Course Information

Course Number/Section: CS/SE 6301.004
Course Title: R for Data Scientists
Term: Spring 2018
Days & Times: MW 1:00 – 2:45pm
Location: ECSS 2.410

Professor Contact Information

Instructor: Dr Bill Semper
E-Mail: WJS130130@utdallas.edu
Office hours: TTHR 2:30 - 4:00 pm, ECSS 4.602
Phone: 972-883-4139
Website: www.utdallas.edu/~wjs130130/

Course Pre-requisites, Co-requisites, and/or Other Restrictions:
Prerequisites: CS/SE/STAT 6313.

Course Description

This course will provide hands on experience with popular computational methods in data science: Linear regression, classification methods (linear/quadratic discriminant analysis, k-nearest neighbor), resampling methods (cross-validation, bootstrap), shrinkage methods (lasso, ridge regression, principal component analysis), nonlinear models (splines), tree-based methods, support vector machines, unsupervised learning (PCA and clustering), time series analysis, recommendation engines, text classification. These techniques will be explored using the R programming language.

Student Learning Objectives/Outcomes

Students will learn to apply the basic techniques of machine learning and gain experience applying various modeling techniques to data sets using the R language.
Textbooks and Materials:

The course does not use a textbook, however all materials will be provided in class, including lecture slides (with links to online resources), R scripts, and datasets.

Academic Calendar:

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 8</td>
<td>Classes Begin</td>
</tr>
<tr>
<td>April 25</td>
<td>Classes End</td>
</tr>
</tbody>
</table>

Grading Policy

Weekly in-class Quizzes/Projects 50%
Final Project 25%
Midterm Project 25%

Please note: No extra credit assignments will be given.

Grading: The base grading scale is given below:

$92 \leq score \leq 100 \quad A \\ 90 \leq score < 92 \quad A - \\ 88 \leq score < 90 \quad B + \\ 82 \leq score < 88 \quad B \\ 80 \leq score < 82 \quad B - \\ score < 80 \quad C$

The Curve: Once all final grades have been calculated using the Policy mentioned above, the A/B cutoff may be adjusted. The A/B cutoff is taken to be approximately the overall mean for the course. Note that if this cutoff is above the base scale cutoff, the base scale cutoff will be used. The cutoff for a B+ is approximately one standard deviation below the mean. The instructor reserves the right to assign B- grades to students with final scores more than 2 standard deviations below the mean score, and to assign C grades to students who are below the B cutoff and score less than 50% on the final project.

Midterm Grade: A midterm grade will be posted after the midterm exam, and will give an accurate assessment of your current grade at that point.

Grade Disputes: If you have a question about the grading of a homework assignment or project, you may submit a request for a re-grade within 48 hours of the time the grade was posted. Any request after this time period may be ignored.

Final Grades: Final course grades will be posted after the final project has been graded. Grades will first be posted to eLearning for you to review, and after 24 hours they will be posted to Orion.

For detailed information about University policies and procedures related to this syllabus, please refer to http://go.utdallas.edu/syllabus-policies.