

CS422:Data Mining

Vijay K. Gurbani, Ph.D.,

Illinois Institute of Technology

参考书籍

- Introduction to Data Mining, 2nd Edition, Pang-Ning Tan, Michael Steinbach, Vipin Kumar; ISBN-13: 978-0133128901
- R For Everyone, 2nd Edition, Jared P. Lander, Addison Wesley, ISBN-13: 978-0134546926
- Mining of massive datasets, 2e, Jure Leskovec, Anand Rajaraman and Jeffrey Ullman, Cambridge University Press, ISBN-13: 978-1-107-07723-2
- Data mining and analysis: Fundamental concepts and algorithms, 2e, Mohammed Zaki and Wagner Meira, Jr., Cambridge University Press, ISBN-13: 978-0521766333
- An introduction to statistical learning with applications in R, 2e, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer, ISBN-13: 978-1-4614-7137-0

R语言参考资料

- <https://ademos.people.uic.edu/index.html>
- <http://www.cookbook-r.com/>
- <https://github.com/matloff/fasteR>
- <http://www.r-tutor.com/r-introduction>
- <https://cran.r-project.org/doc/manuals/R-intro.pdf>
- <https://www.youtube.com/watch?v=s3FozVfd7q4>

作业要求

- 一个pdf文件，可能包含内容说明或计算说明，建议使用LaTeX.
- 一个.Rmd文件，包含基于R编写的可执行文件.
- 一个HTML标签文件，结果输出部分.
- 作业命名并以.zip形式

课程安排

| 周 | 课程主题 | 备注 |
|---|------|----|
|---|------|----|

| 周 | 课程主题 | 备注 |
|----------|--|----|
| 1.08-25 | 课程大纲及期望 介绍数据挖掘(DM, Ch.1) 介绍R语言(RFE, Ch.1, 2.1-2.2, 4, 5, 6) | |
| 2.09-01 | 探索数据(DM, Ch.2.1, 2.2, 2.3) | |
| 3.09-08 | 线性回归(DM, Ch.3) | |
| 4.09-15 | 学习组成部分 监督学习:分类 决策树模型(DM, Ch.3.1-3.3, 3.5, 3.6) | |
| 5.09-22 | 继续学习分类 决策树模型性能评估(DM, Ch.3.4, 3.6, 5.7.2) 替代技术(DM, Ch.4.10, 4.11, 4.12) | |
| 6.09-29 | 继续学习分类 | |
| 7.10-06 | 继续学习分类 关联规则(DMA, Ch.8.1, 8.2, 8.2.1, 8.2.3, 8.3, 12.1; DM, Ch.4.10, 4.11, 4.12) | |
| 8.10-13 | 期中考试(到关联规则) | |
| 9.10-20 | 继续学习关联规则 神经网络:感知型模型(DM, Ch.4.7.1) | |
| 10.10-27 | 继续学习神经网络:感知型模型 继续学习神经网络(DM, Ch.4.7.2-4.7.3) 多层感知器 前向传播 | |
| 11.11-03 | 继续学习神经网络 后向传播 梯度下降 偏差方差权衡(DM, Ch.4.10.3; DMA, Ch.22.3) | |
| 12.11-10 | 无监督学习: 聚类I(DM, Ch.7; DMA, Ch.13-15; SLWR, Ch.10; MMD, Ch.7) 基于K-Means | |

| 周 | 课程主题 | 备注 |
|----------|--|----|
| 13.11-17 | 聚类II 层次聚类、基于密度的聚类和聚类评估 主成分分析(SLWR, Ch.6.3.1, 12.2, 12.2.1, 12.2.2; DMA, Ch7) | |
| 14.11-24 | 感恩节 | |
| 15.12-01 | 推荐系统(SLWR, Ch.9) 基于内容 协同过滤(时间允许) 挖掘社交网络图(将分配阅读材料) | |
| 16.12-08 | 期末考试 | |