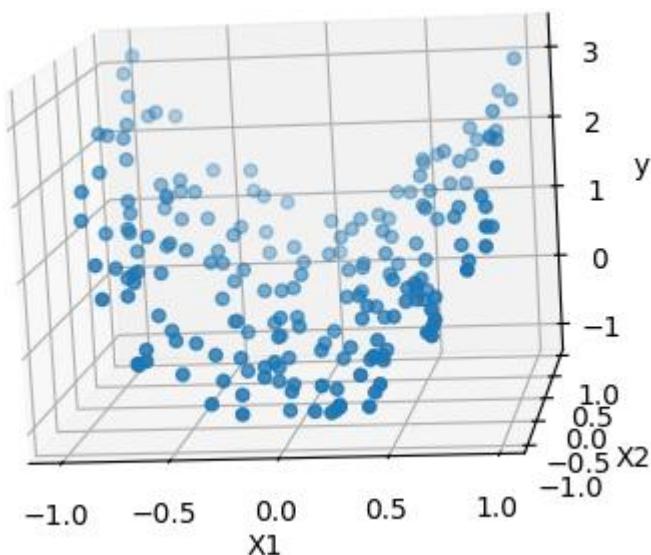
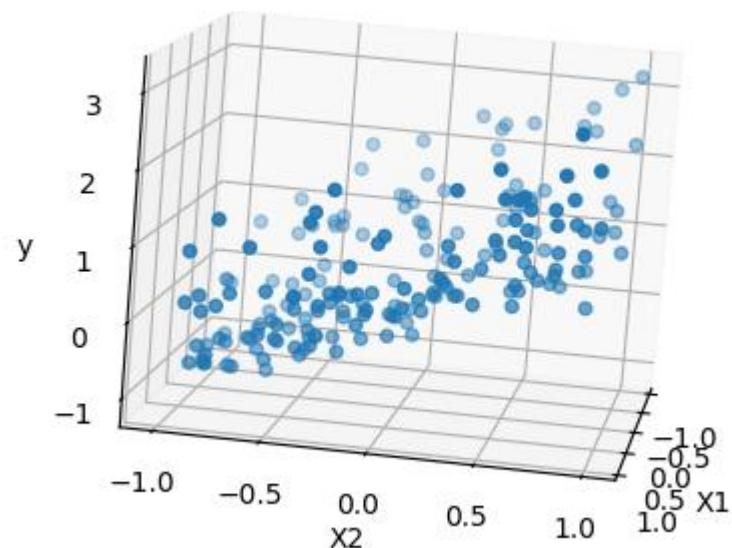


(i) Question i

- a. The data seems to be along a curve when viewed primarily from the first data



. It appears much closer to a linear solution



when viewed from X2

C= 0.001 alpha= 500.0 Intercept: 0.602128856408119 1 : 0.0 X1 : -0.0 X2 : 0.0 X1^2 : 0.0 X1 X2 : 0.0 X2^2 : -0.0 X1^3 : -0.0 X1^2 X2 : 0.0 X1 X2^2 : -0.0 X2^3 : 0.0 X1^4 : 0.0 X1^3 X2 : -0.0 X1^2 X2^2 : 0.0 X1 X2^3 : -0.0 X2^4 : -0.0 X1^5 : -0.0 X1^4 X2 : 0.0 X1^3 X2^2 : -0.0 X1^2 X2^3 : 0.0 X1 X2^4 : -0.0 X2^5 : 0.0	C= 0.01 alpha= 50.0 Intercept: 0.602128856408119 1 : 0.0 X1 : -0.0 X2 : 0.0 X1^2 : 0.0 X1 X2 : 0.0 X2^2 : -0.0 X1^3 : -0.0 X1^2 X2 : 0.0 X1 X2^2 : -0.0 X2^3 : 0.0 X1^4 : 0.0 X1^3 X2 : -0.0 X1^2 X2^2 : 0.0 X1 X2^3 : -0.0 X2^4 : -0.0 X1^5 : -0.0 X1^4 X2 : 0.0 X1^3 X2^2 : -0.0 X1^2 X2^3 : 0.0 X1 X2^4 : -0.0 X2^5 : 0.0	C= 0.1 alpha= 5.0 Intercept: 0.602128856408119 1 : 0.0 X1 : -0.0 X2 : 0.0 X1^2 : 0.0 X1 X2 : 0.0 X2^2 : -0.0 X1^3 : -0.0 X1^2 X2 : 0.0 X1 X2^2 : -0.0 X2^3 : 0.0 X1^4 : 0.0 X1^3 X2 : -0.0 X1^2 X2^2 : 0.0 X1 X2^3 : -0.0 X2^4 : -0.0 X1^5 : -0.0 X1^4 X2 : 0.0 X1^3 X2^2 : -0.0 X1^2 X2^3 : 0.0 X1 X2^4 : -0.0 X2^5 : 0.0
---	---	---

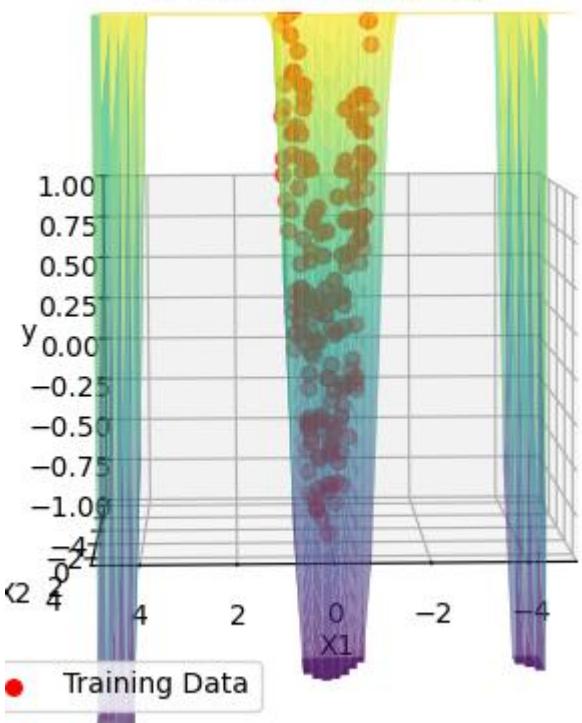
b.

C= 1 alpha= 0.5 Intercept: 0.602128856408119 1 : 0.0 X1 : -0.0 X2 : 0.0 X1^2 : 0.0 X1 X2 : 0.0 X2^2 : -0.0 X1^3 : -0.0 X1^2 X2 : 0.0 X1 X2^2 : -0.0 X2^3 : 0.0 X1^4 : 0.0 X1^3 X2 : -0.0 X1^2 X2^2 : 0.0 X1 X2^3 : -0.0 X2^4 : -0.0 X1^5 : -0.0 X1^4 X2 : 0.0 X1^3 X2^2 : -0.0 X1^2 X2^3 : 0.0 X1 X2^4 : -0.0 X2^5 : 0.0	C= 10 alpha= 0.05 Intercept: 0.0951579778587982 1 : 0.0 X1 : -0.0 X2 : 0.8737999498324229 X1^2 : 1.6513184403380956 X1 X2 : -0.0 X2^2 : 0.0 X1^3 : -0.0 X1^2 X2 : 0.0 X1 X2^2 : -0.0 X2^3 : 0.0 X1^4 : 0.0 X1^3 X2 : -0.0 X1^2 X2^2 : 0.0 X1 X2^3 : -0.0 X2^4 : 0.0 X1^5 : -0.0 X1^4 X2 : 0.0 X1^3 X2^2 : -0.0 X1^2 X2^3 : 0.0 X1 X2^4 : -0.0 X2^5 : 0.0	C= 10 alpha= 0.05 Intercept: 0.0951579778587982 1 : 0.0 X1 : -0.0 X2 : 0.8737999498324229 X1^2 : 1.6513184403380956 X1 X2 : -0.0 X2^2 : 0.0 X1^3 : -0.0 X1^2 X2 : 0.0 X1 X2^2 : -0.0 X2^3 : 0.0 X1^4 : 0.0 X1^3 X2 : -0.0 X1^2 X2^2 : 0.0 X1 X2^3 : -0.0 X2^4 : 0.0 X1^5 : -0.0 X1^4 X2 : 0.0 X1^3 X2^2 : -0.0 X1^2 X2^3 : 0.0 X1 X2^4 : -0.0 X2^5 : 0.0
---	---	---

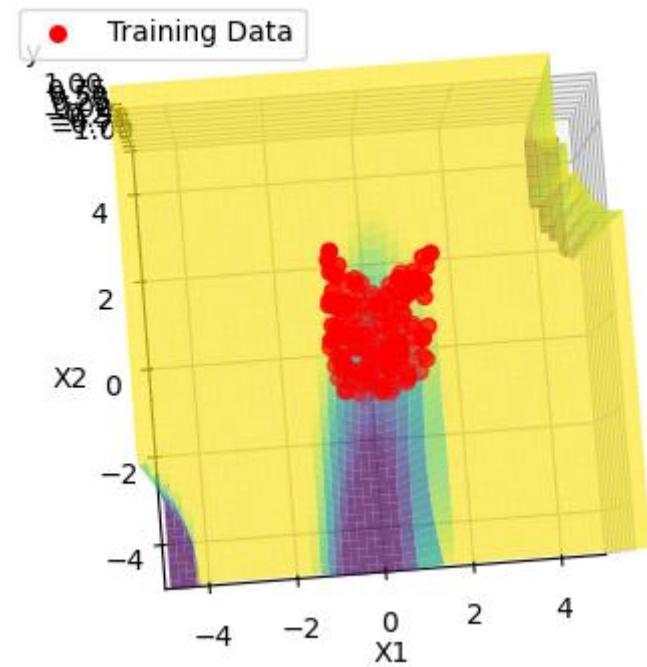
```
C= 1000
alpha= 0.0005
Intercept: -0.0837117528262934
1 : 0.0
X1 : 0.051785941606038846
X2 : 0.9843721567741849
X1^2 : 2.1604743293628177
X1 X2 : 0.13682535077759633
X2^2 : 0.0
X1^3 : -0.04585560919533065
X1^2 X2 : 0.0
X1 X2^2 : 0.0
X2^3 : -0.0
X1^4 : 0.0
X1^3 X2 : -0.255292343159568
X1^2 X2^2 : 0.13277679836358106
X1 X2^3 : -0.016176838723923018
X2^4 : 0.0
X1^5 : -0.0
X1^4 X2 : -0.002742299940941166
X1^3 X2^2 : -0.0
X1^2 X2^3 : -0.0
X1 X2^4 : 0.0
X2^5 : 0.0
```

As the C value increased, and the alpha value lowered, more coefficients came into effect, With C values at or below 1, we can see that it is solely reliant on the intercept, with no non-zero coefficients. Whereas the higher C values have a C value much closer to 0, but use a lot more coefficients, this is great for our model, but may result in overfitting.

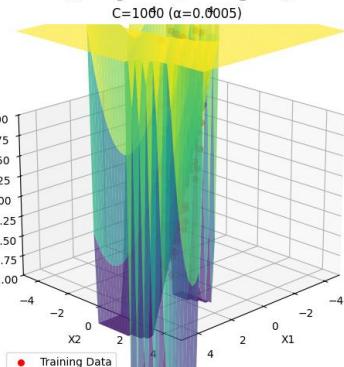
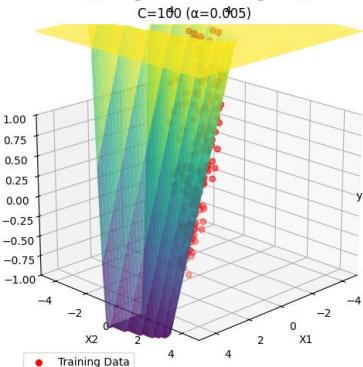
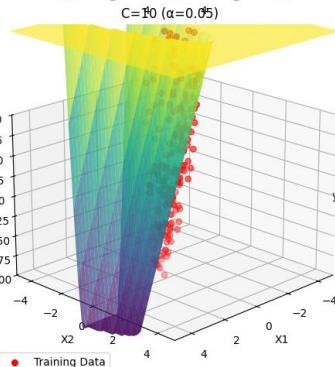
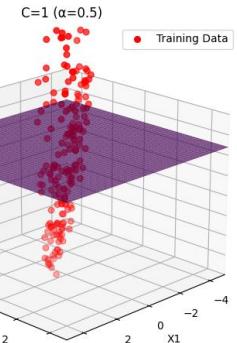
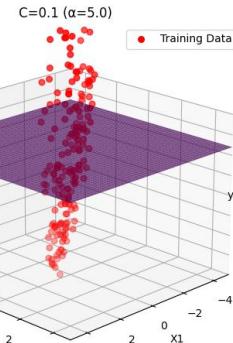
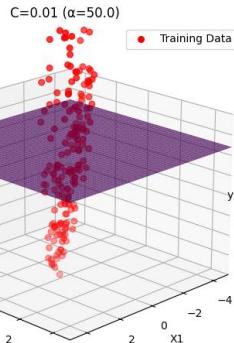
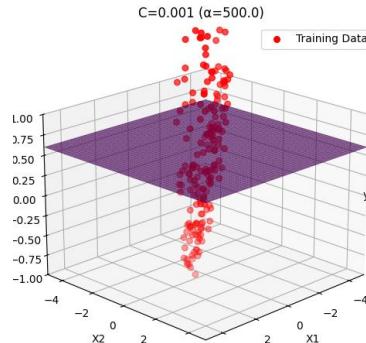
$C=1000 (\alpha=0.0005)$



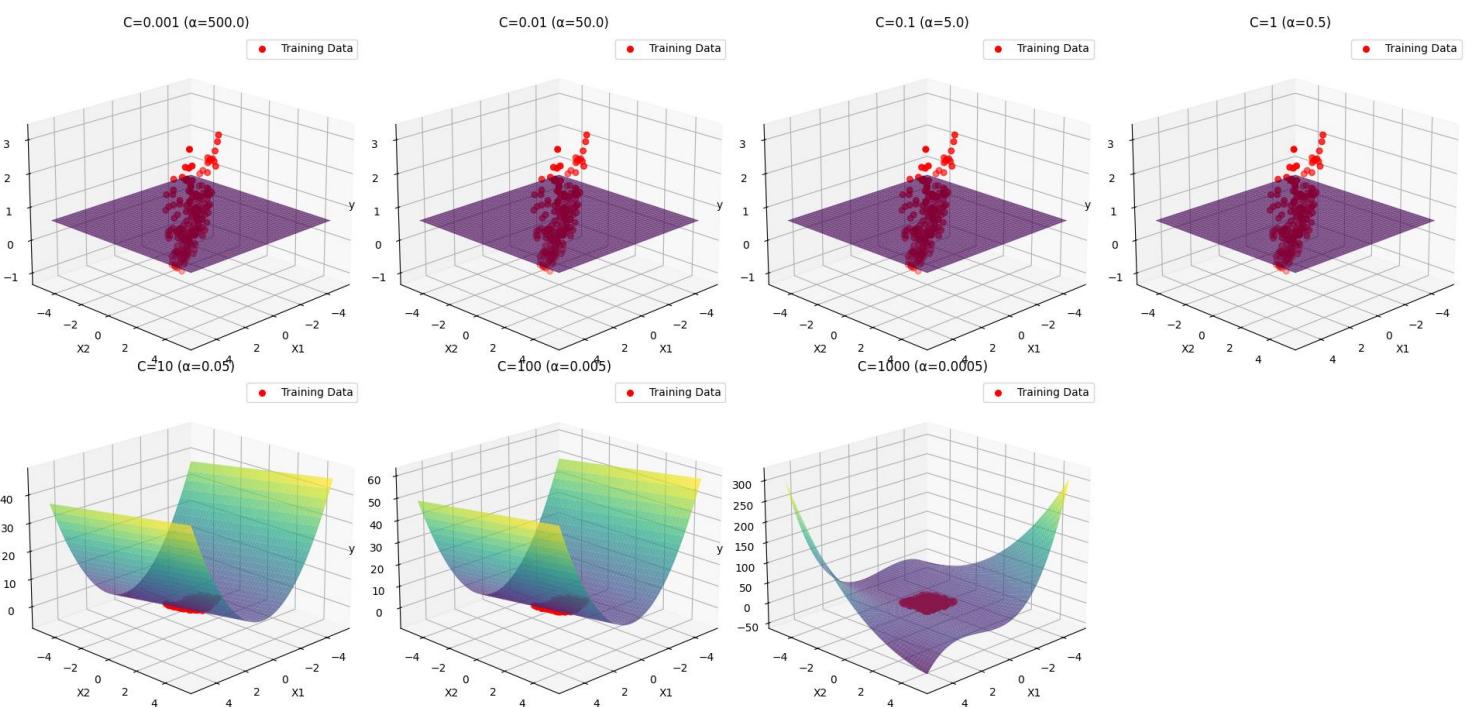
$C=1000 (\alpha=0.0005)$



C.

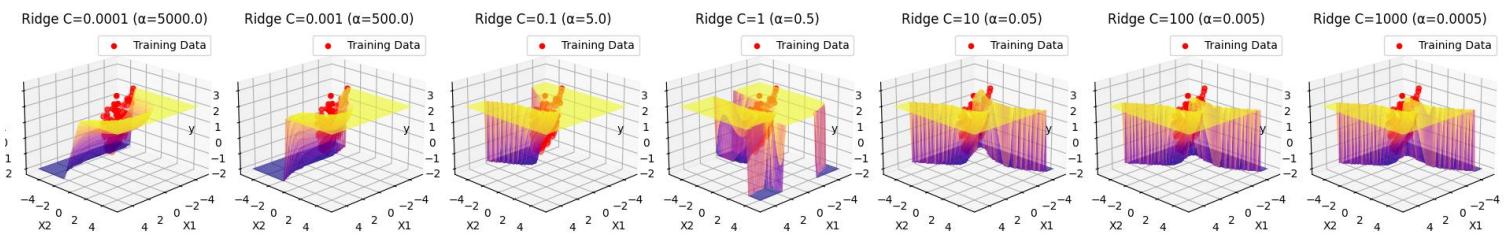
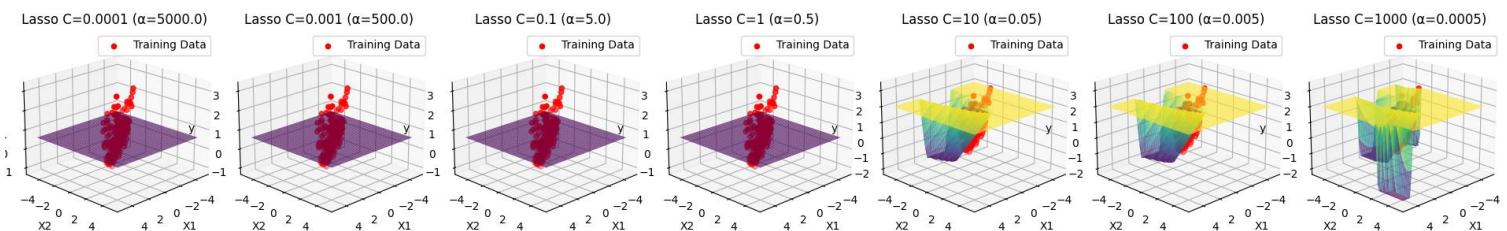


While the surface plots are hard to read at higher  $c$  values, we can see that they are very closely accurate to the training data, and become more jagged and chaotic as it goes on.

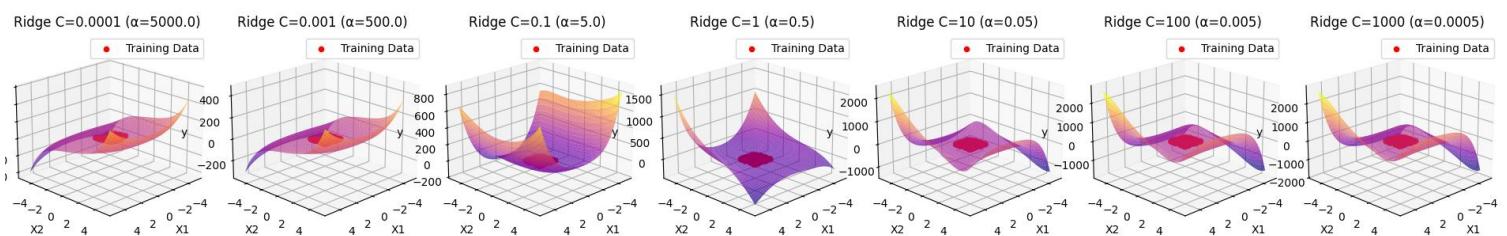
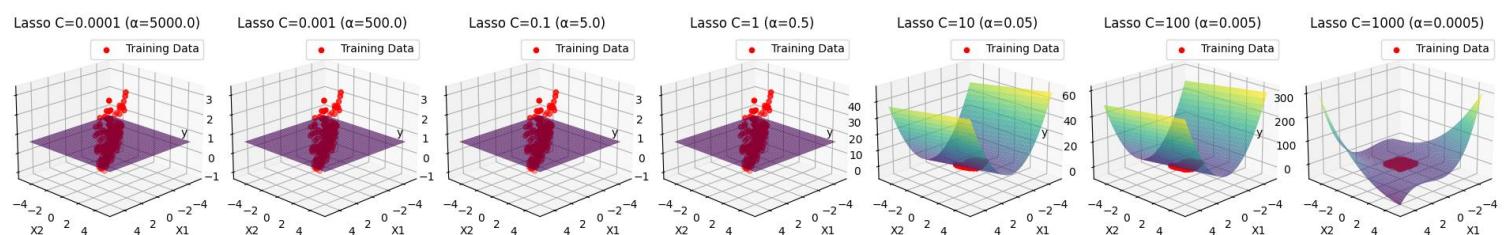


When the graphs are unrestricted and zoomed out we are able to see that they take on much more complex shapes, and that their y values are a lot higher than their X1 and X2 values, this is a clear sign of overfitting with high values of C

- d. By looking at the graphs in question c we can clearly see the y values of the graphs shoot up drastically with high c values(such as 1000), this is not very likely to be accurate to any data outside of what we have received, but is highly accurate to the data we do have, a clear case of overfitting. Inversly, low values of C show a completely flat plane with an atrocious accuracy to the data given, this is a clear case of underfitting, finding a balance to te C value is vital to get a well fitted model



e.



**Ridge Regression - C= 0.0001**  
alpha= 5000.0  
Intercept: 0.5986439014769379  
1 : 0.0  
X1 : -0.0017832656327079418  
X2 : 0.014500226697456937  
X1^2 : 0.008168506705246189  
X1 X2 : 0.00021701975168673102  
X2^2 : -0.00013564497460076912  
X1^3 : -0.0013225993532073332  
X1^2 X2 : 0.004987056027096896  
X1 X2^2 : -0.00033007838767578994  
X2^3 : 0.008775568600736115  
X1^4 : 0.00705052520585937  
X1^3 X2 : -0.000436496923680714  
X1^2 X2^2 : 0.002933979247124109  
X1 X2^3 : -2.9331304889968088e-05  
X2^4 : -0.000216565991318186  
X1^5 : -0.0012566636758364128  
X1^4 X2 : 0.003123120471126489  
X1^3 X2^2 : -0.00019611995816730252  
X1^2 X2^3 : 0.0031944513200731495  
X1 X2^4 : -0.00037150347305411733  
X2^5 : 0.006386191579806327

**Ridge Regression - C= 0.001**  
alpha= 500.0  
Intercept: 0.5688177136159654  
1 : 0.0  
X1 : -0.01314807216041576  
X2 : 0.11968176568506851  
X1^2 : 0.07566008719614753  
X1 X2 : 0.0017468756312059876  
X2^2 : -0.0007793085072267294  
X1^3 : -0.01008970723891265  
X1^2 X2 : 0.040517905598984834  
X1 X2^2 : -0.0014387091647439624  
X2^3 : 0.07113549868108984  
X1^4 : 0.06493021600892634  
X1^3 X2 : -0.004041931645871178  
X1^2 X2^2 : 0.026923261535288432  
X1 X2^3 : -0.0002612335321174325  
X2^4 : -0.001454548013669917  
X1^5 : -0.009891112028176217  
X1^4 X2 : 0.025361145257580306  
X1^3 X2^2 : -0.0007775555266538019  
X1^2 X2^3 : 0.025709857728586184  
X1 X2^4 : -0.002184319665153456  
X2^5 : 0.05118208677617537

```
Ridge Regression - C= 0.1
```

```
alpha= 5.0
```

```
Intercept: 0.09669578157931102
```

```
1 : 0.0
```

```
X1 : 0.018506670723962768
```

```
X2 : 0.7831905339705998
```

```
X1^2 : 1.122196031434751
```

```
X1 X2 : 0.042628087457732805
```

```
X2^2 : -0.02479640521505338
```

```
X1^3 : -0.028997020046993387
```

```
X1^2 X2 : 0.09849525108452249
```

```
X1 X2^2 : 0.02681899310112826
```

```
X2^3 : 0.18869024125726527
```

```
X1^4 : 0.7640885051601497
```

```
X1^3 X2 : -0.08670502375654938
```

```
X1^2 X2^2 : 0.31778737108100846
```

```
X1 X2^3 : 0.008217055239094266
```

```
X2^4 : -0.012605532428787161
```

```
X1^5 : -0.020771974216300577
```

```
X1^4 X2 : 0.009706155694211054
```

```
X1^3 X2^2 : 0.01476462253884801
```

```
X1^2 X2^3 : -0.026070790450066463
```

```
X1 X2^4 : 0.008584632046978563
```

```
X2^5 : 0.03409579846114025
```

```
Ridge Regression - C= 1
```

```
alpha= 0.5
```

```
Intercept: -0.020596118336715707
```

```
1 : 0.0
```

```
X1 : 0.07466469736109828
```

```
X2 : 0.9548147098009786
```

```
X1^2 : 1.6968072892360284
```

```
X1 X2 : 0.14742297331599105
```

```
X2^2 : -0.03913535032177831
```

```
X1^3 : -0.1170099025919902
```

```
X1^2 X2 : 0.15729059879722698
```

```
X1 X2^2 : -0.011243509038852292
```

```
X2^3 : 0.03065938684142042
```

```
X1^4 : 0.4429655817073398
```

```
X1^3 X2 : -0.21651872620443513
```

```
X1^2 X2^2 : 0.28900045146715997
```

```
X1 X2^3 : -0.045410504168113375
```

```
X2^4 : 0.002894574858176119
```

```
X1^5 : 0.05358128046653592
```

```
X1^4 X2 : -0.1379148497967317
```

```
X1^3 X2^2 : -0.0322981697011069
```

```
X1^2 X2^3 : -0.12862662421864493
```

```
X1 X2^4 : 0.02476754108221688
```

```
X2^5 : 0.006745085311525403
```

```
Ridge Regression - C= 10
```

```
alpha= 0.05
```

```
Intercept: -0.07337423953927391
```

```
1 : 0.0
```

```
X1 : 0.148000156845419
```

```
X2 : 0.99965217150856
```

```
X1^2 : 2.0917619640921927
```

```
X1 X2 : 0.2573058493344188
```

```
X2^2 : -0.05635730456615979
```

```
X1^3 : -0.3406161774674386
```

```
X1^2 X2 : 0.349579543022987
```

```
X1 X2^2 : -0.10862022090426456
```

```
X2^3 : -0.23344562774867575
```

```
X1^4 : 0.06871419966718388
```

```
X1^3 X2 : -0.3486630648472574
```

```
X1^2 X2^2 : 0.20612284364565622
```

```
X1 X2^3 : -0.10600929306222906
```

```
X2^4 : 0.0530413512074979
```

```
X1^5 : 0.22225507759165525
```

```
X1^4 X2 : -0.3276677686714913
```

```
X1^3 X2^2 : 0.03293963818438442
```

```
X1^2 X2^3 : -0.14012465235305466
```

```
X1 X2^4 : 0.08837573987231888
```

```
X2^5 : 0.22713821471485415
```

```
Ridge Regression - C= 100
```

```
alpha= 0.005
```

```
Intercept: -0.08182388444362576
```

```
1 : 0.0
```

```
X1 : 0.18447124380834964
```

```
X2 : 1.0213228755730057
```

```
X1^2 : 2.1602680411683854
```

```
X1 X2 : 0.28449067017631535
```

```
X2^2 : -0.06693061549517658
```

```
X1^3 : -0.4555332985229183
```

```
X1^2 X2 : 0.40503005832896494
```

```
X1 X2^2 : -0.1627713050609357
```

```
X2^3 : -0.35924894589052203
```

```
X1^4 : 0.0033037365094292755
```

```
X1^3 X2 : -0.38136616269113083
```

```
X1^2 X2^2 : 0.19040076338330114
```

```
X1 X2^3 : -0.11969956376325538
```

```
X2^4 : 0.07228330835663814
```

```
X1^5 : 0.30859213556794307
```

```
X1^4 X2 : -0.38528609899808647
```

```
X1^3 X2^2 : 0.07713557583293829
```

```
X1^2 X2^3 : -0.1355199211935793
```

```
X1 X2^4 : 0.11958391408912121
```

```
X2^5 : 0.3353076600221713
```

```
Ridge Regression - C= 1000
```

```
alpha= 0.0005
```

```
Intercept: -0.08268950179017576
```

```
1 : 0.0
```

```
X1 : 0.18975224669880114
```

```
X2 : 1.024618562442109
```

```
X1^2 : 2.1674484329318746
```

```
X1 X2 : 0.28779552771541117
```

```
X2^2 : -0.06843756860149738
```

```
X1^3 : -0.4723400480069015
```

```
X1^2 X2 : 0.41128401944014015
```

```
X1 X2^2 : -0.17058777400205868
```

```
X2^3 : -0.3771559429057741
```

```
X1^4 : -0.0034500032186734913
```

```
X1^3 X2 : -0.3853324730110185
```

```
X1^2 X2^2 : 0.1886575151138132
```

```
X1 X2^3 : -0.12131654854798155
```

```
X2^4 : 0.07484459210344474
```

```
X1^5 : 0.32130314351382877
```

```
X1^4 X2 : -0.3919998235127982
```

```
X1^3 X2^2 : 0.08355800054306878
```

```
X1^2 X2^3 : -0.13454329182667837
```

```
X1 X2^4 : 0.12405278355240065
```

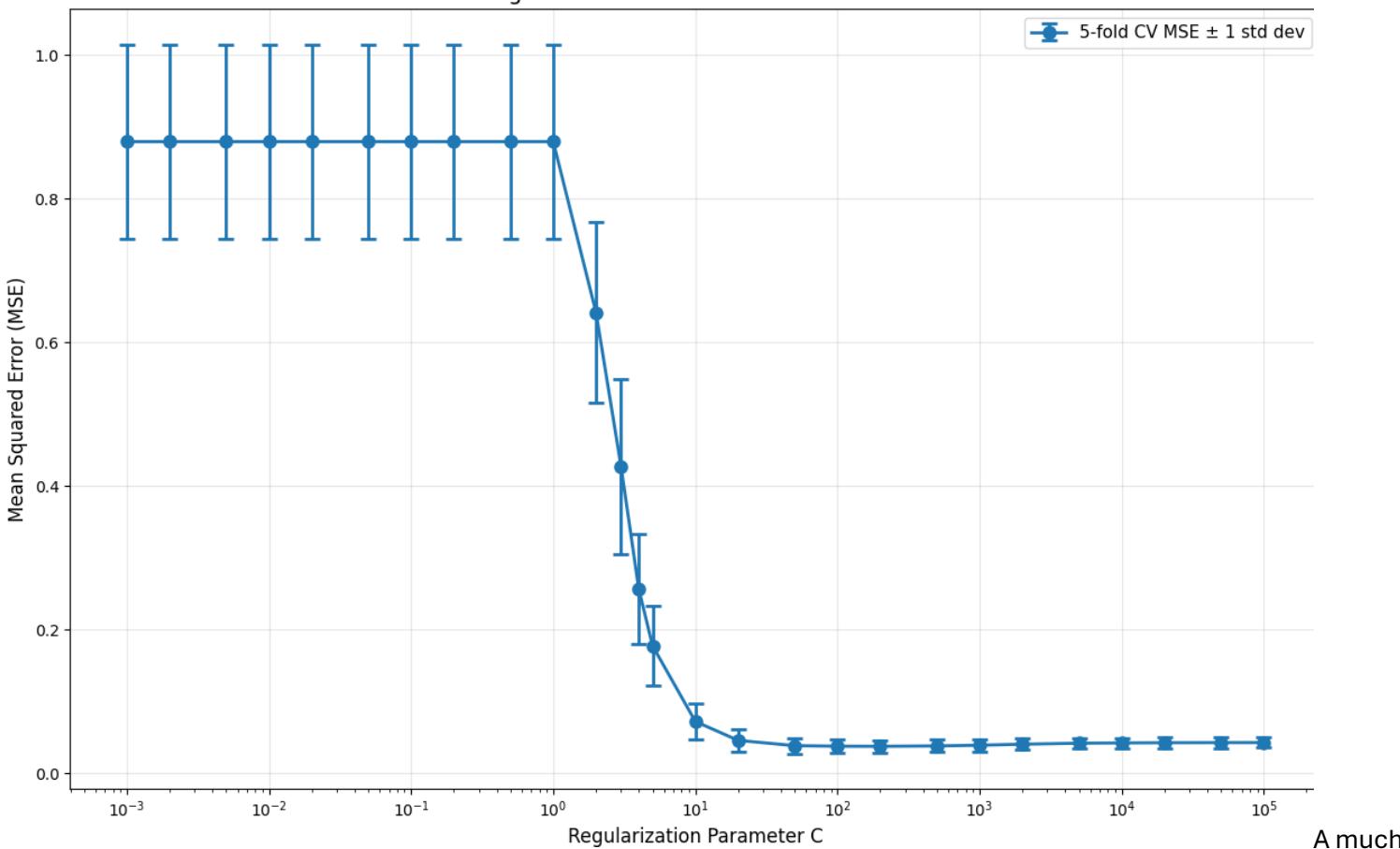
```
X2^5 : 0.3506103349496733
```

The biggest difference between the coefficients of the Ridge model is how there are none at 0, many get very very close (e.g. -0.0007) but none ever reach 0

The most noticeable difference on the graphs is how the low levels of C still have complex suitable forms, fitting the data with surprising accuracy and having a more complex model when seen in whole. Another feature is how the lower c values cause major shift, while the higher ones seem to fall into line, the opposite of the lasso models.

(ii) Cross validation to select C

### Lasso Regression: 5-Fold Cross-Validation Error vs C

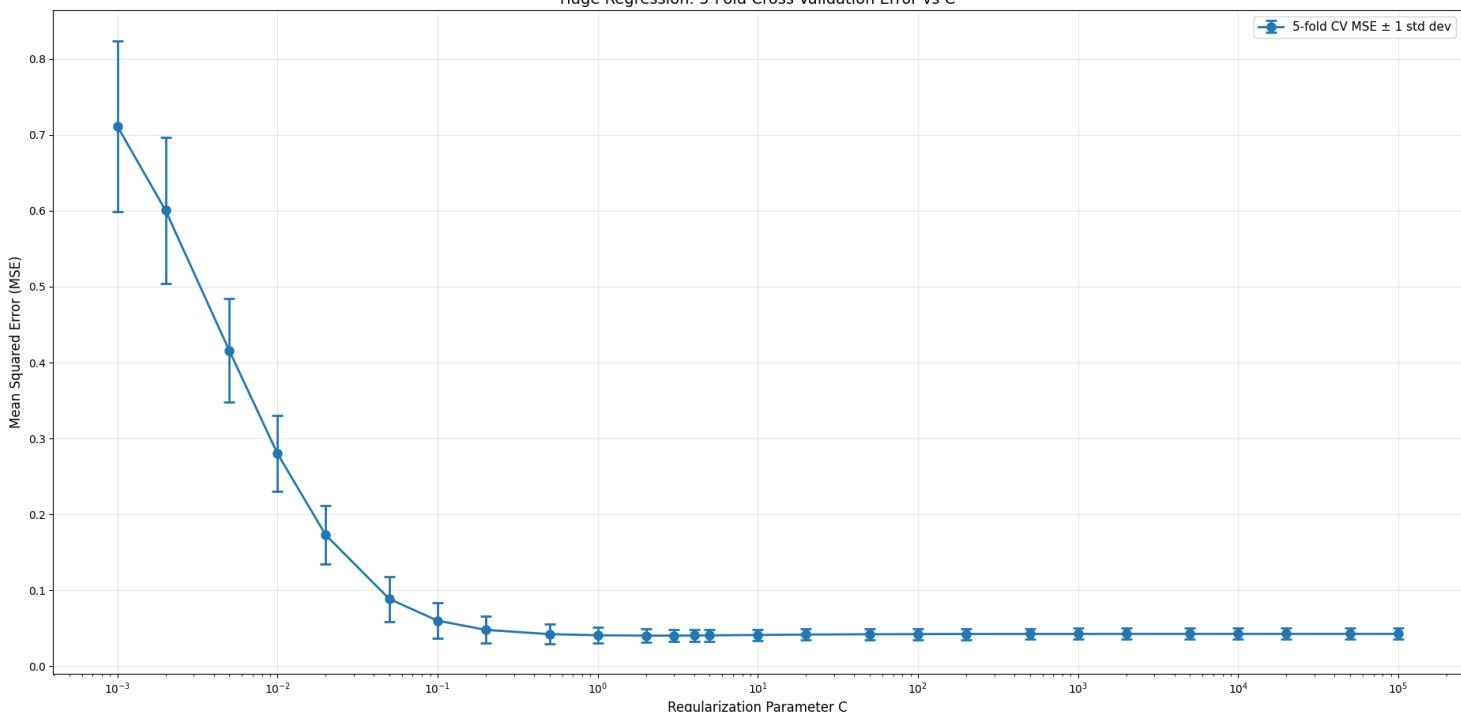


a.

larger range was tested for this, I decided to use a larger, more precise range as the code ran much faster, and the upward trend after the drop was very hard to see without continuing the graph much

C= 0.001	- Mean MSE: 0.8794 +/- 0.1351
C= 0.002	- Mean MSE: 0.8794 +/- 0.1351
C= 0.005	- Mean MSE: 0.8794 +/- 0.1351
C= 0.01	- Mean MSE: 0.8794 +/- 0.1351
C= 0.02	- Mean MSE: 0.8794 +/- 0.1351
C= 0.05	- Mean MSE: 0.8794 +/- 0.1351
C= 0.1	- Mean MSE: 0.8794 +/- 0.1351
C= 0.2	- Mean MSE: 0.8794 +/- 0.1351
C= 0.5	- Mean MSE: 0.8794 +/- 0.1351
C= 1	- Mean MSE: 0.8794 +/- 0.1351
C= 2	- Mean MSE: 0.6415 +/- 0.1252
C= 3	- Mean MSE: 0.427 +/- 0.1217
C= 4	- Mean MSE: 0.2563 +/- 0.077
C= 5	- Mean MSE: 0.1773 +/- 0.0559
C= 10	- Mean MSE: 0.0719 +/- 0.0253
C= 20	- Mean MSE: 0.0456 +/- 0.0152
C= 50	- Mean MSE: 0.0382 +/- 0.0109
C= 100	- Mean MSE: 0.0375 +/- 0.0096
C= 200	- Mean MSE: 0.0373 +/- 0.0088
C= 500	- Mean MSE: 0.0378 +/- 0.0084
C= 1000	- Mean MSE: 0.0388 +/- 0.0083
C= 2000	- Mean MSE: 0.0402 +/- 0.0077
C= 5000	- Mean MSE: 0.0416 +/- 0.0073
C= 10000	- Mean MSE: 0.042 +/- 0.0072
C= 20000	- Mean MSE: 0.0423 +/- 0.0072
C= 50000	- Mean MSE: 0.0425 +/- 0.0073
C= 100000	- Mean MSE: 0.0425 +/- 0.0073

b. I would recommend C=200, it has the lowest MSE and has a decently low standard deviation, and since it is relatively low C value on the graph, it also means that it shouldn't have much risk of overfitting



C.

The graph of Ridge looks very different to the graph of the lasso model, reaching an optimal result at much lower values than the lasso model (C=2 vs C=200),.

```
C= 0.001 - Mean MSE: 0.7115 +/- 0.1126
C= 0.002 - Mean MSE: 0.6006 +/- 0.0964
C= 0.005 - Mean MSE: 0.4159 +/- 0.0681
C= 0.01 - Mean MSE: 0.2803 +/- 0.0496
C= 0.02 - Mean MSE: 0.173 +/- 0.0388
C= 0.05 - Mean MSE: 0.0886 +/- 0.0297
C= 0.1 - Mean MSE: 0.0601 +/- 0.0231
C= 0.2 - Mean MSE: 0.0479 +/- 0.0176
C= 0.5 - Mean MSE: 0.0423 +/- 0.0128
C= 1 - Mean MSE: 0.0408 +/- 0.0103
C= 2 - Mean MSE: 0.0403 +/- 0.0087
C= 3 - Mean MSE: 0.0404 +/- 0.008
C= 4 - Mean MSE: 0.0405 +/- 0.0077
C= 5 - Mean MSE: 0.0407 +/- 0.0075
C= 10 - Mean MSE: 0.0412 +/- 0.0072
C= 20 - Mean MSE: 0.0417 +/- 0.0071
C= 50 - Mean MSE: 0.0422 +/- 0.0072
C= 100 - Mean MSE: 0.0424 +/- 0.0072
C= 200 - Mean MSE: 0.0425 +/- 0.0072
C= 500 - Mean MSE: 0.0425 +/- 0.0073
C= 1000 - Mean MSE: 0.0426 +/- 0.0073
C= 2000 - Mean MSE: 0.0426 +/- 0.0073
C= 5000 - Mean MSE: 0.0426 +/- 0.0073
C= 10000 - Mean MSE: 0.0426 +/- 0.0073
C= 20000 - Mean MSE: 0.0426 +/- 0.0073
C= 50000 - Mean MSE: 0.0426 +/- 0.0073
C= 100000 - Mean MSE: 0.0426 +/- 0.0073
```

The best possible C for Ridge seems to be at 2, which has the lowest MSE

and one of the lowest standard deviation, while also being very early in the possible C values. This performance is lower than Lasso's performance(0.0403 vs 0.0373), this shows that ridge may be less optimal for this dataset at its limits, but its lower C value does suggest that it is better at managing datasets that are dependent on multiple polynomial features, while lasso could be better with datasets without a large amount of polynomial features