

STA 5703 Data Mining Methodology I (Fall 2015)

Class days & times: TT 4:30 to 5:45 PM (CB1 0219)
Office hours: Tuesday 2:00 to 4:30 and Thursday 2:00 to 4:30
Office: Computer Center II, Room 203
Phone: 407-823-2818
e-mail: Chung-Ching.Wang@ucf.edu
Withdraw deadline: November 2, 2015
Holidays: September 7, November 11, November 26 – 28
Football Day: September 3 and November 19
Special note:

- 1) Students who are not officially registered in the class will not have exams graded or returned.
- 2) Possibility of no classes (or substitute instructor) on October 27 or other days when the instructor has other important professional duties such as attending professional conference. Any of these types of events will be announced at least one week in advance.

Required Text:

- 1) “An Introduction to Statistical Learning: with Applications in R,” Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013), Springer.
- 2) “Classification and Regression Trees,” Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984), Chapman and Hall.

Optional Text:

- 1) “Handbook of Statistical Analysis and Data Mining Applications,” Robert Nisbet, John Elder, and Gary Miner (2011), Springer.

About this course:

Statistical learning is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of observational data in order for the data owner to discover meaningful patterns and models. By applying statistical learning techniques, analysts can fully exploit data patterns and behavior, and gain a greater understanding of the inside of the data. The goal of statistical learning applications is to produce new knowledge that decision-makers can act upon. It does this by using techniques such as neural network and decision trees to build a model of the real world based on data collected from a variety of sources including corporate transactions, customer histories and demographics, and external sources such as web logs, credit bureaus and the census. This model produces knowledge that can be used to support decision-making and to predict new business opportunities. This course includes the following twelve lectures written by instructor:

- Lecture 0: Using Enterprise Miner/R (Self Study)
- **Lecture 1: Mining Big Data – An Overview**
- **Lecture 2: Supervised Learning I – Support Vector Machine**
- Lecture 3: Supervised Learning II – Theoretical Foundation of Decision Trees
- Lecture 4: Supervised Learning II – Data Preparation and Exploration using Decision Trees
- Lecture 5: Supervised Learning II – Modeling Using Decision Trees (Boosting Trees and Random Forests)
- Lecture 6: Supervised Learning III – Theory on Regression
- Lecture 7: Supervised Learning III – Theory on Logistic Regression
- Lecture 8: Case Study - Double Target (Continuous and Discrete) Modeling
- Lecture 9, 10, 11: Supervised Learning IV – Neural Network Fundamental
- Lecture 12, 13, 14: Supervised Learning IV – Neural Network Modeling

STA 5703 Data Mining Methodology I (Fall 2015)

- Lecture 15: Case Studies (Optional)

What can I expect to get out of this course?

When you leave this course you should be able to:

- Use SAS/Enterprise Miner/R software effectively to perform data mining exercises
- Use decision tree techniques to build classification and regression models
- Use support vector machine to build classification models
- Use regression techniques to build classification and regression models
- Use neural network techniques to build classification and regression models
- Build models to predict multiple targets simultaneously
- Be capable of using model assessment nodes and decision theory to compare models
- Be capable of performing exploratory data analysis with trees
- Be capable of using trees to perform model validation analyses

How will I learn in this course?

- Reading chapters from the text
- Listening to lectures
- Working on projects
- Working with classmates
- Consulting your instructor if necessary
- Using Internet/e-mail and chat room (if desired)
- Preparing your presentations and reports

How will I be evaluated in this course?

Activity	Points
Assignments	200
Quizzes	300
Mid-Term Exam/Report (October 15)	200
Final Report (December 10)	200
Final Exam/Presentation (December 10)	100
Total	1000

- The proposed grading scale is as follows:

<u>POINTS</u>	<u>GRADE</u>
≥925	A
900 – 924	A-
850 – 899	B+
800 – 849	B
750 – 799	B-
700 – 749	C
≤ 699	F

- **Academic dishonesty on any exam will result in a grade of zero in that exam**

Make-up Exam Policy

Make-up exams or extended assignment due date will be allowed only in extreme instances and with **advanced permission** of the instructor. If any student feels that a sudden illness is sufficiently extreme to warrant a make-up test or to extend the assignment due date, the instructor must be provided with documentation prepared by an appropriate medical authority.