San Jose State University
Department of Computer Science
CS 271, Topics in Machine Learning, Fall 2023

- **Course and Contact information**
  - **Instructor:** Mark Stamp
  - **Office Location:** MH 216
  - **Email:** mark.stamp@sjsu.edu
  - **Text:** for oh ate fore owe to 10-five-too
  - **Office hours:** Friday, 10:00am - 11:30am
  - **Class Days/Times:** Tuesday & Thursday, 9:00am - 10:15am
  - **Classroom:** DH 450
  - **Prerequisites:** CS 149

- **Course Description**
  - Topics in machine learning. The following machine learning techniques and related topics are covered: hidden Markov models (HMM), principal component analysis (PCA), support vector machines (SVM), clustering, data analysis, backpropagation, convolutional neural networks (CNN), recurrent neural networks (RNN), long short-term memory (LSTM), residual networks (ResNet), generative adversarial network (GAN), Word2Vec, transfer learning, ensemble techniques, and a variety of other topics related to deep learning. Illustrative applications of each of the major topics are provided, with most of the applications drawn from the field of information security. Additional machine learning topics will be covered as time permits.

- **Learning Outcomes**
  - The focus of this course will be machine learning, with illustrative applications drawn primarily from the field of information security. After completing this course students should have a working knowledge of a wide variety of machine learning techniques, and have a good understanding of how to apply machine learning to real-world problems.

- **Textbook and Readings**
  - ***Introduction to Machine Learning with Applications in Information Security***, 2nd edition, Mark Stamp, Chapman and Hall/CRC, 2022. Be sure to get the 2nd edition, and do *not* attempt to use the e-book version.

  - Additional relevant material:
    - PowerPoint slides at http://www.cs.sjsu.edu/~stamp/ML/powerpoint
    - Current semester lecture videos are available at http://www.cs.sjsu.edu/~stamp/ML/lectures/CS271_Fall23/. If you are asked to login to access the videos, both the username and password are "infosec". **Note**: The instructor hereby gives students permission to record his lectures (audio and/or video). At least with respect to this class, your instructor has nothing to hide.
    - Class-related discussion will be posted on CampusWire at https://campuswire.com/c/G8E8D46B6/feed. You are strongly encouraged to participate by asking questions, as well as by responding to questions that other students ask. At the start of the semester, you should receive an email asking you to join this discussion group—if not, contact your instructor via email.

  - The applications parts of this course are essentially self-contained, but for additional background information on the security-related topics, the following resources are recommended.

- *Computer Viruses and Malware*, John Aycock, Springer 2006. Many of the applications we discuss are related to malware. Aycock's book is easy to read and in spite of being fairly old, it provides a good foundation for malware research.
- *Information Security: Principles and Practice*, third edition, Mark Stamp, Wiley 2021. If you have not taken CS 265, you should do so. In any case, you can refer to this highly-recommended book if you have questions about security-related topics during this course.
- Open Malware (at http://www.offensivecomputing.net/) includes a large collection of samples of live malware.
- VX Heavens (at http://vx.netlux.org/) is a source for "hacker" type of information on viruses. Malware samples are also available.
- Journal of Computer Virology and Hacking Techniques (at http://www.springer.com/computer/journal/11416) is a journal that is primarily focused on malware-specific research papers. There are also several good conferences that focus on malware and/or machine learning applications in information security.
- Recent masters project reports (at http://www.cs.sjsu.edu/~stamp/cv/mss.html#masters). Most of these projects involve applications of machine learning to malware or other topics in information security.

- **Course Requirements and Assignments**
  - SJSU classes are designed such that in order to be successful, it is expected that students will spend a minimum of forty-five hours for each unit of credit (normally three hours per unit per week), including preparing for class, participating in course activities, completing assignments, and so on. More details about student workload can be found in University Policy S12-3 at http://www.sjsu.edu/senate/docs/S12-3.pdf.

  - Approximate Schedule
    - Week 1 --- Introduction and overview
    - Week 2 --- Hidden Markov Models (HMM)
    - Week 3 --- Data Analysis and experimental design
    - Week 4 --- Principal Component Analysis (PCA)
    - Week 5 --- Support Vector Machines (SVM)
    - Week 6 --- Clustering (K-means and EM clustering)
    - Week 7 --- Classic machine learning mini-topics (k-nearest neighbors, boosting, random forest, LDA, etc.)
    - Week 8 --- Neural networking background
    - Week 9 --- Basic neural networking architectures (MLP, CNN, and RNN)
    - Week 10 --- Backpropagation
    - Week 11 --- Advanced neural networking architectures (LSTM, GAN, RBM, GNN, etc.)
    - Week 12 --- Word embedding techniques (Word2Vec, BERT, etc.)
    - Week 13 --- Deep learning mini-topics (ensembles, regularization, dropouts, attention, explainability, adversarial attacks, etc.)
    - Week 14 --- Trending deep learning topics (time permitting)
    - Week 15 --- Project presentations

  - Homework is due *typewritten* (include source code, but not executable files) by class starting time on the due date. Each assigned problem requires a solution and an explanation and work detailing how you arrived at your solution. Cite any outside sources used to solve a problem. When grading an assignment, your instructor may ask for additional information. Note that a *subset* of the assigned problems will typically be graded.

    Homework must be submitted via email before the start of class on the due date. Be sure to have an extra copy of your homework with you in class, and be prepared to discuss your solutions. Your written solutions must be in a pdf file. Submit any source code or other attachments in separate files (i.e., no code in the solution itself). You must provide enough discussion of your solution so that the

grader can understand your solution, and so that the grader can be sure that you understand your solution. Put your written solution and any relevant source code in a folder named "yourlastname". Then zip your homework folder and submit the file yourlastname.zip via email to ISA.Fall2023@gmail.com. The subject line of your email ***must*** be of the form:

```
CS271HMK assignmentnumber yourlastname last4digitofyourstudentnumber
```

The subject line must consist of the four identifiers listed. There is no space within an identifier and each identifier is separated by a space.

- Assignment 0: Due Tuesday, August 29
  As with all assignments, this must be submitted (via the grader email address given above) prior to the start of class.
  1. Read A Revealing Introduction to Hidden Markov Models (at https://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf) and do the following.
     a. Briefly (1 paragraph) summarize how an HMM is trained.
     b. How is a trained HMM used to score a sequence?
     c. Briefly explain how an HMM and dynamic program differ.
     d. Why is it necessary to scale the values of the matrices A, B, and π when training an HMM?
  2. Read the article "Models will run the world" (your instructor will post a softcopy of the article to Campuswire by the first day of class) and do the following.
     a. In one paragraph, summarize the authors' main points.
     b. Write second paragraph discussing what you most agree with and anything (or any things) that you disagree with in this article.

- Assignment 1: Due Thursday, September 7
  Problems 2.1, 2.2, 2.3, 2.10, 2.11.

- Assignment 2: Due TBD
  TBD

- Assignment 3: Due TBD
  TBD

- Assignment 4: Due TBD
  TBD

- Assignment 5: Due TBD
  TBD

- Assignment 6: Due TBD
  TBD

- Assignment 7: Due TBD
  TBD

- Assignment 8: Due TBD
  TBD

- Assignment 9: Due TBD
  TBD

- Assignment 10: Due TBD
  TBD

- 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 ensure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading."

- **Grading Policy**
  - Test 1, 100 points. Date: TBD.
  - Homework, quizzes, class participation and other work as assigned, 100 points. A subset of the assigned homework problems will be graded.
  - Machine Learning Project, 100 points. You must obtain approval for your project proposal from your instructor (via email) by the close of business on TBD, and you must be prepared to give a brief presentation of your proposed topic on TBD. A written project report is due TBD and project presentations will begin on then.
  - Final, 100 points. Date: Thursday, December 14 from 7:15 - 9:30am. The official finals schedule is here: https://www.sjsu.edu/classes/final-exam-schedule/fall-2023.php
  - Semester grade will be computed as a weighted average of the major scores listed above.
  - ***No*** make-up tests or quizzes will be given and ***no*** late homework or project (or other work) will be accepted.
  - Grading Scale:

Percentage	**Grade**
92 and above	A
90 - 91	A-
88 - 89	B+
82 - 87	B
80 - 81	B-
78 - 79	C+
72 - 77	C
70 - 71	C-
68 - 69	D+
62 - 67	D
60 - 61	D-
59 and below	F

  - Note that "All students have the right, within a reasonable time, to know their academic scores, to review their grade-dependent work, and to be provided with explanations for the determination of their course grades." See University Policy F13-1 at http://www.sjsu.edu/senate/docs/F13-1.pdf for more details.

- **Guest Lectures**
  - TBD
    - Date: TBD
    - Time: TBD
    - Location: TBD
    - Title: TBD
    - Abstract: TBD
    - Short Bio: TBD

- TBD
  - Date: TBD
  - Time: TBD
  - Location: TBD
  - Title: TBD
  - Abstract: TBD
  - Short Bio: TBD

- **Classroom Protocol**
  - Keys to success: Do the homework, complete a good project, and attend class
  - **Wireless laptop is *required***. Your laptop must remain closed (preferably in your backpack and, in any case, not on your desk) until your instructor informs you that it is needed for a particular activity
  - **Cheating** will not be tolerated, but working together is encouraged
  - Student must be respectful of the instructor and other students. For example,
    - No disruptive or annoying talking
    - Turn off cell phones
    - Class begins on time
    - Class is not over until your instructor says it is over
  - Valid picture ID required at all times
  - The last day to drop without a "W" grade is Friday, September 15, and the last day to add is Friday, September 15

- **College and University Policies**
  - Office of Graduate and Undergraduate Programs maintains university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. You may find all syllabus related University Policies and resources information listed on GUP's Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/