Machine Learning in Earth Science

Tuesday & Thursday 2:00 pm – 3:15 pm, ATL 2428

Instructor:

Xin-Zhong Liang (xliang@umd.edu; 301-405-6300)
Office Hours: Tuesday & Thursday 3:30 pm – 4:30 pm by appointment

Required Textbook:

Introduction to Machine Learning with Python, First Edition, 2017
By Andreas C. Müller & Sarah Guido, O'Reilly

Course Description:

This is a comprehensive introductory course designed to prepare undergraduate and graduate students for applying machine learning techniques to solve real-world problems in Earth science. It emphasizes practical solution implementation, providing students with essential hands-on experience using the most popular open-source analytics tools based on Python, a general-purpose programming language. The course works through all steps in machine learning, from problem specification, data analytics to analytical solution, and applies advanced statistical and analytical algorithms to uncover hidden data relationships and transform them into predictive understanding or decision support.

This course has two overarching components: first, students will learn how to program with Python in using Scikit-learn and other major analytics toolkits; second, students will learn how to apply these machine learning tools with basic knowledge of statistics to discover robust signals underlying actual big data in the Earth science domain. These two components will be bridged with homework plus exercise assignments utilizing both analytical and programming skills to examine and interpret Earth's climate/environment variations. These skills learned can be more generally applied to other scientific data with variations in time, space or feature. The lecture notes will be made available after each class at http://www.atmos.umd.edu/~xliang/aosc447/.

Recommended Prerequisite:

Familiarity with basic descriptive statistics, differential/integral calculus (MATH140)

Grades will be based on homework (60%), exercises (25%), and classroom participation (15%). There will be five (5) homework assignments that involve both reasoning and problem solving, and five (5) simpler coding exercises. Graduate students for AOSC 647 will be assigned an extra question on problem sets, and required to do a term project applying machine learning to an earth science data topic of their choosing; final grades will be the regular homework/exercise/participation scores (75%) plus the term project score (25%).

Course Topics:

Python programming, SciPy and Scikit-learn utility, data engineering, visualization, classifiers, regression models, canonical correlation analysis, structural equation models, decision trees, random forests, boosting machines, support vector machines, clustering, dimensionality reduction, principal component analysis, neural networks.

Tentative Course Schedule

The schedule below gives a tentative overview of the topics we will cover in the course. It may be subject to changes as appropriate. Please see my course website for update.

Class #	Date	Topic Covered	Homework
1.	01/26	Machine Learning Basics	
2.	01/28	First Machine Learning Application (Ch 1)	
3.	02/02	Python Programming	
4.	02/04	NumPy Library	EX1
5.	02/09	Pandas Library	
6.	02/11	Pandas Data Manipulation	HW1 set
7.	02/16	Data Visualization	
8.	02/18	Pandas Data Input/Output	HW1 due
9.	02/23	Data Engineering	EX2
10.	02/25	Fundamental Statistics	
11.	03/02	Supervised Learning	HW2 set
12.	03/04	Regression Models (Ch 2)	
13.	03/09	Classifiers (Ch 2)	HW2 due
14.	03/11	Decision Trees (Ch 2)	
		Enjoy Spring Break (March 14 to 21)	
15.	03/23	Random Forests (Ch 2)	
16.	03/25	Boosting Machines (Ch 2)	HW3 set
17.	03/30	Support Vector Machines (Ch 2)	
18.	04/01	Canonical Correlation Analysis	HW3 due
19.	04/06	Structural Equation Models	
20.	04/08	Deep Learning	EX3
21.	04/13	Neural Networks	
22.	04/15	Unsupervised Learning (Ch 3)	EX4
23.	04/20	Dimensionality Reduction (Ch 3)	
24.	04/22	Principal Component Analysis (Ch 3)	HW4 set
25.	04/27	Clustering (Ch 3)	
26.	04/29	Feature Engineering (Ch 4)	HW4 due
27.	05/04	Model Evaluation (Ch 5)	EX5
28.	05/06	Model Improvement (Ch 5)	
29.	05/11	Pipelines and Parallel Computing (Ch 6)	HW5 set
30.	05/17	**No Class but turn in your HW5**	HW5 due