Data Mining (2015 Fall at Nanjing University)

Chong Liu
chongliu@ucsb.edu
7595382

February 2019

1 Introduction

A graduate level course provided for master and PhD students at Nanjing University, Nanjing, China.

2 Instructor

Prof. Lijun Zhang (`https://cs.nju.edu.cn/zlj/`)

3 Website

`https://cs.nju.edu.cn/zlj/Course/DM_15.html`

4 Textbook

Charu C. Aggarwal. Data Mining: The Textbook, Springer, May 2015.

5 My Score

89 / 100

6 Syllabus

1. **Introduction to Data Mining**
   Reference:
   Chapter 1 of the Textbook
   Petersen and Pedersen. The Matrix Cookbook. Technical University of Denmark, 2012.
   Appendices of Boyd and Vandenberghe. Convex Optimization. Cambridge University Press, 2004.
2. **Data Preparation**
   Reference: Chapter 2 of the Textbook
3. **Similarity and Distances**
   Reference: Chapter 3 of the Textbook
4. **Association Pattern Mining**
   Reference: Chapter 4 of the Textbook

5. **Cluster Analysis**

   Reference:

   Chapter 6 of the Textbook

   Belkin and Niyogi. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In NIPS 14, 2001.

   Lee and Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 401: 788-791 1999.

   Xu et al. Document clustering based on non-negative matrix factorization. In SIGIR, 2003.

6. **Outlier Analysis**

   Reference: Chapter 8 of the Textbook

7. **Data Classification**

   Reference:

   Chapter 10 of the Textbook

   Burges. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2): 121-167, 1998.

   Shalev-Shwartz et al. Pegasos: primal estimated sub-gradient solver for SVM. In ICML, 807-814, 2007.

8. **Convex Optimization**

   Reference:

   Boyd and Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

   Nesterov. Gradient methods for minimizing composite functions. Mathematical Programming, 140(1): 125-161, 2013.

   Hazan and Kale. Beyond the regret minimization barrier: an optimal algorithm for stochastic strongly-convex optimization. In COLT, 421-436, 2011.

9. **Data Classification: Advanced Concepts**

   Reference: Chapter 11 of the Textbook

10. **Linear Methods for Regression**

    Reference: Chapter 3 of Hastie, Tibshirani and Vandenberghe. The Elements of Statistical Learning. Springer, 2009.

11. **Mining Text Data**

    Reference: Chapter 13 of the Textbook

12. **Mining Web Data**

    Reference: Chapter 18 of the Textbook

13. **Mining Big Data**

    Reference:

    Hazan and Kale. Beyond the regret minimization barrier: an optimal algorithm for stochastic strongly-convex optimization. In COLT, 421-436, 2011.

    Boyd et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends in Machine Learning, 3(1): 1-122, 2010.

    Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In ICML, 928-936, 2003.

    Hazan et al. Logarithmic regret algorithms for online convex optimization. Machine Learning, 69(2-3): 169-192, 2007.