

About code part:

The version of python is python3.6

There are linear.py, gradient.py and a folder named data

How to run:

You should first enter the current folder in cmd.

That is, you should first go to the directory of 'code part' in cmd.

Make sure .py files and 'data' folder are both in 'code part' folder.

Then you can use following command:

>>python linear.py

or

>>python gradient.py

Full screenshot:

```
C:\2019Spring\Data Mining\Project1\code part>python linear.py
The prediction values of theta are: [[ 0.40679793 -0.17208526 -0.0295757  0.26246741  0.58187472]]
Estimate function is: y = 0.40679793479282894 + ( -0.17208525807795388 ) * x1 + ( -0.029575698828359823 ) * x2 + 0.2624674094814906 * x3 + 0.5818747209357259 * x4
The loss of the estimation function to the training set is  0.048169565780434284
The loss of the estimation function to the testing set is  0.0450385513930922
Accuracy rate is: 0.9666666666666667

C:\2019Spring\Data Mining\Project1\code part>python gradient.py
The prediction values of theta are: [0.9999597964945576, -0.033807975361895, -0.059610323598668226, 0.32140174738367494, 0.4677042339633092]
Estimate function is: y = 0.9999597964945576 + ( -0.033807975361895 ) * x1 + ( -0.059610323598668226 ) * x2 + 0.32140174738367494 * x3 + 0.4677042339633092 * x4
The loss of the estimation function to the training set is: 0.024938655178665472
The loss of the estimation function to the testing set is 0.021857580973334987
Accuracy rate is: 0.9666666666666667
```

Clear screenshot:

```
D:\2019Spring\Data Mining\Project1\code part>python linear.py
The prediction values of theta are: [[ 0.40679793 -0.17208526 -0.0295757  0.26246741  0.58187472]]
Estimate function is: y = 0.40679793479282894 + ( -0.17208525807795388 ) * x1 + ( -0.029575698828359823 ) * x2 + 0.2624674094814906 * x3 + 0.5818747209357259 * x4
The loss of the estimation function to the training set is  0.048169565780434284
The loss of the estimation function to the testing set is  0.0450385513930922
Accuracy rate is: 0.9666666666666667

D:\2019Spring\Data Mining\Project1\code part>python gradient.py
The prediction values of theta are: [0.9999597964945576, -0.033807975361895, -0.059610323598668226, 0.32140174738367494, 0.4677042339633092]
Estimate function is: y = 0.9999597964945576 + ( -0.033807975361895 ) * x1 + ( -0.059610323598668226 ) * x2 + 0.32140174738367494 * x3 + 0.4677042339633092 * x4
The loss of the estimation function to the training set is: 0.024938655178665472
The loss of the estimation function to the testing set is 0.021857580973334987
Accuracy rate is: 0.9666666666666667
```

For 'data' folder:

there are six .txt files:

training: including 3 groups 90 data items as training data(tags are 0 1 2)

testing: including 3 groups 60 data items as testing data(tags are 0 1 2)

training1: including 75 data items(just cut from middle of origin dataset) as training data(tags are 0 1 2)  
testing1: including 75 data items(just cut from middle of origin dataset) as testing data(tags are 0 1 2)

training100: including 3 groups 90 data items as training data(tags are 100 101 102)  
testing100: including 3 groups 60 data items as testing data(tags are 100 101 102)

For .py files:

these two files just use training.txt and testing.txt as default dataset  
if you want to use other dataset, just change it in code