Assignments for Course T-61.6020, part II

February 3, 2008

In each assignment you should implement the algorithm according to the instruction. For each algorithm there will be a stub either in Matlab or Python, although you don’t have to use it. The stubs and the datasets can be obtained from the homepage of the course. The missing pieces of code in the stubs are marked with XXX.

In your report you should explain the algorithm (in roughly 1 page), report the results you achieve. Attach the source code and the commands you used for getting the answers.

The deadline for these homeworks is 21.5.

1 APriori

Implement apriori.py and try it on apriori.dat:

```
python apriori.py apriori.dat <support>
```

Try different values for support.

2 Pagerank / HITS

Implement pagerank.m using eigenvalue decomposition, implement pagerank2.m using power iteration \(^1\). Test the algorithms with coauthor.mat.

Implement hits.m using SVD decomposition, implement hits2.m using iteration given in [Kle98]. Test the algorithms with newsgroups.mat. What are the most significant words?

3 K-means / spectral

Implement kmeans.m and apply it to three gaussians.mat with the cluster number \(K = 3\). Test the algorithm on two circles.mat with \(K = 2\). K-means should fail miserably, why?

Implement spectral.m and test it on two circles.mat. Try the algorithm for different \(s\) parameters.

\(^1\)http://en.wikipedia.org/wiki/Power_iteration
4 FP Tree

Implement `fptree.py` and try it on `fptree.dat`:

```
python fptree.py fptree.dat <support>
```

Try different values for support. You should get the essentially same itemsets as with APriori.

5 SVM

Implement `svm.m` (ignore the soft margin) and try it on `two_gaussian.mat`. Plot $xw + b$. Implement `svm_fast.m` (note that the soft margin is included). Use the dual formulation (Op 3.) and Algorithm 1 in [Joa06]. Plot $xw + b$ again.

References
