sophieclayton / **OEAS805_envdatasci** Public

<> Code ⊙ Issues ⏚ Pull requests ▷ Actions ⊞ Projects ⚠ Security ⬓ Insights

ᛉ master ⌄

OEAS805_envdatasci / homework / **capstone_instructions.md**

🧑 sophieclayton Update capstone_instructions.md 🕐 History

👥 **1 contributor**

≡ 83 lines (56 sloc) | 6.76 KB ···

Capstone Project Instructions

The primary goal of this course is to equip you with skills to manipulate and integrate different data types and to introduce basic machine learning tools to address problems in oceanography. As part of the course, you will undertake an independent project which relates directly to your MS/PhD research. For your project, you will develop a machine learning model to predict values of a target variable that is challenging to measure directly over large areas or at high spatial/temporal resolution. Your model will utilise a set of other, easier to measure (or model) variables to predict your target variable. It is generally a good idea for these other variables to be quantities that are easier to measure directly or remotely than your target variable, or commonly produced as ocean circulation model output (some examples might be SST, salinity, chlorophyll, sea level).

Your project should utilise data from a range of sources, including at least one outside source (i.e. not directly collected by you). I also expect to see some sensitivity analysis of models and model parameters to justify your ultimate choice of model algorithm and structure.

Components of the capstone project

You will each create a github repository for your course project. The project is broken down into the following elements which will build towards your final project report.

1. Capstone Project Proposal (HW5)

(50 points, due **Friday 17th March, 5pm**)
Details on this assignment here.

2. Class Presentations

(15% total course grade, **Tuesday 18th and Thursday 20th April**)
Talk schedule can be found here

Each student will have a total of 15 minutes to present their project and answer questions. Your talk should be 10-12 minutes and cover the following:

- Background on the problem you are trying to solve
- Data that you have used
- Computational tools that you have used
- Results/Implementation
- What would you do differently next time or where will you take the project next?

As a rule of thumb, allow no more than 1 slide per minute of your talk. I **highly recommend** that you practice your talk at least once prior to the in-class presentations. It will allow you to check that you are within the time range, and that your material flows well and make sense. It is usually pretty obvious when a speaker has not practiced their talk.

In terms of presentation, I will expect to see axis labels and titles on any plots that you present and you must reference any external materials (e.g. figures form papers) that you use in your talk. There is no need to provide a reference slide. Avoid crowding your slides with too much text, it is very hard for your audience to read and listen at the same time.

3. Peer Review of Project Repositories

Required but ungraded (if not completed, your project repository will not receive a grade). Due **Friday 14th April**.

Each student will be assigned a peer's GitHub repository to review prior to the Project Repository submission deadline. This is intended to provide useful feedback to improve the accessibility of the code and documentation in your Project Repository by enlisting each other as "beta testers". Review assignments will not be reciprocal, i.e. you will not review the repository of the person who is reviewing your repository.

For this assignment, you will review the requirements listed below for the Project Repository and provide an informal review for your reviewee (and me) based on how well their repository meets the requirements listed below. This can be a simple Slack message, but you should thoroughly go through the repo.

Here are your reviewer assignments for this part of the project.

4. Project Repository on GitHub

(15% total course grade, due **Friday 21st April**)

All of your code should be clearly organised in your project github repository. As a general guideline, you should provide sufficient background information and commenting in your repository and code for someone to replicate your analysis. Some specific requirements:

- Your repository **must** include a README that gives an overview of the project, instructions for using the code, and enough information to allow the user to access the data you used (this can be in an external shared folder, just make sure that you give me access!).
- You should include either a `requirements.txt` or `environment.yml` file with the relevant package/version information.
- Your code should be adequately commented, including descriptions of key variables.
- Don't forget to list which version of python the code is written in as well as any and all packages you have used.

5. Project Report

(20% total course grade, due **Tuesday 25th April, 5pm**)

You will write up the results of your project as a final project report. Please treat this as you would a research paper, including:

- Abstract (250 words max)
- Introduction
- Data
- Methods
- Discussion
- Conclusions

Be sure to properly reference the relevant literature using a citation style of your choice (consistently). The project report should be submitted as a pdf document using the following formatting:

- 11-12 pt font

- Single spaced lines

- Include line and page numbers

- Figures and tables may be in line with the text and must be numbered and include a title.

- **6 page limit** excluding references

[Give feedback](#)