ISE 529 Predictive Analytics
Fall 2019 F 3.30 – 6.30 p.m.
Location: KAP 156

Instructor: Cesar Acosta-Mejia
Office: GER 216
Office Hours: Th 4:30-6 pm
Contact Info: acostame@usc.edu

Teaching Assistant: TBA
Office: TBA
Office Hours: TBA
Contact Info: TBA

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. Generalized linear models are reviewed for a binomial, poisson, negative binomial, and multinomial response. Other methods for classification are also reviewed. 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 difference between a numeric variable and a factor
- To build regression 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 compare the performance of different prediction and classification models.
- To build models to classify observations into two or more classes
- To understand what a principal component is and how can it be used for clustering and prediction.
- To be able to explain the difference between different clustering methods

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
Required software: A programming language is required.

Required Textbook (available on Blackboard)

- VanderPlas, *A Whirlwind Tour of Python*, O'Reilly, 2016 (WTP)

Supplementary Materials (for reference)

- VanderPlas, *Python Data Science Handbook*, O'Reilly, 2017 (PDS)
- McKinney, *Python for Data Analysis, 2nd Ed.*, O'Reilly, 2018 (DSH)
- Muller, *Introduction to Machine Learning with Python,* O'Reilly, 2017 (MLP)

Description and Assessment of Assignments

- **Midterm I** will be in-class based on the schedule and 2 hours length.
- **Final Examination** a two-hour comprehensive exam scheduled by USC.
- **Homework** are assigned every other week. Submit the due date, during the class session. 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

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

A	95-100	B-	80-82	D+	67-69
A-	90-94	C+	77-79	D	63-66
B+	87-89	C	73-76	D-	60-62
B	83-86	C-	70-72	F	59 and below

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

Week	Date	Topics/Daily Activities	Deliverables	Reference
1	**Aug 30**	**Introduction to Analytics** Descriptive, Predictive and Prescriptive Analytics. **Python for Analytics and Machine Learning.**	**None**	WTP p31-36, p84-89.
2	**Sep 6**	**Data Types** Structured, Semi-structured, Unstructured. Pandas library and dataframes. **Data Visualization.** Matplotlib, Seaborn libraries	**HW1**	PDA Ch9
3	**Sep 13**	**Linear Regression I.** OLS vs. linear regression. Targets (labels) and features in dataframes. Categorical variables and preprocessing.		MLP p134, p233.
4	**Sep 20**	**Linear Regression II.** Overfitting, bias, MSE, AIC. Cross validation. Training, test sets, MSPE.	**HW2**	MLP p17, Ch5.1
5	**Sep 27**	**Classification. Logistic Regression vs KNN for classification. KNN for regression.**		MLP Ch2.3
6	**Oct 4**	**Classification. Discriminant Analysis.** Naïve Bayes.	**HW3**	Notes
7	**Oct 11**	**Shrinkage Methods and Regularization.** Ridge regression and the LASSO.		Notes
8	**Oct 16** 6.30 p.m.	**Midterm Exam**		
9	**Oct 25**	**Classification and Regression Trees (CART)** Examples		DSH p421
10	**Nov 1**	**Ensemble Methods**. Random Forest, XGBooting.	**HW4**	DSH p426
11	**Nov 8**	**Support Vector Machines.** Maximal Classifier, Support Vector Classifier, Support Vector Machine.		DSH p405
12	**Nov 15**	**Feature Engineering and pipelines.** How to automate workflows with Pipelines. Data preparation and feature extraction with pipelines.	**HW5**	MLP Ch4.7, Ch6
13	**Nov 22**	**Introduction to Neural Networks (NN).** Data representations for NN, tensors. Layers, Loss functions, optimizers. Introduction to Keras.		Notes
14	**Dec 6**	**Unsupervised Learning** Principal Components, K-Means clustering	**HW6**	DSH p433
15	**Dec 13**	**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.