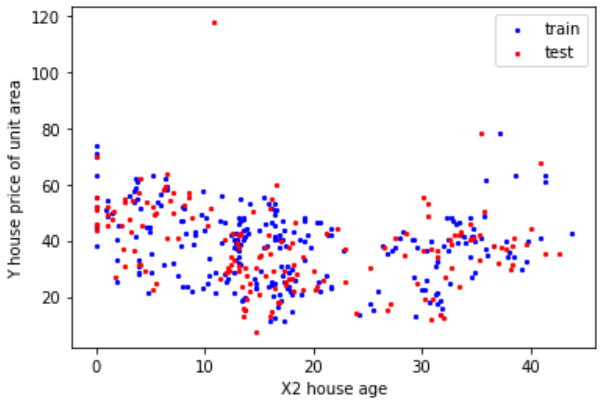
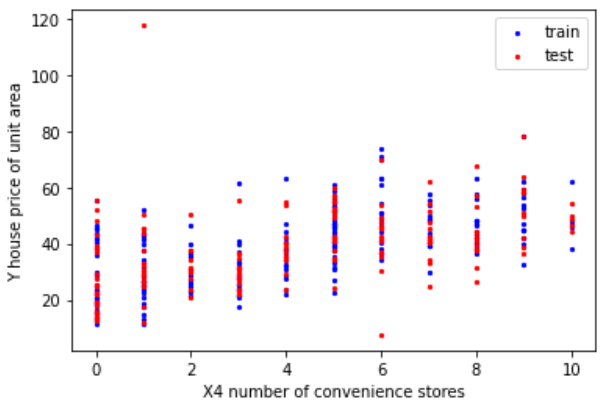
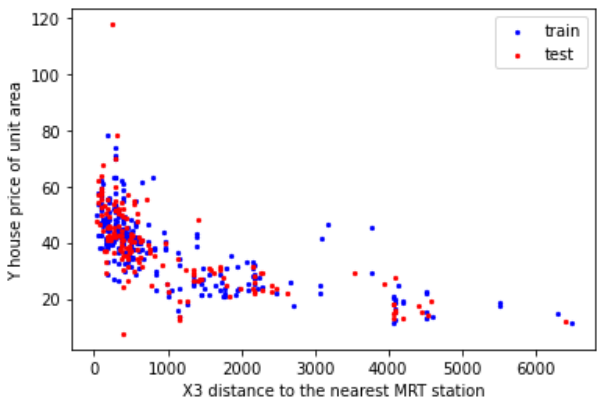
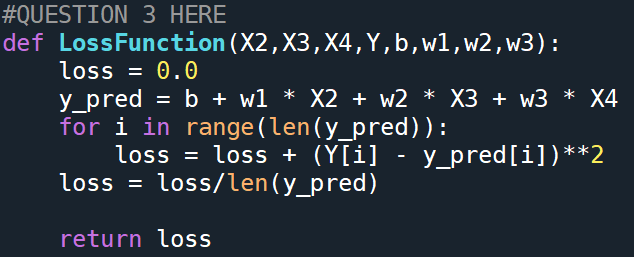
ML HW3 109070032廖品睿

2.

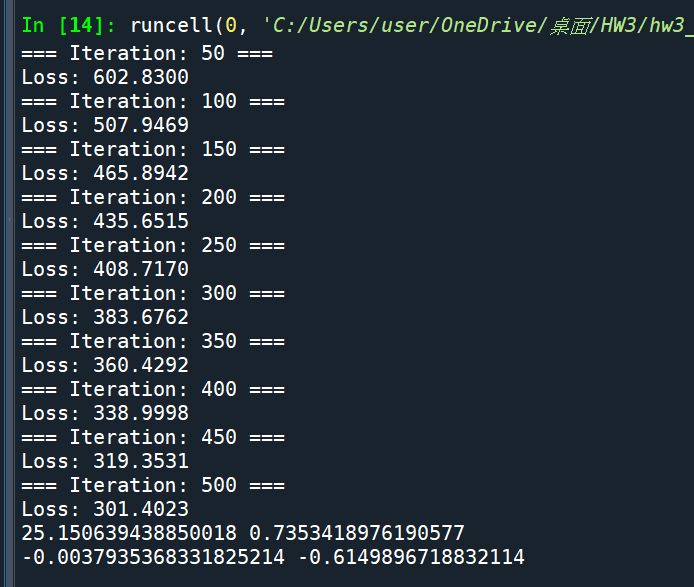
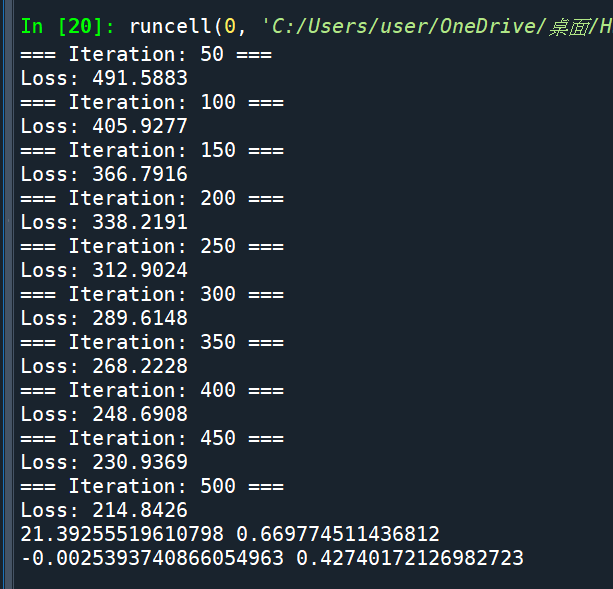
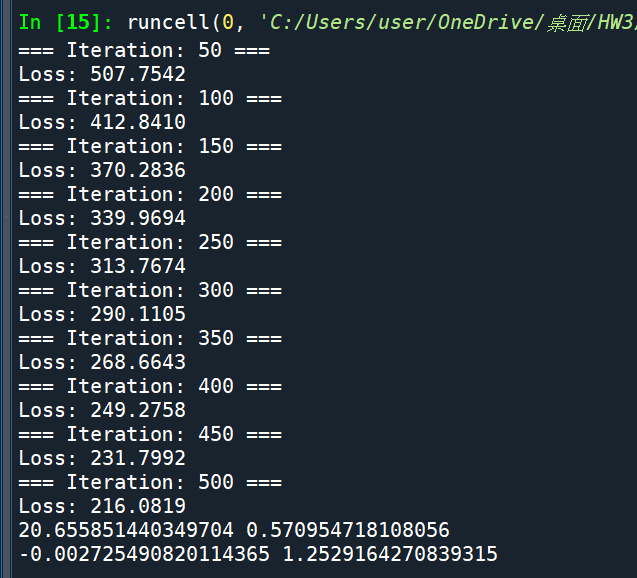
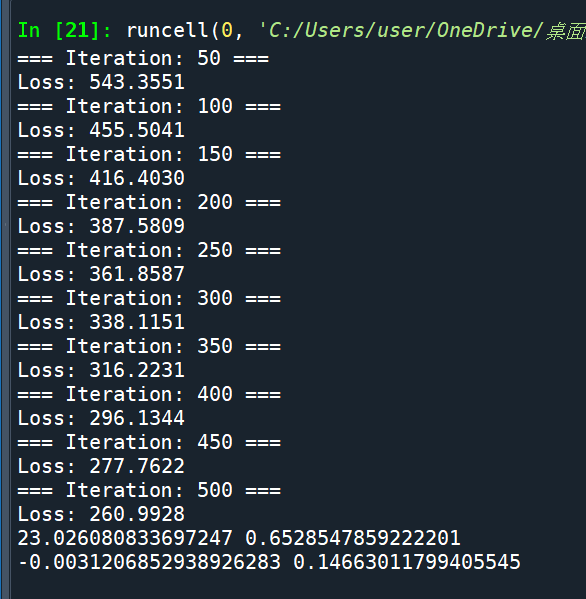
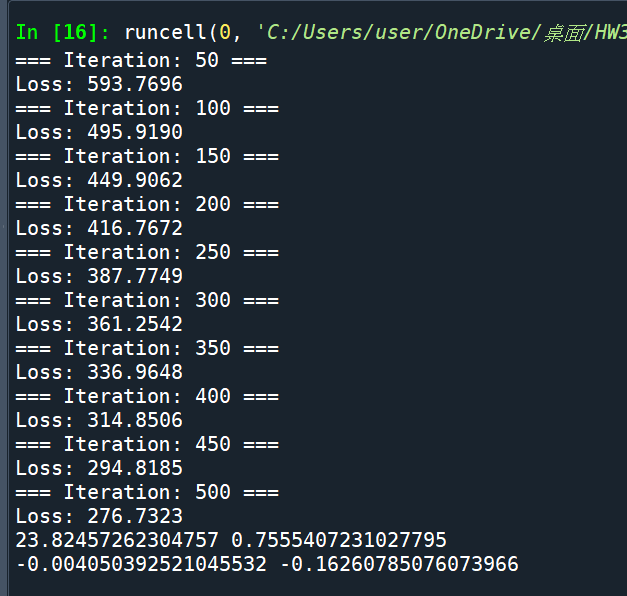
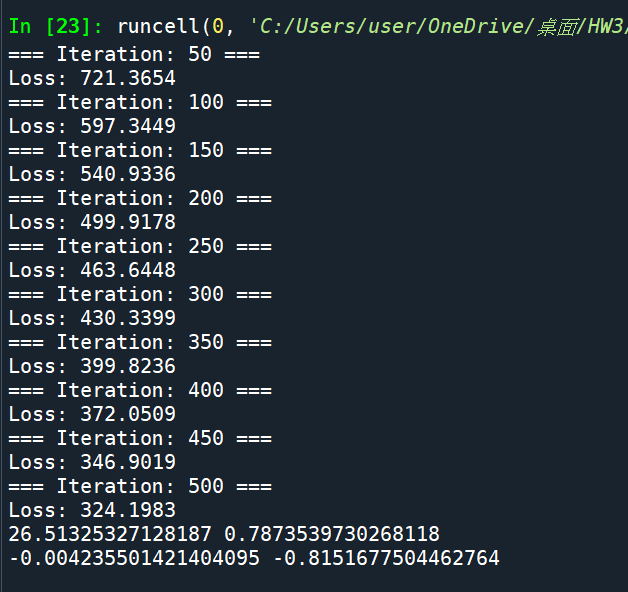
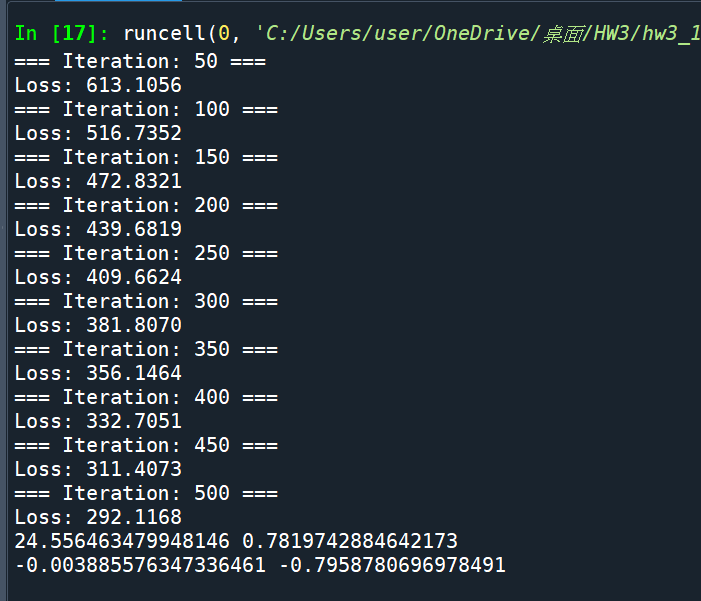
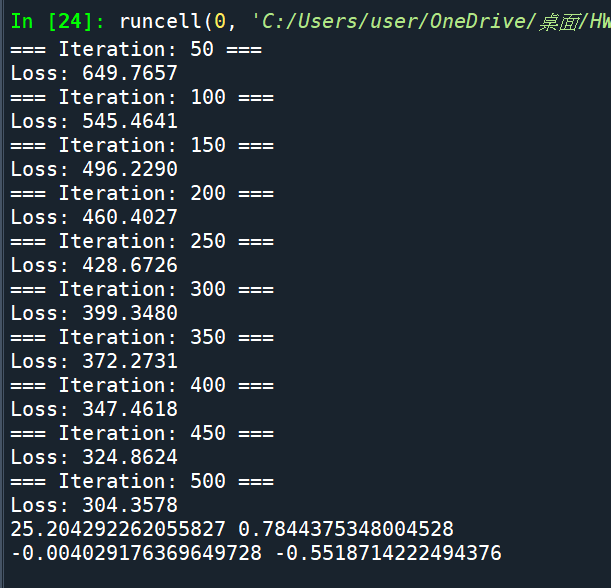
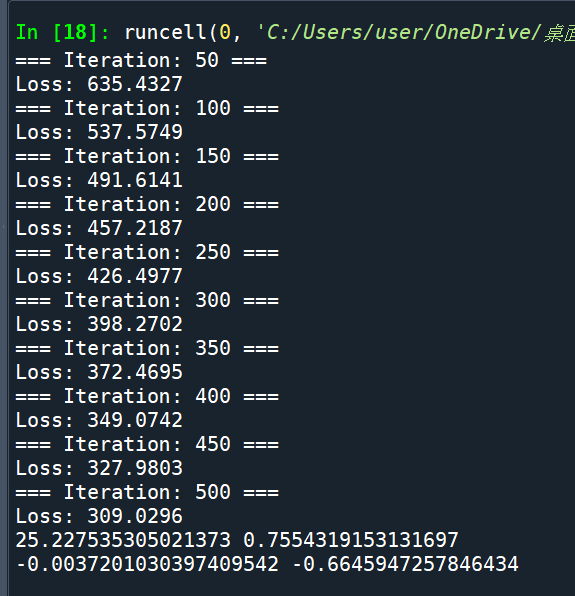
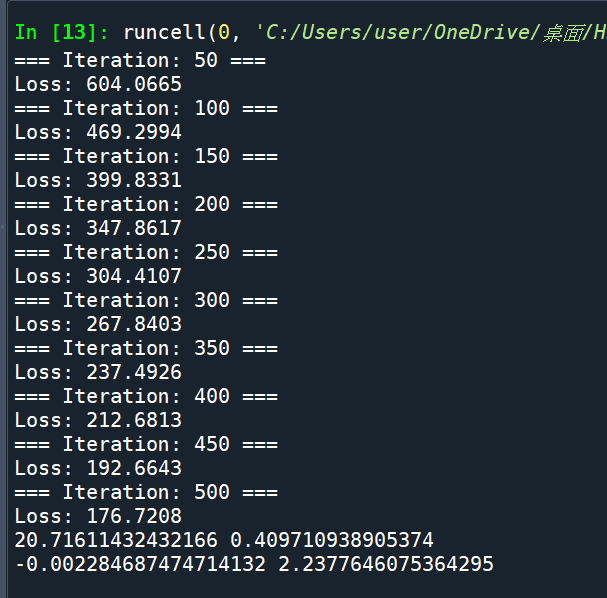




3.



4.

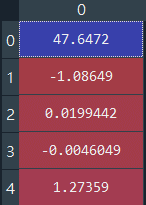


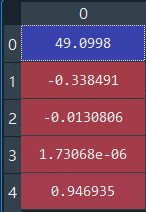
The bottom two lines in every picture indicates the values of B0，B1，B2，B3

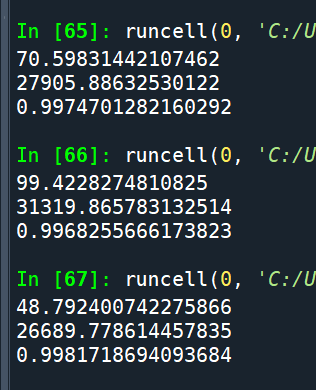
For the ten iterations

5.

 B0:42.2633, B1:-0.3117, B2:-0.004, B3:1.444

 B4:47.6472, B5:1.086, B6:-0.019, B7:-0.004, B8:1.2735

 B9:49.0998, B10:-0.3384, B11:-0.013, B12:0.00000173, B13:0.9469



65 -> the first model

66 -> the second model

67 -> the third model

The first number represents the numerator of Erse, the second number represents

The denominator of Erse, and the third value is R-square

REPORT:

1.

The two set of numbers are different because they use different methods and

They have different losses .

2.

I think the least square method is better because it does not have to iterate through for a lot of times , it puts data in a matrix to produce parameters , unlike gradient method , if the times of iteration is not big enough , like about 500 , it only converges to the local min instead of the global min . Therefore we get an inaccurate answer compared to least square method . Convergence accuracy of forecast model

is assumed to be 0.001 , which is the learning rate , we have to shift the data to near

10 to get a better result , otherwise the loss would be larger after every iteration.

The least square method gets an average loss of 70 , which is much more stable than gradient method , and from the plots below , we can infer that the housing prices is more relevant to data x4 compared to x2 and x3 .

