Introduction to Machine Learning for Engineering Research

Part 1: Curriculum Information

Instructor Information

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

Meeting Times

(In-person Students, UW-Madison) Monday/Tuesday from 4:00 - 5:15 PM central time

(Hybrid remote or in-person) Office Hours Wednesday/Thursday 9:30 - 11:00 AM central time (https://sites.google.com/wisc.edu/informatics-help-desk)

In-person students at UW-Madison are expected to attend one of the two meeting times. For Remote Students In-person lectures will be recorded and can be watched on your own time. You may also attend synchronously via Zoom if preferred. Both Monday and Tuesday meetings will be identical apart from any specific questions / discussion that arise.

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 two 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, work groups, and completion of activities that students will spend ~6-9 hours of time on work related to their participation in the course group each week. This corresponds to approximately a 2 credit hour course if receiving credit for an independent study course such as MSE 401 at UW-Madison.

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, attendance of weekly meetings, and attendance and participation in Skunkworks community events. 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 end of the next week after the activity is introduced (see section 3 for a detailed schedule). For a complete breakdown of the grading see the table below.

Activity	Description	Grading Percentage
Submission of Weekly 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 before losing 3.5% for each one after	35%
Attendance at Weekly Group Meeting (Mon/Tues)	● 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	25%
Final Slide Deck	● 1 slide for each weekly activity following template	25%
Attendance at All-hands Meeting	● Let mentor know if hard conflict with the meeting	15%

Curriculum Structure

The curriculum is divided up into modules with each module taking approximately three to 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 power point presentations, videos, and readings. The second component is a set of tasks for students to complete on their own. 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

- **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 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
  - 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 1.5: 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 2: Introduction to Citrination**
- Learning Objectives
  - Students will learn how Skunkworks is structured and how it will prepare them for undergraduate research projects
  - Students will access the Citrination client, upload data, and create a random forest model.
  - Students learn about how to establish group expectations, dynamics, communication
  - 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
  - Introduction to Citrination
  - Ethical Data Cleaning

- **Module 3: Introduction to MAST-ML**
- 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
  - MASTML Introduction
  - Model Limitations

- **Module 4: Modifying Workflows with MASTML**
- 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 (like Citrination).

- Activities
  - Modifying Workflows in MASTML

- **Module 5: Hyperparameter Optimization**
- Students learn about how model hyperparameters can affect performance and are introduced to some basic ideas on how these hyperparameters can be optimized.
- Students familiarize themselves with a hyperparameter optimization landscape and explore various methods of searching for a global minimum in model errors and avoiding getting trapped in a local minimum.
- Students familiarize themselves with neural networks, what they are, how to use them, and how to optimize them.
- Activities
  - Hyperparameter Optimization Presentation Slides
  - Hyperparameter Optimization activity

Part 3: Overview of Schedule

The education group will have 2 meetings a week, with students expected to attend one of the two meeting time options. (Mon/Tues 4:00 PM - 5:15 PM central time) to introduce each activity, discuss new concepts and ideas, and review the results of the previous week's activity. The table below gives the tentative outline of the structure for the semester. Note that specific dates for modular activities may shift as we will try to steer towards student interest and may spend more or less time on some modules.

In general Activities will be due the Sunday after they are assigned at midnight. Any "attendance" activities will be due Tuesday at midnight as they should be completed during or shortly after the meeting during which they are assigned.

Week	**Machine Learning Activity**	**Professional Development Activities**
1 - 2/06/2023	Introduction to Edu Group	module basics of python
2 - 2/13/2023	Basics of Machine Learning	Group Compact

3 - 2/20/2023	Basics of Machine Learning (cont)	Navigating Roadblocks and Obstacles
4 - 2/27/2023	Intro to Citrination	
5 - 3/6/2023	Intro to Citrination (cont)	Ethical Data Cleaning
6 - 3/13/2023	Introduction to MAST-ML	Reporting Model Limitations
7 - 3/20/2023	Modifying MAST-ML Workflows	
8 - 3/27/2023	Hyperparameter Optimization	Writing professional emails
9 - 4/3/2023	End of Semester Project (Data Cleaning)	
10 - 4/10/2023	End of Semester project (Fit default model)	
11 - 4/17/2023	End of Semester Project (Optimize Model)	
12 - 4/24/2023	End of Semester Project (Final model)	
13.5 - 5/5/2023	All-hands Meeting	
13 - 5/1/2023	Final Projects Due	
14 - 5/15/2023	Final Grades Available (UW-Madison)	

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.