QAC 239: Introduction to Machine Learning for Text, Image/Video, and Audio Analysis

Spring 2019

Instructor:	Jielu Yao	**Time:**	T.R. 1:20PM – 2:40PM
Email:	jyao01@wesleyan.edu	**Location:**	ALLB 204
Office:	PAC 213	**Office Hours:**	M.W. 3:00PM – 4:30PM

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: text analysis, 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 sentiment analysis, face recognition, motion detection, 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
- *Computer Vision with OpenCV and Python 3: Practical examples workbook* by Thileepan Stalin and Divya Vetriselvan, Amazon Digital Services LLC, 2017.
- (Optional) *Introduction to Audio Analysis: A MATLAB Approach* by Theodoros Giannakopoulos and Aggelos Pikrakis, Academic Press, 2014.
- Additional weekly reading materials will be provided through Moodle.

Grading

Component	**% of course grade**
Midterm	25
In-class mini hackathons	10
Four individual assignments	20
Class participation	15
Datacamp online course	5
Final project	25

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-

- Midterm exams (25%): There will be a take-home midterm exam. The exam will be given out (posted on Moodle) five days before the deadline. You can use any course readings, lectures, or materials to complete this exam, but you must not give or receive any help from others. You must direct any questions to the course instructor. Late exams will receive a penalty of 20% per day.
- In-class mini hackathons and individual assignments (30%): We will do mini hackathons in most lab sessions (10%). Students are expected to complete and present their work as a group in class but can turn in their code individually by 10pm Friday. Each student will also do four formal assignments (20%). Assignments are all due at the time specified. Any late assignments will have points reduced by 20% of the total available points per day unless arrangements are made prior to the due date.
- Class participation (15%): We will spend a fair amount of time talking as a class and working in small groups to discuss class materials. I 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 (25%): 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 by April 2 (5%). Students also need to report their preliminary findings and do a lightning talk on May 7 (5%). Final paper (15%) is due at 5pm on May 16. Any late paper will have points reduced by 20% of the total available points per day unless arrangements are made prior to the due date.

Recommendations for Success

- Do the weekly reading and attend lectures. 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.

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

Week 0: Introduction

- Date: Jan 24
- Topic: Motivation and syllabus
- Reading
  - (Optional) Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." Science 349.6245 (2015): 255-260.
  - (Optional) "In the Age of A.I., Is Seeing Still Believing?", Joshua Rothman, the New Yorker, November 12, 2018.

Week 1: Python Tutorial

- Date: Jan 29 / Jan 31
- Lab
  - Python: basic data types, containers, functions, classes
  - Numpy: arrays, array indexing, datatypes, array math, broadcasting
  - SciPy: image operations, distance between points
  - Matplotlib: plotting, subplots, images
  - Jupyter (IPython) Notebook
- Assignment 0: Datacamp course due at 10pm on Feb 1 (Fri).
  - Introduction to Python for Data Science

Week 2: Text Analysis

- Date: Feb 5 / Feb 7
- Topic
  - Text as data
  - Automated content analysis
- Lab
  - Sentiment analysis using Stanfords Large Movie Review Database
- Reading
  - Grimmer, J., Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis, 21(3).
- Assignment 1 due at 10pm on Feb 15 (Fri).

Week 3: Getting Image/Video Data using APIs

- Date: Feb 12 / Feb 14
- Lab
  - Converting speech to text with Google cloud speech API
  - Accessing political ads with Streamlink API

Week 4: Image Analysis I - Image as Data

- Date: Feb 19 / Feb 21
- Topic
  - Encoding images in RGB space
  - Detecting corners in images
  - Using the Scale-Invariant Feature Transform (SIFT)
  - Image convolution
- Lab
  - Introduction to *openCV*
- Reading
  - Stalin and Vetriselvan (2017), Chapter 2

Week 5: Image Analysis II - Face Detection

- Date: Feb 26 / Feb 28
- Topic
  - Viola–Jones object detection framework
- Lab
  - Face detection - Part I
- Reading
  - Stalin and Vetriselvan (2017), Chapter 5
  - (Optional) Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-I). IEEE.

Week 6: Image Analysis III - Pedestrian Detection

- Date: Mar 5 / Mar 7
- Topic
  - Support Vector Machine (SVM)
  - Histograms of Oriented Gradients (HOG)

- Lab
  - Face detection - Part II
  - Pedestrian detection with SVM and HOG
- Reading
  - James, Witten, Hastie, & Tibshirani (2013), Chapter 9.1-9.4
  - (Optional) Dalal, N., Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEEComputer Society Conference on (Vol. 1, pp. 886-893). IEEE.
- Assignment 2 due at 10pm on Mar 8 (Fri).
- Midterm exam

Mid-semester Recess, Mar 9 - Mar 24

Week 7: Image Analysis IV - Face Recognition

- Date: Mar 26 / Mar 28
- Topic
  - Introduction to deep learning and convolutional neural networks
- Lab
  - Introduction to *Face Recognition*
- Reading
  - Convolutional neural networks for visual recognition, Module 1 & 2
  - (Optional) Camastra, F. and Vinciarelli, A. (2015), Machine Learning for Audio, Image and Video Analysis, Springer, Chapter 13

Week 8: Video Analysis I - Motion Detection

- Date: April 2 / April 4
- Topic
  - Video principles
  - Frame differencing and object tracking
- Lab
  - Introduction to *FFmpeg*
  - Detecting motion
- Reading
  - Stalin and Vetriselvan (2017), Chapter 3
  - (Optional) Dietrich, B. J. (2018), Using Motion Detection to Measure Social Polarization in the U.S. House of Representatives, Working Paper.

Week 9: Video Analysis II - Scraping Closed Captioning

- Date: April 9 / April 11
- Topic
  - Video segmentation and keyframe extraction
  - Optical character recognition (OCR)
- Lab
  - Extracting text from images
  - Scraping closed captioning
- Reading
  - (Optional) Camastra, F. and Vinciarelli, A. (2015), Machine Learning for Audio, Image and Video Analysis, Springer, Chapter 14
- Assignment 3 due at 10pm on April 13 (Fri)

Week 10: Audio Analysis I - Audio as Data

- Date: April 16 / April 18 (No class due to the WPSA conference)
- Topic
  - Sound physics
  - Audio acquisition
  - Signal processing fundamentals
  - Fourier transform
- Lab
  - Using *FFmpeg* to extract audio from a given video
- Reading
  - (Optional) Giannakopoulos and Pikrakis (2014), Chapter 2

Week 11: Audio Analysis II - Feature Extraction

- Date: April 23 / April 25
- Topic
  - Short-term and mid-term feature extraction
  - Time-domain audio features
  - Frequency-domain audio feature
- Lab
  - Using *pyAudioAnalysis* to extract and visualize features
  - Shazam algorithm
- Reading

  - (Optional) Giannakopoulos and Pikrakis (2014), Chapter 3-4

Week 12: Audio Analysis III - Recognizing Speech Emotion

- Date: April 30 / May 2
- Topic
  - More classifiers
- Lab
  - Using *pyAudioAnalysis* to classify an unknown audio segment to a set of predefined classes (e.g. music and speech)
  - Using *pyAudioAnalysis* to recognize speech emotion
- Reading
  - James, Witten, Hastie, & Tibshirani (2013), Chapter 8
  - (Optional) Giannakopoulos and Pikrakis (2014), Chapter 5
  - (Optional) Dietrich, B. J., Enos, R. D., Sen, M. (2016). Emotional arousal predicts voting on the us supreme court. Political Analysis.

Week 13: Work on the Final Group Projects

- Date: May 7 / May 9
- Topic: Lightning presentations
- Assignment 4 due at 10pm on May 10 (Fri)

Week 14

- Final project is due at 5pm on May 16 (Wed)