Last Updated August 14, 2016

San José State University
Computer Science Department
CS 185C: Advanced Practical Computing Topics, Sec-04
Python Data Analytics & Prediction, Fall 2016

Course Information

Instructor: Leonard P. Wesley
Department: Computer Science
College of Science, San Jose State University.
Fall Semester, 2016

Course and Contact Information

Instructor:	Leonard Wesley
Office Location:	MH 212
Telephone:	408.924.5287
Email:	Leonard.Wesley@sjsu.edu
Office Hours:	Tuesdays and Thursdays 3:00PM – 4:00PM
Class Days/Time:	Tuesdays and Thursdays 7:30PM – 8:45PM
Classroom:	MH 422
Prerequisites:	Completion of an introductory course in Python such as CS 185C (Python for non CS majors) or Fall 2016 CS223-01 (Python & Bioinformatics) or equivalent, and completion of CS 146 or equivalent, or CS 151 or equivalent, or instructor consent.

Course Description

Computing topics of current interest in industrial practice. Emphasis on effective use and integration of software/hardware. Different topics may be offered at different times in a short-course lecture/lab format and may be repeated for credit.

Expanded Course Description

The Fall 2016 CS 185C-04 Advanced Practical Computing Topics will be a fast-paced introduction to the data analysis process, quantitative and qualitative data analysis methods & techniques, and selected machine learning and prediction algorithms. Instruction will be provided using Python and Python related packages such as Pandas, numpy, scipy, and matplotlib. Students enrolled I n this course will be expected to have the equivalent knowledge and programming skills gained after successfully completing (i.e., B or better) an introductory Python programming class, data structures class, and/or object oriented programming. Students completing this course will have a sound understanding of how Python can be used to carry out data analysis and to use selected

machine learning algorithms to help make predictions in life science and business intelligence contexts.

Learning Outcomes

Upon successful completion of this course, students will become familiar with:

1. SLO 1 the theory, method, and practice of data analysis and machine-learning based predictions.
2. SLO 2 using Python-based modules and packages to carry out data analysis and making predictions using machine-learning algorithms such as
    a. Linear Regression (least square, ridge, Bayesian, logistic, SGD-Stochastic Gradient Descent)
    b. PLS-Partial Least Squares, SVM, kernel funs, Decision Trees, Multiclass classification
3. SLO 3 how to apply and interpret data analysis and machine learning prediction results in life science and business intelligence contexts.

Required Texts

Python for Data Analysis by Wes McKinney, 2013, Publisher O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, ISBN: 978-1-449-31979-3

Other Optional Reading Material

Python Data Analytics by Fabio Nelli, Publisher APRES Media LLC, 2015, ISBN-13 (pbk): 978-1-4842-0959-2, ISBN-13 (electronic): 978-1-4842-0958-5

Additional optional reading material will be provided as appropriate.

Course Requirements and Assignments

Course Logistics

The majority of course lectures will be pre-recorded and available via .mp4 video recordings. Students will be required to review and complete in-class exercises associated with each video. Teams of approximately three students will be formed almost every week to work on and complete in-class exercises. Quizzes, exams, and programming exercises will be assigned as specified below in the section entitled "**Tentative course calendar of assignment…"** below.

Students should expect to spend approximately nine (9) hours per week (on average) outside of the classroom preparing for and completing the assigned course work. This includes reading papers, viewing videos as appropriate, completing homework and programming exercises, and so forth. The amount of time that a student actually spends will depend on individual skills and the

time allocated to the course. The nine (9) hours per week estimate is based on previous experiences of the instructor and students. So please plan and schedule accordingly.

Previously, students have asked for special exception to policies and procedures for this course. An example includes asking the instructor for extra assignments or work to help improve a grade. Even if such a request is reasonable in the view of the instructor, no exception will be given to a student unless it can be made available to the entire class, AND does not constitute significant extra work on the part of students, instructors, graders and so forth. Students should have no concern that other students will receive special exceptions that will not be available to the entire class.

NOTE: that University policy F69-24 at http://www.sjsu.edu/senate/docs/F69-24.pdf states that "Students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading." However, attendance will be required in order to complete and submit many in-class exercises, quizzes, and exams.

Quizzes and Exams

There will be one midterm and one final exam (or project report submission) that will count toward the final grad (percentage wise) as specified in the "Grades" section below. During quizzes and exams, communication with other individuals via any means is strictly prohibited without the express permission of the instructor. Violations will be met with the full impact of SJSU's academic integrity policy and procedures.

Projects

Several life science, business intelligence, and finance related data analytics and prediction project topics will be described near the start of the course. Projects will involve applying one or more of the skills and knowledge learned in the course to the project. Teams of 3-4 students will be formed to work on a selected project topic. Teams will be required to submit a project proposal before starting on a project, and give a project presentation at the end of the course. Individual student scores on a project will be determined by the content and quality of the contribution of each student toward the project.

The score on the course project and project presentation will count toward the final grad (percentage wise) as specified in the "Grades" section below.

Reading and Homework Assignments

Students will need to view videos and/or compete assigned textbook reading or programming assignments as specified in the course calendar of assignments shown below.

Computational Resources

Students are required to make sure that they have access to sufficient UNIX, Windows, or Mac based computational resources (e.g., computers and software) to carryout assignments in the course. An attempt to offer the course in a classroom with sufficient computation resources will be made by the department to support classroom instruction and demonstrations. However, students should be prepared to bring their portable laptops to class.

Tentative course calendar of assignment due dates & exam dates:

(Please note that course calendar below, and its content is "subject to change with fair notice")

Mtg #	Tue	Thur	SUBJECT/TOPIC	Assignment (See Canvas For Due Dates)
0	8/23 No Class	8/25	Introduction To Course Introduction To Data Analysis	Week 0 Assignment Due 9/8
1	8/30	9/1	8/30: Introduction To Data Analysis 9/1: Introduction To numpy, scipy, Pandas, and pandas I/O	Week 1 Assignment Due 9/8
2	9/6	9/8	9/6: numpy, scipy, pandas, and matplotlib 9/8: matplotlib, pandas data manipulation	Week 2 Assignment Due 9/15
3	9/13	9/15	9/13: **In-Class Exercise 1** (numpy, scipy, panadas, and matplotlib) 9/15: Possible Projects Discussion 9/15: ***PREDICTIVE ANALYTICS:*** Linear Regression (least square, ridge, Bayesian, logistic, SGD-Stochastic Gradient Descent)	Week 3 Assignment Due 9/22
4	9/20	9/22	9/20: **Quiz 1:** Covers topics from 8/25 to 9/15 9/20 Linear Regression (least square, ridge, Bayesian, Logistic, SGD)	Week 4 Assignment Due 9/29

Mtg #	Tue	Thur	SUBJECT/TOPIC	Assignment (See Canvas For Due Dates)
			9/22: Linear Regression (least square, ridge, Bayesian, logistic, SGD)	
5	9/27	9/29	9/27:Data Collection, Sampling, and Preprocessing 9/29: Data Collection, Sampling, and Preprocessing	Week 5 Assignment Due 10/6
6	10/4	10/6	10/4: **In-Class Exercise 2** (Linear Regression, Data Collection, Sampling, and Preprocessing) 10/6: PLS-Partial Least Squares, SVM, kernel funs, Decision Trees, Multiclass classification	No Week 6 Assignment
7	10/11	10/13	10/11: **Midterm** Covers topics from 8/25 to 10/6 10/13: PLS-Partial Least Squares, SVM, kernel funs, Decision Trees, Multiclass classification	Week 7 Assignment Due 10/20
8	10/18	10/20	10/18: PLS-Partial Least Squares, SVM, kernel funs, Decision Trees, Multiclass classification 10/20: Feature Selection	Week 8 Assignment Due 10/27
9	10/25	10/27	10/25: **In-Class Exercise 3** PLS, SVM, kernel functions, Decision Trees, Multiclass classification 10/27: Feature Selection 10/27: Evaluating Predictive Models	Week 9 Assignment Due 11/3
10	11/1	11/3	11/1: Evaluating Predictive Models 11/3: **Quiz 2** Covers topics from Midterm to 11/3. 11/3: Descriptive Analytics	Week 10 Assignment Due 11/10

Mtg #	Tue	Thur	SUBJECT/TOPIC	Assignment (See Canvas For Due Dates)
			(Association rules, Sequence rules, Segmentation)	
11	11/8	11/10	11/8: Descriptive Analytics (Association rules, Sequence rules, Segmentation) 11/10: Assessing quality of models	Week 11 Assignment Due 11/17
12	11/15	11/17	11/15: Assessing quality of models) 11/17: Analytics: Putting It All To Work	TBD
13	11/22	11/24 No Class	11/22: In-class exercise work on Project 11/24: Thanksgiving Holiday	TBD
14	11/29	12/1	11/29: In-class exercise on project 12/1: In-class exercise on project	TBD
15	12/6	12/8	12/6: Project Presentations 12/8: Project Presentations	None
Final Project Report Due Date and Time Is Specified In The Course Canvas Shell Final Exam (To Be Published) MH 422				

SCHEDULE FOOTNOTES: (NONE as of Fall 2016)
NONE AS OF JUNE 2016

Grades *

QUIZZES (2 each 50pts)	100 pts
MIDTERM	100 pts
IN-CLASS EXERCISES (3 @ 50pts each)	150 pts
WEEKLY ASSIGNMENTS (10 @ 40pts each)	400 pts
FINAL EXAM/PROJECT RPT	200 pts
PRESENTATION	50 pts
Total Course Points	= 1,000 pts Total

* The total points for each category might change depending on the number of project teams and assignments. The instructor reserves the right to adjust, with sufficient advanced notice, the above point distribution by ± 5 pts. Such adjustments might be based on the difficulty or simplicity of assignments or quizzes or exams.

Grading Percentage Breakdown (NOTE: Ranges might change if point totals change)

Percentage of Total Pts	Pts	Letter Grade
96.66% and above	> 966	A
93.33% - 96.65%	933 - 965	A-
90% - 93.32%	900 – 932	B+
86.66% - 89.99%	866 – 899	B
83.33% - 86.65%	833 – 865	B-
80% - 83.32%	800 – 832	C+
76.66% - 79.99%	766 – 799	C
73.33% - 76.65%	733 - 765	C-
70% - 73.32%	700 – 732	D+
66.66% - 69.99%	666 – 699	D
63.33% - 66.65%	633 – 665	D-
60% - 63.32%	600 - 632	D-
Below 60%	**< 600**	**F**

HOW TO CALCULATE/ESTIMATE YOUR GRADE

If students would like to calculate their numeric grade percentage, the formula is as follows:

Numeric CS 185C Grade Percentage =

$$\frac{Total\ points\ from\ assignments}{Total\ course\ points}\ x\ 100\%$$

There is no guarantee that grades will be curved. If so, it will be done at the end of the semester. The instructor is already aware that graduate students need to maintain an overall GPA of B or better. Just because a student NEEDS a particular grade doesn't mean that the instructor will automatically GIVE the student that grade. Students must EARN a passing grade based on submitted and evaluated course work.

Extra credit options, if available

There are no extra credit assignments in this course except for completing designated "Advanced" assignments. However, homework assignments and exams might contain extra credit options.

Penalty for late or missed work

Late assignments will not be graded and receive a score of zero. There will be partial credit for assignments. Missed assignments (written or programming) will receive a grade of zero.

Receiving An Incomplete (I) Grade

Receiving a grade of Incomplete (I) is not automatic. Students must complete at least 80% of course assignments by the end of the semester to be eligible to receive a grade of incomplete. Students must also provide documentation to support the reason for the request to receive an Incomplete grade. The instructor has the final decision to give an Incomplete grade. If the instructor agrees to give a student an Incomplete grade, the instructor will enter the remaining work to be completed as part of the PeopleSoft grade submission process.

Grade Change Policy

It is a university policy that course grade changes must be made within one semester from the end of the course. Requests for exceptions to this policy must be accompanied with a documented and compelling reason.

University Policies

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs' Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/