Syllabus for CS 303: Introduction to Machine Learning
T/Th 9:30-10:50, Spring 2021
Dr. Joanna Bieri and Dr. Tamara Veenstra

CONTACT INFORMATION

Dr. Joanna Bieri Email: Joanna_Bieri@redlands.edu
Web: joannabieri.com

Dr. Tamara Veenstra Email: Tamara_Veenstra@redlands.edu
Web:http://facweb1.redlands.edu/fac/Tamara_Veenstra/

Prerequisite: Grade of 1.7 or higher in (MATH 241 AND CS 111) --OR-- (MATH 122 AND CS 240) --OR-- permission. Some experience with programming in Python strongly encouraged.

Textbooks:

1) *Machine Learning for Absolute Beginners,* second edition, Oliver Theobald.
2) *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*, second edition, Aurélien Géron.
3) Other course materials will be posted on Moodle as needed.

Course Description: Machine learning is the practice of programming computers to learn and improve through experience. This course provides an introduction to the mathematical underpinnings, algorithms, and practices that enable a computer to learn. Topics include supervised learning, unsupervised learning, and evaluation methodology. Students are required to program in Python. Programming intensive.

Technology: This course will be online. There will be extensive use of technology and a regular internet connection is required. We will meet LIVE from 9:30-10:50 am PST (virtually). Please let your seminar instructor know ASAP if you foresee having any problems with the internet requirements/live virtual meeting component. All classes will be recorded and may be watched asynchronously if needed, but please try to attend class as much as possible.

We will use **Microsoft Teams** for class meetings, small group discussion, and office hours. Please participate in class with video turned on if all possible and contact your professor if you need access to new hardware. **Moodle** will be used extensively for collecting assignments. A course schedule and complete list of assignments will be posted in Moodle, will not, in general, given out in class.. You will be required to submit all assignments via Moodle.

Most homework assignments will be completed using **Python in Google Colab** notebooks. This eliminates the need for downloading software on your own computer. If you want to download Python, we recommend using the Anaconda distribution of Python 3.7 available at https://www.anaconda.com/distribution/ . (This is an open source software package.) Jupyter notebooks may be easily uploaded to Google colab for HW submission, but there are packages that will also have to be installed.

COURSE GOALS

- To understand the basic building blocks and general principles that allow one to design machine learning algorithms
- To become familiar with specific, widely used machine learning algorithms
- To learn methodology and tools to apply machine learning algorithms to real data and evaluate their performance
- To understand the benefits, drawbacks, limitations, applicability and ethics of machine learning algorithms

Course Content will include:

- **Mathematical tools**: probability, matrix and vector manipulation, geometry of machine learning problems, basic optimization techniques including techniques from Calculus
- **Machine learning principles**: problem formulations, notation, underfitting versus overfitting, regularization
- **Methodology**: evaluation, parameter tuning, model selection
- **Applications**: different applications of ML; data preparation: feature engineering and normalization

HOMEWORK: There will be daily homework assignments, consisting of reading assignments, programming exercises, and written problems. Homework is due at the beginning of each class period, by 9:30am. Homework submitted late will be penalized per day late and will not be accepted after two weeks.

EXAMS: There will be five take-home exams. These will be open book and open note. You may consult online programming guides, but you must work on them independently. The exams will occur approximately every three weeks and the last exam will be due at the time of the final exam period.

DAILY CONCEPT CHECKS: There will be short daily quizzes that check your conceptual understanding. These will be completed online prior to attending class and will be due by 8am.

GRADING CRITERIA and SCALE:

Tests: 45%   Daily Homework: 45%   Daily Concept Check: 10%

94-100%: A (4.0)	90-93%: A- (3.7)	87-89%: B+ (3.3)	83-86%: B (3.0)
80-82%: B- (2.7)	77-79%: C+ (2.3)	73-76%: C (2.0)	70-72%: C- (1.7)
67-69%: D+ (1.3)	63-66%: D (1.0)	60-62%: D- (0.7)	<60%: no credit (0.0)

Collaboration: Collaboration on homework assignments is encouraged. However, **every student must write their own code, comment their code in their own words, run their own experiments, and write their own solutions**. Sharing of code or written solutions will be considered a violation of the honor code. Also, we highly encourage each student to first attempt problems on their own, especially for the shorter exercises that are designed to test and reinforce concepts taught in class. Please write the names of all collaborators at the beginning of the written portion of the submission.

Internet sources: The internet is a useful resource when learning to solve computer science problems. In general, it's OK to look at resources for a broad topic (e.g., dynamic programming), but it is not OK to look at solutions for specific problem (e.g., interval scheduling) that is the same or substantially similar to one you are working on for the class. If you are unsure whether something is allowed, ask. **You must cite all online sources used while working on an assignment**. It is always **your responsibility** to learn if a source is allowed.

NOTE: *Incorrect citing of internet sources and lack of authorship assignment to pieces of code that were primarily authored by another source or in collaboration with classmates are instances of academic dishonesty and will be taken very seriously. The first instance will result in a zero on the assignment in question and the second instance will result in failing the course.*

Approximate Schedule of Topics

Week 1

- Introduction – What is Machine Learning?
- Python, numpy, and linear algebra

Week 2

- Importing and Graphing Data, Pandas and Matplotlib
- Data Prep, SKlearn, Train-Test-Validate Split, Scaling

Week 3

- K Nearest Neighbors Algorithm
- PCA and Dimensionality Reduction

Week 4

- Exam 1 due (on Data Analysis Basics , kNN and PCA (Weeks 1-3))
- Multivariate Linear Regression
- Gradient Descent

Week 5

- Introduction to Classification – Logistic Regression
- Logistic Regression and Sklearn intro

Week 6

- Sklearn for Logistic Regression
- Polynomial Regression - Intro to Bias and Variance

Week 7

- Exam 2 due (on Regression --Weeks 4-6)
- Regularization, Bias and Variance

Week 8

- Model Evaluation: cross validation, ROC and precision vs recall curves
- Support Vector Machines

Week 9

- Clustering Algorithms – K Means
- Clustering Algorithms or other Unsupervised Techniques

Week 10

- Exam 3 due (on Regularization, model evaluation SVM, clustering – Weeks 7-9)
- Introduction to Neural Networks, Logic Gates, Keras

- Perceptrons

Week 11

- Multilayer Perceptrons
- Training and Backpropagation

Week 12

- Regression Neural Networks
- Learning rate, Validation loss, and parameter tuning

Week 13

- Exam 4 due (Neural Networks – Weeks 10-12)
- Deep Learning – More complicated Neural Networks
- Convolutional Neural Networks

Week 14

- Ethics in Machine Learning

Friday April 23 by 3 pm: Last (5th) Exam Due (Cumulative)