Introduction to Machine Learning for Engineering Research - Summer 2024

Part 1: Curriculum Information

Instructor Information

Instructors: Benjamin Afflerbach, Dane Morgan
E-mail: bafflerbach@wisc.edu, ddmorgan@wisc.edu

Canvas Course Link:

https://canvas.wisc.edu/courses/408885

If the Canvas course (or assignments in the course) aren't showing up correctly please reach out via Slack to troubleshoot. Thanks!

Group Slack Workspace:

Please join the Informatics Skunkworks Workspace

Join channel: #ml4er-summer24

Meeting Times

This course has no synchronous lecture times. All lectures will be pre recorded and uploaded to Canvas along with each assignment.

Virtual Meeting Link (Same for all discussion times)

To give the opportunity to meet, ask questions, and discuss ideas we will instead host a variety of virtual drop-in discussion sessions throughout the week to give students the opportunity to get their questions answered in real time. From the times below students are required to join 2 of the discussion sessions to complete a short 10-15 minute attendance activity and check-in on progress. Again, you do not have to attend the entire time and to facilitate smaller group sizes you are encouraged to drop-in later during each time block.

ML4ER Discussion Session times:

- Monday/Tuesday 4:00 - 5:15 pm central time
- Wed/Thurs 9:30 - 11:00 am central time

- Friday 1:00- 2:30 pm central time

Curriculum Description

This curriculum provides students an introduction to using machine learning tools and the associated necessary background on machine learning methods and statistical analysis. Throughout the curriculum, students will focus on using software tools (a focus on MAST-ML with scikit-learn) to generate machine learning models. They will learn key ideas for assessing model performance and decision-making skills for how to improve or modify a model.

It is expected that between synchronous (and asynchronous) meetings, and completion of activities that students will spend ~10-15 hours of time on work related to their participation in the course each week.

Learning Outcomes

By the end of the curriculum, students will be able to:

- Build machine learning models from existing data sets.
- Explain the array of output information needed to analyze models.
- Make decisions on how to modify models to improve model quality.
- Have skills to develop models with MAST-ML and other similar python tools.

Curriculum Materials

- A laptop or desktop computer
- An Internet connection

Prerequisite Knowledge

Students are not expected to have any prerequisite technical skills or knowledge, though familiarity with basic programming concepts and python specifically may provide a smoother transition into using the software tools in this curriculum.

Grading

For students receiving course credit grading is primarily assessed via completion of activities associated with each module in the curriculum and attendance of discussion times. Students will create a **weekly slide deck** which summarizes the results of each activity and a **final slide deck** at the end of the semester that summarizes the progress made over the semester. Weekly slide decks are due at the start of the next activity (see section 3 for a detailed schedule) and will generally follow the schedule of one activity being assigned Monday, and the next on Thursday. For a complete breakdown of the grading see the table below.

Activity	Description	Grading Percentage
Submission of Slide Decks	● Graded on completion via Assessment Figures ● For UW students receiving credit (mse 401) submission through Canvas ● 2 missed assignments or 4 late assignments (or 1 missed and 2 late) before losing 3.5% for each one after	40%
Attendance at Discussion sessions	● 2 unexcused absences before losing 2.5% for each absence after 2 ● Attendance tracked via “attendance activity” completed during the meeting and listed in meeting notes document	30%
Final Slide Deck	● 1 slide for each assignment following template in canvas	30%

Curriculum Structure

The curriculum is divided up into modules with each module taking approximately five hours to complete. Each module has two components. The first is information about a topic of machine learning which is delivered in a variety of mediums such as pre-recorded power point presentations, videos, and readings. The second component is a set of hands-on tasks for students to complete on their own. You can think of this as an extended lab activity. These tasks give students hands-on experience with the machine learning topics and the tools used to perform research oriented tasks and seek to reinforce the ideas presented in the first section.

Part 2: Curriculum Outline

Note: Individual modules will be updated throughout the term, if you plan to work ahead please reach out to check if there are any planned updated that haven't been implemented yet!

(optional) Module 0: Machine Learning Predictions Activity

- Learning Objectives
    - Students will familiarize themselves with a general overview of machine learning and how it can be used to make predictions.
- Activities
    - Machine Learning Predictions video
    - Review overview slides from Prof. Morgan

Module 0.5: Python Basics

- Learning Objectives
    - Students will learn some basics of programming in Python.
    - Students will learn how to import and export data for analysis.
    - Students will become familiarized with some of the structures and logic that computers use.
    - Students will learn how to execute code in a jupyter notebook through Google Colab.
- Activities
    - Python Basics

Module 1: Basics of Machine Learning

- Learning Objectives
    - Students are introduced to model types, key outputs, and metrics for assessing performance
    - Students are introduced to effective practices for maintaining a useful record of research
    - Students continue to learn about model types, key outputs, and metrics for assessing performance
    - Students continue discussing effective practices for maintaining a useful record of research
- Activities
    - Basics of Machine Learning
    - Video Introduction to Machine Learning

        - Powerpoint used in Video Introduction (to follow along)
    - Written Introduction to Machine Learning
        - Additional Resources
    - Introduction to Machine Learning for Materials Science Lab Activity
        - Video: Introduction to Machine Learning for Materials Science Lab Activity
    - Research Compact activity
    - Navigating Roadblocks and Obstacles

Module 2: Establishing Research Workflows

- Learning Objectives
    - Students will learn how Skunkworks is structured and how it will prepare them for undergraduate research projects
    - Student will be able to setup and run MAST-ML on google colab
    - Students will think critically about a dataset and identify good or bad data.
    - Students discuss ways to address roadblocks and obstacles that frequently arise in research
- Activities
    - Establishing Research Workflows
    - Ethical Data Cleaning

Module 3: Comparing Model Types

- Learning Objectives
    - Students are introduced to the MAST-ML software and given a few sample workflows to run on their dataset. The models generated here can then be compared against those previously generated with Citrination.
    - Students explore various cross validation methods in MAST-ML
- Activities
    - Comparing Model Types
    - MASTML Introduction
    - Model Limitations

Module 4: Optimizing Model Hyperparameters

- Learning Objectives
  - Students learn how to modify workflows within MAST-ML to employ the software for various applications.
  - Students understand how MASTML can be used to recreate workflows from other machine learning software
- Activities
  - Optimizing Model Hyperparameters

Part 3: Overview of Schedule

The course will host a variety of discussion session times each week (See section 1), with students expected to join 2 discussion times each week to complete the required attendance activity. The table below gives the outline of the structure for the Summer. Note that specific dates for modular activities may shift as we will adapt to student needs.

In general Activities will be assigned Monday and Thursday, the Monday activity will then be due Wednesday by midnight central time, and the Thursday activity due Sunday by midnight. The 2 attendance activities each week will be due Friday at midnight, but are meant to be completed synchronously during discussion sessions.

Assignment #	**Machine Learning Activity**	**Professional Development Activities**
1 - 7/15/2024	Introduction to Edu Group	module basics of python
2 - 7/18/2024	Basics of Machine Learning	Group Compact
3 - 7/22/2024	Basics of Machine Learning (cont)	Navigating Roadblocks and Obstacles
4 - 7/25/2024	Establishing Research Workflows	Ethical Data Cleaning
5 - 7/29/2024	Comparing Model Types	Reporting Model Limitations
6 - 8/1/2024	Hyperparameter Optimization	
7 - 8/5/2024	Computer Vision	Writing professional emails
8 - 8/5/2024	End of Session Project - cut	(Video) how to be successful in undergraduate research
9 - 8/8/2024	End of Session Project (continued) - cut	
10 - 8/11/2024	Final Slide Decks Due	

Materials Contributors and Authors:

Vanessa Meschke, Matthew Stilwell, Wendy Crone, Anne Lynn Gillian-Daniel, Ryan Jacobs, Benjamin Afflerbach

Acknowledgements:

We would like to thank Citrine Informatics for support in development of initial materials by Vanessa Meschke through the Citrine NextGen Fellowship. We would also like to acknowledge funding for additional development of materials from NSF grant #2017072.