QAC 239: Machine Learning Methods for Audio and Video Analysis

Spring 2023

Time: T.R. 1:20PM – 2:40PM **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 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.

Grading

Component	**% of course grade**
Datacamp online course	5
In-class homework	10
Three individual assignments	45
Class participation	10
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 (55%): Students will complete three formal assignments (45%). Additionally, we will give homework, usually to finish some code from class. There will likely be six homework assignments across the semester (10%).
  - Assignments and homework are all due at the time specified on Moodle. Any late assignments will have 20% of their grade deducted per day (any part of 24 hours). No assignments will be accepted more than five days after their due date. Each student is allowed a total of five days worth of extensions across the semester. Students are free to use these as they

see fit (for problem sets and the final project). After the five days are exhausted, no more extensions will be granted. Documented medical and personal emergencies will be handled on a case-by-case basis.

  - 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 (10%): 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. You will receive an email about how to access the course.
- Final project (30%): 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 will also present their findings in class (10%). The final paper (15%) is due at 5pm on May 19. The rules for late final papers are the same as for other assignments described above. Students are encouraged to use Wesleyan Media Project data sets (see here and here) for their final project. Since the focus of the class is on images and audio, the project must involve the use of at least some raw image or audio data (even if, for example, it also has a text analysis component).

Recommendations for Success

- Students should expect to do about 6-8 hours of work per week outside of class to prepare and study. Most of that time for this class should be spent completing homework/assignments, working on the final project, and practicing Python programming.
- Regularly consult Moodle. Class materials (reading materials, slides, code, etc) and assignment deadlines will be posted to 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, and selective reporting.

Recommended Reading Materials

- *An Introduction to Statistical Learning* by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 2021. Free e-book available online at: https://web.stanford.edu/~hastie/ISLRv2_website.pdf
- Machine Learning with PyTorch and Scikit-Learn by Raschka, Liu & Mirjalili, 2022. Packt.

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. This calendar is subject to change. Any changes will be announced and posted on Moodle.

SECTION 1: INTRODUCTION (taught by Markus and Jielu)

January 26: Introduction

- Motivation and syllabus
- Python
  - Jupyter Notebook
  - Google Colab
- Assignment 0: Datacamp course due at 1pm on February 2 (Thu).
  - Introduction to Python for Data Science or
  - Intermediate Python

January 31: Python tutorial

- Python - introduction
  - Basic Python
  - Numpy
  - Pandas
  - Miscellaneous

February 2: Getting Data

- Getting data
  - Tesseract OCR - recognizing characters from images
  - Image scrapers
  - Web scraping
  - wget
  - youtube-dl

February 7: Machine learning - Introduction I

- Intro to ML I
  - Regression & classification
  - OLS & Logistic regression
  - Train, test, and validation datasets
  - Performance metrics

February 9: Machine learning - Introduction II

- Intro to ML II
  - Scikit-learn

February 14: Machine learning - Optimization

- Optimization
  - Model weights & biases
  - Loss function
  - Gradient descent
  - Other optimizers - Newton, Ada*, RMSProp
  - Batching

February 16: Machine learning - Additional models

- Additional models
  - Naive Bayes
  - Random Forest
  - Support Vector Machine (SVM)
- Hyperparameters
- Cross-validation

February 21: Machine learning - Text classification

- Text as data
  - Bag-of-words classification
- Assignment 1: Machine learning due at 5pm on February 28 (Tue)

SECTION 2: IMAGES (taught by Jielu)

February 23: Image as data

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

  - Color spaces
  - Introduction to OpenCV

February 28: Working with images

- Topic:
  - Basic image operations
  - Image features
  - Computer vision techniques

March 2: Simple applications

- Topic:
  - Image segmentation using threshold methods
  - Image comparison using distance-based measures
  - Image classification using machine learning algorithms

March 7: Face detection

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

March 9: Face recognition

- Topic
  - Face recognition algorithms
  - Deep learning I - neural networks

March 11 - 26: Spring break

- No class

March 28: Emotion detection

- Topic
  - Face Emotion Recognizer
  - Deep learning II - convolutional neural networks

March 30: Other pretrained deep learning models

- Topic
  - Recurrent Neural Networks and Image Captioning
  - Generative Adversarial Networks and Synthetic Image Generation

April 4: Image classification

- Topic
  - Image classification overview
  - ImageNet
  - Image classification architectures

April 6: Image classification with Lightning Flash

- Topic
  - Introduction to Lightning Flash
- Assignment 2: Image analysis due at 5pm on April 13 (Thu)

SECTION 3: AUDIO (taught by Markus)

April 11: Introduction to Audio Analysis

- Topic
  - Sound physics
  - Signal processing fundamentals
  - Fourier transform
- Reading:
  - https://jackschaedler.github.io/circles-sines-signals/
  - Read through *Intro, Signals and Sound*, *Sines and Sampling*, *Transforms and Notation* and *Inside the DFT* before class.

April 13: Working with Audio Software

- Topic
  - Audacity
  - FFmpeg
  - Praat
  - Librosa & other Python packages
- Project proposal due at 11:59pm on April 16 (Sun)

April 18: Phonetics

- Topic
  - Human speech production

April 20: Audio features

- Topic

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

April 25: Audio classification I

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

April 27: Audio classification II

- Topic
  - Frame-level, continued
  - Frame to track-level aggregation
  - Audio classification with Lightning Flash

May 2: Speaker Detection

- Topic
  - Speaker embeddings with deep learning
- Assignment 3: Audio analysis due at 11:59pm on May 9 (Tue)

May 4 & 9: Final Group Projects

- Presentations

Final project is due at 5pm on May 19 (Fri)