MACHINE LEARNING IN PYTHON FOR ENVIRONMENTAL SCIENCE PROBLEMS: INTRODUCTION TO MACHINE LEARNING

Short Course Organizers
Ben Toms and Amanda Burke

AMS COMMITTEE ON ARTIFICIAL INTELLIGENCE APPLICATIONS TO ENVIRONMENTAL SCIENCE

SATURDAY, JANUARY 11, 2020

Time	Topic	Speaker
8:30 AM	OPENING, INTRODUCTION, COMPUTER SETUP	Amanda Burke, University of Oklahoma, Norman, OK
8:45 AM	INTRODUCTION TO MACHINE LEARNING AND ARTIFICAL INTELIGENCE	
9:30 AM	COFFEE BREAK	
9:45 AM	DATA SCIENCE FUNDAMENTALS*	Kate Avery, University of Oklahoma, Norman, OK
12:00 PM	LUNCH (INCLUDED)	
12:45 PM	SUPERVISED LEARNING ALGORITHIMS*	Hamid Kamangir, Texas A&M University, Corpus Christi, TX
2:30 PM	COFFEE BREAK	
2:45 PM	INTRODUCTION TO DEEP LEARNING*	Karthik Kashinath, Lawrence Berkeley National Laboratory, Berkeley, CA
4:45 PM	SUMMARY, EVALUATIONS	
5:00 PM	COURSE END	

*Includes an interactive exercise

INTRODUCTION TO MACHINE LEARNING AND ARTIFICAL INTELIGENCE

In this lecture, we will explore different pre-processing techniques applied to data input to machine learning models, including scaling, imputation, and others. We will also introduce the weather dataset used throughout the course.

DATA SCIENCE FUNDAMENTALS

After examining different data pre-processing techniques, we will examine dataset partitioning, covering cross-validation, feature selection and extraction, and other methods. In addition, we will explore parameter tuning for training optimal machine learning models. We will also introduce the python package used in the introductory course, scikit-learn, and discuss linear models. After the lecture, participants will have a chance to train a linear model.

SUPERVISED LEARNING ALGORITHIMS

In addition to a linear model, we will investigate different types of supervised machine learning models, such as decision trees, ensemble methods (e.g. random forests, boosting), neural networks, and support vector machines. We will also cover basic interpretability concepts, such as feature importance and partial dependence plots. This segment includes an exercise where participants will use data-science principles to train and compare different models, to select the best model and parameters on a real-world dataset.

INTRODUCTION TO DEEP LEARNING

As deep learning has grown in popularity amongst the environmental sciences, we will spend time exploring the basic architecture of deep learning models. We will cover topics such as hyperparameter tuning, convolutional neural nets, and regularization and dropout, all fundamentals for training a deep learning model. Also included in this lecture is an interactive exercise where participants will train basic neural nets.