Department of 統計及精算學系
Statistics & Actuarial Science

STAT3612 Data Mining (2017-18 Semester 2)

Course Oultine

Instructor:	Dr. Aijun Zhang (RR224)
Email:	ajzhang@hku.hk
Lecture Hours:	Tuesday 3:30pm – 4:20pm (MW103/RR101)
	Friday 3:30pm – 5:20pm (MW103/RR101)
Tutors:	Dr. Simon K.C. Cheung (RR234) (simonkc@hku.hk)
	Mr. Zebin Yang (RR114) (u3005497@connect.hku.hk)
Tutorial Hours:	TBD (RR101, Starting from Week 2)
Course Website:	http://stat3612.saas.hku.hk & http://moodle.hku.hk

Course Objectives:
This is one of core courses in data science and it provides a comprehensive and practical coverage of essential data mining concepts and statistical models for data mining and machine learning.

Prerequisites:
STAT2602 (Probability and Statistics II) or STAT3902 (Statistical Models).

Intended Learning Outcomes:

1. Understand and apply a wide range of data mining techniques, and recognize their characteristics, strengths and weaknesses.
2. Identify and use appropriate data mining techniques for a data mining project, taking into account both the nature of the data to be mined and the goals of the user of the discovered knowledge.
3. Evaluate the quality of discovered knowledge, taking into account the requirements of the data mining task being solved and the goals of the project.
4. Be proficient with the data scientific languages, in particular R and Python.

Contents and Topics:
Data science, data exploration, linear methods, variable selection, regularization, model assessment, decision trees, ensemble methods, support vector machines, neural networks, deep learning, principal component analysis, clustering, sparse coding.

Assessment:

Assignments:	Three assignments (Hands-on data analysis and modeling)	30%
Tests:	Two in-class tests (Problem solving and calculations)	40%
Group Project:	Project proposal, oral presentaiton and written report	30%

Software/Programming:
R (R Studio), Python (IPython), TensorFlow/Keras.

References:

1. James, G., Witten, D, Hastie, T. and Tibshirani R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer, New York. `http://www-bcf.usc.edu/~gareth/ISL/`

2. Hastie, T, Tibshirani, R. and Friedeman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Second Edition, Springer, New York. `https://web.stanford.edu/~hastie/ElemStatLearn/`

3. Abu-Mostafa, Y. S., Magdon-Ismail, M. and Lin, H. T. (2012). *Learning From Data: A Short Course.* `http://www.amlbook.com/`

4. Wickham, H. and Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly. `http://r4ds.had.co.nz/`

5. Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, OReilly. `https://github.com/ageron/handson-ml`

6. Chollet, F. (2018). *Deep Learning with Python.* Manning. `https://www.manning.com/books/deep-learning-with-python`

For future data scientisits: