

**ISE 529 Predictive Analytics**

Fall 2021 Th 4:00 – 6:30 p.m.

Location: GFS 101**Instructor:** Cesar Acosta-Mejia**Office:** GER 216**Office Hours:** W 5-6 p.m.**Contact Info:** acostame@usc.edu**Teaching Assistant:** Jieying Hu**Office:** TBD**Office Hours:** TBD**Contact Info:** hujieyin@usc.edu**IT Help:****Hours of Service:****Contact Info:****Course Description**

This course focus on building models for prediction and classification with the objective to predict a numeric value or a category. 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 used to evaluate the performance of these models.

The course also focuses on models for classification. Logistic regression, KNN, Trees and ensembles of trees (random forests, bagging, and gradient boosting), support vector machines, and artificial neural networks are reviewed. For these models the prediction is a category.

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)
- 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

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

A	94-100
A-	90-93.9
B+	87-89
B	83-86.9
B-	80-82.9
C+	77-79
C	73-76.9
C-	70-72.9
D+	67-69
D	63-66.9
D-	60-62.9
F	59.9 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

	Date	Topics/Daily Activities	Deliverables	slides	Files
1	Aug 26	Introduction to Analytics Descriptive, Predictive and Prescriptive 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	Sep 2	Pandas library, data structures. Most Common Data Operations. Pivot tables and cross tabulation.	HW1 due HW2 Pandas	Pandas .ppt	Example3.ipynb
3	Sep 9	Data Visualization and Web scraping with the pandas-datareader library.	HW2 due HW3 Financial Analytics	finance .ppt fanalytics.ppt	Project3.ipynb
4	Sep 16	Linear Regression. OLS vs. linear regression. Libraries sklearn and statsmodels	HW3 due	slr.ppt mlr.ppt	slr2.ipynb finished3.ipynb
5	Sep 23	Linear Regression. Categorical variables. Interaction terms. Label encoding and one-hot encoding	HW4 Regression with categorical variables	categoricals.ppt	plots2.ipynb part2c.ipynb homes_sk.ipynb example1b.ipynb example2.ipynb
6	Sep 30	Overfitting and Cross validation, Training/test sets, mean square prediction error (MSPE).	HW4 due	cv2.ppt	Polynomial4.ipynb feature-cv3.ipynb
7	Oct 7	Classification Problems. Logistic Regression vs KNN.	Midterm exam	classification2.ppt logistic2.ppt	cancerlogistic.ipynb iris2.ipynb
8	Oct 14	Fall recess			
9	Oct 21	Shrinkage Methods and Regularization. Ridge regression and the LASSO.		rr2.ppt	ridge5.ipynb
10	Oct 28	Classification and Regression Trees (CART) - Examples	HW5 Regularization	trees2.ppt categ.ppt	regression3.ipynb cart3.ipynb
11	Nov 4	Ensemble Methods. Random Forest, Bagging, and Boosting.	HW5 due	ensembles2.ppt	ensemblereg.ipynb ensembcancer2.ipynb polyboosting3.ipynb
12	Nov 11	Support Vector Machines. Maximal Classifier, Support Vector Classifier, Support Vector Machine.	HW6 Ensemble Methods	review.ppt svm.ppt	svm.ipynb function5.ipynb optdigits.ipynb
13	Nov 18	Introduction to Neural Networks (NN). Data representations for NN, tensors. Layers, Loss functions, optimizers.	HW6 due	nn3.ppt	perceptron4.ipynb multilayerp3. Ipynb
14	Nov 25	Thanksgiving Break			
15	Dec 2	NN Aplications with Library Keras. NN for binary classification. K-fold validation. NN for regression.			mnist. Ipynb
	Dec 9	Final Exam 4:30 p.m.			

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.