ISE 529 Predictive Analytics
Summer 2021 MTh 4:00 – 6:00 p.m.
Location: online

Instructor: Cesar Acosta-Mejia
Office: online
Office Hours: TBD
Contact Info: acostame@usc.edu

Teaching Assistant: Jieying Hu
Office: on-line
Office Hours: TBD
Contact Info: hujieyin@usc.edu

IT Help:
Hours of Service:
Contact Info:

Catalog description

Analytics for supervised and unsupervised statistical learning. Generalized linear models, discriminant analysis, support vector machines. Nonparametric classification, trees, ensemble methods, k-nearest neighbors. Principal components, clustering.

Course Description

This course focus on building models for prediction, classification, and clustering. For the first two cases the objective is to predict a numeric value or a category. For clustering the objective is to group observations in clusters that share some common attributes.

The standard multiple linear regression model is the basic prediction model. This model is extended to shrinkage models (ridge and lasso regression) for improved accuracy and dimension reduction. Overfitting, bias, cross validation, and AIC are reviewed to help evaluate the performance of these models.

The course also focuses on regression models for a categorical response. Trees and ensembled trees (random forests, bagging, and boosting), discriminant analysis, and support vector machines. For these models the prediction is a category.

Clustering focuses on discovering relationships between variables when a response is not available. Principal components analysis and clustering (K-means clustering and hierarchical clustering) are reviewed.

Learning Objectives and Outcomes

- To understand the Data Analytics levels: Descriptive, Predictive, and Prescriptive Analytics.
- To understand the difference between supervised and unsupervised learning methods.
- To learn the most common data aggregation operations (cross tabulation and pivot tables).
- To build models for prediction and classification.
- To understand key concepts for predictive analytics (overfitting, shrinkage, regularization, R^2, adjusted R^2, VIF, mean square prediction error, Cross-validation).
- To estimate the performance of Analytics models.
- To compare the performance of different prediction and classification models.
- To build models to classify observations into two or more classes (categories).

Prerequisite(s): An undergraduate course on Statistics and knowledge of a programming language (any).

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
Required software: Python programming language is used throughout the course. Jupyter Notebook is used as the main interface for documenting the scripts and results.

Supplementary Materials (for reference)

- Heydt M., *Learning Pandas, Packt,* 2017, ISBN 978-1-78712-313-7 (LP)
- VanderPlas, *Python Data Science Handbook*, O'Reilly, 2017 (PDS)
- McKinney, *Python for Data Analysis, 2nd Ed.*, O'Reilly, 2018   (DSA)
- Muller, *Introduction to Machine Learning with Python,* O'Reilly, 2017 (MLP)

Description and Assessment of Assignments

All assignments and examinations are on-line. Unless otherwise noted the assignments are individual. All homework assignments are released and distributed one-week in advanced. Dates are shown in the Course schedule on page 4. Submit on to Blackboard by the due date. No late homework will be accepted.

Grading Policy

Assignment	Points	% of Grade
Homework	100 each (6 homework assignments)	30
Midterm	100	30
Final	100	40
TOTAL		100

Assignment Submission Policy
Assignments should be typewritten and clean. Email submissions and late submissions are not allowed. No make-up exams are considered.

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

	Date	Topics/Daily Activities	Deliverables	slides	Files
1	July 1	**Introduction** to Analytics. Python and Jupyter Notebook (JN) setup. Python review, Numpy library.	HW1 MS Excel	overview.ppt analytics.ppt python.ppt	Intro.ipynb numpy.ipynb
2	July 5	Independence Day			
3	July 8	Pandas library, data structures. Most Common Data Operations. Pivot tables and cross tabulation.	HW1 due HW2 Pandas	Pandas .ppt	Example3.ipynb
4	July 12	Financial Analytics. Data Visualization Web scraping with the pandas-datareader library	HW2 due HW3 Financial Analytics	finance .ppt fanalytics.ppt	Project3.ipynb
5	July 15	**Linear Regression.** OLS vs. linear regression. Libraries sklearn and statsmodels	HW3 due	slr.ppt mlr.ppt	slr2.ipynb finished3.ipynb
7	July 19	**Linear Regression.** Categorical variables. Interaction terms. Label encoding and one-hot encoding	Midterm released	categoricals.ppt	plots2.ipynb part2c.ipynb homes_sk.ipynb example1b.ipynb example2.ipynb
8	July 22	**Overfitting and Cross validation**, Training/test sets, mean square prediction error (MSPE).	Midterm due	cv2.ppt	Polynomial4.ipynb feature-cv3.ipynb
9	July 26	**Classification Problems.** Logistic Regression, KNN.		classification2.ppt logistic2.ppt	cancerlogistic.ipynb iris2.ipynb
10	July 29	**Shrinkage Methods and Regularization.** Ridge regression and the LASSO.		rr2.ppt	ridge5.ipynb
12	Aug 2	**Classification and Regression Trees (CART) -** Examples	HW5 Regularization	trees2.ppt categ.ppt	regression3.ipynb cart3.ipynb
13	Aug 5	**Ensemble Methods**. Random Forest, Bagging, and Boosting.	HW5 due	ensembles2.ppt	ensemblereg.ipynb ensembcancer2.ipynb polyboosting3.ipynb
	Aug 9	**Final Exam**			

Date and time of the final for this class shown in the USC *Schedule of Classes* at classes.usc.edu/.

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.