CERI 7703/8703 - Fall 2021
Seminar in Earth Science:
Machine Learning with Python

Instructors: Thomas Goebel, thgoebel@memphis.edu, Eunseo Choi, echoi2@memphis.edu
Lectures: TR, 9:40 a.m. - 11:10a.m., CERI
Location: Earthquake Ctr house 1, room 110
Office Hours: T, 2-4 p.m., Earthquake Ctr house 2, room 107

Course Description: Artificial intelligence and machine learning (ML) - driven by the availability of exceedingly large amounts of data - have led to a rapid acceleration in tech and science breakthroughs. ML is not only useful for google-searches and movie selections on Netflix but also for fault mapping, earthquake detection, phase picking, elastic stress calculations, failure predictions, aftershock forecasting and much more.

The goal of this class is to expose students to the newest developments in machine learning and applications in engineering and science. We will discuss basics of unsupervised and supervised learning, classification and regression problems, model training and performance evaluation as well as ensemble and deep learning.

The class will combine lectures, student discussions and hands-on coding activities. The final weeks of the class will focus on working on a group-project with respective topics provided by the instructors or selected by the students. The ultimate goal is to finish a complete ML project and draft a manuscript that can be used as building block for a scientific paper.

Course Objectives and Learning Goals:

1. Understand fundamental concepts and theoretical framework of machine learning (ML) algorithms
2. Implement basic algorithms in Python and Matlab with application to nominal, ordinal and numerical datasets
3. Identify best ML algorithms for different classes of science problems
4. Create training data, hyper-parameters optimization, evaluate model performance and compare to base-algorithms
5. Apply deep-learning to classification and regression problems
6. Discuss opportunities and limitations of ML to resolve fundamental physical processes
7. Solve final group project and draft publication-quality manuscript

Prerequisites: Students should have taken Data Analysis (CERI7104) or Programming Tools (CERI7002/CIVL7102) prior to signing up for this course. Contact instructor for potential exceptions.

Books and Resources

- **Python machine learning**. Raschka, Sebastian. Packt publishing ltd, 2015.
- **Intro. Lectures**. Ng, Andrew. https://www.coursera.org/learn/machine-learning
- **Deep Learning**. Online course materials by Yann LeCun at NYU: www.atcold.github.io/pytorch-Deep-Learning/
- **Deep Learning**. online textbook: www.deeplearningbook.org/
- **Papers with Code project**. www/paperswithcode.com/

Hardware / Software Requirements

Personal laptops or access to mac-lap computers is required. Install Anaconda3 on your computer (make sure to install **Anaconda3 for python 3.7**) .
Anaconda: https://www.anaconda.com/distribution/
PyCharm: https://www.jetbrains.com/pycharm/
Atom: https://atom.io/packages/ide-python
I highly recommend using an Integrated Development Environment (IDE). Examples are: **spyder, pycharm, eclipse, atom, IDLE and visual studio**. You can start with using a jupyter notebook but more complex projects are easier managed through an IDE. IDEs naturally promote better coding habits and facilitate code sharing.

Grading

Grading is based on class participation, presentation and final project. There is no final exam. The class includes 4 different homework problem sets that have to be solved and submitted. Homework problems can be solved in groups of 2 students. One group will be selected to present their solution in class. In addition, each student will present a recent paper on machine learning applications in geophysics and seismology. The final project includes a presentation and final paper.

Grade break down: 20% paper presentation; 30% homework; 50% final project.

Final Project

The final project consists of solving a data analysis/machine learning problem related to geoscience or engineering. Students can identify an interesting science problem themselves or chose from the list below. The problems can be solved in groups of up to three students. Each students must equally contribute to project design, algorithm or code development, manuscript writing and presentation on the final day of class.

Final Project Topics

Seismology

1. Induced seismicity detection based on waveform characteristics or earthquake clustering
2. (Un)supervised learning applied to seismicity data: identify earthquake sequences
3. Classify fore-main-aftershocks, earthquake swarms and background events based on CNN / RNNs trained on forward ETAS models
4. Magnitude estimates or predictions based on differently windowed synthetic and real earthquake waveforms
5. (Many-to-Many) Earthquake sequence (total number of events and Mmax) predictions based on origin time and magnitude or origin-time, location and magnitude - an ML approach to operational aftershock forecasting
6. Seismic phase picking and event association for laboratory and natural earthquakes
7. Predict mainshock location and magnitude based on number and spatial distribution of aftershocks
8. Detect false picks in laboratory experiments or local seismic deployments

Geodynamics

1. Uncertainty identification and enhancement of tomographic images with deep learning
2. Mantle convection inversions with sequence-to-sequence modeling

Preliminary Schedule of Classes

Week	Date	Topic	Instruc.	Homework/Present	chapter
		Intro. Supervised Learning			
1	08/23-08/27	Intro. and Terminology, Data fitting vs. learning,	Goebel		1
2	08/30-09/03	Perceptron, Adaline, Linear Regression	Goebel	HW#1 - model fit	2, 10
3	09/06-09/10	Logistic Regression and SVM	Goebel		3
4	09/13-09/17	Decision Trees, Random Forest, K-nearest neighbor	Goebel	SCEC meeting	3
		Data prep., validation and testing			
5	09/20-09/24	Data prep., normalization etc.	Choi	present #1 (ERC-Tectonic workshop)	4
6	09/27-10/01	Dimensionality Reduction, PCA, Validation, Model Performance, Parameter Tuning	Goebel	present #2	5
7	10/04-10/08	Multi-Layer Perceptrons for classification and regression	Goebel	present #3 , HW#2 - MLP regress	6
8	10/11-10/15	Ensemble Learning, Unsupervised Learning	Goebel	present #4 (Fall break, Mon-Tue)	7, 10, 11
		Deep Learning			
9	10/18-10/22	Parallelizing with Tensorflow	Choi	HW#3 - CNN event classification	13, 14
10	10/25-10/29	Convolutional and Recurrent Neural Nets	Choi	HW#4 - Event Assoc.	15, 16
11	11/01-11/05	Building layered NN architecture	Choi	**Discuss Term Projects**	
12	11/08-11/12	Adversarial Networks, Reinforcement Learning, Final Project	Choi		17,18
		Final Project			
13	11/15-11/19	Final Project			
14	11/22-11/26	Final Project		Thanksgiving	
15	11/29-12/03	Final Project		Wed. last day	
16	12/06-12/10	Final Presentations		Final Present.	