This course focuses on modern regression techniques. The enormous increase in computing power has greatly influenced applied statistical methods, and especially data modeling. In a variety of scientific and business fields and applications, the relationship between one or more variables (the response or output) and a candidate group of explanatory variables (the regressors, predictors or inputs) is under study. The goal is usually modeling this relationship so that the process is understood and possibly used for predicting future values of outputs.

The traditional statistical approach is to use linear regression for fitting and prediction. Even though linear regression has been an overall successful data modeling tool, also due to its computational simplicity, it superimposes mathematical models to the data that may be oversimplified. It may also be unable to capture the complexity of the underlying relationships and it cannot be applied when the number of variables exceeds the number of observations.

In this course, we will study nonparametric regression and extensions of the linear model. Linear regression will be the starting point to provide the essential framework for the course. Topics will include nonparametric density estimation, nonparametric regression smoothing techniques (e.g., kernel regression, nearest neighbor regression, local polynomial fitting), additive models, penalized regression and classification, support vector machines, regression trees, and dimension reduction methods. In addition, extracting information and building inferential procedures from plots associated with regression models, i.e., regression graphics, will be presented.
LEARNING OUTCOMES
Students will learn how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms. At the completion of this course students should be able to (i) apply modern regression/statistical learning methods to build predictive models, (ii) select and validate statistical learning models, (iii) assess model fit and error and (iv) use the R language for modern regression and data analysis.

COURSE OUTLINE

1. Linear regression analysis
2. Model Selection and Regression Shrinkage Methods:
   a. Partial F-test, AIC, BIC, Stepwise
   b. Ridge Regression
   c. LASSO and Related Methods
3. Nonparametric density estimation
4. Nonparametric regression and smoothing techniques
   a. Kernel regression
   b. Smoothing splines
   c. Local polynomial regression
   d. Additive models
5. Dimension reduction
   a. Principal Component Analysis and Principal Component Regression
   b. Partial Least Squares
   c. Sufficient Dimension Reduction (lecture notes)
6. Classification
   a. Logistic Regression
   b. Linear Discriminant Analysis
   c. Quadratic Discriminant Analysis
   d. K-nearest-neighbor
   e. Support Vector Machines (if time allows)
7. Tree-Based Methods
   a. Regression and Classification Trees
   b. Random Forests (if time allows)
9. Regression Graphics (also as a dimension reduction technique; lecture notes)

RECOMMENDED TEXTBOOKS

*Elements of Statistical Learning* by Hastie, Tibshirani & Friedman.

*Introduction to Statistical Learning with applications in R* by James, Witten, Tibshirani & Hastie.

*Statistics for High-Dimensional Data* by Buhlmann & van de Geer.

PREREQUISITES
Regression at the level of Stat 2118 and preferably 6214, linear algebra, some probability theory and statistical theory and inference (Stat 6201-2).
HOMEWORK
There will be 4-5 homework assignments. The homework will comprise of theoretical problems and analysis of data sets. Late assignments will not be accepted for any reason, medical or otherwise. Homework assignments that are collected and graded should be treated as individual and not collective efforts.
The provided LaTeX template should be used to report your work. You will submit the write-up and code scripts via Blackboard. You also have to submit a printed copy of the solutions (not the code!) in class by the due date.

EXAMS
There will be an in-class midterm exam and a take-home final data analysis project/exam. Students are encouraged to suggest problems of interest to them.

GRADING
Your final grade will be determined by a weighted average of homework and exam scores: Homework 50%; Midterm 20%, Final 30%

COMPUTING RESOURCES
The statistical software we will be using is R. You can download the R software and get related information from the R home page. Alternatively, RStudio offers a GUI R platform.
A tutorial in R is housed at Clarkson University Department of Mathematics and a more detailed one at http://wwwmaths.anu.edu.au/~johnm/r-book/3edn/xtras/daagur-ohp.pdf
Here is a short reference card.
A list of books using R is given at http://www.r-project.org/doc/bib/R-publications.html
The R Manuals can be found by clicking here http://www.r-project.org/ and selecting Manuals under Documentation on the left column.

A list of nonparametric regression programs and tools is given at J. Fox's webpage.
Arc, particularly helpful for regression graphics, can be downloaded from the Arc software page of the School of Statistics at the University of Minnesota. More information about Arc can be found by clicking here.

BLACKBOARD REGISTRATION
Course information and material, including notes, grades, and details about course assignments will be posted in Blackboard periodically. It is the student’s responsibility to check the Stat 242 Blackboard website frequently for up-to-date information about assignments. Once enrolled in the course, you should automatically be registered on Blackboard. Log into the course website at: https://blackboard.gwu.edu/webapps/portal/frameset.jsp
ACADEMIC INTEGRITY
I personally support the GW Code of Academic Integrity. It states: “Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information.” For the remainder of the code, see: http://www.gwu.edu/~ntegrity/code.html

SUPPORT FOR STUDENTS OUTSIDE THE CLASSROOM
DISABILITY SUPPORT SERVICES (DSS)
Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 in the Marvin Center, Suite 242, to establish eligibility and to coordinate reasonable accommodations. For additional information please refer to: http://gwired.gwu.edu/dss/

UNIVERSITY COUNSELING CENTER (UCC) 202-994-5300
The University Counseling Center (UCC) offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include:
- crisis and emergency mental health consultations
- confidential assessment, counseling services (individual and small group), and referrals
  http://gwired.gwu.edu/counsel/CounselingServices/AcademicSupportServices

SECURITY
In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.