MTH 5195 Course Syllabus
2020 / FALL
Predictive Modeling

Instructor: Jacob Turner, Ph.D.
Department: Mathematics and Statistics
Email: turnerja2@sfasu.edu
Phone: 936-468-1692
Office: 342 NM
Office Hours:

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:30pm-4:30pm</td>
<td>10:00am-11:00am</td>
<td>3:30pm-4:30pm</td>
<td>10:00am-11:00am</td>
<td></td>
</tr>
</tbody>
</table>

Class meeting time and place: TBD. Zoom meetings

Text and Materials:

<table>
<thead>
<tr>
<th>Introduction to Statistical Learning</th>
<th>James, Witten, Hastie, &amp; Tibshirani</th>
<th>Free Online</th>
<th>Springer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elements of Statistical Learning</td>
<td>Hastie, Tibshirani, Friedman</td>
<td>Free Online</td>
<td>Springer</td>
</tr>
<tr>
<td>R statistical software</td>
<td></td>
<td>Free Download Online</td>
<td></td>
</tr>
</tbody>
</table>

Course Details:

Course Objectives:
The majority of the course will consist of a broad overview of various statistical methods with an emphasis in predictive modeling and R programming. Topics will include statistical learning concepts and terminology, classification, advanced feature selection methods such as regularization, bootstrap and cross validation procedures, tree algorithms, unsupervised learning. With time permitting, we will use some of the methods covered to highlight Bayes Theorem and its usefulness in predictive models as well the general Bayesian paradigm for building models. Regardless of time constraint a great emphasis will be placed on how these algorithms are conducted in practice, rules of thumb, and their own advantages and disadvantages.

Course outline (Approximate time spent)

Course Outline: Approximate time spent

- What is statistical learning?
  - Exploratory Analysis 15%
  - Key Definitions
  - Need for Validation
  - The Bias/Variance Trade Off
- Multiple Linear Regression 15%
  - Estimation and assumptions
- Pitfalls and Shortcomings
- K-Nearest Neighbors
- GAMs / Splines

- Classification 25%
  - Logistic Regression
  - Linear Discriminate Analysis
  - Support Vectors
  - Intro to ROC Curves

- Resampling Methods 15%
  - Cross Validation
  - Bootstrap
  - Feature Selection Using LASSO

- Tree Based Methods 15%
  - CART models
  - CART vs Regression
  - Bagging and Random Forrest

- Unsupervised Learning and Data Visualization 15%
  - Clustering methods
  - Intro to PCA
  - Visualizations for Unsupervised learning

**Grading:** The final average will be computed using the following weights:

- **Homework and additional to dos** 30%
- **1-2 data analysis projects** 40%
- **Cheat Sheet assignments (Serves as Final)** 30%

**Homework**

Homework assignments will be given out of the text and potentially some standalone work that I will provide. The point of these will be to reinforce concepts and get acquainted with key functions in R to effectively implement the methods under discussion.

**Cheat Sheet Assignment**

During the first few weeks of class, I will distribute a “cheat sheet” word document template. For each method discussed, the student will fill out various information about the method in a structured way. This will provide a mechanism for the student to organize their thoughts on each method and also provide them with a document that they can take with them in the real world and reference it if they find themselves needing a refresher. This will serve as the final for the course. Grading will be based on correctness and conciseness of explanations and summaries.

**Projects:**

There will be one or two projects during the course. These will be larger data sets that I will provide that will give students the opportunity to gain experience with real analysis workflows from data cleaning, feature engineering, application and comparison of model builds.
Grading Scale:

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>% Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90-100</td>
</tr>
<tr>
<td>B</td>
<td>80-89</td>
</tr>
<tr>
<td>C</td>
<td>70-79</td>
</tr>
<tr>
<td>D</td>
<td>60-69</td>
</tr>
<tr>
<td>F</td>
<td>below 59</td>
</tr>
</tbody>
</table>

Expectations and Class Rules:
ALL cell phones are to be turned OFF and put away. If someone violates this rule (texting, playing games, etc.), then the person must leave the class room for that class period without any further delay according to the professor’s instructions. No food in the classroom. Don’t leave the classroom in the middle of the lecture.

Attendance Policy:
Be on time. There is no need for an attendance policy in advanced mathematics courses. They are self-regulating.

Formalities Which Must Be Included in This Document:

Academic Integrity (A-9.1)
The penalty for violating Academic Integrity policy A-9.1 at any time during this semester is failure of the course. No exceptions. No grade will be calculated for a student who violates the policy. They will be asked to sign an academic dishonesty form after evidence of their violation has been provided to them and they will receive an F in the course no matter what grades have been accumulated to that point in the semester.

Academic integrity is a responsibility of all university faculty and students. Faculty members promote academic integrity in multiple ways including instruction on the components of academic honesty, as well as abiding by university policy on penalties for cheating and plagiarism.

Definition of Academic Dishonesty
Academic dishonesty includes both cheating and plagiarism. Cheating includes but is not limited to (1) using or attempting to use unauthorized materials to aid in achieving a better grade on a component of a class; (2) the falsification or invention of any information, including citations, on an assigned exercise; and/or (3) helping or attempting to help another in an act of cheating or plagiarism. Plagiarism is presenting the words or ideas of another person as if they were your own. Examples of plagiarism are (1) submitting an assignment as if it were one's own work when, in fact, it is at least partly the work of another; (2) submitting a work that has been purchased or otherwise obtained from an Internet source or another source; and (3) incorporating the words or ideas of an author into one's paper without giving the author due credit.

Please read the complete policy at http://www.sfasu.edu/policies/academic_integrity.asp

Withheld Grades Semester Grades Policy (A-54)
Ordinarily, at the discretion of the instructor of record and with the approval of the academic
chair/director, a grade of WH will be assigned only if the student cannot complete the course work because of unavoidable circumstances. Students must complete the work within one calendar year from the end of the semester in which they receive a WH, or the grade automatically becomes an F. If students register for the same course in future terms the WH will automatically become an F and will be counted as a repeated course for the purpose of computing the grade point average.

The circumstances precipitating the request must have occurred after the last day in which a student could withdraw from a course. Students requesting a WH must be passing the course with a minimum projected grade of C.

Students with Disabilities

To obtain disability related accommodations, alternate formats and/or auxiliary aids, students with disabilities must contact the Office of Disability Services (ODS), Human Services Building, and Room 325, 468-3004 / 468-1004 (TDD) as early as possible in the semester. Once verified, ODS will notify the course instructor and outline the accommodation and/or auxiliary aids to be provided. Failure to request services in a timely manner may delay your accommodations. For additional information, go to http://www.sfasu.edu/disabilityservices/