Homework4

Xunzhi Li

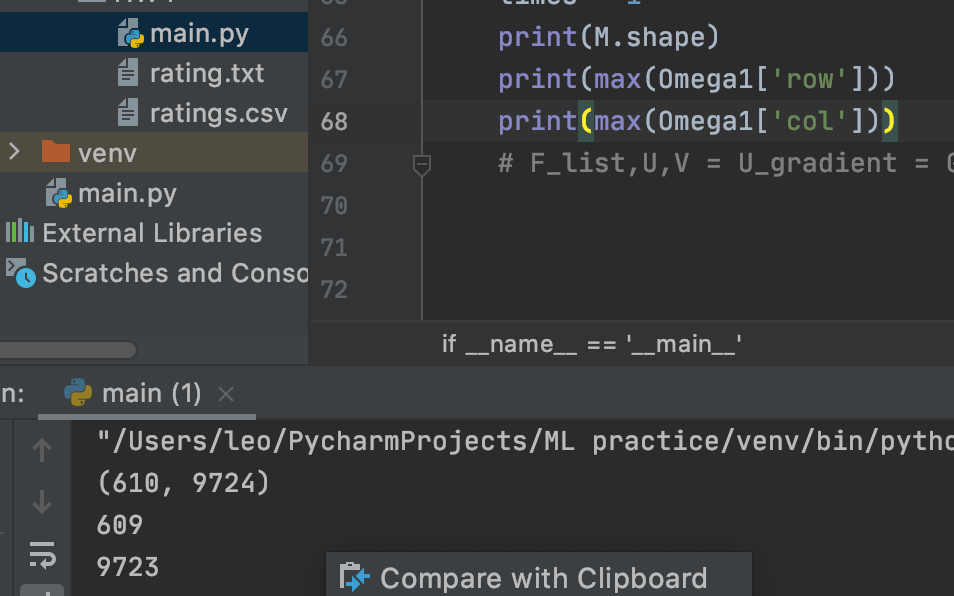
**1 Data**

**Text

Description automatically generated**

we split 90% raw dataset as training set.

It’s a 610 x 9724 matrix.



**2 Learning**

1) gradient :

Text, letter

Description automatically generated

2) we choose learning rate as 0.1

Text

Description automatically generated

detains are in my .py file.

3)

Chart

Description automatically generated

My conclusion is if λ =1 ,Gradient Descent works good. Normally, we only need 5 times iterations and F(U,V) can reach its lowest point.

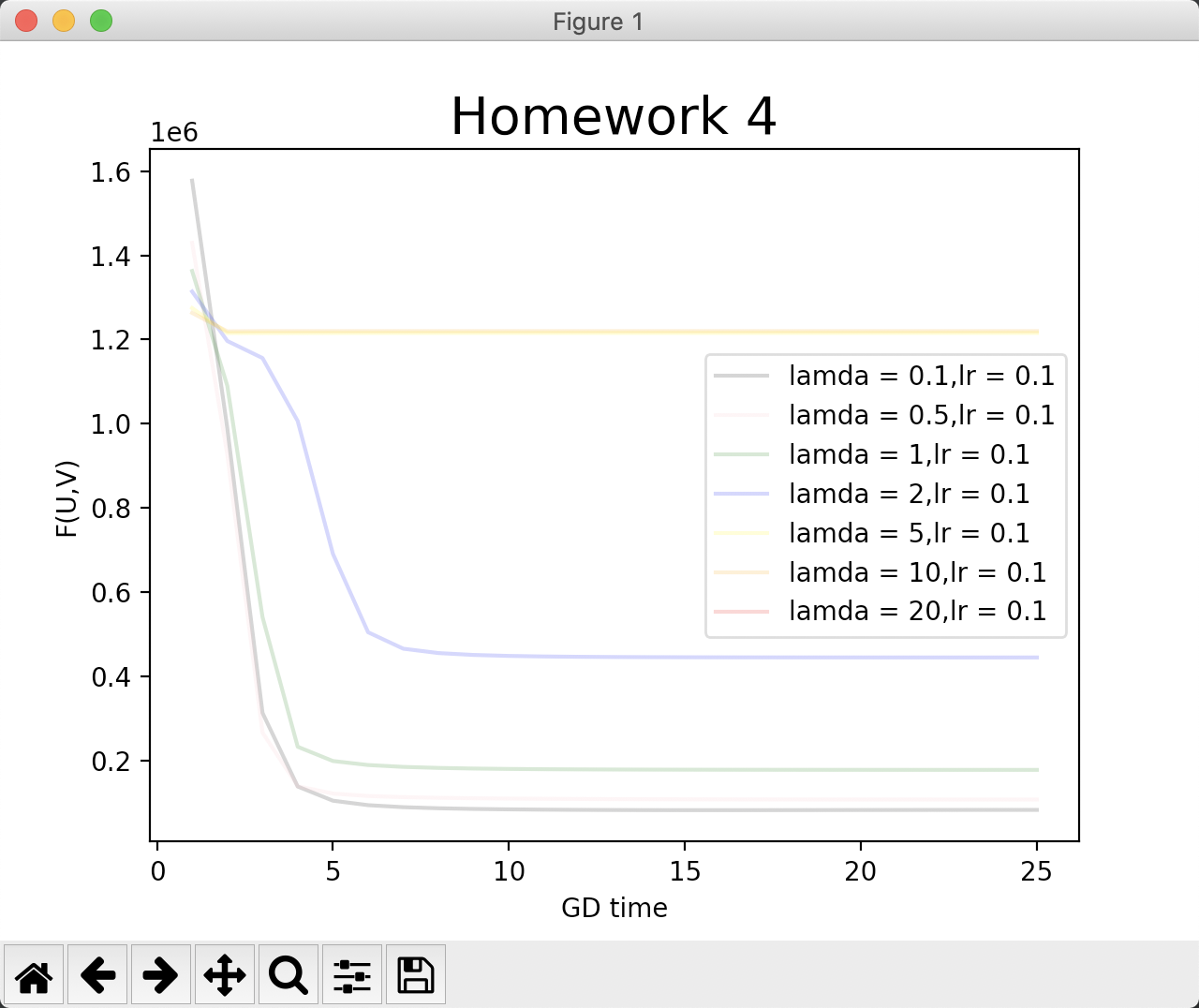
**3.Evaluation**

1. RMSE for λ =1 : 152.85007182892136

Text

Description automatically generated

2.



Graphical user interface

Description automatically generated Graphical user interface

Description automatically generated

For GD and λ , Conclusion is :

A picture containing logo

Description automatically generated

These two Norm(F) terms are to prevent overfitting U and V (user\_matrix and movie\_matrix). So, if λ is too small(0.001,0.1,0.5,1) GD would perform unbelievable perfect.

However, if λ= 2,it works not good. If =5 or even bigger , it performs terrible .

As for RMSE and λ :

RMSE for λ as 0.1, 0.5, 1, 2, 5 , 10, 20 =

[135.3492328386075, 128.93005182519866, 151.2574235000474, 223.32630349421808, 366.63873930421727, 366.63878681885257, nan] Graphical user interface, text

Description automatically generated

However, I have no idea for even bigger λ, my program throw out “RuntimeWarning” warning. So I didn’t process the program based on λ = 50, 100, 500, 1000 and even the smallest 10−6. I think maybe my computer is too old .

Chart

Description automatically generated

My finding is ,RMSE’s value decreases in the beginning (λ <=0.5) then increases while λ arises (>=0.5 <=5), but it remains stable after λ=5. So when λ = 0.5, I get the best result.