F 3:30 – 6:20 p.m. Room KAP 134

Profesor : Cesar Acosta, Ph.D.
Office : GER 216
Office hours: Th 4:00pm-5:00pm
e-mail : acostame@usc.edu

Teaching Assistant: Michael Hintlian
Office : GER 309
Office hours : TBD
e-mail : hintlian@usc.edu

Pre-requisites: The DSO 529 pre-requisite does not apply. This course has no course prerequisite beyond a first course in probability and statistics, and matrix algebra.

Course Objectives: The course provides an overview of data analytics methods and their application to engineering and decision problems. The tools and methods are classified as supervised and unsupervised statistical learning. The learning process starts analyzing past history and/or discovering relationships among the variables of interest. Most of the methods to be reviewed are drawn from the Statistics and Computer Science literature. To apply these methods some computational tool is needed. In this course *R* will be extensively used.

Course Details

- 40% course is made of computing sessions
- Datasets are expected to be multivariate and high dimensional
- A laptop is required during class sessions
- No textbook is required. References will be the choice of the student.

References

An Introduction to Statistical Learning with Applications in R. James, Witten, Hastie, and, Tibshirani. Springer, 2015.
The Elements of Statistical learning, Hastie, Tibshirani, Friedman. Springer, 2013.
Applied linear Statistical Models 5th ed., Kutner, Nachtsheim, Neter, Li. McGraw Hill, 2005.
Applie Multivariate Statistical Analysis, 6th ed., Johnson and Wichern, 2012, PHI.
An Introduction to R, Venables, Smith. Network Theory Ltd., 2004.

Software

The *R language and environment for statistical computing and graphics* is the main computational tool. Many libraries (known as *R* packages) are expected to be used in this course. A WiFi connection is required to download and install them
In addition *R studio* is usually the most preferred interface but not required.
CSV files are sometimes the file type from which the data is loaded on to *R*.
Students will use their own laptop during exams and class sessions.

Grading Policy: participation 10%, assignments 30%, midterm 25%, final exam 35%.

Academic Integrity. The Viterbi School of Engineering adheres to the University's policies and procedures governing academic integrity as described in SCampus (www.usc.edu/dept/publications/SCAMPUS/). Students are expected to be aware of and to observe the academic integrity standards described in SCampus, and to expect those standards to be enforced in this course.

Students with Disabilities. Any Student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301and is open 8:30 a.m. - 5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776

Week	Topic	Exams
Aug 25	Introduction	
	- Definitions	
	- Supervised vs Unsupervised learning	
	- Regression vs Classification – examples	
	- inference vs prediction	
	- R introduction & examples	
Sep 1	Linear Regression I	
	- normal equations	
	- outliers	
	- categorical variables	
	- interaction terms	
Sep 8	Linear Regression II	
	- model accuracy (bias vs MSE, AIC, BIC)	
	- cross validation	
	- variable selection	
Sep 15	Classification - Logistic Regression	
	- logistic model	
	- estimating parameters	
	- making predictions	
Sep 22	Classification - Linear discriminant analysis	
	- Bayes theorem for classification	
	- estimating the Bayes classifier	
	- quadratic discriminant analysis	
Sep 29	Linear Regression III	
Oct 6	- Subset Selection	
	- Shrinkage methods (Ridge regression, lasso)	
	- Dimension Reduction (PCR, PLS)	
	Midterm Exam	Oct 13
Oct 20	Nonlinear Regression	
	- Polynomial regression	
	- Regression Splines	
	- Smoothing Splines	
	- Local regression	
Oct 21	Generalized Additive models	
	- GAM for regression	
	- GAM for classification	
Oct 28	Tree methods	
	- Regression trees	
	- Classification trees	
Nov 4	Bagging and boosting	
	- Bagging and Bootstrap method	
	- Random forest	
	- Boosting	
Nov 11	Support Vector Machines	
	- Max Margin classifier	
	- SV classifier	
	- Computing the SVM	
Nov 18	Unsupervised Learning - PCA	
Dec 2	Unsupervised Learning – Clustering methods	
	- K-means clustering	
	- Hierarchical clustering	
	FINAL EXAM	Dec 11 (2 pm)