CS 3840
Applied Machine Learning

Course Introduction

Lingwei Chen

2022-01-10

WRIGHT STATE
UNIVERSITY

About Me

Research interests:

Trustworthy machine learning

Machine learning on security issues

More information: https://lgchen.org/

Introduce Yourself

Your name and major?

Which year you are in right now?

Are you familiar with machine learning?

If yes, what experience do you have?

Syllabus

This is **CS 3840 Applied Machine Learning**

Syllabus has been uploaded on Pilot

You are responsible for reading it

If you have any concerns or issues about the information

You should contact me during the first week of class

Class Time and Location

Face-to-Face Lectures

- Monday/Wednesday 4:40pm – 6:00pm
- Russ 346
- Instructor: Lingwei Chen
- Please follow the University-wide mask mandate
- Skipping the class is highly discouraged

Lecture Recording

- Face-to-face lectures will be recorded
- The lecture recording links will be sent only to those students that have excused absences from the corresponding classes (e.g., unavoidable health reasons)

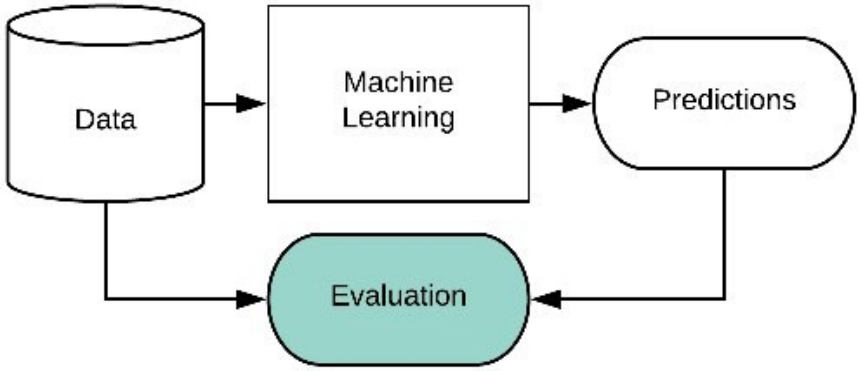
Contact Information

Email:	lingwei.chen@wright.edu
Office:	Russ 338
Office hours:	3:00pm - 4:00pm, M/W
	Or by appointment
	(In person or Webex)
Office phone:	937-775-5023
Course website:	https://pilot.wright.edu
	(Log in using your CaTS username-w000xxx and password)

Course Overview

Course description

This course explores applied machine learning, focusing more on the techniques and methods that can automatically learn and generalize from data to accomplish important tasks than on the statistics behind these methods



Course Overview

Course schedule

- Terms and Principles of Machine Learning
- Feature Representation
- Evaluation Metrics
- Model Training
- Logistic Regression
- K-Nearest Neighbors
- Support Vector Machines
- Decision Trees
- Ensemble Learning
- Neural Networks
- K-Means Clustering
- Anomaly Detection
- Main Challenges of Applied Machine Learning

Course Overview

Course objectives

- Help you understand key terms and principles of machine learning, machine learning workflow, a variety of specific machine learning models including supervised and unsupervised learning models, and how to use them with useful tools (i.e., Python with scikit-learn)
- At the end of the course, you should be able to implement basic machine learning tasks, measure and optimize the effectiveness of your machine learning model, detect and avoid issues in applied machine learning (e.g., overfitting)
- More importantly, apply the learned machine learning to solve real-world problems and get ready for more advanced machine learning techniques

Textbook

- There is **no required textbook**. The course web page leads to lecture notes for this course. Some selected reading materials may be also assigned.
- If you wish to go into more depth in certain areas, contact me

- **Recommended**: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd Edition, by Aurélien Géron, ISBN 978-1492032649

Prerequisites

Have knowledge of basic computer science principles and skills at a level sufficient to write a reasonably non-trivial computer program

Be familiar with Python programming language

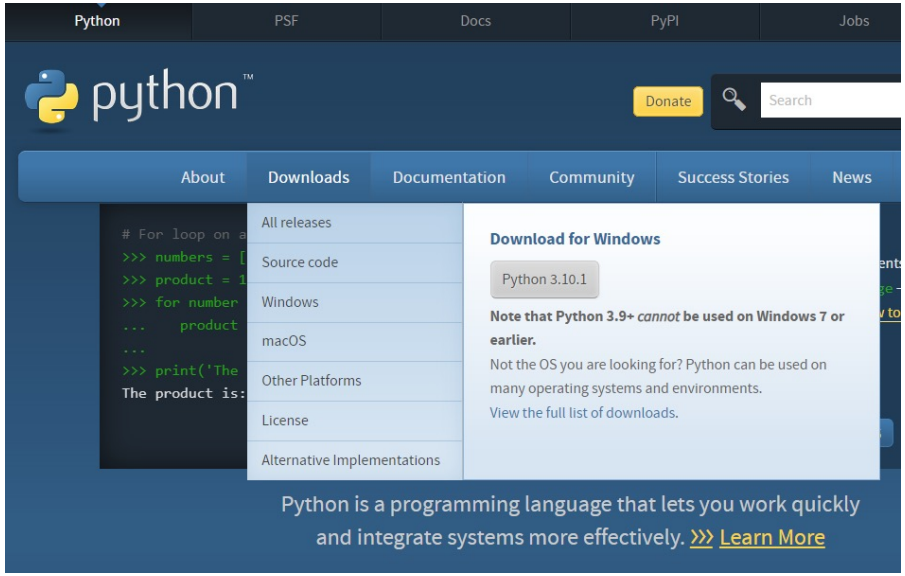
- Good understanding
- Some coding

Basic knowledge

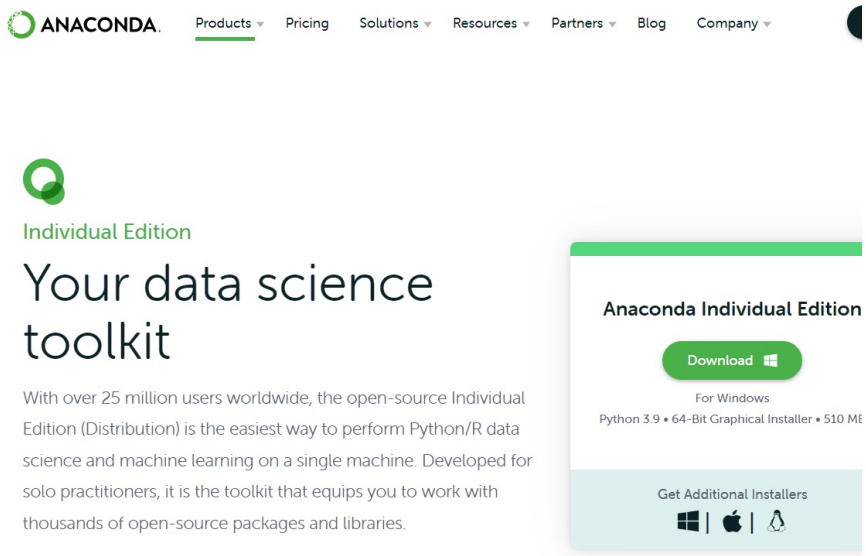
- Probability theory
- Calculus

Learning Environment

https://www.python.org/

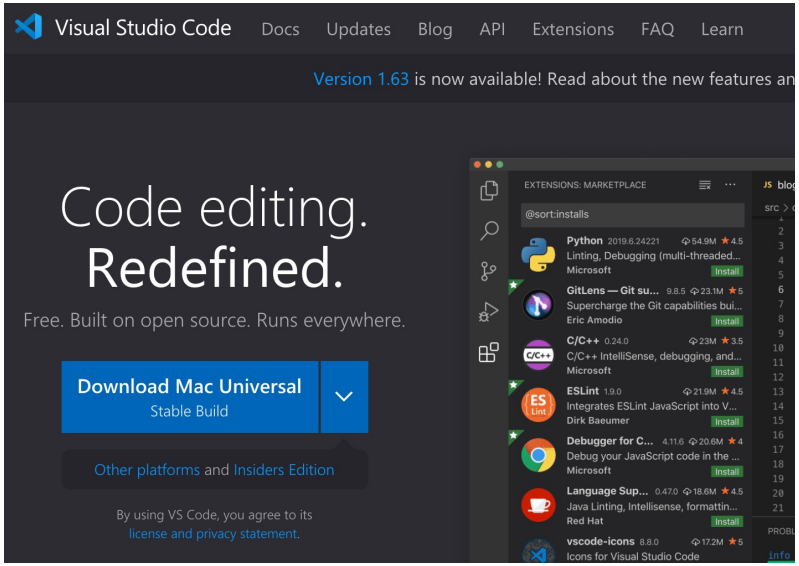


https://www.anaconda.com/products/individual



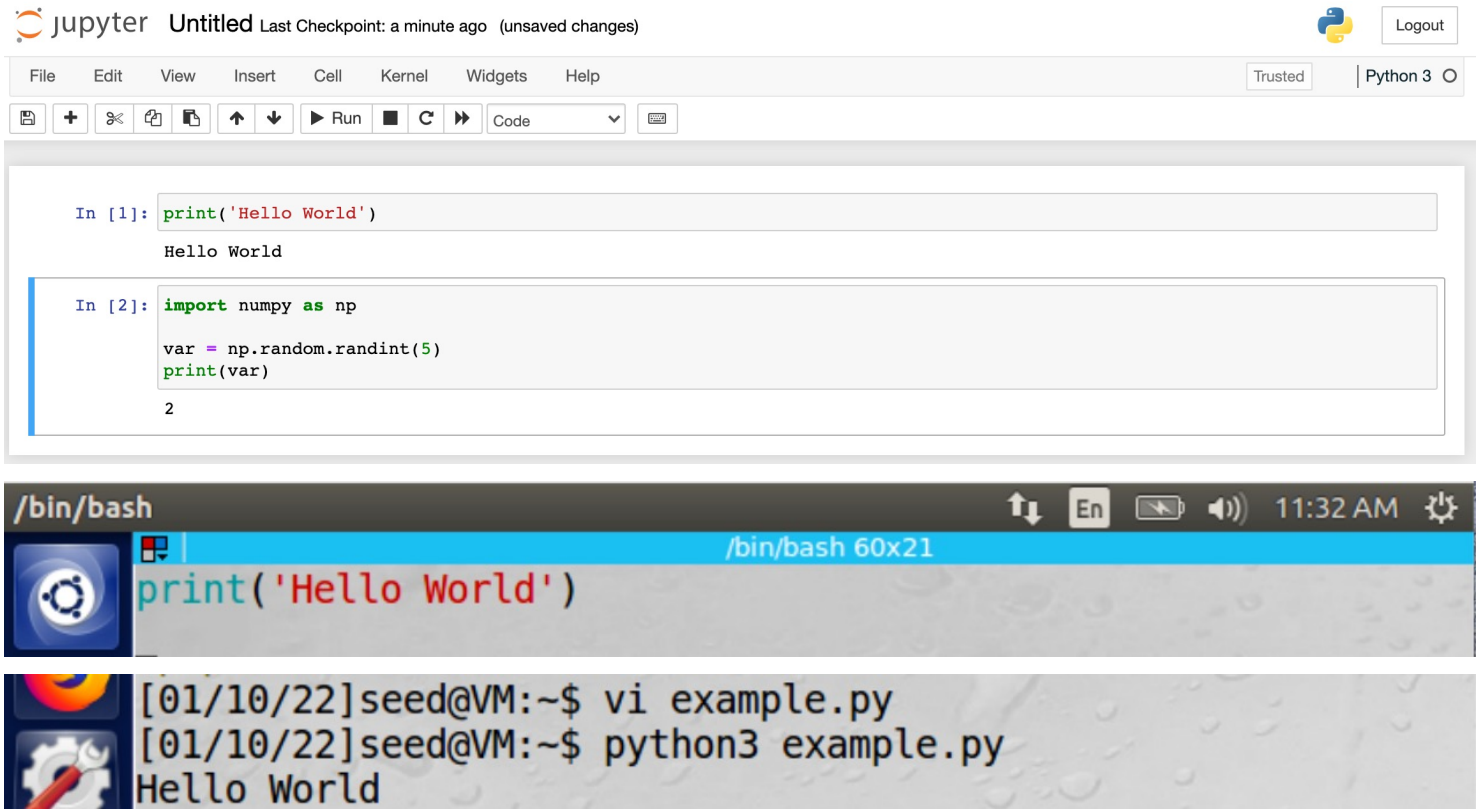
- You may need to have a laptop/desktop to install Python 3 for this course
- You can directly install Python 3, or install it using Anaconda
- After installing Python 3, you should also install libraries we are going to use: jupyter, matplotlib, numpy, pandas, scipy, scikit-learn

Learning Environment



- After install Python 3 and libraries, you can either use IDEs such as VS Code, PyCharm, or Jupyter, or directly use command lines to perform Python programming

Learning Environment



- After install Python 3 and libraries, you can either use IDEs such as VS Code, PyCharm, or Jupyter, or directly use command lines to perform Python programming

Grading

Total (100%)

- Lab assignments (30%)
- Quizzes (20%)
- Midterm (20%)
- Final project (30%)

Course grades will be assigned according to the university standard grading scheme

- 90-100% = A; 80-89% = B; 70-79% = C; 60-69% = D; 0-59% = F
- Depending on the overall distribution of final grades, the boundaries may be adjusted by the course instructor to ensure consistency from term to term.

If you have any complaints about you grades, you have a week from when the assignment grades have been returned to you to contact me. I may not consider to regrade for any assessment if you have failed to do that.

Lab Assignments

Each lab assignment includes some hands-on tasks closely related to the class topic and content, where you may need to do some programming, run the programs, describe the observations with screenshots, and submit the lab report.

For this assignment, you will be using the Breast Cancer Wisconsin (Diagnostic) Database to create a classifier that can help diagnose patients. First, read through the description of the dataset (below).

```
import numpy as np
import pandas as pd
from sklearn.datasets import load_breast_cancer

cancer = load_breast_cancer()

#print(cancer.DESCR) # Print the data set description
```

The object returned by `load_breast_cancer()` is a scikit-learn Bunch object, which is similar to a dictionary.

```
cancer.keys()
```

Question 0 (Example)

How many features does the breast cancer dataset have?

Quizzes

In-class quizzes: frequently given in class for class participation and attendance checking

- Exercise questions
- Grades based on submissions

Please log in Pilot right now. Go to Assessment, choose Quizzes & Exams, and click Quiz 1: In-Class Survey to complete your first quiz

Out-of-class quizzes:

- Only given out of class after a big machine learning topic for knowledge testing.
- 4-5 short answer questions
- Grades based on the answers

Midterm Exam

Midterm will be given during Week 7-Week 8 in the form of true-or-false questions, short-answer questions, and analysis questions to test the detailed knowledge to ensure that you keep up with the lecture materials and have mastered the course.

1. How would you define Machine Learning?
2. Can you name four types of problems where it shines?
3. What is a labeled training set?
4. What are the two most common supervised tasks?
5. Can you name four common unsupervised tasks?

Final Project

Final project is to solve a specific real-world problem using machine learning methods. The problem with the dataset can be decided by yourself. The submission includes implementation code, dataset, project report, and presentation slides. The final project presentation will be conducted in class in the last week (Week 15).

Project report should include

- Title
- Introduction (background, problem to be solved, and how you want to solve it)
- Dataset (data statistics, features, and data split for training and testing)
- Methods (machine learning methods and detailed steps)
- Results (experimental results using evaluation metrics)
- Discussion (analysis and observations from results)
- Challenges and future steps
- Conclusion

Final Project

- You can work on the final project by yourself, or form a group with another student to work together (with clear individual contribution)
- Please start your project early
- You may need to submit your project plan right after midterm exam (week 8-week 9)
- Contact me anytime for the comments and feedback on projects
- Some data source: http://archive.ics.uci.edu/ml/index.php
- Some data source: https://www.kaggle.com/

Late Assignments

- Project reports must be submitted by 11:59 pm (EST) on the due date.
- Late assignments will be accepted, but 10% of the total available points will be deducted for each day late.
- No points will be awarded for assignments turned in more than 5 days late.

Academic Integrity

- You may discuss course contents with other students.
- It is expected that lab assignments will be completed on an individual basis. Each student is to write up solutions in their own words.
- If the submitted work is copied from any source or the same work is turned in by two or more students, all parties involved will be held equally accountable for violation of academic integrity and will not be given credit.
- In addition to the policy stated above, students are expected to comply with the Wright State University Code of Student Conduct and in particular the portions pertaining to Academic Integrity at all times.

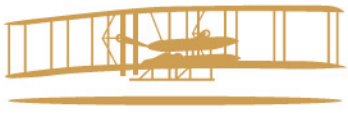
Additional Need

- Students with disabilities please register with Office of Disability Services and notify me of your eligibility for reasonable accommodations.
- Students with any other additional needs are encouraged to set up an appointment with me to discuss any accommodations that may be necessary.

Next Class

(2022-01-12)

Terms and Principles of Machine Learning



WRIGHT STATE UNIVERSITY