AMATH 482/582: Computational Methods for Data Analysis

Course Syllabus - Winter 2022

Department of Applied Mathematics
University of Washington

Course Description:
This course offers an overview of various computational tools for analysis of data sets and extraction of meaningful information. Increasingly we all deal with large amounts of data in our lives but raw data is not useful until we can make sense of it. This course offers you a collection of tools for this purpose with many opportunities for you to obtain hands-on experience in using them. We will not go deep into the theory but include some discussion to inform you of how and why different approaches work. The course is split into two parts: Linear methods, which covers a lot of the classical approaches to data analysis that are broadly based on linear analysis and approximation theory; and nonlinear methods, which covers more advanced methods that rely on nonlinear transformations and approximation theory.

Learning Outcomes:
- Identify suitable computational tools for analysis of data sets.
- Implement, prototype and validate new data analysis algorithms and tools.
- Obtain foundational understanding of how and why algorithms work.

Prerequisites: *Familiarity with a scientific programming language (MATLAB, Python etc) and college-level coursework in linear algebra, or permission of instructor*

Class Time and Place:

AMATH 582: MWF 830am - 9:20am LOW 216
AMATH 482: MWF 9:30am - 10:20am MEB238

Course Instructor:

Bamdad Hosseini
Office: Lewis Hall 220
Email: bamdadh@uw.ed

Teaching Assistants:

Katherine Owens
Email: lacyk3@uw.edu

Roman Levin
Email: rilevin@uw.edu

Office Hours:

Three hours per week held over zoom.

Times and Links:

- Instructor office hours: **3 - 4 pm on Wednesdays, Jan 19 - Mar 16**

  Zoom link: https://washington.zoom.us/j/97376054146?pwd=bWQ3b0U5M3FmTm1EbWRGVFlBUXpHUT09
  Pass code: **482582**

- TA office hours:

  - Katherine Owens: **1:30-2:30 pm on Thursdays, Jan 20 - Mar 17**
    Zoom link:
    https://washington.zoom.us/j/93825273159?pwd=Vk1nb0hocURUdzE0WERkNWN6TENIdz09
    Pass code: **482582**

  - Roman Levin: **5-6 pm Fridays, Jan 21 - Mar 18**
    Zoom link: https://washington.zoom.us/j/97503319463?pwd=UWdpaUlPMGxENkNhT2JzM1ozY2N2QT09
    Pass code: **482582**

Assessment Rubric:

Your grade will be determined entirely from your homework (100%). There will be 5 homeworks over the quarter. See below for instructions and further information about these homeworks.

Suggested Texts and References:

Instructor notes are the best resource for the course as I will be compiling those from numerous sources. I will also provide further reading materials pertaining to different lectures.

With that in mind, if you would like a textbook to have nearby then I suggest "Data-driven modeling & scientific computations" by Nathan Kutz.

Course Website or Learning Management System:

We will use Canvas to manage the course. Lecture videos will be uploaded.

You can use the discussion section to post questions and seek help/advice from your peers. **Note the discussions section is meant for you to talk to each other and will be minimally moderated by the TAs.**

I will upload recorded lectures to Canvas in case you want to follow them at your own pace or simply review the material. There will be two different versions, one that I will record myself outside of class and another Panopto recording of the class lectures. The former will be higher quality but the latter includes in class questions from students.

*You may watch and download videos of the in person lectures from this class's Panopto Video Folder. See Viewing Panopto Recordings for more information on using Panopto. If you need help with Panopto issues, please contact UW-IT Support: help@uw.edu

Attendance and Participation:

I will not take attendance however I ask that you attend the lectures assigned to your section (i.e. no 482 students in 582 and vice versa) in order to manage classroom capacity.

All lectures will be uploaded online and you are welcome to follow them at your own convenience.

COVID Policy

Up to date UW policy for students: https://www.washington.edu/coronavirus/students/

Academic Misconduct

Students at the University of Washington are expected to practice high standards of professional honesty and integrity as described in the Student Academic Responsibility statement. **See also Collaboration Policy under Homework Instructions and Advice below.**

Accommodation

Your experience in this class is important to me. If you have already established accommodations with Disability Resources for Students (DRS), please communicate your approved accommodations to me at your earliest convenience so we can discuss your needs in this course.

Inclusivity

Among the core values of the university are inclusivity and diversity, regardless of race, gender, income, ability, beliefs, and other ways that people distinguish themselves and others.

Religious Accommodation

Washington state law requires that UW develop a policy for accommodation of student absences or significant hardship due to reasons of faith or conscience, or for organized religious activities. The UW's policy, including more information about how to request an accommodation, is available at Religious Accommodations Policy. Accommodations must be requested within the first two weeks of this course using the Religious Accommodations Request form.

Incompletes

An incomplete grade (I) is given only when a student has done satisfactory work up until the last two weeks of the quarter but cannot complete the remaining work because of illness or other circumstances beyond the student's control.

Course Calendar (Tentative)

Note this is a tentative schedule and will most likely change as we make our way through the course. It is presented here to give you an idea of what we plan to do.

	Mon	Wed	Fri
Linear Methods			
Week 1 (Jan 3-7) (Signal Processing)	Intro + review of prob and stat	Fourier transform and series	Fourier transform demo
Week 2 (Jan 10-14)	Windowed fourier transform	Wavelets	Multi-Resolution Analysis

(Signal Processing)			and the Wavelet Basis **(HW1 in out)**
Week 3 (Jan 17-21) (Signal Processing)	**MLK day**	Linear filtering and denoising	Filtering demo
Week 4 (Jan 24-28) (Principle Comp. Analysis)	Linear algebra review + SVD	PCA + PCA Demo	PCA + POD + ICA + KL ***HW1 is due** **(HW2 is out)**
Week 5 (Jan 31-Feb 4) (Machine Learning)	PCA + POD + ICA + KL	Introduction to ML	ML Demo Linear discriminant analysis
Week 6 (Feb 8-11) (Kernel Methods)	Intro to kernels	Kernels and feature maps	Kernel methods demo ***HW2 is due** **(HW3 is out)**
Nonlinear Methods			
Week 7 (Feb 14-18) (Kernels + Graphs)	Kernels and Gaussian Processes	Spectral clustering	Spectral clustering demo
Week 8 (Feb 21-25) (Graphs)	**President's day**	Graphical semi-supervised learning	SSL demo ***HW3 is due** **(HW4 and 5 are out)**
Week 9 (Feb 28-Mar 4) (Promoting sparsity)	Basics of CS	Image reconstruction with CS	Advanced topics in CS
Week 10 (Mar 7-11) (Neural networks)	Introduction to NNs	Image Classification with CNNs	ResNets and ODEs ***HW4 is due**
Week 11 (Mar 14-18)			***HW5 is due**

Homework Instructions and Advice

During the quarter, you will receive 5 homeworks that you will turn in via Canvas. These 5 homeworks are equally weighted and worth 100% of your grade. This homework should be written as if it were an article/tutorial being prepared for submission. I expect a high level of professionalism on these reports. The following is the expected format for homework submission:

- **A PDF file of a maximum of 6 pages including references. Minimum font size of 10pts and margins of 1 in on A4 or standard letter size paper.**
- You **do not need to** include your code in the report but need to upload it separately along with the PDF report.
- Your report should be formatted as follows:
    - Title/author/abstract: Title, author/address lines, and short (100 words or less) abstract. This is not meant to be a separate title page.
    - Sec. 1. Introduction and Overview
    - Sec. 2. Theoretical Background
    - Sec. 3. Algorithm Implementation and Development
    - Sec. 4. Computational Results
    - Sec. 5. Summary and Conclusions
    - Acknowledgements (See Collaboration Policy below)

    - References
- I suggest you use LaTex (Overleaf is a great option) to prepare your reports. A LaTex template will be provided as a starting point. You are also welcome to use Microsoft Word or any other software that **properly** typesets mathematical equations.

Your homework will be graded based upon how completely you solved it as well as neatness and little things like: did you label your graphs and include figure captions. **Each homework is worth 20 points.** 10 points will be given for the overall layout, correctness and neatness of the report, and 10 additional points will be for specific technical things that the TAs will look for in the report itself.

Homework Advice:

The report does not have to be long. But it does have to be complete.

This report is not for me but for yourself. So write a nice report so that you could reproduce the results if you need the methods addressed here in the future. Alternatively, view the reports as practice for publishing scientific research in the future.

A few more things to keep in mind:

- Use a professional grade word processor (Latex or Microsoft word, for example).
- For equations: Latex is state-of-the-art, but Word also has a decent equation editor.
- Label all of your graphs, figures and tables and include a brief caption. All graphs, figures, and tables should be referenced and discussed in the main text.
- Make sure axis labels and legends on your figures have sufficiently large fonts to be readable without zooming.
- Label/number all equations that are referenced in the text. If an equation is not referenced later in the text it can be left without a label/number.
- Provide references to scientific documents such as papers, books, arXiv preprints where appropriate. Avoid online references such as Wikipedia.
- Always spell check twice before submitting.

Collaboration Policy

I encourage collaborations however **everything that is handed in, both your report and your code, should be your work**. You are welcome to discuss your assignments with your peers and seek their advice but **these should be clearly stated in the acknowledgements section of your reports.** This also includes any significant help or suggestions from the TAs or any other faculty in the university. You don't need to give all the details of the help you received, just a couple of sentences.

Coding Advice:

Python

I strongly suggest using Python as it is more widely used in data science and machine learning. It is the best language for your future prospects in industry and academia. It also has a lot of tools that are not available in MATLAB.

Python 3: If you are going with Python make sure you download the latest version of Python 3. We will mostly use **Numpy, Scipy, and Scikit-learn**.

Anaconda: I also suggest you download and install anaconda. It is an excellent environment management system for python. It will make your life a lot easier. I can also provide conda files for demos and assignments.

Jupyter Notebooks: I will be using Jupyter notebooks for our python demos. It is very convenient, light, and eliminates the need for an IDE. You can hand in your homework codes as a notebook.

Running code on the cloud: There are many excellent services out there that allow you to run your Jupyter notebooks online without needing to install anything on your computer. Two examples include Google Colaboratory and Microsoft Azure. These have a lot of the packages you may require pre-installed. More examples can be found here.

MATLAB

A MATLAB license is provided to UW students for free. If you decide to use MATLAB you can download it through the UW portal at Mathworks.

LaTeX

I suggest you use LaTeX to prepare your homework reports. This is not mandatory but knowing LaTeX is a very useful skill especially if you need to prepare documents with a lot of mathematical equations.

Overleaf: Everything you need for the course can be done on Overleaf. It eliminates the need for downloading and installing your own LaTeX distribution which can be large and time consuming. They also have one of the best LaTeX tutorials around.

Version Control

This is totally optional and not at all required but you may set up a GIT repository and upload your project codes to keep track and maintain them for the future. Here is an excellent video tutorial on GIT provided by colleagues at UW.