CSCI 4750/5750 - Spring 2020
Machine Learning
Jie Hou
Saint Louis University
Monday, 13 January 2020

CSCI 4750/5750
Machine Learning

Contents

1 Overview 3
   1.1 Catalog Description ........................................... 3
   1.2 Prerequisites ..................................................... 3
   1.3 Student Learning Outcomes .................................... 3
   1.4 Topical Outline .................................................. 3

2 Course Administration 4
   2.1 The Staff ....................................................... 4
   2.2 Class Meetings .................................................. 4
   2.3 Textbook .......................................................... 5

3 Online Resources 5
   3.1 CSCI 4750/5750 Web Page: cs.slu.edu/~jhou4/5750 ............ 5
   3.2 Email with Instructor ........................................... 5
   3.3 Electronic Assignment Submission .............................. 6

4 Graded Work 6
   4.1 Homework Assignments (20%) ................................... 6
   4.2 Quizzes (20%) ...................................................... 6
   4.3 Course projects (20%) .......................................... 6
   4.4 Exams (35%) ...................................................... 7
   4.5 Course Participation (5%) ....................................... 7
   4.6 Course Grades ..................................................... 7
   4.7 Extra Credit ....................................................... 7
   4.8 Late Policies ..................................................... 8

5 Academic Integrity and Collaboration Policy 8
   5.1 Academic Integrity ................................................ 8
   5.2 Collaboration Policy ............................................. 9
6 Additional Information
6.1 Title IX Statement
6.2 Supporting Student Success
6.3 Disability Services
1 Overview

1.1 Catalog Description

This course introduces students to the field of machine learning with emphasis on the probabilistic models that dominate contemporary applications. Students will discover how computers can learn from examples and extract salient patterns hidden in large data sets. The course will introduce classification algorithms that predict discrete states for variables as well as regression algorithms that predict continuous values for variables. Attention will be given to both supervised and unsupervised settings in which (respectively) labeled training data is or is not available. Emphasis is placed on both the conceptual relationships between these different learning problems as well as the statistical models and computational methods used to employ those models.

1.2 Prerequisites

For undergraduate students:

1. Data Structures (CSCI 2100)
2. Calculus III (MATH 2530)

1.3 Student Learning Outcomes

After successfully completing this course, students will be able to:

1. select a machine learning algorithm and model appropriate for a given problem, and apply an existing implementation to a real dataset
2. formulate an appropriate evaluation scheme in order to tune model parameters and evaluate a solution;
3. implement at least two machine learning algorithms from scratch
4. (CSCI 5750) apply machine learning techniques to a solve a research problem in the student’s major field

1.4 Topical Outline

Major themes within the course will include

- What is Machine Learning?
• Decision Trees
• Nearest Neighbors
• Linear models and the Perceptron
• Non-linear methods (kernel methods, SVMs)
• Probabilistic models
• Ensemble Methods
• Clustering algorithms
• Expectation Maximization
• Introduction to Deep Learning
• Recurrent Neural Networks
• Convolutional Neural Networks

2 Course Administration

2.1 The Staff

Instructor: Dr. Jie Hou
Email: jie.hou@slu.edu
Web: http://cs.slu.edu/~jhou4/
Office: Ritter Hall 217
Phone: 314-977-6600
Office hours: Tuesday 10:00am–11:00am
Thursday 09:00am–10:00am
or by appointment

Please make sure to take advantage of office hours, as they offer a wonderful opportunity for individual attention.

2.2 Class Meetings

The material will be presented in three weekly lectures. Attendance is expected and class participation is most welcome. These meetings will offer learning opportunities that cannot be re-created purely from readings. In fact, you will get the most out of the meetings if you read the appropriate material in the textbook before the class meeting in which it will be discussed, and come prepared with any questions that you have.

Information on the lecture topic can be found on the course schedule web page.

Time: Mon/Wed/Fri, 10:00-10:50pm
Place: Ritter Hall 327
2.3 Textbook

The textbook upon which we will organize the class is

Title: Introduction to Machine Learning, third Edition
Authors: Alpaydin, Ethem
Publisher: Cambridge, Massachusetts : The MIT Press, [2014]
Companion website: https://www.cmpe.boun.edu.tr/ ethem/i2ml3e/

Note well that this is the Third Edition of the book, and there were very significant changes between editions so it really is the third edition that is relevant.

There are also two required textbooks for the course. The first one is Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems by Aurélien Géron. The SLU bookstore has copies available, or you can order the book from Amazon or directly from the publisher (O'Reilly). The Géron book comes with a series of excellent Jupyter notebooks that mirror the content of the book. I'll ask you to walk through and sometimes tweak these notebooks for homework as we make our way through the book. The second one is A Course in Machine Learning by Hal Daumé III, and is freely downloadable from website.

For those of you who choose to use the lab computers, please read the department and university policies on appropriate use of computer systems.

3 Online Resources

3.1 CSCI 4750/5750 Web Page: cs.slu.edu/~jhou4/5750

With the exception of the first day’s printed handout, most of the information for this course will be distributed only by means of the course web page. This web site will contain all assignments, a schedule of lectures, detailed lecture notes and links to many other sources of information.

The web page contains some information (e.g. solutions, submitted assignments, individual grades) that is more sensitive and will be password protected. The username and password will be provided in class.

3.2 Email with Instructor

Face-to-face contact in class and in office hours is most desirable. Yet email is a convenient form of communication as well. I try to respond to email promptly, including at least once each evening when possible.

If your question involves your progress on a current programming assignment, my response will be more informative if you can point out the specific problem you have encountered, and if I am able to see all of your source code. Therefore I strongly suggest
that you either attach all relevant files to the email or submit preliminary versions of such files through our online system.

3.3 Electronic Assignment Submission

To allow the student and instructor to exchange electronic files for programming assignments, we will rely on a version control system known as git, and a web-based system known as gitlab and available at git.cs.slu.edu. For further documentation of the use of this system, please see the course webpage.

4 Graded Work

4.1 Homework Assignments (20%)

There will be a series of homework and reading assignments during the semester, approximately once per week or two. These will include both questions that involve writing code and working with a dataset. I’m happy to look over your solutions in office hours to make sure you’re on the right track.

We expect there to be about 6–8 such assignments during the course. At the end of the semester, we will throw away your lowest of the homework grades, with the remaining scores contributing equally to this portion of the grade.

4.2 Quizzes (20%)

We will have biweekly quizzes to verify that you’re doing the work outside of class. There will be six such quizzes, given at the beginning of class on Mondays (the exact dates are given on the course schedule), and each will cover the material from the two preceding weeks. The quizzes are usually true/false, multiple choice, and some short answer, and only take about 10-15 minutes. I’ll drop your lowest quiz score, but I will not allow you to make up quizzes that you miss because of absence or if you arrive late for class. Together the quizzes make up 20% of your final grade.

4.3 Course projects (20%)

You will also be asked to do a semester software project related to some topic we cover in the course, accounting for 20% of your final grade. I’ll give you some ideas as we approach the middle of the semester. Since we cover a lot of different things, this is a good opportunity for you to explore some particular topic in greater depth. Students enrolled in CSCI 5750 are (1) required to work on a problem in their major area, and (2) their writeup should take the form of a publication-ready research paper, with an introduction, survey of related work, experiments, results, and a bibliography.
4.4 **Exams (35%)**

- **Midterm Exam (15%)**, tbd, 10:00am-10:50pm
- **Final Exam (20%)**, tbd

4.5 **Course Participation (5%)**

Attendance in class and interaction in or out of class will be counted for 5%. Let me emphasize that just coming to class is not enough to be successful in this course. First of all, the only way to become a skilled Machine Learning practitioner is by actually working with real datasets, and that will be work mostly done outside of class. Secondly, there isn’t enough time for me to cover all of the theoretical details in 40 class periods. You are responsible for understanding all of the material in the reading and homework assignments, even stuff not covered explicitly in lecture. Again, if there is material you are struggling with, come to office hours or ping me online – I’m super friendly.

Some of the assignments are required to complete with group members. Group members will be assigned and approved by instructor. Group members need contribute equally to the assignments.

4.6 **Course Grades**

Letter grades will be based on each student’s overall percentage of awarded points according to the following formula.

- Student percentage above 90% will result in a grade of A or better.
- Student percentage above 87% will result in a grade of A- or better.
- Student percentage above 83% will result in a grade of B+ or better.
- Student percentage above 80% will result in a grade of B or better.
- Student percentage above 77% will result in a grade of B- or better.
- Student percentage above 73% will result in a grade of C+ or better.
- Student percentage above 70% will result in a grade of C or better.
- Student percentage above 67% will result in a grade of C- or better.
- Student percentage above 60% will result in a grade of D or better.
- Student percentage below 60% will result in a grade of F.

Any modification to this scale at the end of the year will be *in favor of the students*. That is we may later decide to award an A to a student who is slightly below the above cutoff, but we certainly will not deny an A from someone who is above the cutoff.

4.7 **Extra Credit**

Assignments will generally include a small extra credit challenge. Please notice, however, that the actual extra credit given for these challenges is relatively insignificant. Students
who are seriously concerned about improving their overall grade would be best advise to focus all efforts on doing as well as possible on the required work and in preparing for exams.

Our true reason for including these opportunities is to provide some fun and encouragement for students who wish to dig a bit deeper than was required in an assignment. For those students, the chosen extra credit challenges provide a good next step.

4.8 Late Policies

All exams must be taken at the required time. Requests for rescheduling an exam will only be considered if the request is made prior to the start of the exam, or else in an “emergency” situation with appropriate documentation.

All lab writeups must be submitted at the required time (typically 24 hours after the lab period).

For homework assignments, we wish to allow students to continue to work comfortably beyond the official deadline when a little more time will result in more progress, while at the same time discourage students from falling significantly behind pace and jeopardizing their success on future assignments. Our solution is the following exponentially decaying late formula (some have suggested that we should offer extra credit to anyone who fully understands this formula).

We will consider an assignment submission “complete” when any part of the assignment is last submitted or modified. Any assignment which is not complete promptly by its due date and time will be assessed a penalty based on the formula $S = R \cdot e^{-h/173}$, where $S$ is the grade given, $R$ is the grade the work would have received had it been turned in on time, and $h$ is the amount of time (in hours or fractions thereof) that the work was late. Examples:

- work turned in 1 hour late receives over 99.6% of its original credit
- work turned in 5 hours late receives over 97% credit
- work turned in one full day late receives less than 88%
- work turned in two full days late receives less than 76%
- work turned in five days late receives less than 50%

The above policies will be waived only in an “emergency” situation with appropriate documentation.

5 Academic Integrity and Collaboration Policy

5.1 Academic Integrity

*Academic integrity is honest, truthful and responsible conduct in all academic endeavors.*

The mission of Saint Louis University is “the pursuit of truth for the greater glory of God and for the service of humanity.” Accordingly, all acts of falsehood demean and compromise the corporate endeavors of teaching, research, health care, and community
service via which SLU embodies its mission. The University strives to prepare students for lives of personal and professional integrity, and therefore regards all breaches of academic integrity as matters of serious concern. The governing University-level Academic Integrity Policy was adopted in Spring 2015, and can be accessed on the Provost’s Office website.

Additionally, each SLU College, School, and Center has adopted its own academic integrity policies, available on their respective websites. All SLU students are expected to know and abide by these policies, which detail definitions of violations, processes for reporting violations, sanctions, and appeals. Please direct questions about any facet of academic integrity to your faculty, the chair of the department of your academic program, or the Dean/Director of the College, School or Center in which your program is housed. Specific College of Arts and Sciences Academic Honesty Policies and Procedures may be found at: www.slu.edu/arts-and-sciences/student-resources/academic-honesty.php

In addition to those general statements, we wish to discuss our policy in the context of this course. When it comes to learning and understanding the general course material, you may certainly use other reference materials and you may have discussions with other students in this class or other people from outside of this class. This openness pertains to material from the text and practice problems.

However, for work that is submitted for this course, you are not to use or search for any direct assistance from unauthorized sources, including but not limited to:

- other texts, books, or solution manuals
- online information other than that referenced by course materials
- other students in this class (other than when collaboration is explicitly allowed, as described below)
- students or acquaintances who are not in this course

Acceptable sources of information include consultations with the instructor, teaching assistants, or members of organized tutoring centers on campus, as well as any materials explicitly authorized in an assignment. Even in these cases, if you receive significant help you should make sure to document both the source of the help as well as the extent.

Any violations of these policies will be dealt with seriously. Penalties will apply as well to a student who is aiding another student. Any such violations will result in a minimum penalty of a zero on the given assignment which cannot be dropped, and severe or repeated violations will result in an immediate failing grade in the course. Furthermore all incidents will be reported in writing to the Department and/or the Dean, as per the College procedure.

5.2 Collaboration Policy

On certain programming assignments, we will explicitly allow students to work in pairs. In this case, conversations between partners is both permissible and required. Furthermore, all students are expected to contribute significantly to the development of the submitted work. It is unethical to allow a partner to “sign on” to a submission if that partner did not significantly contribute to the work.
6 Additional Information

6.1 Title IX Statement

Saint Louis University and its faculty are committed to supporting our students and seeking an environment that is free of bias, discrimination, and harassment. If you have encountered any form of sexual misconduct (e.g. sexual assault, sexual harassment, stalking, domestic or dating violence), we encourage you to report this to the University. If you speak with a faculty member about an incident of misconduct, that faculty member must notify SLU’s Title IX coordinator, Anna R. Kratky (DuBourg Hall, room 36; akratky@slu.edu; 314-977-3886) and share the basic fact of your experience with her. The Title IX coordinator will then be available to assist you in understanding all of your options and in connecting you with all possible resources on and off campus.

If you wish to speak with a confidential source, you may contact the counselors at the University Counseling Center at 314-977-TALK. To view SLU’s sexual misconduct policy and for resources, please visit the following web address: www.slu.edu/here4you.

6.2 Supporting Student Success

In recognition that people learn in a variety of ways and that learning is influenced by multiple factors (e.g., prior experience, study skills, learning disability), resources to support student success are available on campus. The Student Success Center, a one-stop shop, which assists students with academic and career related services, is located in the Busch Student Center (Suite 331) and the School of Nursing (Suite 114). Students can visit www.slu.edu/success to learn more about:

• Course-level support (e.g., faculty member, departmental resources, etc.) by asking your course instructor.
• University-level support (e.g., tutoring services, university writing services, disability services, academic coaching, career services, and/or facets of curriculum planning).

6.3 Disability Services

Students with a documented disability who wish to request academic accommodations must contact Disability Services to discuss accommodation requests and eligibility requirements. Once successfully registered, the student also must notify the course instructor that they wish to access accommodations in the course.

Please contact Disability Services, located within the Student Success Center, at Disability_services@slu.edu or 314-977-3484 to schedule an appointment. Confidentiality will be observed in all inquiries. Once approved, information about academic accommodations will be shared with course instructors via email from Disability Services and viewed within Banner via the instructor’s course roster. Note: Students who do not have a documented disability but who think they may have one are encouraged to contact Disability Services.