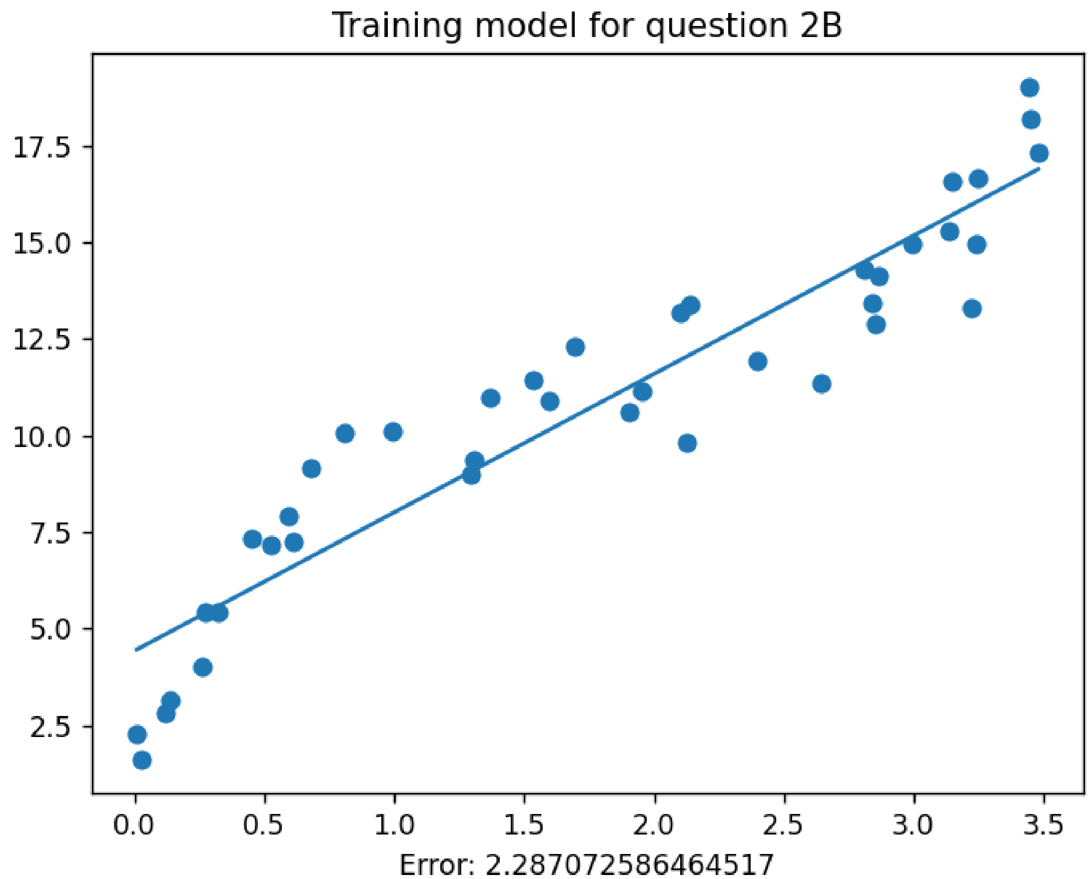
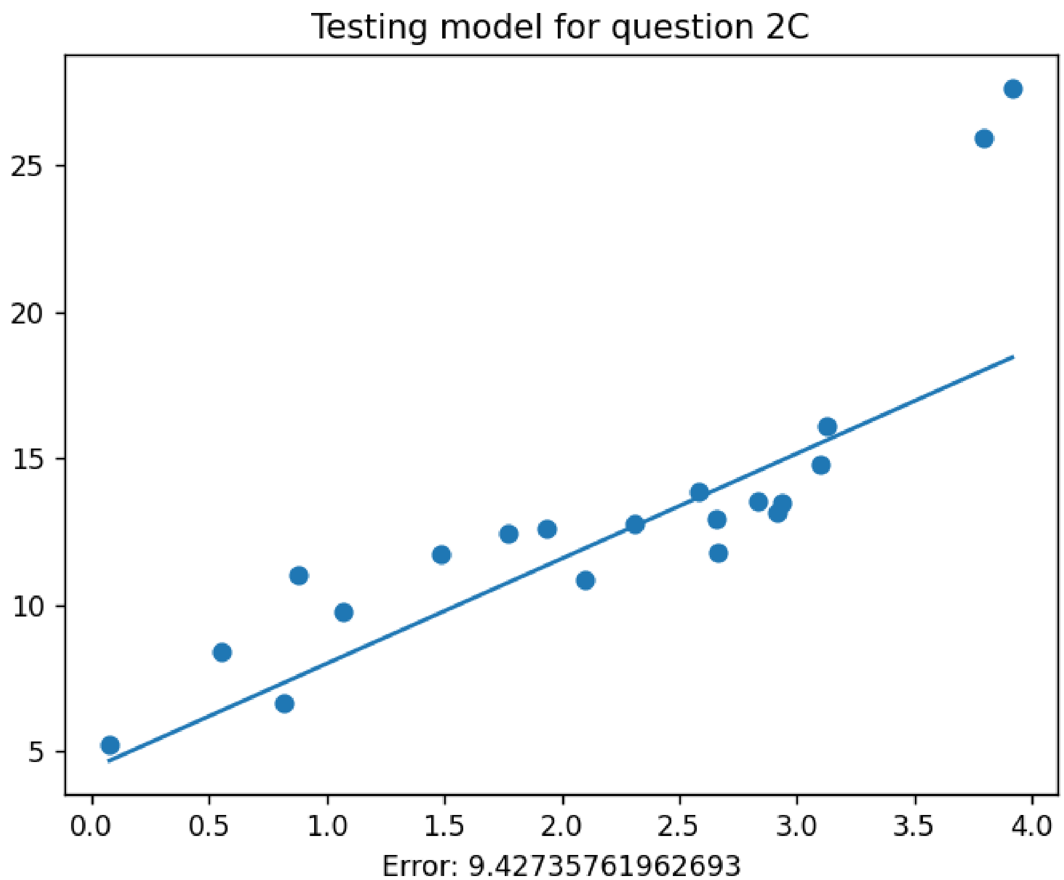
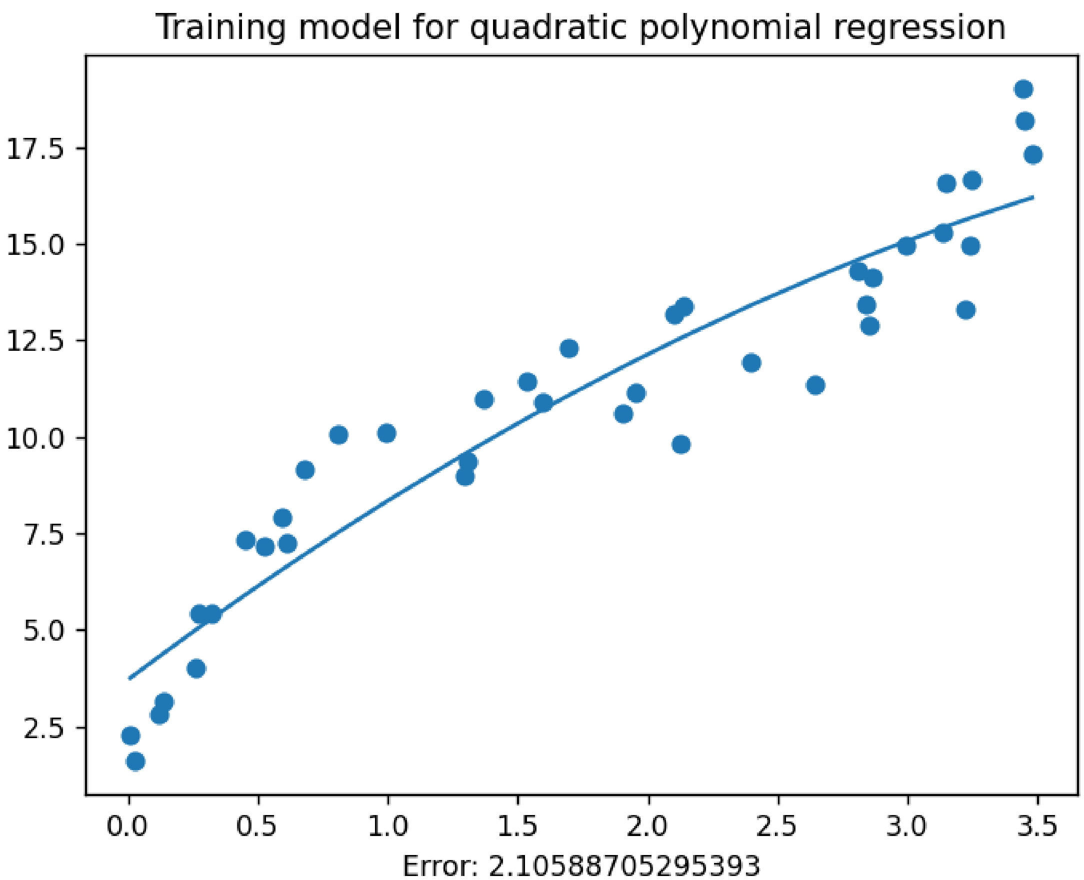
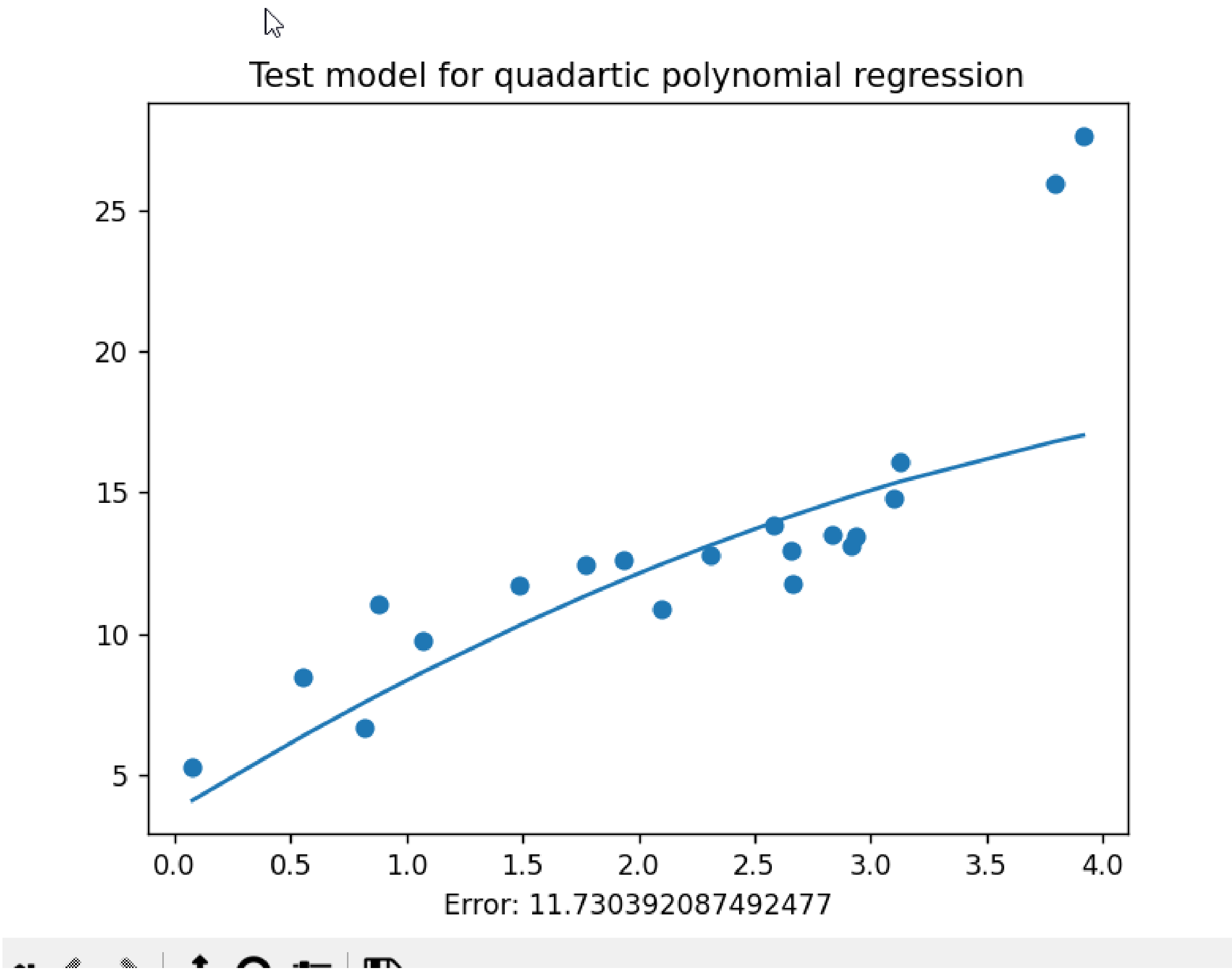


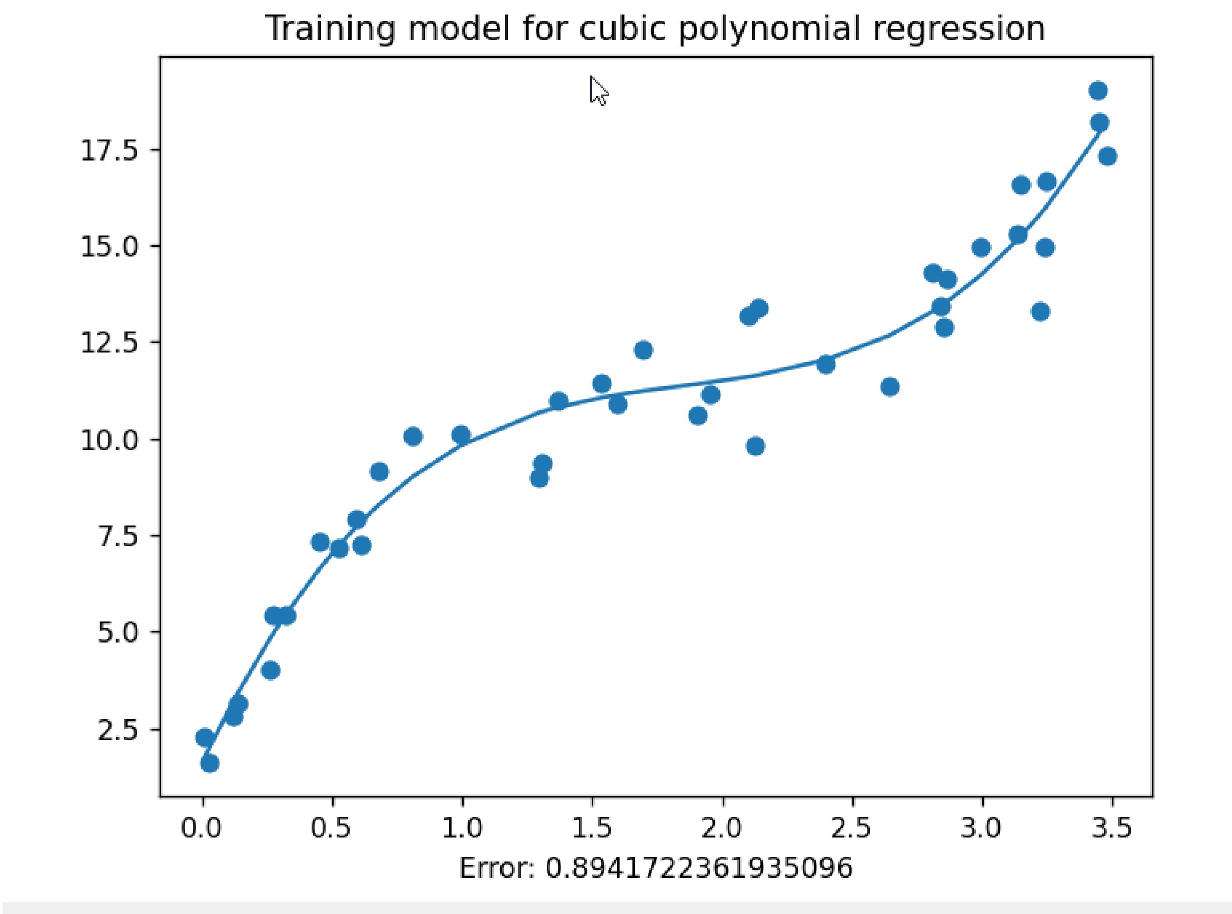
Figure 1
Test Data
25
20
15
10
5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

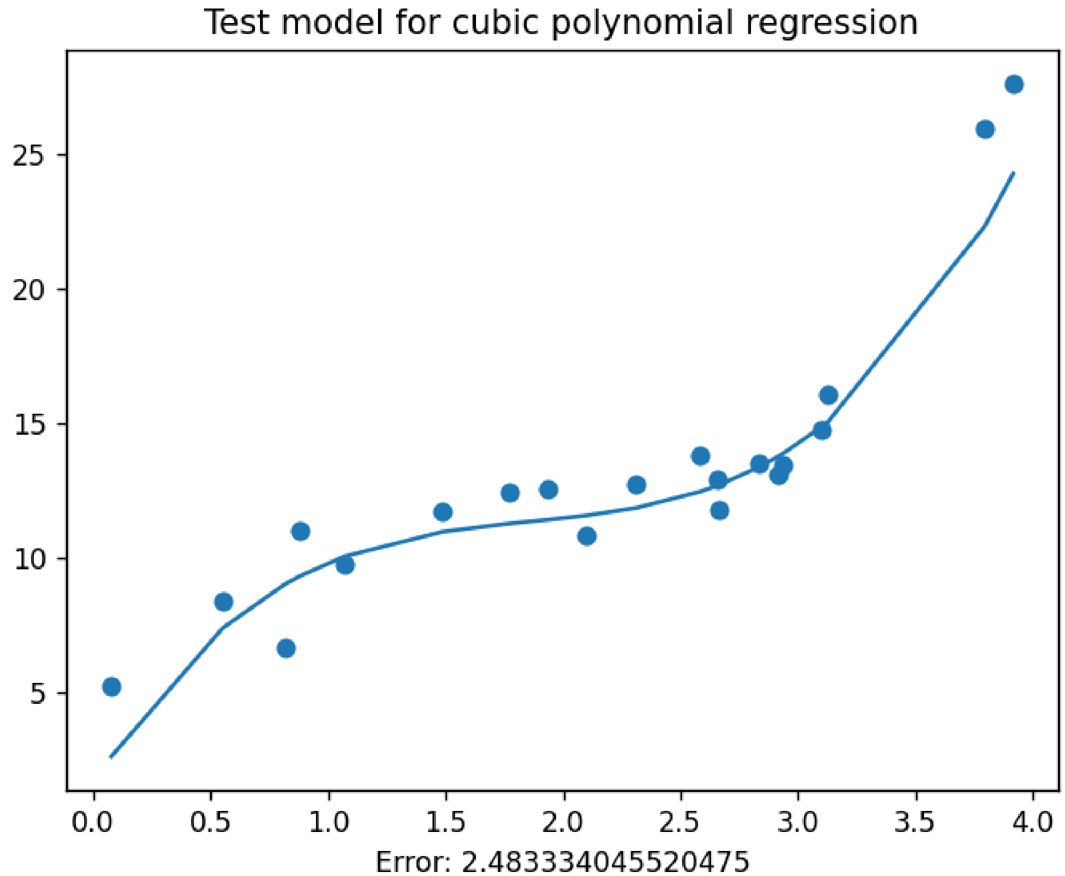

Training model for question 2B
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Error: 2.287072586464517

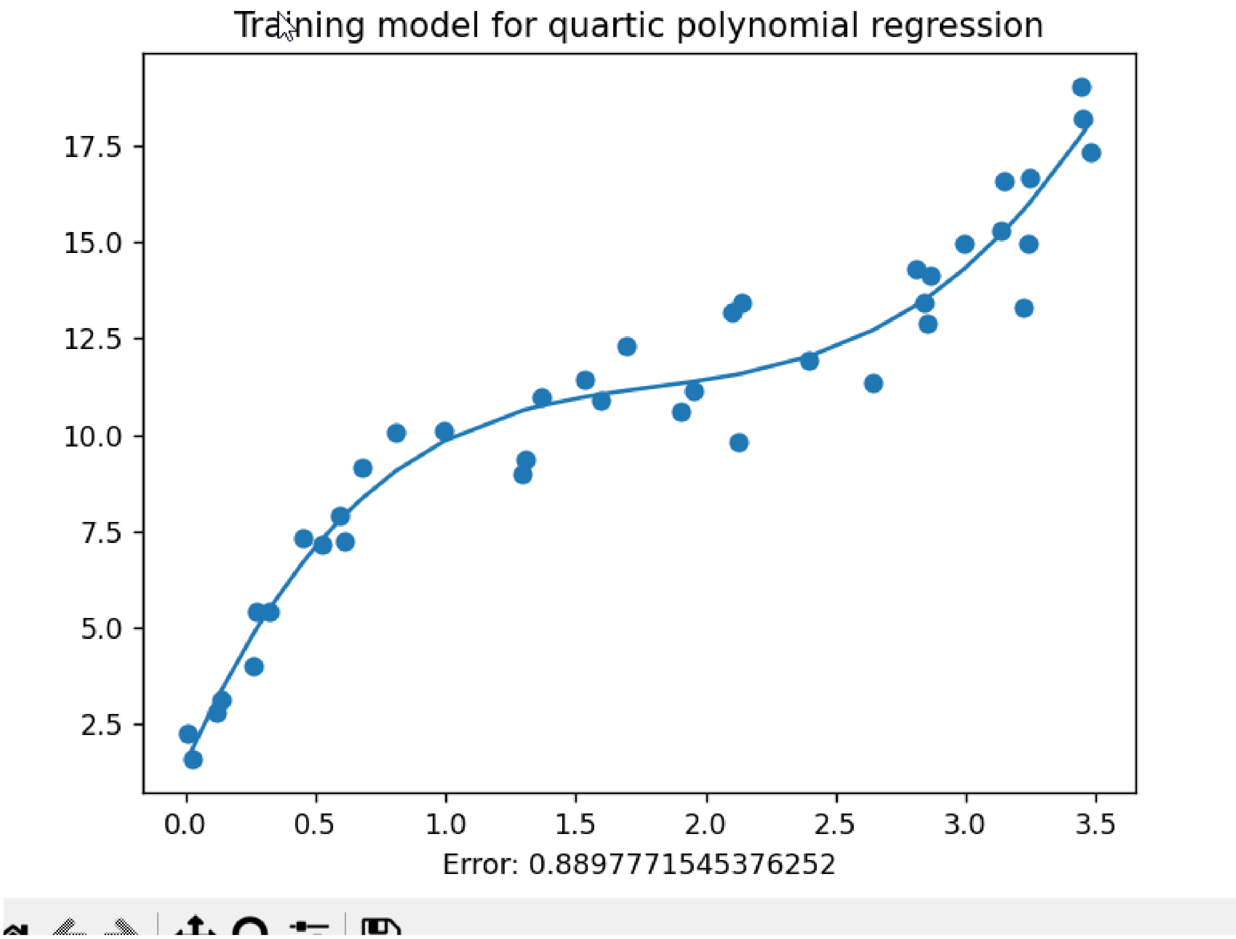

Testing model for question 2C
25
20
15
10
5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Error: 9.42735761962693

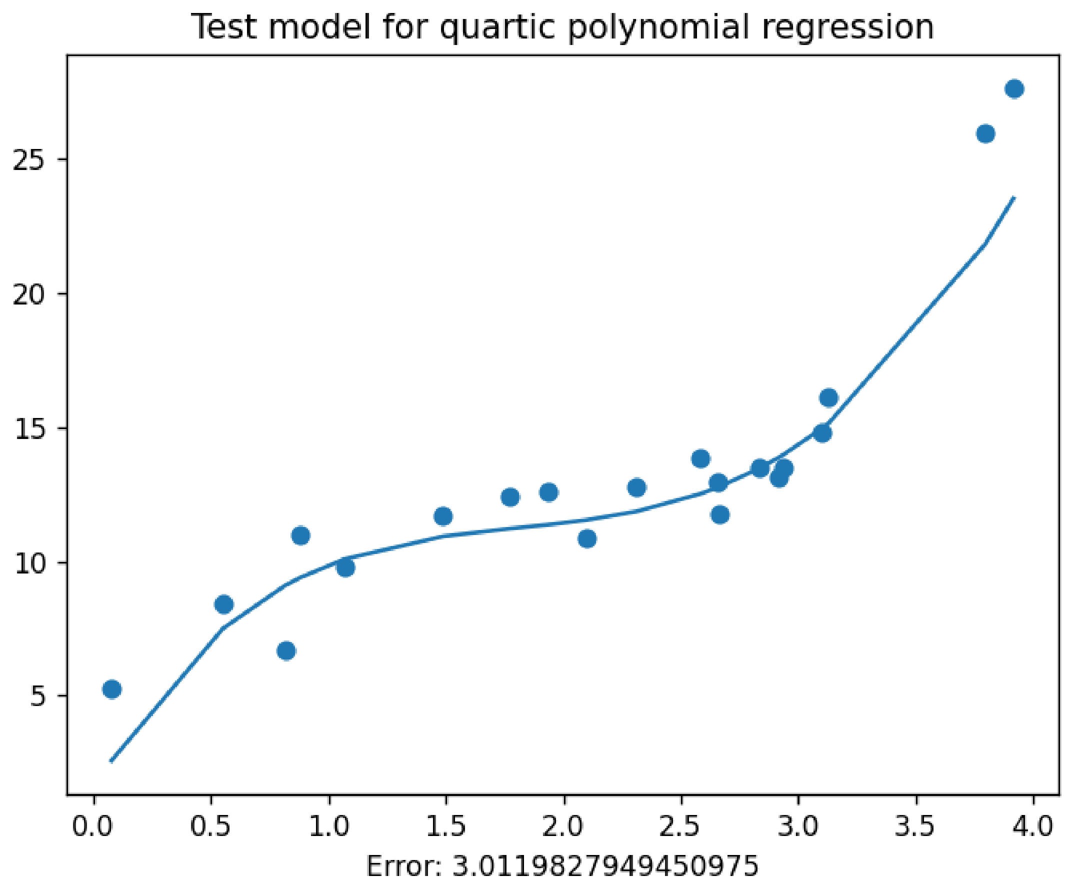

Training model for quadratic polynomial regression
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Error: 2.10588705295393


Test model for quadartic polynomial regression
25
20
15
10
5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Error: 11.730392087492477


Training model for cubic polynomial regression
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Error: 0.8941722361935096


Test model for cubic polynomial regression
25
20
15
10
5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Error: 2.483334045520475


Training model for quartic polynomial regression
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Error: 0.8897771545376252

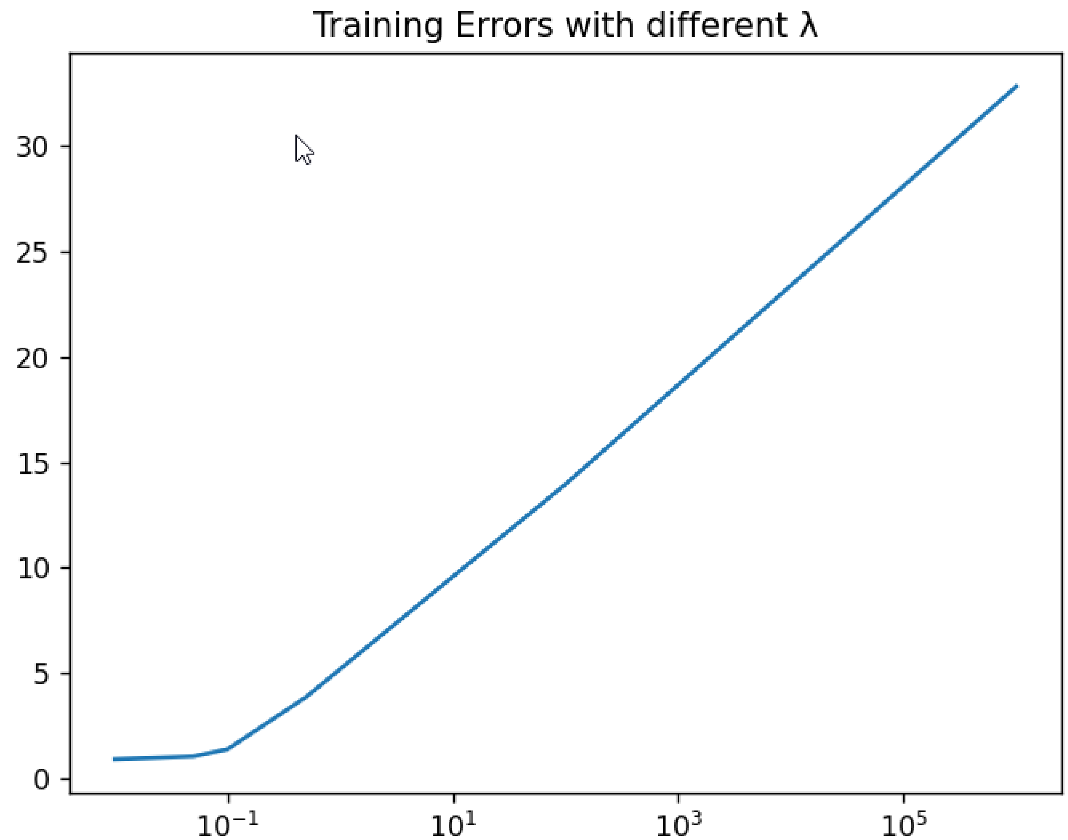


Out of all of different degree of polynomial regression listed above, the best fit will the one with the smallest error value for the testing and training mode. The best equation in the equations above is cubic followed by quartic, linear and quadratic. Even though the quartic have a lower training error then cubic, the actual test error is 0.6 higher than the cubic one which means overfitting.

Question 3A

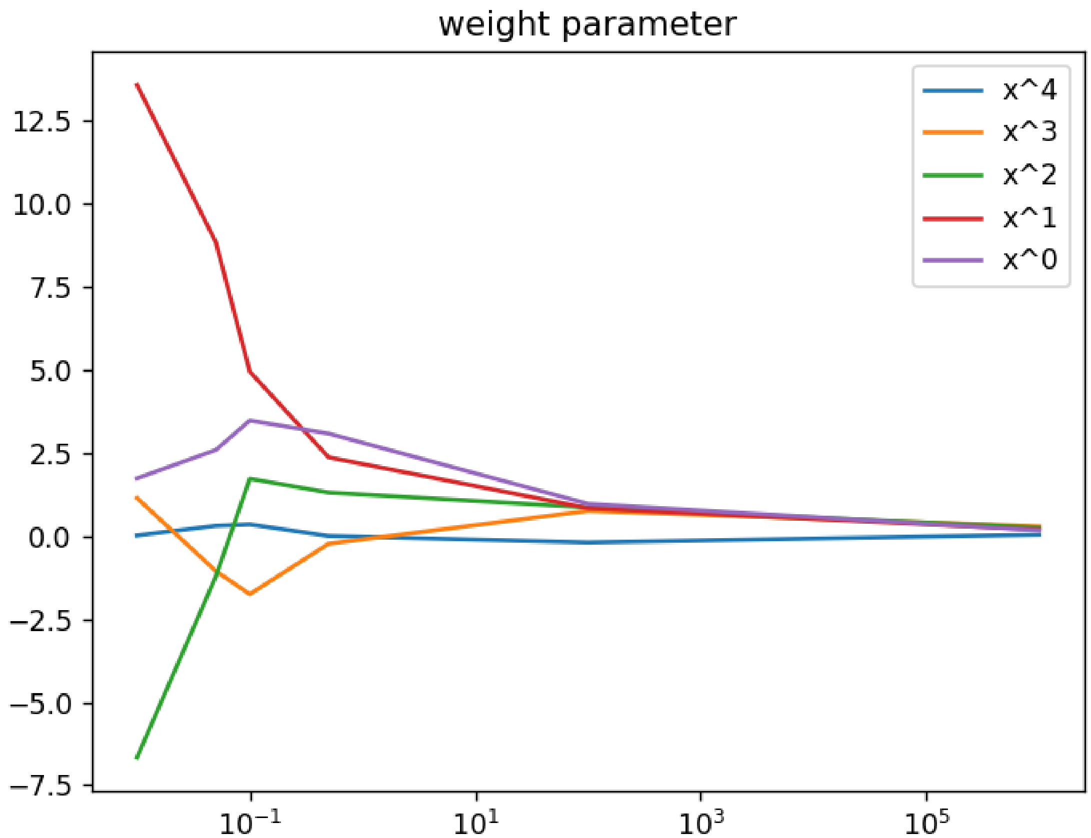
λ	Training Error	Test Error
0.01	0.897482895514694	2.5058009642404486
0.05	1.032312555684975	1.4947374059092655
0.1	1.366437903365513,	2.223431462310436
0.5	3.8514214385291874	13.230161041169646
100	13.948781079351482	25.293945460628255
10^6	32.82941976048144	34.44555027735635

The best λ for fitting the data would be 0.05 even though 0.01 had a lower training error then 0.05, the test error is a lot higher as well. Summing both errors will results in 0.05 having a overall lower error rate.



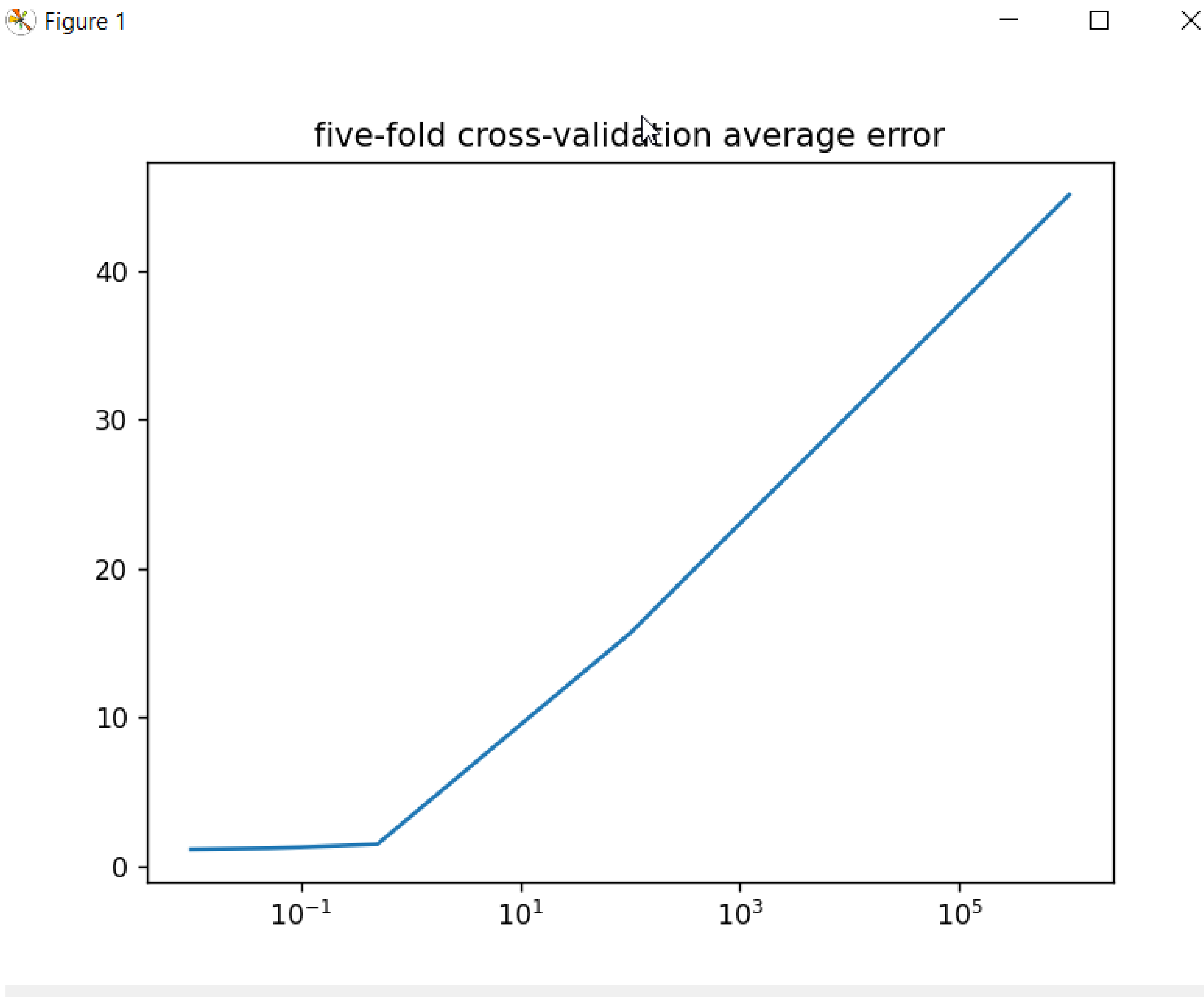
Question 3B

	λ = 0.01	λ = 0.05	λ = 0.1	λ = 0.5	λ = 100	λ = 10^6
W0	1.432e-02	3.035e-01	3.4491e-01	-2.852e-03	-1.975e-01	3.631e-02
W1	1.135	-1.059	-1.754	-2.360e-01	7.524e-01	2.865e-01
W2	6.659	-1.216	1.718	1.301	8.557e-01	2.490e-01
W3	1.355e+01	8.828	4.936	2.363	8.334e-01	1.864e-01
W4	1.729	2.584	3.470	3.076	9.618e-01	1.667e-01



Question 3C

	λ = 0.01	λ = 0.05	λ = 0.1	λ = 0.5	λ = 100	λ = 10^6
Average test error	1.154	1.227	1.294	1.509	15.667	45.101



The best average error is λ = 0.01 which is different from 3a but 0.01 had the lowest training error in 3a so the five-fold cross-validation can determine the best regularization parameter

Best fit (λ = 0.01)