Catalog Description


Learning Objectives and Outcomes

- Develop an advanced level or proficiency with the primary classes of predictive modeling used by data scientists.
- Develop skills in using the Python programming environment and the primary packages and tools currently used by data scientists.
- Understand key concepts for measuring the performance of analytical models and key techniques for enhancing their performance.

Prerequisite(s): An undergraduate course on statistics.

Recommended Preparation: ISE 225 (Engineering Statistics I) or equivalent, working knowledge of a programming language.

Course Notes
The course material is available on Blackboard.

Technological Proficiency and Hardware/Software Required
The course makes extensive use of the Python programming language and several of its key data science packages. These are all open source and can be downloaded by the student for no cost.
Textbooks

There are no mandatory texts for this class. It is intended that the PowerPoint materials presented in class cover all of the content required. However, students are encouraged to consult the texts that the material is based on for clarification and elaboration.

The theoretical material in the course is drawn from the following texts:
- James, et. al., *An Introduction to Statistical Learning with Applications in R*, Springer, 2017 (ISLR)
- Scmueli, et. al., *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*, Wiley, 2017 (DMBA)

In addition, the following texts will be used as references for Python programming:
- Muller, *Introduction to Machine Learning with Python*, O’Reilly, 2017 (MLP)

Description and Assessment of Assignments

- **Eleven homework assignments (one per module)** - 50% of final grade
- **Mid-term exam** – 20% of final grade (covering Modules 1 – 5)
- **Final exam** - 30% of final grade

Grading Scale (Course final grades will be determined using the following scale)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>95-100</td>
</tr>
<tr>
<td>A-</td>
<td>90-94</td>
</tr>
<tr>
<td>B+</td>
<td>87-89</td>
</tr>
<tr>
<td>B</td>
<td>83-86</td>
</tr>
<tr>
<td>B-</td>
<td>80-82</td>
</tr>
<tr>
<td>C</td>
<td>77-79</td>
</tr>
<tr>
<td>C+</td>
<td>73-76</td>
</tr>
<tr>
<td>D</td>
<td>67-69</td>
</tr>
<tr>
<td>D+</td>
<td>63-66</td>
</tr>
<tr>
<td>D-</td>
<td>60-62</td>
</tr>
<tr>
<td>F</td>
<td>59 and below</td>
</tr>
</tbody>
</table>

Class Participation: Up to 2 additional points on the overall course grade may be awarded to students based on active class participation.

Assignment Submission Policy
Assignments will all be prepared and submitted using R Markdown unless otherwise directed. They should be submitted via backboard by the due date. Email submissions and late submissions are not allowed.

Timeline and Rules for submission
Assignments are to be returned the week after submission. Solutions will be released soon after the homework submission date.
## Course Schedule: A Weekly Breakdown

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topics/Daily Activities</th>
<th>Assignments</th>
<th>References</th>
</tr>
</thead>
</table>
| 1    | 8/26       | Module 1: Introduction to Predictive Analytics  
Introduction to Python, Jupyter Notebook | Module 1 HW Assigned  | ISLR, Chapter 1  
DMBA, Chapter 2 |
| 2    | 9/2        | Module 2: Data Preparation and Visualization  
Data integration, cleaning, reduction, enhancement  
*Tools: NumPy, Pandas* | Module 1 HW Due | DMBA, Chapters 3, 4 |
| 3    | 9/9        | Univariate/bivariate analysis, data quality assessment  
*Tools: Pandas datareader, Matplotlib* | Module 2 HW Assigned |               |
| 4    | 9/16       | Module 3: Modeling Introduction.  
Statistical learning, modeling types, model assessment and selection | Module 2 HW Due  
Module 3 HW Assigned | ISLR, Chapter 2  
DMBA, Chapter 5 |
| 5    | 9/23       | Module 4: Linear Methods for Regression  
*Tools: sklearn, statsmodels* | Module 3 HW Due  
Module 4 HW Assigned | ISLR, Chapter 3  
DMBA, Chapter 6 |
| 6    | 9/30       | Module 5: Linear Methods for Classification  
Logistic regression, linear discriminant analysis, and tree-based methods | Module 4 HW Due  
Module 5 HW Assigned | ISLR, Chapter 4  
DMBA, Chapter 8, 10 |
| 7    | 10/7       | Module 6: Resampling Methods  
Mid-Term (90 minutes) | Module 5 HW Due | ISLR, Chapter 5 |
| 8    | 10/21      | Module 7: Linear Model Selection and Regularization | Module 6/7 HW Assigned | ISLR, Chapter 6 |
| 9    | 10/28      | Module 8: Moving Beyond Linearity  
Generalized additive models, generalized linear models, nonparametric logistic regressions | Module 6/7 HW Due  
Module 8 HW Assigned | ISLR, Chapter 7 |
| 10   | 11/4       | Module 9: Tree-Based Methods  
Decision trees, forests, gradient boosting | Module 8 HW Due  
Module 9 HW Assigned | ISLR Chapter 8  
DMBA, Chapter 9 |
| 11   | 11/11      | Module 10: Support Vector Machines | Module 9 HW Due  
Module 10 HW Assigned | ISLR Chapter 9  
DMBA, Chapter 11 |
| 12   | 11/18      | Module 11: Neural Networks | Module 10 HW Due  
Module 11 HW Assigned | ISLR Chapter 9  
DMBA, Chapter 11 |
| 13   | 12/2       | Final Exam Review and Course Wrapup | Module 11 HW Due |               |
| Final| 12/9       | Final Exam (7:00PM – 9:00PM)          |                      |                |
Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the Office of Equity and Diversity http://equity.usc.edu or to the Department of Public Safety http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. The Center for Women and Men http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage http://sarc.usc.edu describes reporting options and other resources.

Support Systems:

Student Health Counseling Services - (213) 740-7711 – 24/7 on call engemannshc.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 – 24/7 on call engemannshc.usc.edu/rsvp
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) | Title IX - (213) 740-5086
equity.usc.edu, titleix.usc.edu
Information about how to get help or help a survivor of harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following protected characteristics: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations.

Bias Assessment Response and Support - (213) 740-2421
studentaffairs.usc.edu/bias-assessment-response-support
Avenue to report incidents of bias, hate crimes, and microaggressions for appropriate investigation and response.
The Office of Disability Services and Programs - (213) 740-0776
dsp.usc.edu
Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Support and Advocacy - (213) 821-4710
studentaffairs.usc.edu/ssa
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101
diversity.usc.edu
Information on events, programs and training, the Provost’s Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu, emergency.usc.edu
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu
Non-emergency assistance or information.