Research Seminar on Machine Learning in Business Analytics
Part I: Method and Part II: Application

Kunpeng Zhang – Part I
Jessica Clark – Part II
Fall 2020

Class time: Friday 8:30 – 11:30PM EST, 3 hours in duration
Location: https://umd.zoom.us/j3206804434
Office hours: by email and meet via Zoom

Learning Goals
By the end of the session, students should be able to:

1. Understand, assess, and review the use of basic machine learning in business analytics research.
2. Implement common machine learning techniques in Python.
3. Formulate research questions which either apply machine learning techniques to novel applications, or develop new techniques.
4. Formulate, conduct, and position machine learning research for publication in business school journals.
5. Conduct and position research using machine learning techniques.

Course Description
The goal of this course is to give students the necessary foundation to evaluate and conduct research in machine learning for business analytics applications. This course relies on knowledge of the methods covered in term A, with an increased emphasis on discussing the use of ML and DS techniques in published research from top journals. We will also discuss frameworks of requirements for how to position data science research. Students will:

- Be exposed to numerous applications from business school journals from a variety of fields.
- Evaluate the use of machine learning techniques in papers.
- Apply methods learned in term A.
- Complete the research project proposed at the end of term A, ideally one which could lead to a research paper.

Recommended Textbooks:

1. Machine Learning: A Probabilistic Perspective (MLaPP)

https://doc.lagout.org/science/Artificial%20Intelligence/Machine%20learning/Machine%20Learning_%20A%20Probabilistic%20Perspective%20%5BMurphy%202012-08-24%5D.pdf

2. An Introduction to Statistical Machine Learning with Applications in R (SMLA)

https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf

3. Deep Learning (DL)

http://faculty.neu.edu.cn/yury/AAI/Textbook/DeepLearningBook.pdf

Deliverables and grading

- Final project: Should be the completion of the work proposed at the end of Term A. Worth 50% of the final grade.
- Weekly work: There will be papers assigned every week. Students must write a short (1-page) reaction to each assigned reading. Students must write a longer review of at least 2 of the assigned readings. Worth 20% of the final grade.
- Participation: This is primarily a discussion-based class. Students should come prepared to discuss the assigned readings. Worth 30% of the final grade.

Course Schedule

[Part I]

- **Week 0**: Machine learning background (done by yourself)
  - Introduction to ML - MLaPP Chp 1; SMLA Chp 1, 2.1
  - Probability – MLaPP Chp 2, DL Chp3
  - Python

- **Week 1**: Supervised learning
  - Parameter estimation
  - Decision tree
  - GLM and Probit regression

  Readings:
  1. *SMLA Chp 4, 8*
  2. *MLaPP Chp 9.1 - 9.4*

  [Assignment 1: implement Probit regression (Python)]

- **Week 2:** Unsupervised learning
  - Measuring (dis)similarity and evaluating the output of clustering methods
  - Traditional clustering methods: K-means
  - Non-negative Matrix factorization

  Readings:
  1. *SMLA Chp 10*
  2. *Daniel Lee and H. Seung, Algorithms for Non-negative Matrix Factorization*

- **Week 3:** Language modeling - 1

- N-gram language models
- Word2Vec

Readings:

1. *N-gram Language Models: https://web.stanford.edu/~jurafsky/slp3/3.pdf*
2. *Mikolov et al., Efficient Estimation of Word representations in Vector Space: https://arxiv.org/abs/1301.3781*

[Assignment 2: Using word2vec to identify similar documents (Python)]

- **Week 4:** Language modeling - 2
  - Topic modeling (LDA)

Readings:

1. *David et al., Latent Dirichlet Allocation, Journal of Machine Learning 3 (2003) 993-1022.*
2. *MLaPP Chp 27.3*
3. *https://ldabook.com/parameter-estimation.html*

[Assignment 3: Derive posterior distribution for Gibbs sampling in LDA]

- **Week 5:** Recommender systems
  - Collaborative filtering
  - Evaluation of recommender systems

Readings:

1. *Badrul et al., Item-Based Collaborative Filtering Recommendation Algorithms. WWW 2010, p285-295.*
2. *Shani, Guy & Gunawardana, Asela. (2011). Evaluating Recommender Systems.*

- **Week 6:** Deep learning - 1
  - Deep neural nets (DNN)
  - Convolutional neural nets (CNN)
  - Sequence learning (recurrent neural nets, RNN)

Reading:

1. *DL Chp 6, 9, & 10*

- **Week 7:** Deep learning - 2
  - Representation learning - Autoencoder

Reading:

1. *DL Chp 14*

[Assignment 4: Image classification using deep learning]

Project proposal due
* All assignments are optional.

[Part II]

- **Week 8**: Discussion: a framework and typology of ML/AI/DS in business research
- **Week 9**: Developing and positioning new ML techniques for business journals
- **Week 10**: Unsupervised learning (how apply it correctly, believable evaluation techniques)
- **Week 11**: Data acquisition: active learning, mechanical turk

- **THANKSGIVING**
- **Week 13**: Language modeling applications
- **Week 14**: Recommender systems
- **Week 15**: Deep learning and other cutting-edge ML techniques