CSE 538 Fall 2023

CSE 538 Fall 2023

Natural Language Processing

This will be a research and project-based course that introduces you to research topics in Natural Language Processing.

Syllabus

Detailed syllabus with lecture topics and dates will be made available in the first week of class. The topics covered will depend on the specific set of topics chosen by the students. The goal is to cover some of the following:

1. Word Representations (Word2Vec, GLoVE)
2. Sentence Representations (DAN, RNN, CNN, Transformers)
3. Language Modeling (BERT, GPT, T5 etc.)
4. Language Generation
5. Prompting, In-Context and Instruction-based Learning
6. Efficiency Considerations and Long Context Modeling
7. Applications: Question Answering and Machine Translation

Schedule

- Will follow soon

Requirements

H[illegible] is a list of things that would be useful for this class. I won't be able to respond to individual requests on whether your background is suitable. Please use the following to make your own determination.

The following are critical. If you are completely unaware of the following then you will likely have difficulties following material in class.

CSE 538 Fall 2023

Strongly Recommended:

- Basic probability and statistics (joint and conditional probabilities, Bayes rule etc)
- Basic linear algebra (vector and matrix operations)
- Basic calculus (differential calculus)
- Machine learning basics (classification, basic ml recipe)

The following are useful but some of these will be covered in class and with some effort you can pick these up as we go along.

Can be picked up as we go along:

- Deep learning basics (neural networks, feed-forward, sequence models, etc.)

Books and Reading Material

I am not going to follow any book. Material presented in class will be in slides. Your best bet for learning is to attend class and follow up on pointers I give out to reading materials in class.

That said, there are many books and reading material out there on the web. Here are a few you may find helpful:

- Jacob Eisenstein's NLP Book (pdf)
- Dan Jurafsky's SLP3 (multiple pdfs)

Course Structure

This will be a project and research-based course. There will be some introductory lectures in the first part of the course. The students are expected to execute a research project through the course. I will assist with the selection of topics and design. We will read papers on topics related to the projects chosen. The students will be expected to present and lead discussions. You will also write summaries of the papers discussed.

At the end of the course, you will write a project report, do a demo, and present the results.

- **Project (60%)**
- **Midterms (20%)**
- **Presentation (10%)**
- **Paper Summaries (10%)**

Accessibility

If you have a physical, psychological, medical, or learning disability, please contact the Department of Student Affairs. They will determine with you what accommodations if any, are necessary and appropriate. All information and documentation of disability is confidential.

We will make every effort to support accessibility needs for all parts of the course. Please contact me via email to make specific arrangements.

Critical Incident Management

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Student Conduct and Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn.

Grading CSE 538 Fall 2023

I will likely make adjustments to the grading scheme based on the overall performance of the class. Here is a tentative grading rubric:

A: 90 and above

A-: 80 or more but less than 90

B+: 75 or more but less than 80

B: 70 or more but less than 75

B-: 65 or more but less than 70

Five point intervals for lower letter grades.

Academic Integrity

In this class, we encourage collaboration. Whenever possible we will clearly state what forms of collaboration are allowed and what aren't. Of course, it is near impossible to list all forms of unethical or dishonest behavior. You can consult the SBU website on Academic Integrity for more information.

While a masters program can be an excellent opportunity, it can also be difficult and exhausting, here are some of my thoughts on ways in which you can structure your expectations for your MS program that might be of help to you.

Grades are hardly the point of grad classes. Please don't cheat.

- It is hardly worth the risk.
- It is often very easily detected.
- Part of your training is to learn how to make ethical decisions.
- If you are under difficult circumstances of any kind, come talk to me about it.
- When in doubt, cite the sources from which you got content/code/ideas and give credit to people who you worked with.
- When in doubt, ask the instructor or the TAs before engaging in any specific forms of collaboration or use of outside material.

Here is the official statement from SBU on academic integrity, which I endorse and will follow for this class:

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty is required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at http://www.stonybrook.edu/commcms/academic_integrity/index.html