Name: Dong Ao

ZID: z5305320

Q1:

(a):.

Thus, and . We have and .

. Since is the prediction made by the model, they did not change. Thus, for , it is still the same as from the formula provided.

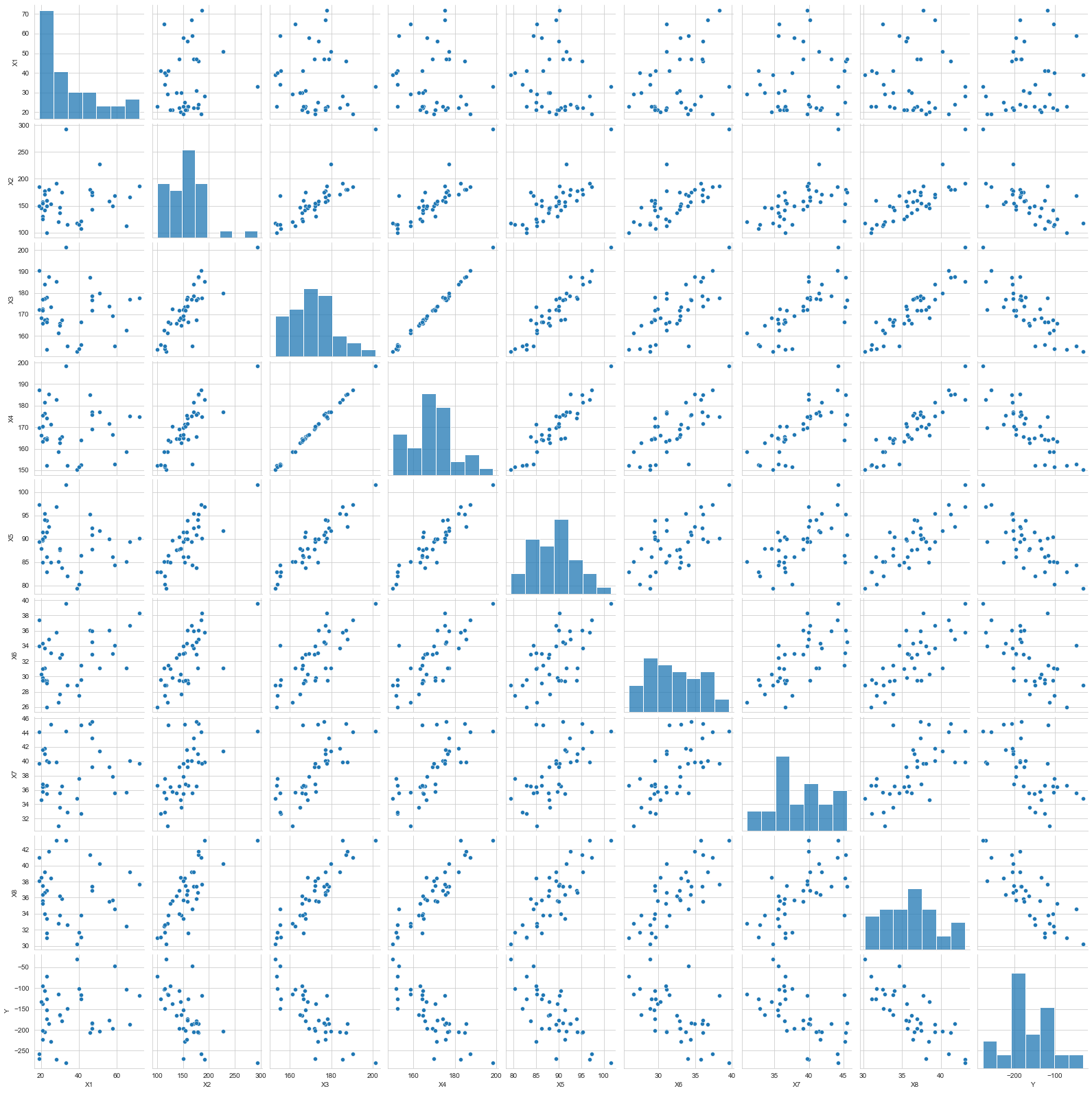
, and .

(b): We know that and for any , can only be 1 since it is treatment group. Thus, . Similarly, because it is placebo group without receiving a dose of drug. We say that and are respectively the number of data points of treatment group and non-treatment group. Then, and .

Thus, and .

Q2

(a):



From the pairs plot, Y has a negative correlation respectively with . Moreover, there are no particular relationships between and ,and is proportional to . In linear model regression, will possibly contribute nothing and feature 3 and 4 may be considered as one feature.

(b):

: X1 38.0

: 38.00000000000001

: 38.00000000000002

: 37.99999999999999

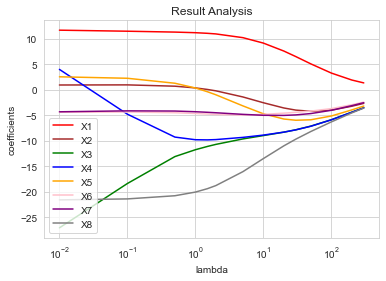
: 38.0

: 38.0

: 38.00000000000001

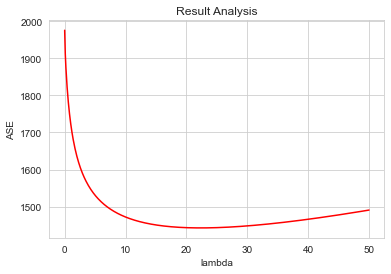
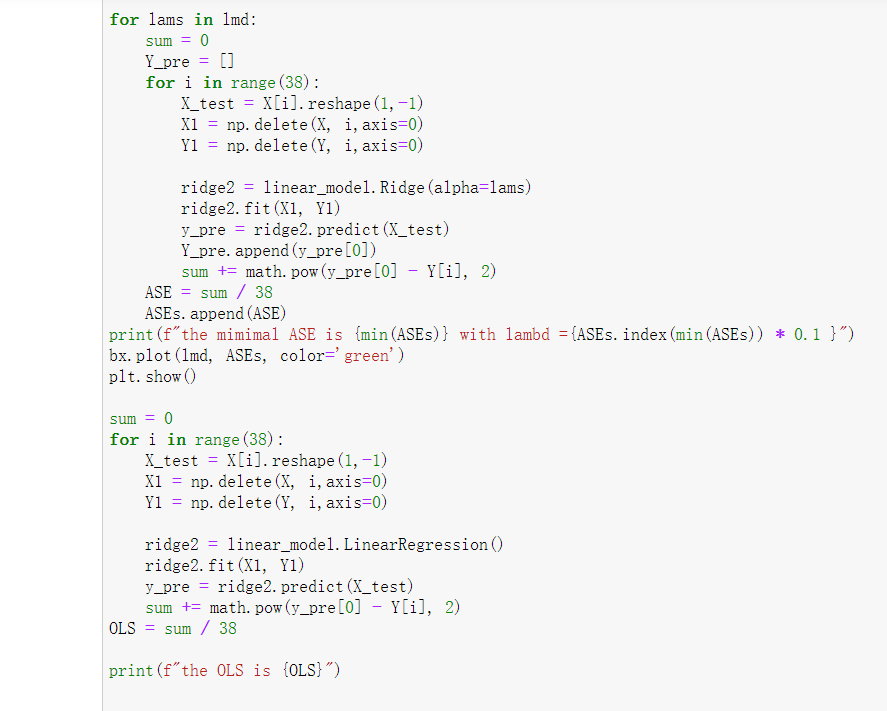
: 37.999999999999986

(c):

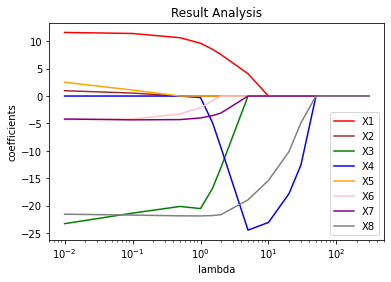
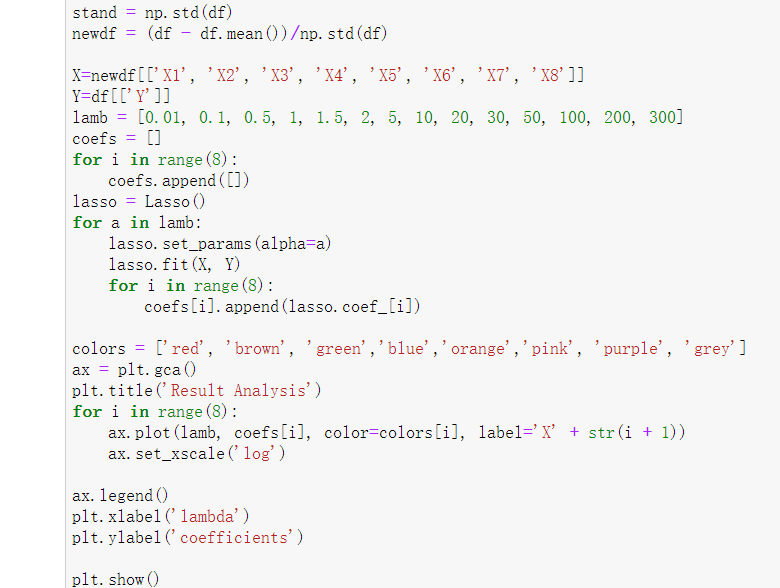


I saw that when lambda get larger, the coefficients all tend to converge to 0 even it has not been reached in the plot. Feature 3 and 4 shows some strong relationship when , when is increasing, decrease in almost the same speed. Feature 3, 4 and 5 changes their slope almost at the same time.

(d):

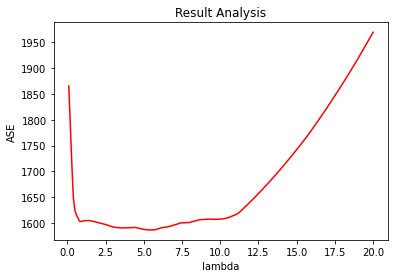


The least MSE from using ridge (1442.6982227952915 with lambda 22.3) is much less than the OLS one (1975.4147393421724).

(e):

When lambda get extremely large, all the coefficients tend to 0. When lambda is around 5, the absolute value of the coefficient of becomes very large, which is increasing. Also, and have relationships from the graph. When around , one of and is chosen in Lasso and the other one changes immediately, which means they are likely to be reduced to one single variable.

(f):



The least error from using lasso (1442. 1586.6715081806428

with lambda 5.4) is much less than the OLS one (1975.4147393421724).

(g):

The MSE from Ridge and Lasso are quite different. The MSE of Lasso almost remain still when lambda is in range of 1 and 10. Moreover, the MSE of Lasso grows extremely fast when lambda is greater than 10 but Ridge’s grows relatively slow.

I prefer to choose Ridge. Using the predicted model we got, the least MSE of Ridge is less than Lasso’s. From the definition of Lasso, it has an ability to select the variates and exclude those from our model by giving very small coefficient. But from the pairs plot we got, we found that only is irrelated to Y, which means unselect some variates in this circumstance is not that efficient. We have 7 out of 8 features in our model which influenced out model much, that is why we choose Ridge here.

Q2

(a):

We suppose that when , is the max. Thus, .

(b):

It reaches the minimum when is minimal. We know and only when , we can reach the minimum. Thus, is a solution of Lasso problem.

(c):

For any , .

We know that from (b), when is sufficiently large, . Moreover, we .

Thus,

End of proof.