Pre-requisites: A first course in Probability and Statistics, and matrix algebra. Knowledge of a computer programming language is desirable.

Course Objectives: The course provides an overview of data analytics methods and their application to engineering and decision problems. The tools and methods are classified as supervised and unsupervised statistical learning. The learning process starts analyzing past history and/or discovering relationships among the variables of interest. Most of the methods to be reviewed are drawn from the Statistics and Computer Science literature. To apply these methods some computational tool is needed. In this course R will be extensively used.

Course Details
- 40% course is made of computing sessions
- Datasets are expected to be multivariate and high dimensional
- A laptop is required during class sessions and all exams.

References
An Introduction to Statistical Learning with Applications in R. James, Witten, Hastie, and, Tibshirani. Springer, 2015.
The Elements of Statistical Learning, Hastie, Tibshirani, Friedman. Springer, 2013.

Software
The R language and environment for statistical computing and graphics is the main computational tool. Many libraries (known as R packages) are expected to be used in this course. A WiFi connection is required to download and install them. In addition R studio is usually the most preferred interface but not required. CSV files are sometimes the file type from which the data is loaded on to R. Students will use their own laptop during exams and class sessions.

Grading Policy: attendance 10%, assignments 30%, midterm 25%, final exam 35%.

Academic Integrity. The Viterbi School of Engineering adheres to the University's policies and procedures governing academic integrity as described in SCampus (www.usc.edu/dept/publications/SCAMPUS/). Students are expected to be aware of and to observe the academic integrity standards described in SCampus, and to expect those standards to be enforced in this course.

Students with Disabilities. Any Student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301and is open 8:30 a.m. - 5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.
<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Exams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug 24</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Definitions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Supervised vs Unsupervised learning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Regression vs Classification - examples</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- inference vs prediction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- R introduction &amp; examples</td>
<td></td>
</tr>
<tr>
<td>Aug 31</td>
<td>Linear Regression I</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- normal equations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- outliers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- categorical variables</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- interaction terms</td>
<td></td>
</tr>
<tr>
<td>Sep  7</td>
<td>Linear Regression II</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- model accuracy (bias vs MSE, AIC, BIC)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- cross validation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- variable selection</td>
<td></td>
</tr>
<tr>
<td>Sep 14</td>
<td>Classification - Logistic Regression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- logistic model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- estimating parameters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- making predictions</td>
<td></td>
</tr>
<tr>
<td>Sep 21</td>
<td>Classification - Other Regression models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Poisson Regression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Negative Binomial Regression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Multinomial Regression</td>
<td></td>
</tr>
<tr>
<td>Sep 28</td>
<td>Shrinkage methods</td>
<td></td>
</tr>
<tr>
<td>Oct  5</td>
<td>- Ridge regression</td>
<td>Oct 12</td>
</tr>
<tr>
<td></td>
<td>- Lasso regression</td>
<td></td>
</tr>
<tr>
<td>Midterm Exam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct 19</td>
<td>Neural Networks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Feed Forward and Feedback Networks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Multilayer and Deep learning Neural Networks</td>
<td></td>
</tr>
<tr>
<td>Oct 26</td>
<td>Generalized Additive models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- GAM for regression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- GAM for classification</td>
<td></td>
</tr>
<tr>
<td>Nov  2</td>
<td>Tree methods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Regression trees</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Classification trees</td>
<td></td>
</tr>
<tr>
<td>Nov  9</td>
<td>Bagging and boosting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Bagging and Bootstrap method</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Random forest</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Boosting</td>
<td></td>
</tr>
<tr>
<td>Nov 16</td>
<td>Support Vector Machines</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Max Margin classifier</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- SV classifier</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Computing the SVM</td>
<td></td>
</tr>
<tr>
<td>Nov 30</td>
<td>Unsupervised Learning - PCA</td>
<td></td>
</tr>
<tr>
<td>Dec  7</td>
<td>Unsupervised Learning - Clustering methods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- K-means clustering</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Hierarchical clustering</td>
<td></td>
</tr>
<tr>
<td>FINAL EXAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec 14</td>
<td>FINAL EXAM</td>
<td>Dec 14 (2 pm)</td>
</tr>
</tbody>
</table>