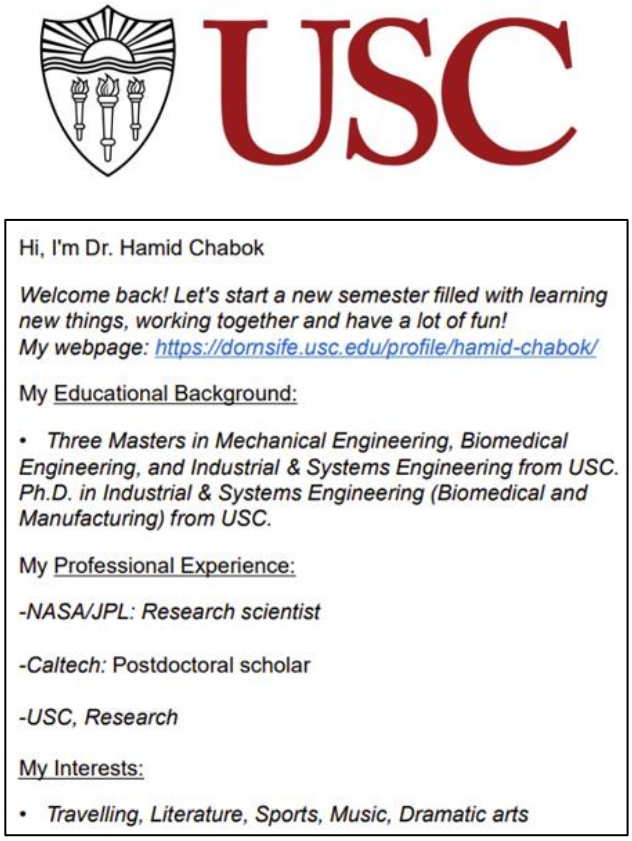

ISE 529 Predictive Analytics
Units: 4
Summer 2024 TTH 10:00 am – 2:10 pm
Location: GFS 207

Instructor: Hamid Chabok
Office: SHS 365
Office Hours: TBD
Contact Info: chabok@usc.edu

Teaching Assistant: TBD
Office: TBD
Office Hours: TBD
Contact Info: TBD

IT Help:
Hours of Service:
Contact Info:

Catalog Description

Supervised learning. Linear Regression, Cross Validation, Ridge and Lasso regression, Classification, Logistic regression, KNN, Decision Trees, Random Forest, Neural Networks.

Course Description

The course focuses on developing models for regression and classification. It begins with the standard multiple linear regression model and then extends to include shrinkage models such as ridge and lasso regression, which enhance accuracy and reduce dimensionality. Tree models and neural networks are also part of the curriculum.
To assess model performance, concepts like overfitting, bias, and cross-validation are utilized. Furthermore, the course covers classification models like logistic regression, KNN, and decision trees, along with an examination of their respective performance metrics.

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 learn how to apply cross validation to models with hyperparameters.
- To learn how to derive the loss function for shrinkage models.
- To compare the performance of different prediction and classification models.
- To build models to classify observations into two or more classes (categories).

Prerequisites: None

Recommended Preparation: An undergraduate course on Statistics, 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. Jupyter Notebook is used as the main interface for documenting the scripts and results.

Supplementary Materials (for reference)

- James G., *Introduction to Statistical Learning with python,2021* (ISLP)
- 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

Unless otherwise noted the assignments are individual. Dates are shown in the Course schedule. Submit on to Blackboard by the due date. No late homework is to be accepted.

Grading Policy

Assignment	Points	% of Grade
Homework	100 each (6 homework assignments)	60
Final Project	100	40
TOTAL	700	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.

Final Project
As part of the course, students undertake a regression analysis project.
The project involves:

- Applying the regression models taught during the semester (including multiple linear regression, ridge and etc.).
- Utilizing appropriate evaluation metrics (such as RMSE or R-squared) to assess model performance.
- Implementing cross-validation to validate the model's robustness.
- Properly handling train-test splits and performing necessary data preprocessing.

Course Schedule: A Weekly Breakdown

Week	Dates	Topics/Daily Activities	Deliverables	slides	Files
1	16-May	Introduction to Analytics Descriptive, Predictive and Prescriptive Analytics. Python and Jupyter Notebook setup.	-	overview.ppt analytics.ppt	Intro.ipynb HW1 Intro to Analytics
		Python data structures. Numpy library. Operations on Numpy arrays		python.ppt Numpy.ppt	Python.ipynb NumPy.ipynb
2	21-May	Pandas library. Data structures. Merging data frames. Pivot tables, Common Data Operations.	-	Pandas.ppt	Pandas.ipynb
	23-May	Data Visualization - part 1 (matplotlib)	HW1	matplot.ppt	Matplotlib.ipynb
		Data Visualization - part 2 (Seaborn)		seaborn.ppt	HW2 Visualization and prediction
3	28-May	Preprocessing	-	Preprocces.ppt	Preprocess.ipynb
		Simple Linear Regression. OLS, regression assumptions, prediction, confidence intervals. Performance measures.		SLR.ppt	SLR.ipynb
	30-May	Multi Linear Regression. Feature selection	HW2	MLR.ppt	MLR.ipynb
		Regression applications. Price Prediction.		PricePred.ppt	Example1.ipynb HW3 Price Prediction
4	4-Jun	Overfitting, Cross validation	-	CV.ppt	CV.ipnby
		Polynomial Regression		PolyReg.ppt	Polynomial.ipynb
	6-Jun	Regression trees	HW3 & Project Phase 1	tree1.ppt	tree.ipynb
		Ensemble of Regression Trees. Random Forest		RF.ppt	RF.ipynb HW4 Polynomial, Trees and RF
5	11-Jun	Classification Problems. Logistic Regression, KNN	-	classification.ppt	classification.ipynb
	13-Jun	Classification Trees.	HW4	tree2.ppt	tree1.ipynb
		Performance Measures for classification trees		Performance.ppt	Performance.ipynb HW5 Classification
6	18-Jun	Ridge and Lasso Regression	-	ridge.ppt lasso.ppt	ridge.ipynb lasso.ipynb
	20-Jun	Introduction to Neural Networks (NN), Layers, Loss functions, optimizers.	HW5	NN.ppt	NN.ipynb HW6 NN, Ridge, lasso
7	25-Jun	Final Review	HW6	Review.ppt	-
	26-Jun	Final Project Due	Final Project Report.rar	-	-

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.

Sure, here's the weekly breakdown with dates included, starting from Wednesday, May 15, 2024:

Week #	Dates	Topics/Daily Activities	Deliverables	Slides	Files
1	May 15 - May 21	Introduction to Analytics: Descriptive, Predictive, and Prescriptive Analytics. Python and Jupyter Notebook (JN) setup. Python data structures. Numpy library. Operations on numpy arrays (indexing, subarrays, shaping, ufuncs, aggregation, counting). Pandas library. Data structures. Merging data frames. Most Common Data Operations. Pandas library. Pivot tables and cross-tabulation.	HW1 MS Excel, Pandas	overview.ppt, analytics.ppt, python.ppt, Pandas .ppt	Intro.ipynb, numpy.ipynb, Example3.ipynb
2	May 22 - May 28	Data Visualization library matplotlib. Web scraping with the pandas-datareader library. Data Visualization with pandas. Linear Regression: OLS, regression assumptions, prediction, confidence and prediction intervals, ANOVA decomposition. Performance measures. Linear Regression examples with libraries sklearn and statsmodels.	HW1 due HW2 Financial Analytics	matplot.ppt, finance.ppt, fanalytics.ppt, slr.ppt, mlr.ppt	matplot.ipynb, Project3.ipynb, slr2.ipynb, finished3.ipynb
3	May 29 - June 4	Linear Regression with categorical variables. Label encoding and one-hot encoding. Interaction terms. Linear Regression applications: Time Series forecasting. Overfitting. Cross-validation strategies. Training/test sets, mean square prediction error (MSPE). Linear Regression applications: Polynomial regression, feature selection, scaling the data.	HW3 MLR and categorical variables	categoricals.ppt, cv2.ppt	plots2.ipynb, part2c.ipynb, homes_sk.ipynb, example1b.ipynb, example2.ipynb, Polynomial4.ipynb, feature-cv3.ipynb
4	June 5 - June 11	Classification Problems: Logistic Regression, K-nearest neighbor (KNN), hyperparameter search. Regularization and Overfitting: Ridge regression and the LASSO, hyperparameter tuning.	HW4 Regularization HW3 due	classification2.ppt, logistic2.ppt, knn.ppt, rr2.ppt	cancerlogistic.ipynb, iris2.ipynb, ridge5.ipynb
5	June 12 - June 18	Trees based Methods: Predictors Space strategy, Tree pruning, Feature Selection. Regression trees, Classification Trees, Performance Measures for classification trees (gini index, cross-entropy). Ensemble of Regression Trees: Random Forest, Bagging, and Gradient Boosting. Ensemble of Classification Trees: Applications and Examples.	HW4 Regularization HW3 due	trees2.ppt, categ.ppt, ensembles2.ppt	regression3.ipynb, cart3.ipynb, ensemblereg.ipynb, ensembcancer2.ipynb, polyboosting3.ipynb
6	June 19 - June 25	Support Vector Machines: Maximal Classifier, Support Vector Classifier, Support Vector Machine. Support Vector Machines for regression. Introduction to Neural Networks (NN): Data representations for NN, tensors. Layers, Loss functions, optimizers. NN Applications: Library Keras, NN for binary classification, K-fold cross-validation, NN for regression.	HW5 Ensemble Methods HW4 due	review.ppt, svm.ppt, svmreg.ppt, nn3.ppt	svm.ipynb, function5.ipynb, optdigits.ipynb, perceptron4.ipynb, multilayerp3.ipynb, mnist.ipynb
7	June 26 - July 2	Final Review			
8	July 3	Final Exam			

This breakdown includes the dates for each week, starting from May 15, 2024, and ending with the final exam on July 3, 2024. Adjustments can be made based on specific requirements or constraints.