**梯度下降对比试验**

**Data set:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Trainging  n | Features  d | 稀疏性 | Worker |
| Rcv1 | 677,399 | 47,236 | 0.16% | 8 |
| Susy | 2,181,583 | 18 |  |  |

**实验１：对比　SVRG,Momentum,Mini-Batch SGD,Adagrad**

说明:

1.下面是多次实验，可以取得的最好的实验效果

2.每次实验都需要调节参数比如：SVRG的内部更新次数，采样比例等

3.时间单位 S

数据集1RCV

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Method | 1e-7 | 1e-6 | 1e-5 | 1e-4 | 1e-3 | 1e-2 | 1e-1 |
| SVRG | 2.6 | 2.5 | 2.5 | 2.1 | 2.0 | 1.5 | 1.5 |
| Momentum | 4.7 | 3.9 | 3.6 | 2.7 | 2.6 | 1.5 | 1.3 |
| SGD | 4.2 | 2.6 | 1.8 | 1.2 | 1.2 | 1.3 | 1.1 |
| Adagrad | 不收敛 | - |  |  |  |  |  |

数据集2Susy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | 1e-7 | 1e-6 | 1e-4 | 1e-2 |
| SVRG | 7.8 | 6.8/6.1/7.4/8.0 | 6.6 | 5.4/5.2 |
| Momentum | 8 | 6.5/6 | 4 | 3.4 |
| SGD | 10 | 7.9 | 4.3 | 3.4 |
| Adagrad | 24 | 25 | 22 | 20 |

说明：

1似乎SVRG 在 1e-6 波动挺大,但是在高精度范围内,效果还是好于SGD,在高精度收敛要求下,比SGD快 20%-30%

2:动量法在数据集RCV上不如,SGD,在数据集Susy上优于SGD

3:Adagrad在RCV数据集上不收敛,在Susy上,效果很差.