Research Seminar: Machine Learning and Data Analytics

CSCI 480

Spring 2019

COURSE DESCRIPTION

CSCI 480 Research Seminar (Machine Learning and Data Analytics): Python programming with Numpy, Pandas, Matplotlib, and Scikit-learn for Data Analytics and Machine Learning

PROFESSOR/CLASS INFORMATION

Dr. Shieu-Hong Lin

Course Title: Research Seminar (Machine Learning and Data Analytics)
Term: Spring, 2019
Location: Lim 041
Office Phone: 562 903-4741
Office Hours: MW 1:30-3:30pm T Th 3:00-5:00pm
Meetings with Professor: Reserving a slot by email in advance is encouraged

E-Mail: shieu-hong.lin@biola.edu
University Website: www.biola.edu
Course Code/#: CSCI 480
Class Days/Time: T Th 1:30-2:45pm
Credit Hours/Units: 3
Office Location: Lim 137
Admin Assistant: Jerrianne Smith, x4741
Class Website: http://csci.biola.edu/math333/

DISABILITY SERVICES

Disability Services exists to assist any student who thinks he or she may need such assistance. Students desiring accommodations for this class on the basis of physical learning, psychological and/or emotional disabilities are to contact The Learning Center that houses both learning assistance and disability services. The Learning Center is located in the Biola Library, Upper Level, Room U-137, and this department can be reached by calling 562.906.4542 or by dialing extension #4542 if calling from on campus.

BIOLA UNIVERSITY MISSION STATEMENT

TRUTH. TRANSFORMATION. TESTIMONY.

The mission of Biola University is biblically-centered education, scholarship, and service; equipping men and women in mind and character to impact the world for the Lord Jesus Christ.

COURSE ALIGNMENT WITH PROGRAM LEARNING OUTCOMES

CSCI 480 Research Seminar: This upper-division elective course is offered to CS majors designed to be taken within the junior or senior year of the program. Successful completion of this course (see next section) will prepare students to demonstrate a developing proficiency toward the accomplishment of PLO #1: analysis, modeling, and problem solving.

COURSE OBJECTIVES AND STUDENT LEARNING OUTCOMES

By the completion of this course including class participation, class assignments (referred to as "Tasks"), class readings and group interaction, the following objectives and learning outcomes will be assessed and demonstrated:

IDEA Objective #4: Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course (Essential emphasis).

STUDENT LEARNING OUTCOMES (The learner will demonstrate that he or she has satisfactorily fulfilled IDEA Objective #4 by being able to):

- Establish the in-depth understanding of the conceptual framework of machine learning and its applications.
- Cultivate the problem-solving and data-analysis capability using Python, NumPy, Matplotlib, Pandas, and ScikitLearn.

REQUIRED TEXTS & STUDY RESOURCES

Required Textbooks, Web sites, other media and technology sources (Each of the following texts and/or study resources are required and will be used in this course):

- Jake VanderPlas, *Python Data Science Handbook*, O'Reilly Media, 2016. (Free online)
- Andreas Müller and Sarah Guido, *Introduction to Machine Learning with Python*, O'Reilly Media, 2016. (Available as an e-book through the Biola Library account)
- I.Witten & E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufmann, 2011. (Full-text contents available online through your Biola Library account)
- Online documentation and resources for learning Python, NumPy, Matplotlib, Pandas, and ScikitLearn.

LEARNING TASKS (Assignments) & ASSESSMENT (Grading)

Task 1: Weekly Reading and Progress Report

Due Date: Wednesday of the week
Weighting: 15%
Possible Points: 4 points each.
Effort (2 points):

Report the (i) a numerical estimate of the amount of time he/she spent for the reading, (ii) a numerical percentage regarding the percentage of stuff in the reading actually read and understood, and (iii) whether the student has come to the class this week.

Reflection on the reading (2 points):
The student need to put down 1 to 2 paragraphs of his/her thoughts about the reading such as new insight you gained, interesting things encountered, questions of things you don't understand, and so forth.

Assessment:
For the effort part,
the student is expected to **(i)** have attended the class this week at least once (0.5 point), and **(ii) have either** gained a good understanding of **80% or more of the contents or** have spent at least **three hours** in the reading (1.5 points).

For the reflection part,
the student is expected to show substantial evidence of understanding or effort of trying to understand the contents in the reading.

Task 2: Homework, Labs, and Programming Assignments

Due Date: Wednesday of the week
Weighting: 30%
Possible Points: 6 points each for programming assignments. 1-6 points for each Lab and Homework depending on the requirement in it.

Description: Hands-on programming assignments require the student to incrementally develop modeling, analytical, and programming skills based on the concepts of artificial intelligence learned in the class.

Integrity rules for programming assignments:

- **Peer discussion is encouraged**: Peer discussion is encouraged to cultivate an open learning environment in the class, but you should carefully read the guidelines below to avoid any dishonest behavior and never step over the guidelines explicitly described in the following.
- **Never use code written by others**: Any copy-and-paste of code from other people's programs or from websites is viewed as cheating and you will get 0 points for the assignment.
- **Never circulate your code to others**: Peer discussion of code shown on the screen is acceptable for debugging purpose and for explanation of ideas. But you should never pass around your code (electronically or on paper) to others except for the TA and the instructor. Violating this rule is viewed as cheating in the class and the provider will receive 0 points for the assignment.
- **Never provide false or exaggerated results of test cases**: You need to describe results of test cases in the self-evaluation report**.** Providing false or exaggerated results of test cases in the report is viewed as cheating and you will receive 0 points for the assignment.
- **Demonstrate the credibility of your authorship of the work**: When you submit your code as your own work for points, you should make sure that you are able to explain your code

and reconstruct your code from scratch without any outside help when requested. If you are not able to do that on your own when requested, you will get 0 points for the assignment and there will be an investigation.

- **Consequence of cheating in the class**: Cheatings end in 0 points for the assignments followed by discipline actions described in the student handbook.

Assessment: The student needs to submit (all related .cpp and .h files) together with a self-evaluation report. The self-evaluation report should describe results from sufficient test cases that are verified by a peer reviewer. We'll grade each programming assignment in a 0-6 scale based on the following rubric.

0. Nothing done **or missing the self-evaluation report or missing the integrity review** in the report.
1. Source code is completed but the code fails to compile successfully.
2. Source code can compile and do something required, but has serious bugs or miss a couple of key features.
3. Source code can compile and do most of the features required, but has many minor bugs or miss a key required feature.
4. Source code can compile and do all the features required, nearly fully functional, only a couple of minor bugs.
5. Source code can compile and do all the features required, fully functional, no bugs.
6. In addition to the points received above, get one more point if

   a. **the self-evaluation report contains sufficient descriptions of test cases used (0.25 point)**, and
   b. **the self-evaluation report indicates the results of the test cases were verified by a peer reviewer (0.25 point)**, and
   c. **the source code is well indented and commented to make it visually very readable (0.5 point)**.

Task 3: **Study Project**

Weighting: 15%

Description: The student conducts a study project on assigned subjects to develop their own problem-solving approach based on the machine learning concepts explored in the class.

Assessment: The submitted work done for the study project will be examined to determine whether a satisfactory level of study has been done based on the breadth, depth, and clarity of the submitted work.

Task 4: Exams (Quizzes, midterm and final exams)

Due Date: Exams in the midterm and final exam week

Weighting: 40%

Description: The exams have both written components which mainly test the conceptual understanding of AI and machine learning, and the programming components which test the skills of implementation.

Assessment: The written component will be graded based on the answers provided while the programming component will be graded based on the same rubric provided for the weekly programming assignments.

CLASS INFORMATION

Class Attendance and Attendance Policy:

Include a brief paragraph explaining your philosophy about attendance and detailing your specific policy—how many absences without penalty, excused absences, tardies, etc.

Assignments:

- **Penalty for late submission within 2 days after the due date**:
  You may get a deduction of 20% of total points for being late if they are submitted within 2 days of the due date.
- **No submission accepted more than 2 days after the submission due date**: No submission will be accepted more than 2 days after the submission due date, except for extremely exceptional situations such as a serious disabling health problem with evidence from the doctor.

Incomplete Grade:

A temporary mark of "IN" (Incomplete Grade) will be issued in special cases when approved by the Associate Provost of Academic Administration for undergraduate students or the dean of the respective graduate school. “IN” grades course assignments are normally completed no later than five weeks after the end of the term. In the event of the inability of a student to complete the coursework by the approved deadline, the Office of the Registrar will assign the grade which the student has earned by the end of term.

To read more about Biola’s policies and procedures regarding absences, view https://studenthub.biola.edu/undergraduate-student-handbook-absences-attendance

Academic Honesty:

Biola University is committed to ethical practice in teaching, scholarship, and service. As such, plagiarism and other forms of academic dishonesty will not be tolerated. Please see the undergraduate/graduate student handbook and/or the departmental/program/school policy on academic honesty. It is imperative that you present all written, oral, and/or performed work with a clear indication of the source of that work. If it is completely your own, you are encouraged to present it as such, taking pleasure in ownership of your own created work. However, it is also imperative that you give full credit to any and all others whose work you have included in your presentation via paraphrase, direct quotation, and/or performance, citing the name(s) or the author(s)/creator(s) and the source of the work with appropriate bibliographic information. To do otherwise is to put oneself in jeopardy of being sanctioned for an act or acts of plagiarism that can carry serious consequences up to and including expulsion from the university.

To read more about Biola’s policies and procedures regarding academic integrity, view https://studenthub.biola.edu/undergraduate-student-handbook-academic-integrity.

Another helpful resource is Plagiarism.org.

Non-Discrimination Policy:

As Christian scholars we are keenly aware of the power of language, and believe in treating others with dignity. As such, it is important that our language be equitable and prejudice free. Good writing and speech do not make unsubstantiated or irrelevant generalizations about personal qualities such as age, disability, economic class, ethnicity, marital status parentage, political or religious beliefs, race, sex, or sexual orientation. Respectful use of language is particularly important when referring to those outside of the religious and lifestyle commitments of those in the Biola community. By working toward precision and clarity of language, we mark ourselves as serious and respectful scholars, and we model the Christ-like quality of invitation.

Avoid the use of stereotypes or terminology that demeans persons or groups based on age, disability, ethnicity, gender, race, language or national origin. Avoid drawing attention to irrelevant identifiers of race or gender. Avoid gender-specific language when referencing people in general. Avoid terms that assume the universality of human experience, and in particular presume the normativity of the socially dominant group.

Confidentiality and Sexual Misconduct:

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I also have a responsibility in my role as a faculty member to share information I hear regarding sexual harassment, sexual assault, domestic violence, dating violence, stalking, sexual exploitation, and gender/sex-based discrimination with the Title IX Coordinator and/or the Campus Safety Response Team. Confidential resources available to students on campus include the Biola Counseling Center (562-903-4800) and the Student Health Center (562-903-4841). Both the Title IX Coordinator and the Campus Safety Response Team understand the sensitive nature of these situations and can provide information about available on and off-campus resources, such as counseling and psychological services, medical treatment, academic support, university housing, safety measures and other forms of assistance. More information about confidential resources on and off-campus, additional resources, and the University's Sexual Misconduct Policy is available at **https://www.biola.edu/title-ix**.

Additional University and/or Department Policies:

All university and departmental policies affecting student work, appeals, and grievances, as outlined in the Undergraduate Catalog and/or Department Handbook will apply, unless otherwise indicated in this syllabus.

Computation of Final Grade:

Weekly Progress Report	15 %
Homework and Programming Assignments	30%
Study Project	15%
Exams	40 %
Total	**100%**

Final grades will be awarded on the following point system:

A	93%
A-	90%
B+	87%
B	84%
B-	80%
C+	77%
C	74%
C-	70%
D+	67%
D	64%
D-	60% to pass class

COURSE CALENDAR

While the course calendar is intended to provide you with an overview of the semester's schedule, the professor reserves the right to make adjustments to the schedule to responsively meet the needs of this class.

Tentative Schedule

- Weeks 1-3 Intro to Python, NumPy, and Matplotlib
- Weeks 4-6 Intro to Pandas
- Weeks 7-9 Machine learning: concepts and key techniques
- Weeks 10-12 Basics of ScikitLearn
- Weeks 13~14 Advanced features of ScikitLearn