**Coursework (2) for Introductory Lectures on Optimization**

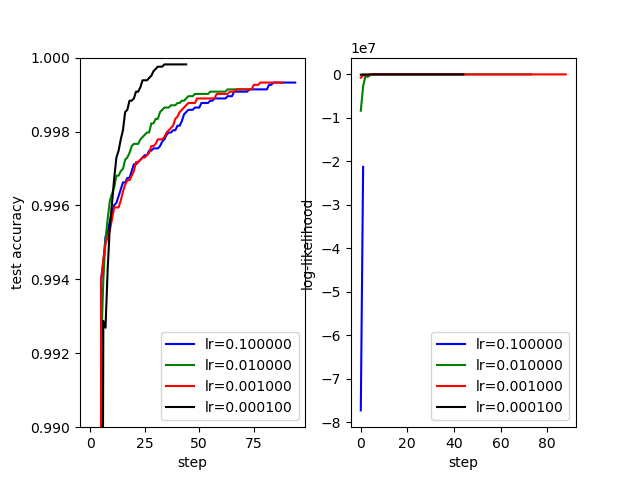
Yu Chen

12321098

1. 针对Gradient Descent算法，分别测试了lr=0.0001，0.001，0.01，0.1，结果如下表所示：

|  |  |
| --- | --- |
| Lr | accuracy |
| 0.0001 | 0.999816 |
| 0.001 | 0.999324 |
| 0.01 | 0.999140 |
| 0. 1 | 0.999324 |

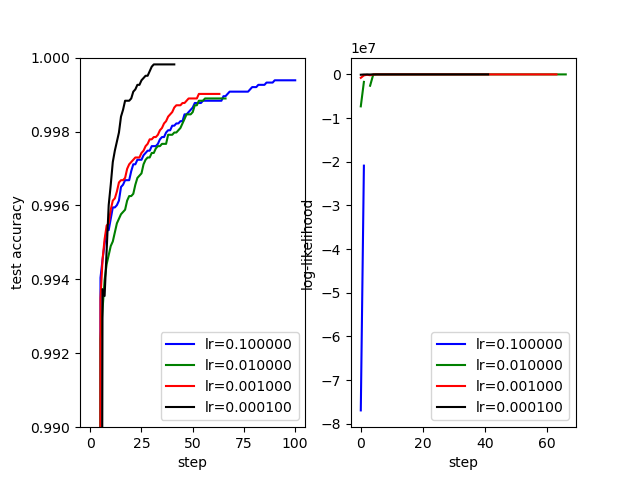
四种学习率的测试集正确率和对数似然变化如下图：



1. 针对Accelerated Gradient Descent算法，分别测试了lr=0.0001，0.001，0.01，0.1，结果如下表所示：

|  |  |
| --- | --- |
| Lr | accuracy |
| 0.0001 | 0.999816 |
| 0.001 | 0.999017 |
| 0.01 | 0.998894 |
| 0. 1 | 0.999386 |

四种学习率的测试集正确率和对数似然变化如下图：



可以得到：

1. 四种学习率得到的最优正确率相差不大，但lr越小，正确率越高，收敛的越快；
2. 学习率越大，收敛性能下降，计算出的对数似然越小。