STT 481 Capstone in Statistics

Term:	Spring 2019	**Instructor:**	Dr. Chih-Li Sung
Time:	Tu Th 3:00 - 4:20 PM	**Office:**	C418 Wells Hall
Room:	A232 Wells Hall	**Phone:**	(517) 353-2963
Credit Hours:	3	**Email:**	`sungchih@msu.edu`

Office Hours: 1:30 - 3:00 PM Tu Th and by appointment.

Textbook: *An Introduction to Statistical Learning with Applications in R* by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. The textbook is available online at Gareth James' website `(link)`, as well as available for purchase at Amazon and other retailers.

Prerequisites: (STT 442 or approval of department) and completion of Tier I writing requirement

Software: R and Rstudio

Website: `https://sites.google.com/view/2019-spring-stt481`

Course Description: Statistical capstone experiences are essential for statisticians to perform an in-depth analysis of real-world data. Capstone experiences can develop statistical thinking by engaging in a consulting-like experience that requires skills outside the scope of traditional courses: defining a complex problem, analyzing data, building a strong team, and communicating effectively. In this course, selected readings and projects will be given to illustrate special problems encountered by professional statisticians in their roles as consultants, educators, and analysts.

This course covers the following topics:

1. problem formulation
2. advanced statistical modeling, preliminary data analysis, and machine learning
3. Statistical software (R)
4. thorough and elaborate statistical analyses of data
5. final presentation and data visualization

Grade Policy: Final grades will be based on one final project (20%), five homework assignments (40%), and two midterms (20%+20%). The final grade would be based on your total grade percentage and will be determined roughly as:

Percentage	90-100	85-89	80-84	75-79	70-74	65-69	60-64	0-59
Grade	4.0	3.5	3.0	2.5	2.0	1.5	1.0	0

Homework: Homework assignments include conceptual and applied exercises. Typesetting your reports/solutions in Latex or R markdown is strongly encouraged (you will receive 5 extra credit points). Unreadable handwriting is subject to zero credit.

Assignments are due at the beginning of class on the day that they are due. You will be allowed 2 total late days without penalty for the entire semester. Please use these wisely, and plan ahead for conferences, travel, deadlines, etc. Once those days are used, you will be penalized according to the following policy:

- Homework is worth full credit at the beginning of class on the due date.
- is worth half credit for the next 48 hours.
- It is worth zero credit after that.

This class abides by `SpartanCodeofHonor`. Unless otherwise specified, homework will be done individually and each student must hand in their own assignment. It is acceptable, however, for students to collaborate in figuring out answers and helping each other understand the underlying concepts. When collaborating, the "whiteboard policy" is in effect: You may discuss assignments on a whiteboard, but, at the end of a discussion the whiteboard must be erased, and you must not transcribe or take with you anything that has been written on the board during your discussion. You must be able to reproduce the results solely on your own after any such discussion. Finally, you must write the names of the students you collaborated with on each homework.

Project: The project is an opportunity for you to deeply explore one (or several) of the techniques covered in class and apply them to a real problem about predicting house prices. In this problem, you are given the house dataset on Kaggle and the goal is to predict the final price of each home, based on 79 explanatory variables that describes (almost) every aspect of residential homes in Ames, Iowa. Kaggle is a open platform for predictive modeling and analytics competitions, where companies and researchers can post their data and problems for users to solve. In this competition, you are challenged to provide an accurate prediction for the final prices of the houses in Ames, Iowa.

Your prediction result will be calculated and posted on the website. However, as a statistician/data scientist, your role is not only providing an accurate prediction. In many cases, you also need to provide a model interpretation which can turn into useful suggestions for decision making. Thus, in this project, the main threads and their gradings are:

1. Accurate Prediction - RMSE on 05/02 5:45pm (10%)
2. Model Interpretation & Final Presentation (10%)

For more details regarding the project, please see the course website.

Midterm: Midterm exams include one written exam (20%) and one midterm report (20%).

- Written exam: There will be a closed-book midterm but you can bring 2 double-sided pages of the cheat sheet. The exam will be in class.
- Midterm report: This is the detailed report of your project. You should re-state the problem you are solving and your approaches and findings, and summarize your results.

Class Participation: Although class participation is not explicitly graded, I will use your class participation to determine whether your grade can be lifted in case you are right on the edge of two grades. Participation means attending classes, participating in class discussions, asking relevant questions, volunteering to provide answers to questions, and providing constructive criticism and creative suggestions that improve the course.

Course Schedule: The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts you should grasp after each week, and also as a study guide before each exam. Each exam will test on the material that was taught up until 1 week prior to the exam.

Week 01, 01/07 - 01/11: ..... Introduction to R
Week 02, 01/14 - 01/18: ..... Introduction to Statistical Learning & KNN
Week 03, 01/21 - 01/25: ..... Linear Regression
Week 04, 01/28 - 02/01: ..... Multiple Linear Regression
Week 05, 02/04 - 02/08: ..... Logistic Regression
Week 06, 02/11 - 02/15: ..... Multinomial Logistic Regression
Week 07, 02/18 - 02/22: ..... Linear Discriminant Analysis & Midterm 1
Week 08, 02/25 - 03/01: ..... Quadratic Discriminant Analysis
Week 09, 03/04 - 03/08: ..... Spring Break
Week 10, 03/11 - 03/15: ..... Resampling Methods
Week 11, 03/18 - 03/22: ..... Linear Model Selection and Regularization
Week 12, 03/25 - 03/29: ..... Nonparametric Regression & Midterm 2
Week 13, 04/01 - 04/05: ..... Tree-Based Methods
Week 14, 04/08 - 04/12: ..... Support Vector Machines
Week 15, 04/15 - 04/19: ..... Principal Components Analysis
Week 16, 04/22 - 04/26: ..... Clustering Methods & Final Presentation
Week 17, 05/02 (5:45pm - 7:45pm): ..... Final Presentation

Tips: Test each method for your project once you learn a new technique in class.

Important Dates:

Monday, 01/07 ..... Classes Begin
Tuesday, 02/19 ..... **Midterm 1**
Wednesday, 02/27 ..... Middle of Semester
Monday, 03/04 - Friday, 03/08 ..... Spring Break
Thursday, 03/28 ..... **Midterm 2**
Friday, 04/26 ..... Classes End
Thursday, 05/02 (5:45pm - 7:45pm) ..... **Final Presentation**
Friday, 05/03 - Sunday, 05/05 ..... Commencements

Policy:

- Academic Honesty: The Department of Statistics and Probability adheres to the policies of academic honesty as specified in the General Student Regulations 1.0, Protection of Scholarships and Grades, and in the All-University of Integrity of Scholarship and Grades which are included in Spartan Life: Student Handbook and Resource Guide. Student who plagiarize will receive a grade 0.0 on the assignment.

- Make-up tests will be given only when you have a verifiable excuse; otherwise the exam score will be 0.
- Attendance: You are expected to attend all meetings of the class. If you miss a class for whatever reason, you are responsible for all materials, assignments and deadlines missed. While office hours provide an opportunity for further clarification of materials covered in class, office hours will not substitute for classes. See more regarding attendance policy at `https://reg.msu.edu/ROInfo/Notices/Attendance.aspx`.
- ADA: To arrange for accommodation a student should contact the Resource Center for People with Disabilities (353-9642) `http://www.rcpd.msu.edu/`.

Disclaimer: The instructor reserves the right to make any changes he considers academically advisable. Changes will be announced in class, it is your responsibility to keep up with any changed policies.