QAC 239: Machine Learning Methods for Audio and Video Analysis

Spring 2021

Time: T.R. 1:00PM – 2:20PM **Location:** ALLB 204

Instructors:

Jielu Yao
Email: jyao01@wesleyan.edu
Office hours: Wednesday 4-6pm

Markus Neumann
Email: mneumann01@wesleyan.edu
Office hours: Thursday 4-6pm

Course Description

In this course, students will learn machine learning techniques to analyze text, image, video, and audio data. The course consists of three parts: general techniques, image/video analysis and audio analysis. Each part will first introduce how these non-traditional data can be converted into mathematical objects suitable for computer processing and, particularly, for the application of machine learning techniques. Students will then learn a selection of supervised, unsupervised, and deep learning algorithms that are effective for text, image/video and audio analysis. Finally, the course will introduce major applications of these techniques such as object detection, face recognition, image classification, speaker detection, speech recognition, etc. The course also provides opportunities to apply machine learning techniques to the Wesleyan Media Project data sets.

Reading Materials

- *An Introduction to Statistical Learning* by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 2013. Free e-book available online at: `http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf`
- Additional weekly reading materials will be provided through Moodle.

Grading

Component	**% of course grade**
Datacamp online course	5
In-class homework	10
Four individual assignments	40
Class participation	15
Final project	30

The grading scale for this class is as follows: 95-100 = A; 91-94.9 = A-; 88-90.9 = B+; 85-87.9 = B; 81-84.9 = B-; 78-80.9 = C+; 75-77.9 = C; 71-74.9 = C-; 68-70.9 = D+; 65-67.9 = D; 60-64.9 = D-

- Assignments and homework (50%): Students will complete four formal assignments (40%). Additionally, we will give homework, usually to finish some code from class (10%). Assignments and homework are all due at the time specified on Moodle. Any late assignments and homework will have points reduced by 20% of the total available points per day unless arrangements are made prior to the due date. Students may work together on their assignments and homework, but we expect you to turn in your work individually, and your code and answers should differ from your collaborators.
- Class participation (15%): We will spend a fair amount of time talking as a class and working in small groups to discuss class materials. We expect everyone to participate in these endeavors.
- Online Python course (5%): Working knowledge of programming in Python is expected. An online course from DataCamp.com is used as a refresher on Python.
- Final project (30%): Students are encouraged to use Wesleyan Media Project data sets for their final project. The project should be an original work, with thorough exploration of a chosen question, using Python. The project can be individual- or group-based. If it's a group project, the size of the group should be less than or equal to four. Students should settle on a topic, write a one-page proposal, and discuss it with the instructor (5%). Students also need to report their preliminary findings and do a lightning talk (5%). The final paper (20%) is due at 5pm on May 21. Any late paper will have its points reduced by 20% of the total available points per day unless arrangements are made prior to the due date.
- Students who have selected the CR/U grading mode will receive a CR if and only if a) they have completed all course work and have not received less than 70% in any assignment/course component, and b) their weighted average is greater than 77%. Please note that a pass/fail course cannot be used to satisfy requirements for the data analysis minor or the applied data science certificate programs.

Recommendations for Success

- Do the weekly reading and come to class. Students should expect to do about 6 hours of work per week outside of class to prepare and study. Most of that time for this class should be spent completing the readings, working on group assignments, and practicing Python programming.
- Regularly consult Moodle. Class materials (reading materials, slides, code, etc) and assignment deadlines will be posted to the Moodle.
- During in-class discussions and group activities, make an effort to contribute to the group. This will make the class more interesting for you and will ensure that you receive full credit.
- Make sure to stay up on the deadlines.
- Academic integrity is essential to your success. The rules of science should be carefully upheld in everything that you do. The following behavior is absolutely unacceptable: plagiarism, data fabrication, selective reporting, omission, suppression or distortion.

Accessibility Services

Wesleyan University is committed to ensuring that all qualified students with disabilities are afforded an equal opportunity to participate in and benefit from its programs and services. To receive accommodations, a student must have a documented disability as defined by Section 504 of the Rehabilitation Act of 1973 and the ADA Amendments Act of 2008, and provide documentation of the disability.

Since accommodations may require early planning and generally are not provided retroactively, please contact Accessibility Services as soon as possible.

If you believe that you need accommodations for a disability, please contact Dean Patey in Accessibility Services (accessibility@wesleyan.edu), located in North College, Room 021, or call 860/685-5581 for an appointment to discuss your needs and the process for requesting accommodations.

Calender

The calendar below gives the dates of exams and other important deadlines for the course. Readings should be completed **prior to** the start of the class. This calendar is subject to change. Any changes will be announced and posted on Moodle.

SECTION 1: INTRODUCTION (taught by Markus and Jielu)

February 9: Introduction

- Topic: Motivation and syllabus
- Assignment 0: Datacamp course due at 5pm on February 19 (Fri).
  - Introduction to Python for Data Science or
  - Intermediate Python

February 11: Python tutorial I

- Topic
  - Google Colab
  - Jupyter Notebook
  - Python - introduction

February 16: Python tutorial II

- Topic: Python - intermediate

February 18: Python tutorial III

- Topic: Python - intermediate, part 2

February 23: Getting data: image scrapers, OCR, and OpenCV (Jielu)

- Topic:
  - Tesseract OCR - recognizing characters from images
  - Image scrapers
  - Twint
  - Introduction to OpenCV

February 25: Getting data: web scraping & youtube-dl (Markus)

- Topic:

  - Web scraping
  - wget
  - youtube-dl

March 2: Machine learning (Jielu)

- Topic:
  - Supervised and unsupervised learning
  - Regression and classification models

March 4: Machine learning - Text (Markus)

- Topic:
  - Bag-of-words
- Assignment 1: Machine learning due at 5pm on March 12 (Fri)

SECTION 2: IMAGES (taught by Jielu)

March 9: Getting started with images

- Topic:
  - Pixels and image as a matrix
  - Image channels
  - Color spaces

March 11: Image as data

- Topic:
  - Basic image operations
  - Detecting edges and corners
  - SIFT and SURF

March 16: Matching

- Topic:
  - Template matching
  - Feature matching

March 18: Face detection

- Topic:
  - Viola-Jones object detection framework
  - Histograms of oriented gradients (HOG)

March 23: Spring Break

- No class

March 25: Face recognition

- Topic
  - Deep learning and convolutional neural networks
  - *Face Recognition*
- Assignment 2: Image analysis due at 5pm on April 2 (Fri)

March 30: Image classification with PyTorch I

- Topic
  - Image classification overview
  - ImageNet
  - LeNet and LeNet-5
  - Types of layers

April 1: Image classification with PyTorch II

- Topic
  - Image classification architecture
  - AlexNet / VGG-16 / GoogleNet / Resnet
  - Creating your own architecture

April 6: Object detection with PyTorch

- Topic
  - Object detection overview
  - Single stage object detection
  - Two stage object detection
- Assignment 3.1: Image analysis due at 5pm on April 9 (Fri)

SECTION 3: AUDIO (taught by Markus)

April 8: Speech Recognition

- Topic
  - Speech recognition with Google Cloud API

April 13: Introduction to Audio Analysis

- Topic

  - Sound physics
  - Signal processing fundamentals

April 15: Working with Audio Software

- Topic
  - Audacity
  - FFmpeg
  - Praat
  - Librosa & other Python packages

April 20: Phonetics

- Topic
  - Human speech production
- Assignment 3.2: Audio analysis due at 11:59pm on April 25 (Sun)

April 22: Audio features

- Topic
  - Time and frequency-domain audio features
  - Mel frequency cepstral coefficients

April 27: Audio classification (Frame-level)

- Topic
  - Using an SVM to classify sounds frame by frame

April 29 Audio classification (Track-level)

- Topic
  - Classifying audio at the track level

May 4: Speaker Detection

- Topic
  - Speaker embeddings with deep learning
- Assignment 4: Audio analysis due at 23:59pm on May 9 (Sun)

May 6 & 11: Final Group Projects

- Presentations

Final project is due at 5pm on May 21 (Fri)