CS/STAT 6301.0U2 – R for Data Scientists

Course Information

Course Number/Section: CS/SE 6301.0U2
Course Title: R for Data Scientists
Term: Summer 2017
Days & Times: F 1:00 – 5:00pm
Location: ECSS 2.410

Professor Contact Information

Instructor: Dr Bill Semper
E-Mail: WJS130130@utdallas.edu
Office hours: MW 3:00 - 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 the basics of statistical 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>June 2</td>
<td>Classes Begin</td>
</tr>
<tr>
<td>August 4</td>
<td>Classes End</td>
</tr>
</tbody>
</table>

Grading Policy

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

Please note: No extra credit assignments will be given.

Grading will be on a curve, with the median point for the class being used as the B+/A- cutoff point. Cutoffs for A and B+ grades will be roughly 2% from the mean. If the mean score is above 90%, the following scale is used:

\[
\begin{align*}
92 \leq score & \leq 100 \quad A \\
90 \leq score & < 92 \quad A - \\
88 \leq score & < 90 \quad B + \\
82 \leq score & < 88 \quad B \\
80 \leq score & < 82 \quad B - \\
\text{score} & < 80 \quad C
\end{align*}
\]

The instructor reserves to the right to assign B- grades to students with final scores less than 2 standard deviations below the median score. The instructor reserves the right to assign C grades to students who score less than 50% on the final.

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