Term Project
Part II

Assigned: Wednesday November 5, 2014

Due: 3:00 PM Wednesday November 19, 2014

Instructor Wei Ding

K-Means (Jia, Shaohua; Koh, Do Hyong; Kopsiaftis, Elias;

- K-Means by Joydeep Ghosh and Alexander Liu (http://www.cs.umb.edu/~ding/classes/470_670/homework/K-means.pdf)
- Assignment : Implement K-Means using a language of your choice. Evaluate the algorithm using the standard benchmark Iris dataset (available online from the UCI dataset repository)
- Oral presentation (one page):
  - Jia: Prove that the value of the k-means objective function converges when k-means is run
  - Koh: Describe three advantages and three disadvantages of k-means compared to other clustering methods
  - Kopsiaftis: describe or plot a two-dimensional example where k-means would be unsuitable for finding clusters.

PageRank (Chen, Si)

- PageRank by Bing Liu and Philip S. Yu (http://www.cs.umb.edu/~ding/classes/470_670/homework/PageRank.pdf)
- Assignment : Implement PageRank using a language of your choice. Evaluate the algorithm using Question 5 of the paper.
- Oral presentation (one page):
  - Explain Question 1 of the paper. Explain P and Equation (6.7)

KNN (Zhuang, Yong)

- Chapter 8 kNN: k-Nearest Neighbors by Michael Steinbach and Pang-Ning Tan (http://www.cs.umb.edu/~ding/classes/470_670/homework/ebook_Top%20Ten%20Algorithms%20in%20Data%20Mining.pdf)
- Assignment : Implement KNN using a language of your choice. Evaluate the algorithm using Question 1 of the paper.
- Oral presentation (one page):
  - Explain Question 5 of the paper.

Apriori (Donthula, Naresh)

- Chapter 4 Apriori by Hiroshi Motoda and Kouzou Ohara (http://www.cs.umb.edu/~ding/classes/470_670/homework/ebook_Top%20Ten%20Algorithms%20in%20Data%20Mining.pdf)
- Assignment : Implement Apriori using a language of your choice. Evaluate the algorithm using Question 6 of the paper.
- Oral presentation (one page):
  - Explain the Apriori property of the algorithm (~~Question 6 of the paper~~).

SVM (Singleton, Charles A)

- Chapter 3 SVM by Hui Xue, Qiang Yang, and Songcan Chen (http://www.cs.umb.edu/~ding/classes/470_670/homework/ebook_Top%20Ten%20Algorithms%20in%20Data%20Mining.pdf)
- Assignment : Implement SVM using a language of your choice. You may use external libraries for optimization. Evaluate the algorithm using a dataset from the UCI Machine Learning Repository.
- Oral presentation (one page):
  - Question 1 of 3.8 Exercises. The class must be able to understand how 1(a), 1(b), and 1(c) are calculated.

Non-negative Matrix Factorization (Kang, Tianyu) Parallel and GPU-accelerated BiPart (Mohebbi, Hamidreza)

- Weekly research meetings
- Oral Presentations: Use one slide to cleary explain one core component of the algorithm to the class.