Machine Learning HW7

Aaron Myers

November 25, 2014

1 Parallel PageRank

1. Parts a-d have been completed. Results for all methods on the livejournal and friendster data sets are in tables 1 and 2 below, respectively. My results on Galois and Spark seem to be odd.

Run Time(s, unless specified): 20 iter: livejournal			
Method	1 Thread	8 Thread	16 Thread
Graphlab	86.2	15.1	9.4
Openmp	19	9	5.8
Galois	17.6	7.8	4.3
Mahout	13m50s	-	-
Spark	38m	-	-

Run Time(s,unless specified): 20 iter: friendster			
Method	1 Thread	8 Thread	16 Thread
Graphlab	193.2	164.2	103.5
Openmp	517	169	127
Galois	102	98	45
Mahout	53m	-	-
Spark	21h	-	-

2 Distributed ALS

1. Mahout downloaded and built.

2. I will not be reporting any results for Mahout since I kept getting connections errors.

3 Newton Method for Logistic Regression

1. All algorithms have been implemented. Output for current algorithm below. Showing results for covtype dataset. My accuracy doesn't appear to increase significantly throughout the algorithm.

```
Wall Time: 14.4917 Func. Val: 506034 Train acc: 49.5032
Wall Time: 14.3107 Func. Val: 507156 Train acc: 51.8364
Wall Time: 14.3171 Func. Val: 508288 Train acc: 52.1056
```

```
Wall Time: 14.3189 Func. Val: 509438 Train acc: 52.3211
Wall Time: 14.3157 Func. Val: 510600 Train acc: 54.1667
Wall Time: 14.3179 Func. Val: 511774 Train acc: 54.5230
Wall Time: 14.3175 Func. Val: 512957 Train acc: 55.0174
Wall Time: 14.3162 Func. Val: 514148 Train acc: 55.2136
Wall Time: 14.3198 Func. Val: 515344 Train acc: 55.9045
Wall Time: 14.3153 Func. Val: 516543 Train acc: 55.9365
```