Linear Regression

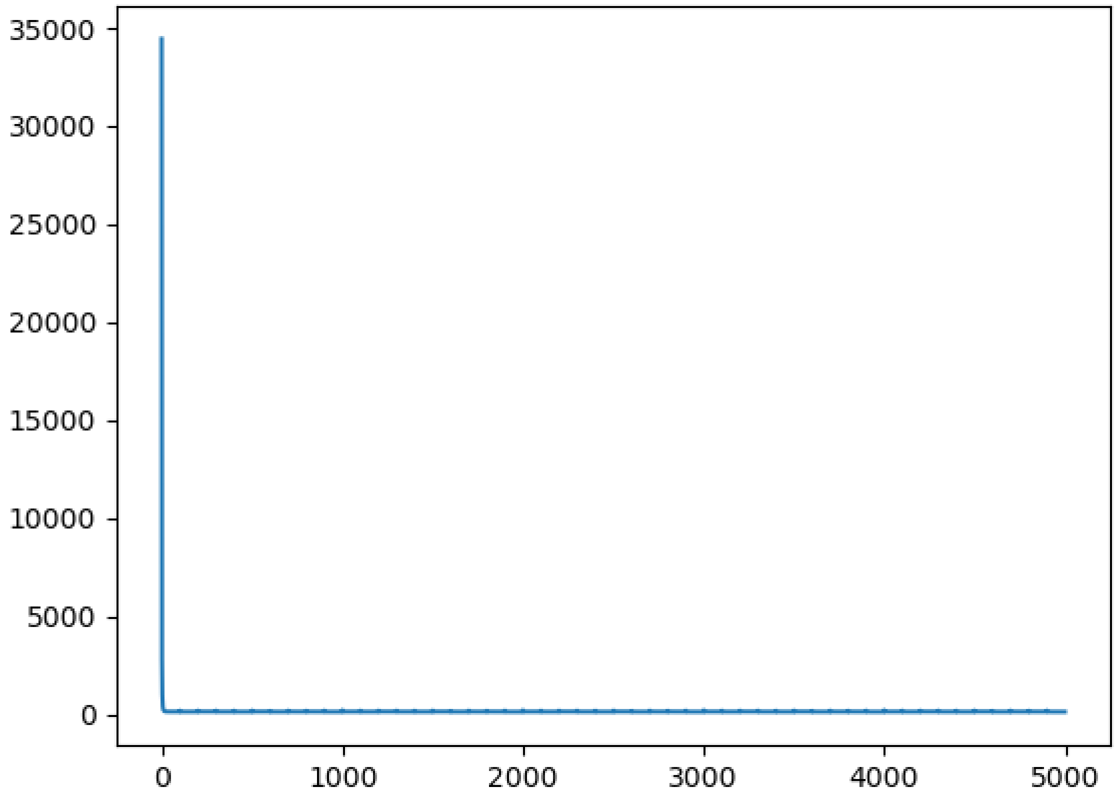
刘军腾 520030910185

ii)

code is in file `hw1.py`

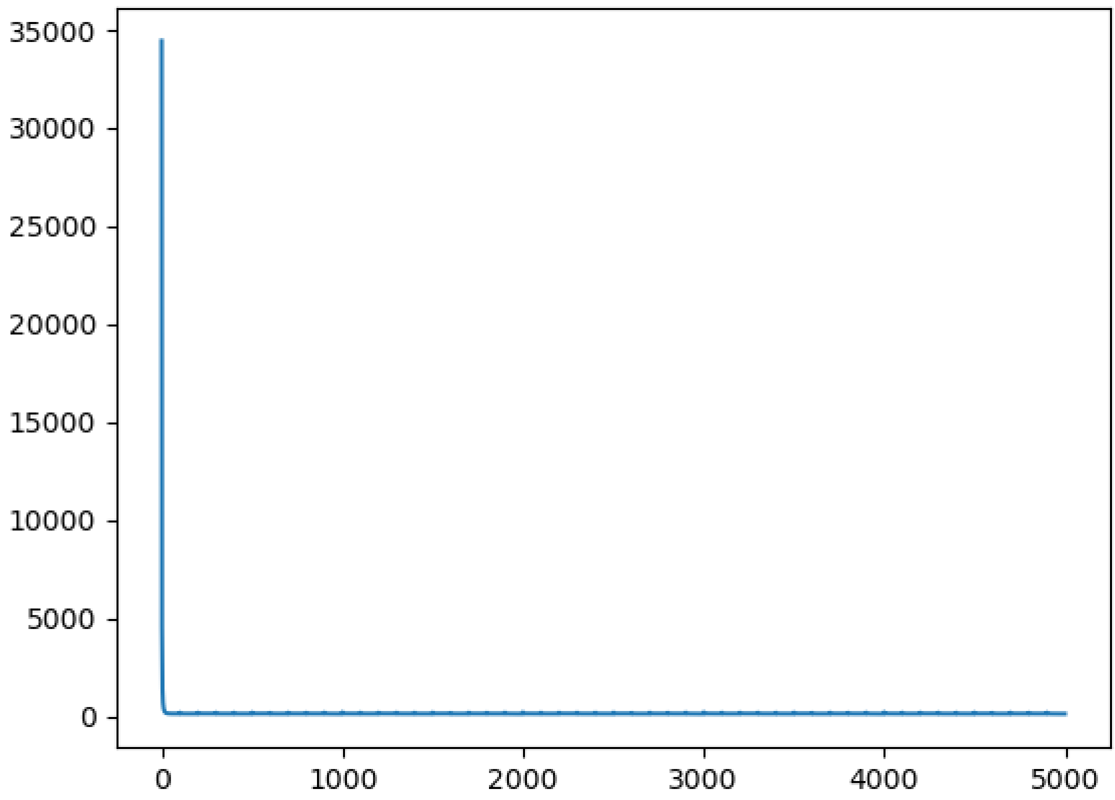
I choose $\gamma = 0.01$ for the experiment.

We first show the convergence result of choosing $n = 100, m = 200$. We set hyper-parameter $num\_epochs = 5$ and we set iteration counts 100 for every gradient descent. The result of training loss during process is draw below.



As the figure show, the objective value descent quickly in the first training.

We set step_size or learning rate constant =0.001,the result is shown below.



We can see that it also has a good performance.

iii)

code is in the file `hw2.py`

We let $abs(W' - W).sum()/abs(W).sum()$ to represent the differences.

m=120 differences is 0.16011091528128138

m=200 differences is 0.07153494739234209

m=300 differences is 0.041042675735039204

m=500 differences is 0.03207245944853902

We can easily find that the more data we has, the better performance we get.