Machine Learning and Data Analytics

CSCI 480

SEMESTER: Fall 2018

COURSE DESCRIPTION

Concepts and techniques of artificial intelligence, representation, search strategies, control, communication and perception, and applications.

PROFESSOR/CLASS INFORMATION

Dr. Shieu-Hong Lin

Term: Fall 2018
Class Location: Busn 210
Office Hours: See announcements at http://csci.biola.edu/

E-Mail: shieu-hong.lin@biola.edu
Class Days/Time: T R 12:00-1:15pm
Credit Hours/Units: 3 hours
Office: Lim 137
Meetings with Professor: Make Appt by email
Admin Assistant: Jerrianne Smith, x4741
Course Website: http://csci.biola.edu/

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

This course is an elective course for computer science majors designed to be taken within the junior year of the program. Successful completion of this course will prepare students to demonstrate proficiency toward the accomplishment of

PLO #1 on Analysis, modeling and problem solving (Demonstrate the ability to model and analyze problems, to devise problem-solving schemes accordingly, and to validate the correctness and effectiveness of these schemes) and

PLO #2 on Knowledge of the theory and practice of computing (Demonstrate an understanding of the theoretical and operational underpinnings of modern computing infrastructure that enables effective utilization of the whole spectrum of the infrastructure, including elements in programming environments, operating systems, and computer networks).

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

- Jake VanderPlas, *Python Data Science Handbook*, O'Reilly Media, 2016.
- *Explorations in Artificial Intelligence and Machine Learning*, 2018, CRC Press. (Online free book)
- 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: 45%
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: 25%

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

1. Class Attendance:

Attendance: You are expected to attend the class regularly since we will examine details of C++ programs using the computers in the lab. Missing the class can seriously hamper your understanding of many key concepts and programming skills critically needed in your programming assignments. Class attendance is counted toward points for the weekly reading report.

2. Turning in Assignments:

Assignments are expected to be electronically submitted under the Canvas system. Due dates are all on Wednesdays. The submission link under Canvas may remain open for 2 more days after the due date as grace period.

3. Late Policy:

1 point will be deducted for late submission within 2 days of the due date while the submission link is still open. **You will receive no points after the submission on canvas is closed** unless it is something like a serious health issue with statements from the doctor as proof.

4. Computation of Final Grade:

Weekly Progress Report	15 %
Homework and Programming Assignments	45%
Study Project	15%
Exams	25 %
Total	**100%**

5. 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%

The GPA System used by the University Registrar's Office is:

A = 4.0	B = 3.0	C = 2.0	D = 1.0
A- = 3.66	B- = 2.66	C- = 1.66	D- = 0.66
B+ = 3.33	C+ = 2.33	D+ = 1.33	F = 0.0

6. Method of Instruction:

The following methods of instruction will be included in this course:

- Lecture
- Written Reports
- Programming Assignments
- Labs
- Reading

7. **Posting of Grades**:

Grades for individual assignments will be posted under Biola's Canvas system. To access the records online, log on to http://canvas.biola.edu/ to make sure the records are accurate.

8. Report Delayed:

In virtually every case that students do not meet the course requirements and when required course tasks are not submitted to the professor, such students should anticipate receiving a failing grade. In rare and unusual situations (e.g., serious illness of the student or illness or death in a student's immediate family), the student may formally request a report delay (RD) through the Vice Provost's Office. Details can be found in the student handbook.

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

9. 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.

10. 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.

11. 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.

COURSE CALENDAR

Tentative Schedule

- Weeks 1-3 Intro to Python, NumPy, and Matplotlib
- Weeks 4-6 Intro to Pandas
- Weeks 7-8 Machine learning: concepts and key techniques
- Weeks 9-12 Basics of ScikitLearn

- Weeks 13~15 Advanced features of ScikitLearn; Kera and Deep Learning
- Week 15 Project presentation