**Modules (including HW, key links, notes, etc.)**

How a Module X generally works (although there may be deviations)

1. HW X: Usually reading, video, maybe software tutorial/lab. Due at start of discussion X class.
2. Quiz XA: Simple questions based on HW X mostly to encourage us to do it. Due at start of discussion X class.
3. Discussion X: Discuss the HW and topic of X to lock in key information and answer questions.
4. Lecture X: Lecture on X to share my perspective and further explore aspects of X.
5. Lab X: Hands on lab on X to be started in class and maybe finished for HW. Due ~1 week after end of lecture.
6. Quiz XB: Final harder quiz on X. Due ~1 weeks after lab is due.

Note that above creates a staggered set of interleaved deadlines that can be a pain.

1. Module 1: Introduction to machine learning in materials
   1. Learning goals
      1. Learn the major components of applying machine learning in materials.
      2. Understand a typical workflow for a basic prediction of a physical property for a material.
   2. HW 1:
      1. Install Python Integrated Development Environment (IDE) on your laptop. I suggest Pycharm: <https://www.jetbrains.com/pycharm/> (~20min)
      2. Learn basic Python commands if you do not know it. I suggest the following short course for beginners with video and exercises: <https://www.udemy.com/course/pythonforbeginnersintro/> (~2h).
      3. Read papers about machine learning in materials science (~4.5h). pdfs filenames are listed below and they are located here: <https://drive.google.com/drive/folders/12Yz7ArgEqCn372cRzAruAuQeSD6G9G-I>
         1. Agrawal\_APL Mat\_16\_Mat Informatics 4th Paradigm.pdf
         2. Morgan\_AnnRevMatSci\_20\_Opp\_Challenges\_ML4MSE.pdf
         3. Ramprasad\_NPJ\_Comp\_Mat\_18\_Machine\_Learning\_Materials.pdf
   3. Quiz 1A: None to give us some time to get used to things.
   4. Lecture 1: Gives overview of key parts of ML for MSE and a basic materials property prediction workflow.
   5. Lab 1: Ben’s lab on band gaps – more info coming ???
   6. Quiz 1B: See canvas.
2. Module 2: Getting data
   1. Learning goals
      1. Be familiar with modern data sharing practices and some data resources that are available.
      2. Be able pull data from major databases in an automated fashion.
   2. HW 2:
      1. Read papers (~3h). pdfs filenames are listed below and they are located here: <https://drive.google.com/drive/folders/13VkD6J0UGVu4yf1ylrklSWN_dAxgdQdD>
         1. Scheffler\_Nature\_22\_FAIR Data and MSE
      2. Watch video on extracting data from online databases: <https://www.youtube.com/watch?v=hzG1UNvfbFU>
   3. Quiz 2A: See canvas.
   4. Lecture 2: None.
   5. Lab 2: Nanohub lab on data extraction, maybe updated with chemical database and some ideas on FAIR models and shared models. See Nanohub tool here: <https://nanohub.org/resources/34288>
   6. Quiz 2B: See canvas.
3. Module 3: Featurization.
4. Module 4: Model assessment, optimization, and Uncertainty Quantification.
5. Module 5: Deep learning (molecular properties)
6. Module 6: Image analysis (classification, object detection)
7. Module 7: Text analysis (NLP)
8. Module 8: Machine learning potentials