CS 686-03 – Machine Learning
Course Syllabus
Spring 2019

SUMMARY INFORMATION

David Guy Brizan
Office: Harney 440A
Office Hours: 10.30-12.00 Mon & 10.30-11.30 Weds
Office Phone: 415.422.4813
Email Address: `dgbrizan@usfca.edu`

Paul Intrevado
Office: McLaren 103
Office Hours: 16.45-17.45 Weds & 13.30-14.30 Fri
Office Phone: 415.422.2527
Email: `pintrevado@usfca.edu`

Class Location: 307 LoSchiavo
Class Time: Monday/Wednesday/Friday 14:50-16:40

ON COURSE GOALS. Any student who successfully completes this course should:

- **ON PYTHON**
  - Understand the strengths and weaknesses of using Python;
  - Be able to download, import and maintain Python packages;
  - Be comfortable with Git and manage her or his own GitHub account;
  - Use Python, Jupyter Notebooks, Markdown and LaTeXto write code, generate analysis, interpret results and produce reports for consumption by a general audience;
  - Understand Python data types;
  - Confidently create and manipulate basic Python data structures;
  - Be comfortable with using both the `NumPy` and `pandas` packages to create and manipulate both Series and Data Frame objects;
  - Be comfortable creating graphs in `ggplot2`; and
  - Be able to import `csv` data files.
- **ON STATISTICS**
  - summarizing categorical and quantitative data for one and two variables using tables and graphical displays;
  - measures of central tendency and variability;
  - measures of distribution shape, relative location, and outliers;
  - five-number summaries and box-plots;
  - measures of association between two variables;
  - the normal probability distribution;
  - the central limit theorem;
  - confidence intervals and hypothesis testing for the population mean; and
  - Bayes's Theorem and it's applications.

- **ON MACHINE LEARNING**
  - Understand the structural forms of classical simple and multiple linear regression models, as well as the assumptions underlying these models;
  - Formulate and test hypotheses, as well as use linear regression models for both prediction and explanation;
  - Be able to use Python and the `statsmodels` package to load and manipulate data, fit regression models, and generate various outputs such as ANOVA tables, confidence intervals for parameters, diagnostic assessments, etc.;
  - Be able to verify/test whether or not fitted residuals conform to the assumptions that underlie classical regression;
  - Understand and address multicollinearity;
  - Understand how to use indicator (or dummy) variables;
  - Understand the fundamental concepts and applications of binomial and multinomial logistic regression, ridge regression, LASSO, and elastic nets (time permitting);
  - Understand and interpret general criteria used to evaluate the quality of results produced by various machine learning techniques using `scikit-learn`;
  - Understand details of machine learning classification algorithms, such as k-Nearest Neighbor (kNN), Decision Trees, Boosting and Random Forests;
  - Understand how to generate and interpret output from classification techniques such as logistic regression, kNN and Random Forest machine learning algorithms;
  - Understand how to generate and interpret output from clustering techniques such as the k-means clustering algorithm;
  - Determine feature importance in a classification or regression model;
  - Be able to communicate the results of complete and well-reasoned analysis of data, employing machine learning techniques, both orally and in writing;
  - Create, train and evaluate deep neural networks (DNNs) for classification or regression using `pytorch`; and
  - Create, train and evaluate DNN-based applications, such as recommendation systems, autoencoders, etc.

ABOUT DAVID. My name is David Guy Brizan. Please call me David. I'm an Assistant Professor in the MS in the Computer Science department in the College of Arts & Sciences. I did my Ph.D. work in Computer Science, specifically in Natural Language Processing (NLP).

ABOUT PAUL. My name is Paul Intrevado. Please call me Paul. I'm an Assistant Professor of Data Science in the Department of Mathematics & Statistics, in the College of Arts & Sciences. I hold a Ph.D. in Operations Management.

ABOUT THIS CLASS. We will meet to discuss machine learning from Mondays, Wednesdays and Fridays, January 23^{rd}, 2019 through Thursday, May 8^{th}, 2019. The course will be serially co-taught by both professors: Paul will teach the first five weeks, and David will teach the remaining.

ON COMMUNICATION. All formal course material such as the course syllabus, course notes, and data sets will be posted online. All grades will be posted exclusively on Canvas. All electronic homework will be submitted via Canvas or Github, owing to the application. All other forms of **formal** communication with the instructors or TA will occur via Slack. You are required to check Slack daily, and are responsible for any clarifications, changes and/or updates.

The Slack channel for this class aims to facilitate general discussion of course material with your peers and instructors. Be aware that instructors will not always able to reply immediately to queries on Slack; Slack should therefore not be used for formal, urgent or time-sensitive issues. Although general discussion of course material with your peers on Slack is encouraged, sharing detailed information, including answers to homework problems, is considered cheating.

ON TEXTBOOKS. The material presented in the course is sourced from multiple locations. There is therefore no formal required course textbook, although relevant reference material will be presented in class.

ON USE OF SPECIFIC COMPUTER LANGUAGES. The mastery of Python (this includes all relevant packages), Jupyter Notebooks, Markdown and LaTeX are course objectives. **You are not permitted to use any other languages for this course when completing and submitting deliverables**.

ON ATTENDANCE. Formal attendance will not be taken, nor will it be required. You are all graduate students and are expected to mature enough to manage your time intelligently. If you miss lecture(s), you need not explain or excuse yourself formally to us. Our objective as course instructors is to ensure that you understand the material to be covered in this course. If you are already familiar with the material or choose to learn it on your own time, that is your prerogative.

ON GRADING. Part our job as instructors is to assign grades fairly and in a manner that reflects the high academic standards at the University of San Francisco and the Computer Science department. Your grade in this course will be computed according to the following weights:

Component	Weight
Labs	5%
Homework	50%
Data Challenge	15%
Final Project	30%
	100%

ON HOMEWORK. You must work on all homework assignments **individually** and **submit your own individualized write-up**. You may consult with external parties, including other students in the class, regarding homework, but each student should complete all parts of the assignments successfully without assistance. Significant differences between homework scores and test scores may be subject to investigation.

ON DELIVERABLES. All deliverables are subject to the following rules:

1. When making reference(s) to *summary* results, include all relevant output in text of the deliverable where it is being discussed, not in an appendix at the back of the deliverable.

2. Do not include a copy of the raw data unless there is a compelling reason.

3. Software can generate hundreds of graphs and statistical output extremely easily. Only include *relevant* graphs and output in the deliverable. All graphs and statistical output included in the deliverable should be referenced in the text of the deliverable. Everything should be orderly and easy for the grader to read.

4. All deliverables are required to be typed and all graphs and statistical output generated using course-mandated software. Deliverables with *any* handwritten elements will not be accepted and will receive a grade of zero. You are required to submit all homework in the course-mandated software.

5. **All code should be commented** in a neat, concise fashion, explaining the objective(s) of individual lines of code.

6. **We will not accept late deliverables under any circumstance.**

ON GRADES. There is no predefined letter grade distribution or quota to which we must adhere. The A grade range indicates distinguished performance and competence; the B grade range indicates strong to adequate performance and competence; the C grade range demonstrates weak understanding of the material. A grade of C- is the minimal passing grade. A grade of F is given for performance that insufficiently demonstrates academic competence.

ON ACADEMIC INTEGRITY. As a Jesuit institution committed to *cura personalis* – the care and education of the whole person – USF has an obligation to embody and foster the values of honesty and integrity. USF upholds the standards of honesty and integrity from all members of the academic community. All students are expected to know and adhere to the University's Honor Code. You can find the full text of the code online at `http://myusf.usfca.edu/academic-integrity/`. The policy covers:

- Plagiarism – intentionally or unintentionally representing the words or ideas of another person as your own; failure to properly cite references; manufacturing references.
- Working with another person when independent work is required.
- Submitting work written by another person or obtained from the internet.
- The penalties for violation of the policy may include a failing grade on the assignment, a failing grade in the course, and/or a referral to the Academic Integrity Committee.

ON DISABILITIES. If you are a student with a disability or disabling condition, or if you think you may have a disability, please contact USF Student Disability Services (SDS) at (415) 422-2613 within the first week of class, or immediately upon onset of disability, to speak with a disability specialist. If you are determined eligible for reasonable accommodations, please meet with your disability specialist so they can arrange to have your accommodation letter sent to me, and we will discuss your needs for this course. For more information, visit `http://www.usfca.edu/sds`.

ON LAPTOPS. Please bring a laptop to lecture and have the course-mandated software installed on it. You will be expected to use your laptops in a lecture setting for in-class examples and labs. I expect you to use your laptops judiciously, refraining from surfing the web or engaging in any other distracting behavior during lecture.

ON BEHAVIORAL EXPECTATIONS. All students are expected to behave in accordance with the Student Conduct Code and other University policies (see `http://www.usfca.edu/fogcutter/`). Students whose behavior is disruptive or who fail to comply with the instructor may be dismissed from the class for the remainder of the class period and may need to meet with the instructor or Dean prior to returning to the next class period. If necessary, referrals may also be made to the Student Conduct process for violations of the Student Conduct Code.

ON COUNSELING AND PSYCHOLOGICAL SERVICES (CAPS). CAPS provides confidential, free counseling to student members of our community.

For more information, see https://myusf.usfca.edu/student-health-safety/caps.

ON CONFIDENTIALITY, MANDATORY REPORTING, AND SEXUAL ASSAULT.
For information and resources regarding sexual misconduct or assault visit:

- The Title IX coordinator website (https://myusf.usfca.edu/title-ix)
- USF's Callisto website (https://usfca.callistocampus.org).