

# Homework 3

---

17343147 张涵健

## Exercise 1

(a)

本题多变量线性回归模型

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

初始值

```
alpha = 0.00015 # learn rate
theta0 = 0.0
theta1 = 0.0
theta2 = 0.0
```

梯度下降迭代

repeat until convergence: {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (\mathbf{h}_{\theta}(\mathbf{x}^{(i)}) - \mathbf{y}^{(i)}) \cdot \mathbf{x}_0^{(i)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (\mathbf{h}_{\theta}(\mathbf{x}^{(i)}) - \mathbf{y}^{(i)}) \cdot \mathbf{x}_1^{(i)}$$

$$\theta_2 := \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^m (\mathbf{h}_{\theta}(\mathbf{x}^{(i)}) - \mathbf{y}^{(i)}) \cdot \mathbf{x}_2^{(i)}$$

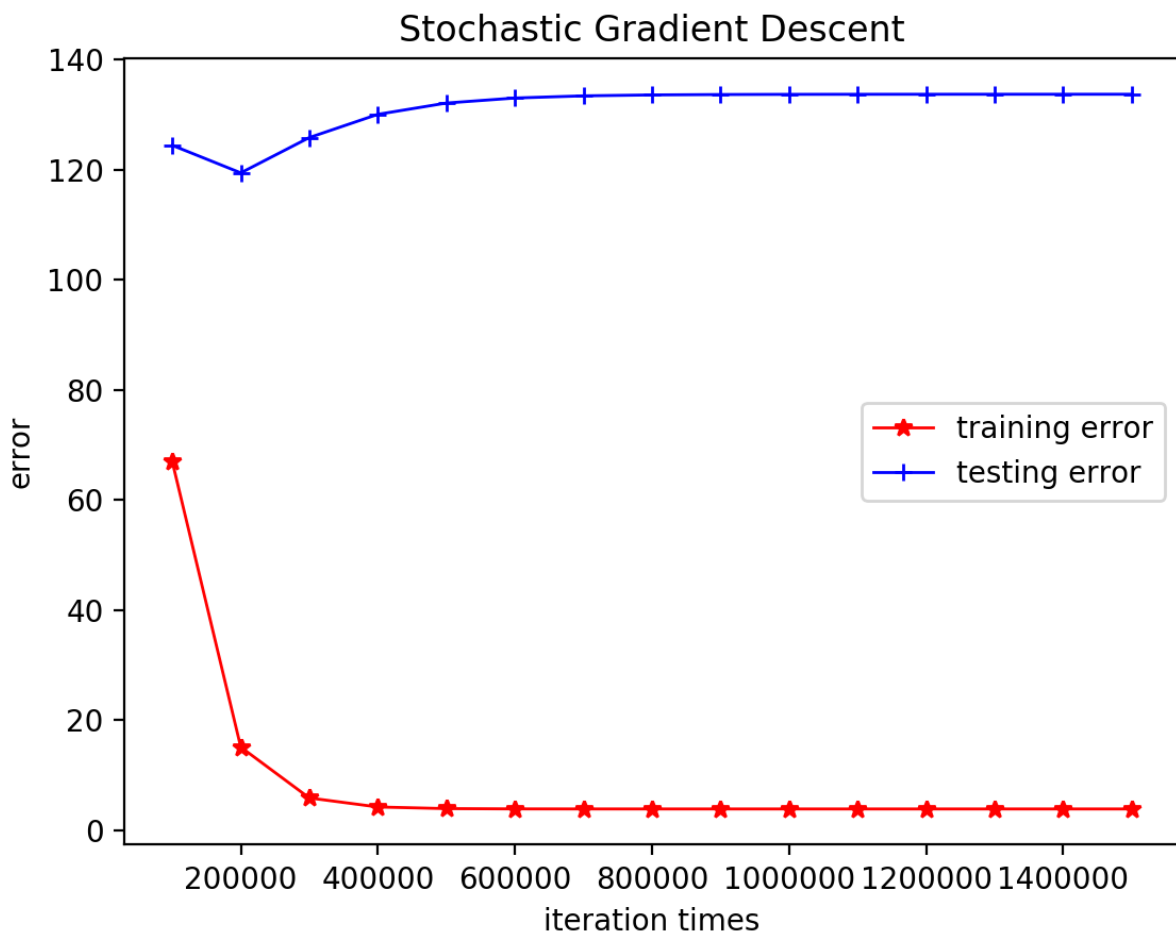
...

}

误差计算

$$E_{\theta} = \frac{1}{2m} \sum (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

| number of iterations | theta0            | theta1             | theta2             | training_error     | testing_error      |
|----------------------|-------------------|--------------------|--------------------|--------------------|--------------------|
| 100000               | 46.32960907409016 | 7.089390249476694  | -72.75988620414542 | 67.00770433342761  | 124.46329858931466 |
| 200000               | 65.48714313030769 | 6.90006422168188   | -72.54075315993076 | 15.059483400506053 | 119.45523365857287 |
| 300000               | 73.56819957267346 | 6.820202465355145  | -72.44831816837659 | 5.816168547023758  | 125.87200620863314 |
| 400000               | 76.9769615267465  | 6.786515073305167  | -72.4093271175553  | 4.1714756596405875 | 130.0963802843273  |
| 500000               | 78.41485001848125 | 6.772305012879157  | -72.39287986074687 | 3.878830138343642  | 132.14835040996874 |
| 600000               | 79.02138205049656 | 6.766310906330022  | -72.38594205705064 | 3.8267587748738925 | 133.06196409214206 |
| 700000               | 79.27723019725956 | 6.763782464349798  | -72.38301554348719 | 3.8174935488708424 | 133.45589532783626 |
| 800000               | 79.38515240210221 | 6.762715913597816  | -72.38178107765161 | 3.81584495726178   | 133.6235851646329  |
| 900000               | 79.43067628978632 | 6.76226601974406   | -72.38126035363443 | 3.8155516180283118 | 133.694590993324   |
| 1000000              | 79.44987923651175 | 6.7620762449215395 | -72.38104070113839 | 3.815499423230369  | 133.72459092904776 |
| 1100000              | 79.45797944898648 | 6.761996193854025  | -72.380948047031   | 3.815490136041301  | 133.73725411151614 |
| 1200000              | 79.46139629136175 | 6.7619624266065    | -72.38090896355249 | 3.815488483541725  | 133.7425972368003  |
| 1300000              | 79.46283758834896 | 6.761948182861316  | -72.38089247730768 | 3.815488189507201  | 133.7448513523211  |
| 1400000              | 79.463445558156   | 6.76194217454583   | -72.38088552305807 | 3.815488137188594  | 133.74580223459776 |
| 1500000              | 79.46370201278938 | 6.761939640110196  | -72.38088258960731 | 3.815488127879413  | 133.7462033456204  |



#### 分析与结论：

使用了4个参数来训练该线性回归模型，其中学习率为 0.00015，其余为0.0。根据迭代计算结果，参数收敛在[ 79.46 6.76 -72.38 ]左右，误差也逐渐稳定降低。

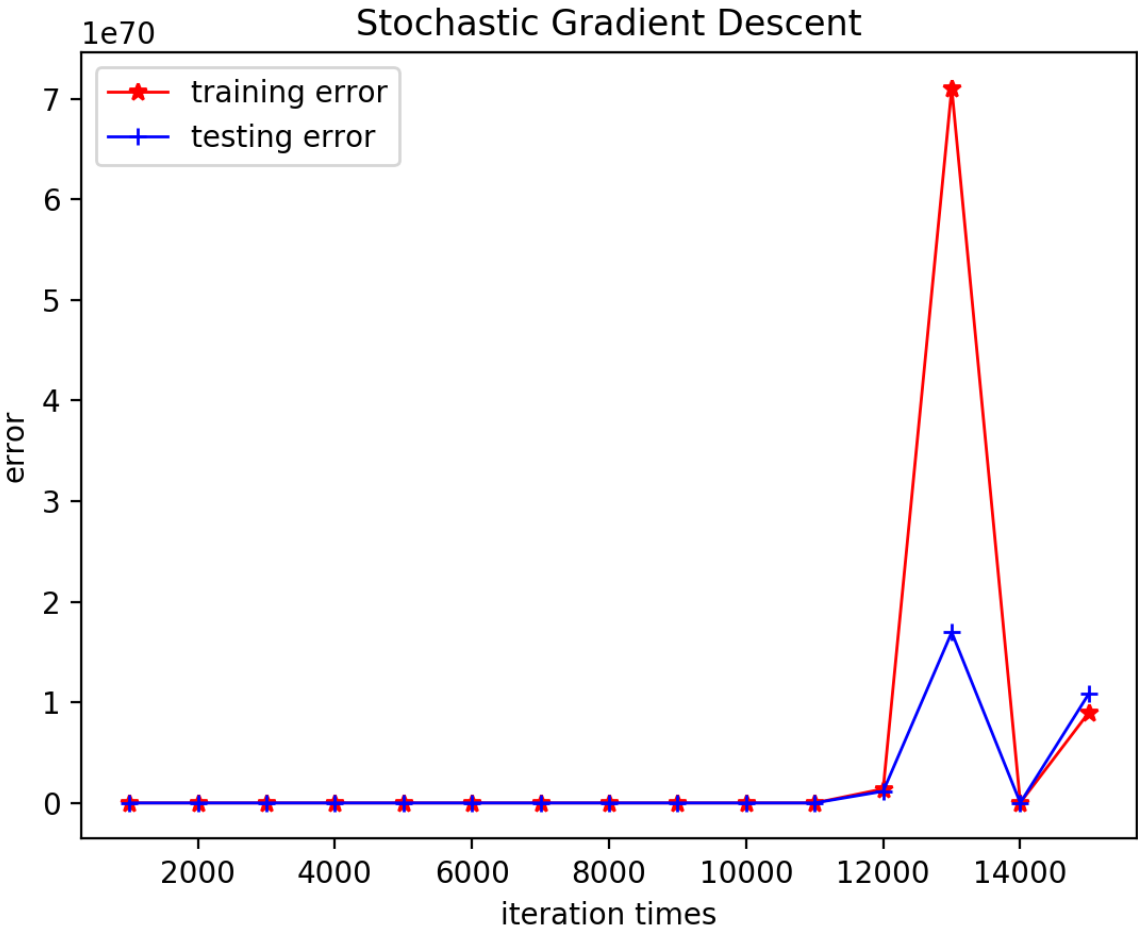
(b)

当保持相同的迭代次数时，将学习率设置为 0.0002，程序数据过大而越界，无法得到理想的训练结果。

把迭代次数减少，使得数据能正常

# zhanghanjian @ CoderAtMBP in ~/github/DataMining/hw3 on git:master x [20:35:23]  
\$ python3 main.py

| number of iterations | theta0                  | theta1                  | theta2                  | training_error         | testing_error          |
|----------------------|-------------------------|-------------------------|-------------------------|------------------------|------------------------|
| 1000                 | -1747.092728282745      | 131.7958507256941       | -13599.995235543165     | 4674704255.004636      | 1538813586.113474      |
| 2000                 | 51436.014528618136      | -1668787.4037725977     | -2716174.8003212786     | 3.5437945554783068e+16 | 2.8357602183455696e+16 |
| 3000                 | 202387865.0383042       | 14756348.215457141      | 1969394858.6123743      | 1.6894932081755318e+20 | 7.17873281947429e+19   |
| 4000                 | 2452706584.5497084      | -489456826.4368838      | 5465044143.748105       | 5.8599010255726576e+20 | 8.06910246464213e+20   |
| 5000                 | -1386934847759261.0     | 1.2570051056519264e+16  | 2499552024712733.0      | 1.7480581799851765e+36 | 1.463898196040892e+36  |
| 6000                 | 3.96625180332535e+16    | 8527347279727604.0      | -5.58894082317717e+16   | 4.230671683943286e+35  | 4.6709781673527046e+35 |
| 7000                 | 1.6433349672571274e+17  | 6.999276275629832e+16   | -5.7522380393950176e+17 | 2.184088979250372e+37  | 2.6162941775394873e+37 |
| 8000                 | -1.7626500014429325e+17 | -3.916079087362513e+18  | 1.7439749218446905e+19  | 1.0248526582280742e+41 | 1.0289891882556515e+41 |
| 9000                 | -1.9575725251527967e+21 | -2.0053860164092135e+21 | -4.53118495841687e+22   | 2.1345939235573072e+47 | 1.194375794616432e+47  |
| 10000                | 3.2580836445075433e+28  | 4.202168148587722e+28   | -1.6902161595334898e+29 | 1.2624757062624578e+61 | 1.2442213123943994e+61 |
| 11000                | -2.138901906816152e+29  | -6.748896073187024e+28  | 1.2301807044033262e+30  | 1.0087025336838211e+61 | 1.1437410571706484e+61 |
| 12000                | -2.040177208487852e+31  | -3.503790950709059e+32  | -1.6577575537485006e+32 | 1.399919628331892e+69  | 1.1621244410547459e+69 |
| 13000                | -1.9576155727364825e+33 | 2.001728671140339e+33   | -7.558395696270693e+34  | 7.104707957607205e+70  | 1.696487841554995e+70  |
| 14000                | -9.039300711241737e+32  | 9.147054651323816e+31   | -1.9695873227141808e+33 | 2.6593315856179417e+67 | 1.22643225625218e+67   |
| 15000                | 3.895424271491892e+33   | 1.4279774035998565e+33  | -1.1978678059974384e+34 | 8.99330895104758e+69   | 1.0855819371629072e+70 |



结果也是不能正确收敛。

结论：学习影响多变量线性回归模型的收敛，学习率越大，在梯度下降中影响越大，应把学习率适当调小，加快模型收敛，或者将数据进行归一化，统一数据的规模。

(c)

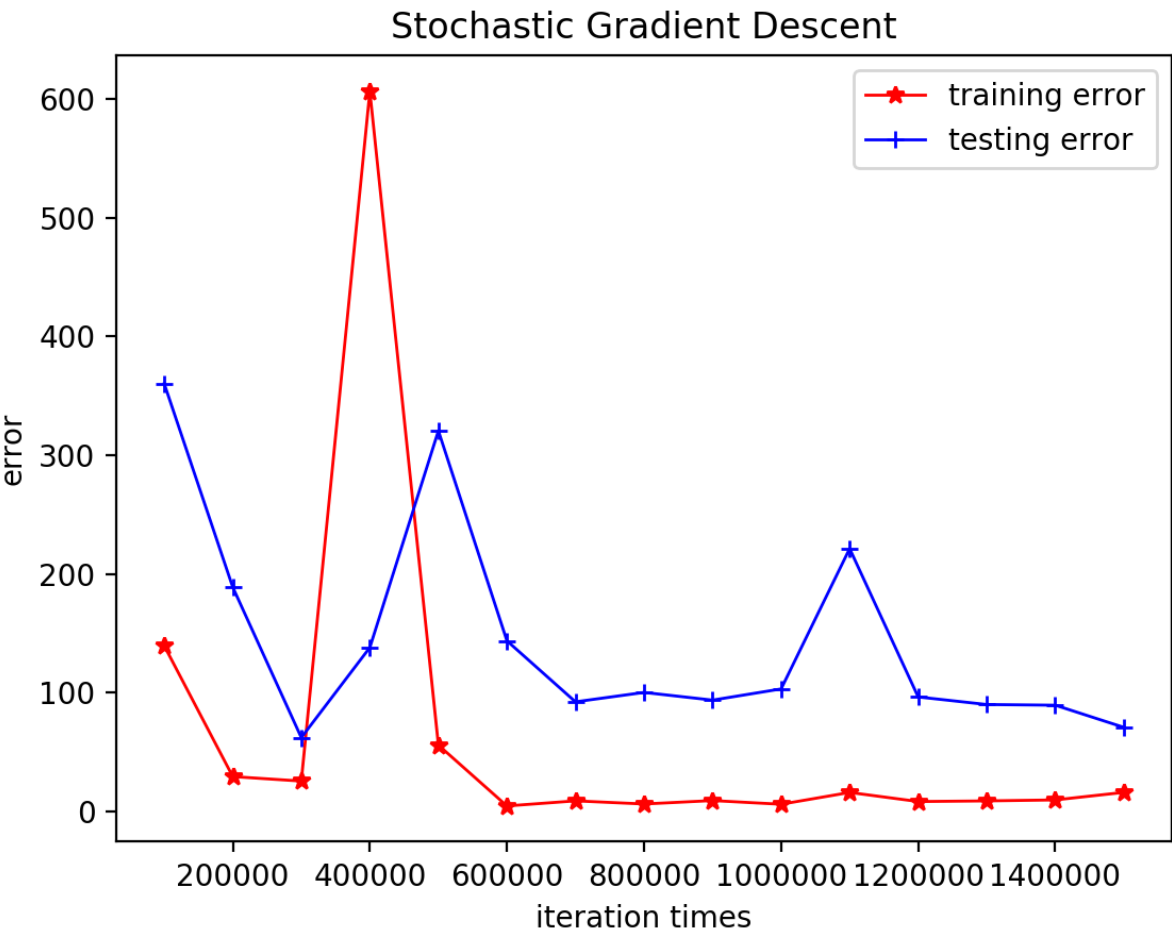
随机梯度下降迭代

$$\theta_j := \theta_j - \alpha (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

第一次实验

```
# zhanghanjian @ CoderAtMBP in ~/github/DataMining/hw3 on git:master x [19:37:05]
$ python3 main.py
```

| number of iterations | theta0            | theta1            | theta2             | training_error     | testing_error      |
|----------------------|-------------------|-------------------|--------------------|--------------------|--------------------|
| 100000               | 47.53275356332223 | 7.184177513826893 | -73.21044226976132 | 139.40754237320385 | 360.2513702418428  |
| 200000               | 66.23739455757273 | 6.913466788937517 | -72.21467472624671 | 29.183690447155097 | 188.7550287531466  |
| 300000               | 73.79847050315364 | 6.7879178859813   | -72.68875209310113 | 25.39185356849647  | 62.02875908366562  |
| 400000               | 77.27579711700503 | 6.545142753250406 | -72.34699769036621 | 606.6351112895893  | 137.95789312612104 |
| 500000               | 78.580834959409   | 6.836525857454467 | -72.32728676044837 | 55.71261196940427  | 320.54609313342223 |
| 600000               | 78.72470849399069 | 6.771274549700968 | -72.2909023818609  | 4.435231682476635  | 143.56039452715004 |
| 700000               | 79.03916467750359 | 6.745896567270808 | -72.40514159703906 | 8.718213802662914  | 92.16218140025892  |
| 800000               | 79.25985383719525 | 6.744056176368504 | -72.27017627482333 | 6.096219021400686  | 100.09260236870466 |
| 900000               | 79.38953161583946 | 6.744421557640727 | -72.45046189727157 | 8.962778342268143  | 93.61141155158026  |
| 1000000              | 79.26327229484556 | 6.747004380493538 | -72.31843421644423 | 5.873675604905121  | 103.03240553708966 |
| 1100000              | 79.47426836418322 | 6.800168456826186 | -72.48420457078659 | 15.818539309947871 | 221.27269425153779 |
| 1200000              | 79.53519099056251 | 6.74383752990372  | -72.43064229273239 | 8.150004373548603  | 96.28158332507724  |
| 1300000              | 79.50625052539027 | 6.737356975646466 | -72.3198850014816  | 8.688270727328131  | 89.89914454522278  |
| 1400000              | 79.73852207582391 | 6.736293688165116 | -72.37677130074702 | 9.509649122648861  | 89.33030518431887  |
| 1500000              | 79.61349823071782 | 6.725191458464088 | -72.33981539121214 | 15.87204620333893  | 70.68113095202317  |

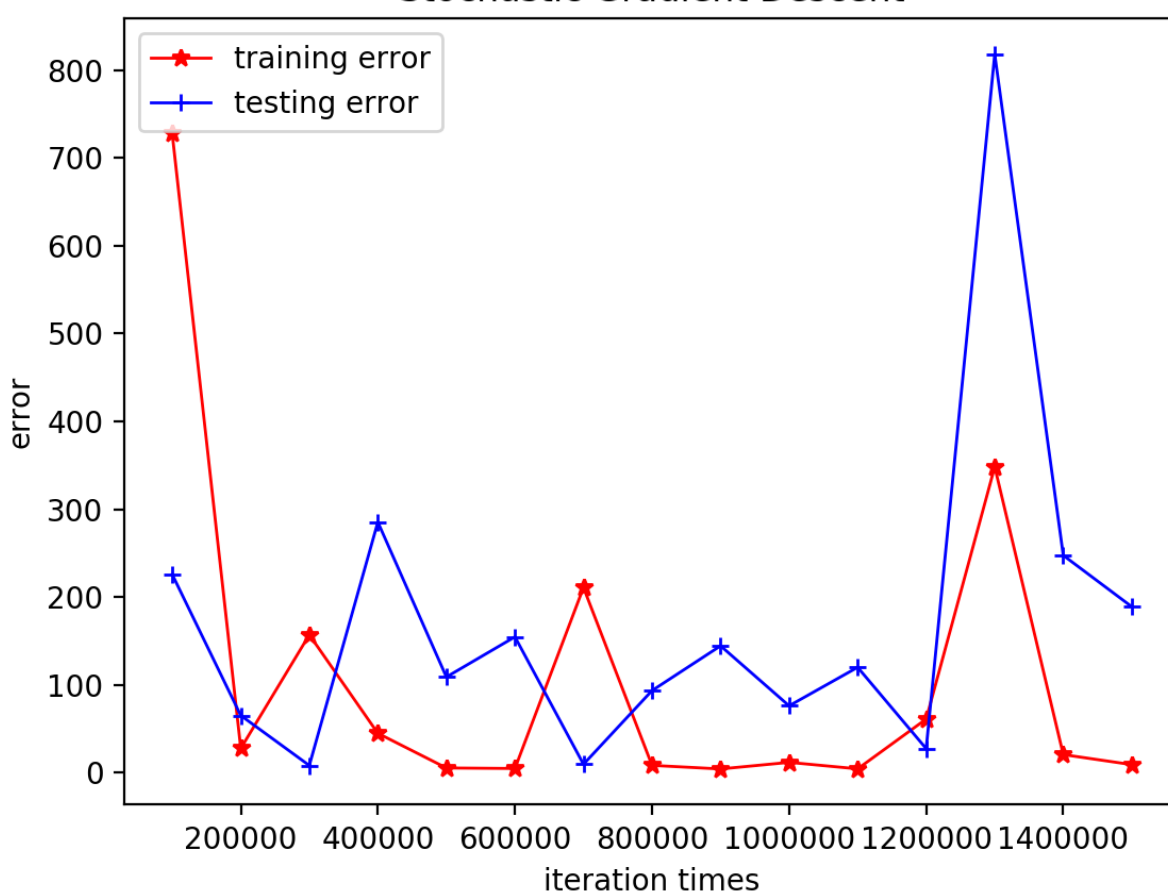


第二次实验

```
# zhanghanjian @ CoderAtMBP in ~/github/DataMining/hw3 on git:master x [19:41:10]
$ python3 main.py
```

| number of iterations | theta0            | theta1            | theta2             | training_error     | testing_error      |
|----------------------|-------------------|-------------------|--------------------|--------------------|--------------------|
| 100000               | 45.85960252126788 | 6.898903407153634 | -73.75987521888949 | 727.5731758288935  | 225.85117097572916 |
| 200000               | 65.03092510001713 | 6.872988564904757 | -72.57715188561805 | 27.753538646312208 | 65.243233784745    |
| 300000               | 73.43426328552881 | 6.796867754328899 | -72.51523876119572 | 157.0999240753341  | 8.350857247674377  |
| 400000               | 76.83931345106463 | 6.841865885828547 | -72.26832796622891 | 44.96951032851007  | 285.41621364615054 |
| 500000               | 78.10241874902766 | 6.765606364359882 | -72.44034060027492 | 5.496214741007151  | 109.06696127677097 |
| 600000               | 78.74044883387606 | 6.77879305858796  | -72.38434189009806 | 4.921106010686469  | 154.74264987896933 |
| 700000               | 78.9895101617303  | 6.625481541605875 | -72.32439977568212 | 210.92540356371256 | 9.729871733636493  |
| 800000               | 79.27303361285716 | 6.744319505925608 | -72.4060008950449  | 8.457518142608723  | 93.58189478619256  |
| 900000               | 79.52597518716244 | 6.763255065955384 | -72.2832052017285  | 4.37620262126744   | 144.52405753462867 |
| 1000000              | 79.46419040526574 | 6.726662192311295 | -72.21494731473604 | 11.876661123606691 | 76.25217728615245  |
| 1100000              | 79.34351022353054 | 6.758640179080688 | -72.44727125067293 | 4.495327553848098  | 120.0967529768261  |
| 1200000              | 79.57780002380296 | 6.693858672092825 | -72.49221811905828 | 61.185348907412376 | 27.5018806815618   |
| 1300000              | 79.60248167805605 | 6.942578537053386 | -72.45869227727611 | 348.2815659542219  | 818.0016998729815  |
| 1400000              | 79.58914674677673 | 6.812958116692476 | -72.64530437557977 | 20.713216952718525 | 247.49934124190884 |
| 1500000              | 79.68383215363977 | 6.783876641904582 | -72.41168840349357 | 9.29754607796619   | 189.06177224871794 |

Stochastic Gradient Descent



误差波动较大，但随着迭代次数的增加，整体呈现下降趋势，但仍然存在特别大的波动。随机梯度下降对于训练样本有较大的随机波动性，存在重复样本过多的可能性，导致数据与理想的批量梯度下降有区别。优点是训练时间较快。

## Exercise 2

(a)

- Sigmoid 函数（逻辑函数）

$$g(z) = \frac{1}{1 + e^{-z}}$$

- 逻辑回归模型（概率模型+线性回归）

$$P(y = 1|x; \theta) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

- y符合二项分布，对应条件对数似然函数

$$LCL = \sum \log p^{(i)} + \sum \log(1 - p^{(i)})$$

(b)

- 对  $w_0$  求导后

$$\frac{\partial}{\partial w_0} LCL = \sum (y^{(i)} - p^{(i)})$$

- 对  $w_j$  求导后

$$\frac{\partial}{\partial w_j} LCL = \sum (y^{(i)} - p^{(i)}) x_j^{(i)}$$

- 进行梯度下降迭代（theta 换 w, p为关于w, x的函数）

$$w_j := w_j + \alpha \frac{\partial}{\partial \theta_j} LCL = w_j + \alpha \sum (y^{(i)} - p^{(i)}) x_j^{(i)}$$

(c)

由于训练数据比ex1大的多，因此使用随机梯度下降减少训练时间。

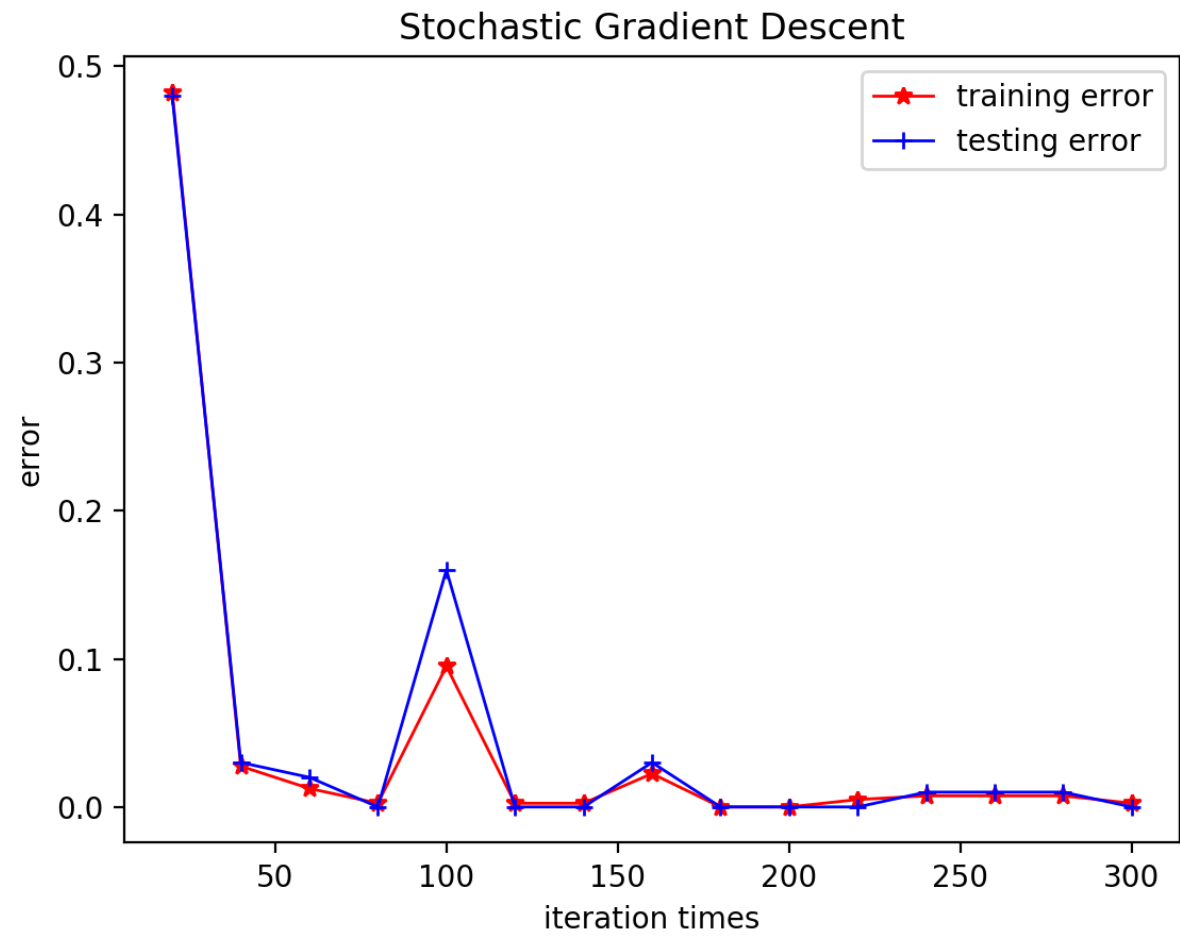
$$w_j := w_j + \alpha (y^{(i)} - p^{(i)}) x_j^{(i)}$$

训练结果使用错误率表示

错误分类/样本总数

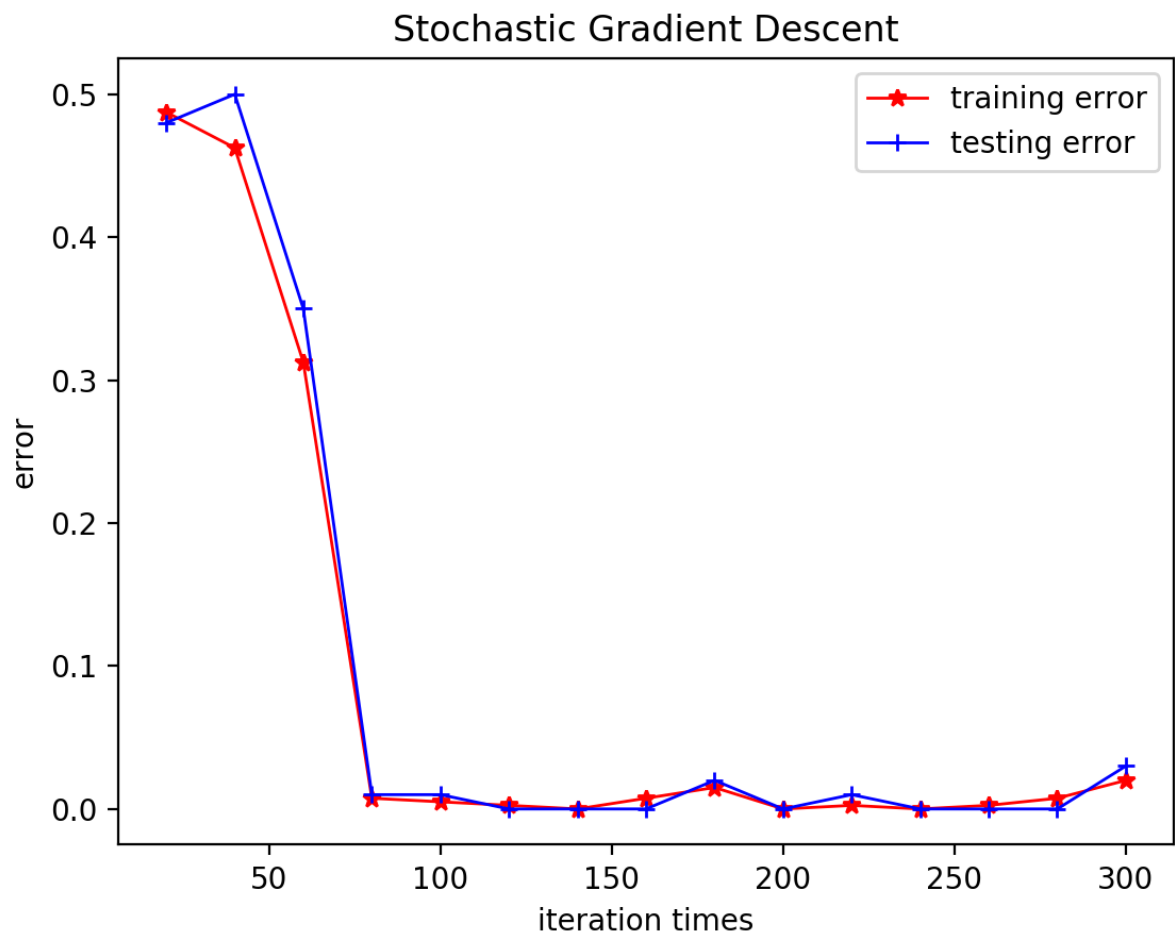
• 测试参数

迭代次数以 20 次迭代为单位记录一次，共记录15次，学习率调整为0.002



| number of iterations | theta0  | theta1  | theta2 | theta3  | theta4 | theta5  | theta6  | t0-rate               | t1-rate              |
|----------------------|---------|---------|--------|---------|--------|---------|---------|-----------------------|----------------------|
| 20                   | 0.0003  | -0.0001 | 0.0005 | -0.0002 | 0.0006 | -0.0002 | 0.0001  | 0.48250000000000004   | 0.48                 |
| 40                   | 0.0002  | -0.0005 | 0.0008 | -0.0005 | 0.0010 | -0.0004 | 0.0001  | 0.02749999999999997   | 0.030000000000000027 |
| 60                   | 0.0003  | -0.0007 | 0.0013 | -0.0009 | 0.0014 | -0.0006 | 0.0001  | 0.012499999999999956  | 0.020000000000000018 |
| 80                   | 0.0001  | -0.0012 | 0.0015 | -0.0013 | 0.0017 | -0.0011 | 0.0001  | 0.0024999999999999467 | 0.0                  |
| 100                  | -0.0003 | -0.0019 | 0.0017 | -0.0018 | 0.0018 | -0.0016 | 0.0000  | 0.09499999999999997   | 0.16000000000000003  |
| 120                  | 0.0002  | -0.0019 | 0.0023 | -0.0019 | 0.0025 | -0.0016 | 0.0001  | 0.0024999999999999467 | 0.0                  |
| 140                  | 0.0003  | -0.0022 | 0.0026 | -0.0022 | 0.0029 | -0.0019 | -0.0000 | 0.0024999999999999467 | 0.0                  |
| 160                  | 0.0001  | -0.0027 | 0.0029 | -0.0027 | 0.0031 | -0.0023 | -0.0001 | 0.022499999999999964  | 0.030000000000000027 |
| 180                  | 0.0007  | -0.0027 | 0.0036 | -0.0027 | 0.0038 | -0.0022 | 0.0002  | 0.0                   | 0.0                  |
| 200                  | 0.0008  | -0.0029 | 0.0040 | -0.0030 | 0.0042 | -0.0023 | 0.0002  | 0.0                   | 0.0                  |
| 220                  | 0.0005  | -0.0034 | 0.0040 | -0.0036 | 0.0044 | -0.0030 | 0.0001  | 0.005000000000000044  | 0.0                  |
| 240                  | 0.0013  | -0.0033 | 0.0048 | -0.0035 | 0.0053 | -0.0028 | 0.0004  | 0.007499999999999951  | 0.010000000000000009 |
| 260                  | 0.0015  | -0.0036 | 0.0053 | -0.0037 | 0.0058 | -0.0029 | 0.0006  | 0.007499999999999951  | 0.010000000000000009 |
| 280                  | 0.0017  | -0.0039 | 0.0058 | -0.0040 | 0.0062 | -0.0031 | 0.0005  | 0.007499999999999951  | 0.010000000000000009 |
| 300                  | 0.0015  | -0.0042 | 0.0060 | -0.0044 | 0.0065 | -0.0035 | 0.0004  | 0.0024999999999999467 | 0.0                  |





| number of iterations | theta0  | theta1  | theta2 | theta3  | theta4 | theta5  | theta6  | t0-rate              | t1-rate              |
|----------------------|---------|---------|--------|---------|--------|---------|---------|----------------------|----------------------|
| 20                   | 0.0039  | -0.0012 | 0.0060 | -0.0014 | 0.0068 | -0.0004 | 0.0019  | 0.4875000000000004   | 0.48                 |
| 40                   | -0.0050 | -0.0103 | 0.0074 | -0.0112 | 0.0065 | -0.0101 | -0.0034 | 0.4625               | 0.5                  |
| 60                   | -0.0039 | -0.0154 | 0.0127 | -0.0150 | 0.0108 | -0.0144 | -0.0024 | 0.3125               | 0.35                 |
| 80                   | 0.0071  | -0.0150 | 0.0227 | -0.0151 | 0.0221 | -0.0115 | 0.0002  | 0.00749999999999951  | 0.01000000000000009  |
| 100                  | 0.0080  | -0.0187 | 0.0290 | -0.0180 | 0.0269 | -0.0154 | 0.0008  | 0.005000000000000044 | 0.01000000000000009  |
| 120                  | 0.0089  | -0.0235 | 0.0347 | -0.0221 | 0.0324 | -0.0189 | 0.0002  | 0.002499999999999467 | 0.0                  |
| 140                  | 0.0060  | -0.0280 | 0.0368 | -0.0282 | 0.0356 | -0.0225 | -0.0016 | 0.0                  | 0.0                  |
| 160                  | 0.0050  | -0.0330 | 0.0406 | -0.0345 | 0.0392 | -0.0279 | -0.0026 | 0.00749999999999951  | 0.0                  |
| 180                  | 0.0042  | -0.0358 | 0.0447 | -0.0398 | 0.0445 | -0.0336 | -0.0053 | 0.015000000000000013 | 0.020000000000000018 |
| 200                  | 0.0129  | -0.0368 | 0.0533 | -0.0407 | 0.0547 | -0.0329 | -0.0027 | 0.0                  | 0.0                  |
| 220                  | 0.0176  | -0.0376 | 0.0594 | -0.0426 | 0.0617 | -0.0333 | -0.0004 | 0.002499999999999467 | 0.01000000000000009  |
| 240                  | 0.0127  | -0.0449 | 0.0633 | -0.0495 | 0.0633 | -0.0406 | -0.0047 | 0.0                  | 0.0                  |
| 260                  | 0.0098  | -0.0501 | 0.0665 | -0.0544 | 0.0676 | -0.0457 | -0.0066 | 0.002499999999999467 | 0.0                  |
| 280                  | 0.0089  | -0.0554 | 0.0702 | -0.0591 | 0.0706 | -0.0498 | -0.0084 | 0.00749999999999951  | 0.0                  |
| 300                  | 0.0062  | -0.0620 | 0.0732 | -0.0638 | 0.0728 | -0.0557 | -0.0115 | 0.020000000000000018 | 0.030000000000000027 |

(d)

在迭代次数足够多时为0

(e)

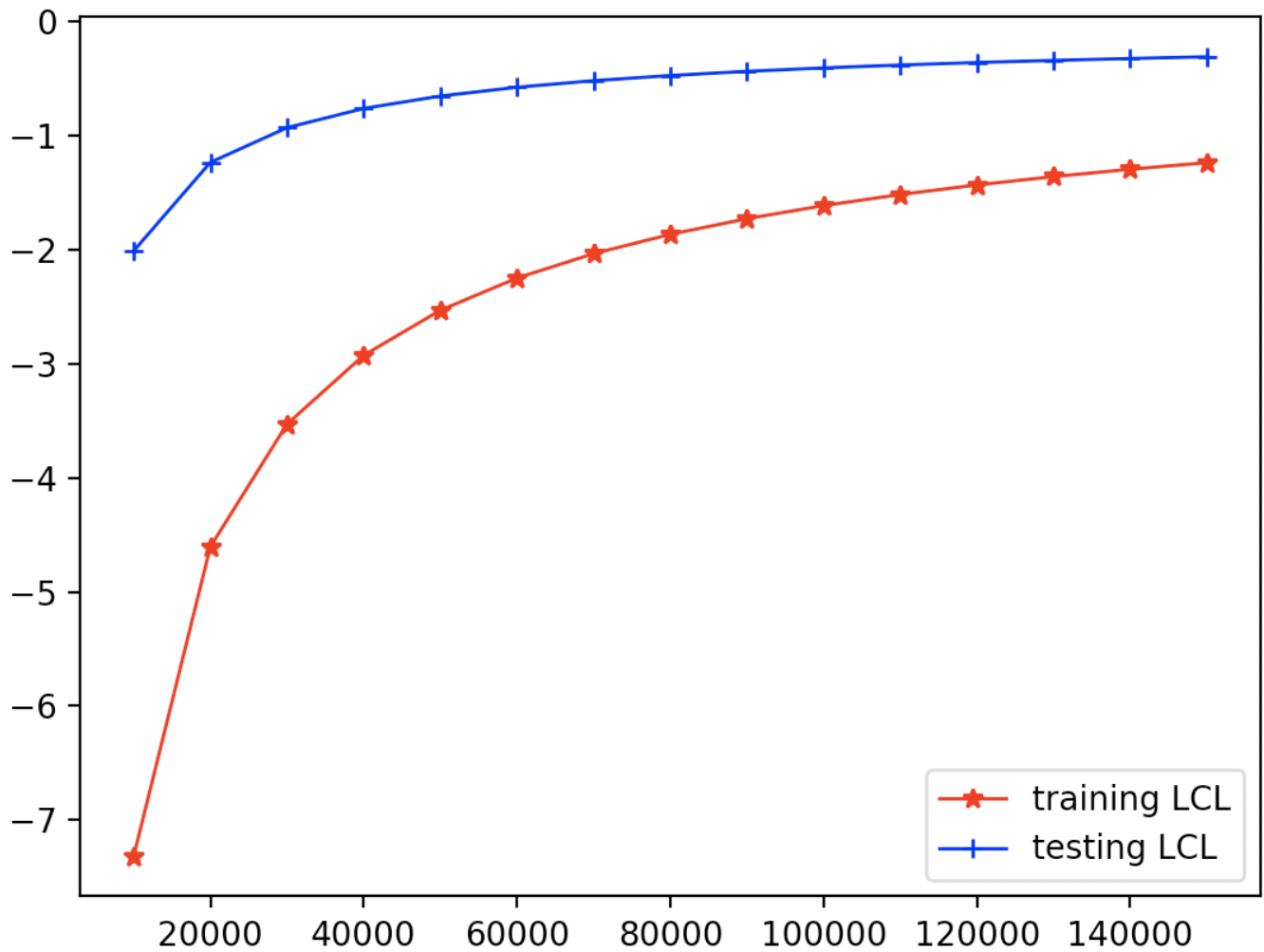
收敛至少需要 30000 次迭代



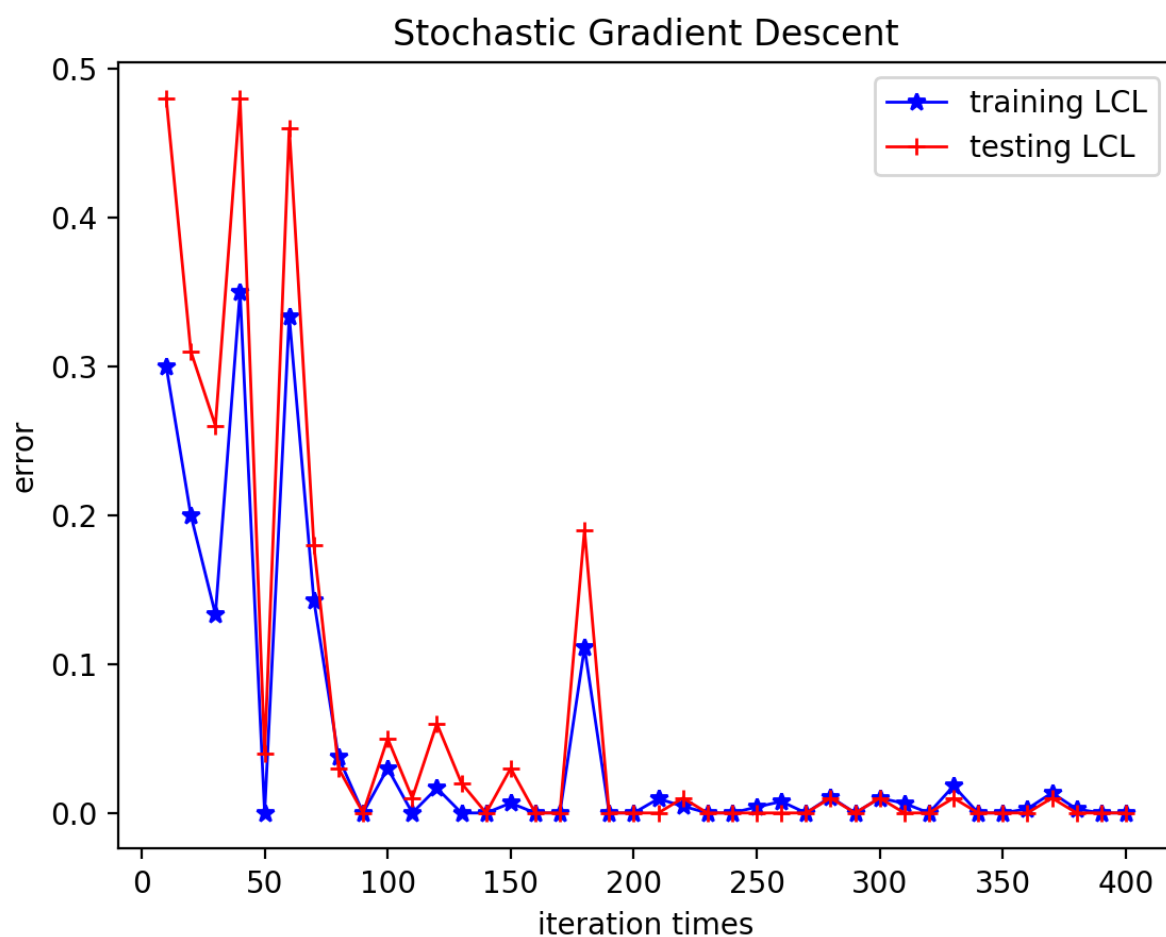
```
# zhanghanjian @ CoderAtMBP in ~/github/DataMining/hw3 on git:master x [16:57:09]
$ python3 ex2.py
```

| number of iterations | theta0 | theta1  | theta2  | theta3  | theta4 | theta5  | theta6 | t0-rate             | t1-rate              |
|----------------------|--------|---------|---------|---------|--------|---------|--------|---------------------|----------------------|
| 10000                | 0.8815 | -4.1809 | 5.7424  | -4.2273 | 5.4712 | -3.3599 | 0.1282 | -7.316995434449817  | -2.0134179956173317  |
| 20000                | 1.0181 | -4.9595 | 6.9149  | -5.0341 | 6.4778 | -3.8936 | 0.1523 | -4.605964204015925  | -1.2380751044884846  |
| 30000                | 1.0958 | -5.4492 | 7.6823  | -5.5520 | 7.1053 | -4.2098 | 0.1543 | -3.530992301819622  | -0.9338329927512855  |
| 40000                | 1.1494 | -5.8137 | 8.2692  | -5.9440 | 7.5691 | -4.4350 | 0.1489 | -2.9282177112021937 | -0.7654955258374675  |
| 50000                | 1.1899 | -6.1071 | 8.7509  | -6.2639 | 7.9398 | -4.6097 | 0.1408 | -2.5336426111080734 | -0.6566866269008187  |
| 60000                | 1.2221 | -6.3542 | 9.1626  | -6.5366 | 8.2501 | -4.7522 | 0.1316 | -2.251353600064071  | -0.5797251638382727  |
| 70000                | 1.2487 | -6.5686 | 9.5240  | -6.7757 | 8.5175 | -4.8723 | 0.1224 | -2.0373650678783375 | -0.5219764609049863  |
| 80000                | 1.2712 | -6.7586 | 9.8472  | -6.9897 | 8.7531 | -4.9759 | 0.1133 | -1.8684059433702604 | -0.4767934441214429  |
| 90000                | 1.2906 | -6.9297 | 10.1403 | -7.1839 | 8.9640 | -5.0668 | 0.1047 | -1.7309020702670181 | -0.4403217202928711  |
| 100000               | 1.3077 | -7.0856 | 10.4090 | -7.3622 | 9.1550 | -5.1477 | 0.0966 | -1.6163512143077792 | -0.4101613721333954  |
| 110000               | 1.3228 | -7.2292 | 10.6575 | -7.5273 | 9.3298 | -5.2205 | 0.0890 | -1.519130452061745  | -0.38473443306239585 |
| 120000               | 1.3363 | -7.3623 | 10.8888 | -7.6815 | 9.4910 | -5.2865 | 0.0818 | -1.4353580069881087 | -0.3629576712850479  |
| 130000               | 1.3486 | -7.4866 | 11.1054 | -7.8262 | 9.6408 | -5.3469 | 0.0751 | -1.362259278382012  | -0.34406099365363063 |
| 140000               | 1.3598 | -7.6033 | 11.3094 | -7.9628 | 9.7806 | -5.4024 | 0.0688 | -1.2977931818637083 | -0.3274808183084083  |
| 150000               | 1.3700 | -7.7134 | 11.5021 | -8.0923 | 9.9118 | -5.4538 | 0.0629 | -1.2404214633468333 | -0.3127944976242205  |

## Stochastic Gradient Descent



(f)



统一参数和迭代次数下，随着测试样本的增加，错误率逐渐降低，但由于样本的随机选取性，存在错误率提高的波形，但整体还是下降趋势的。