

CS3332 Machine Learning Introduction, Syllabus for Fall 2024.

Catalog description

A survey of methods to process, classify and analyze large amounts of data using machine-learning methods.

Big-data topics examine the obstacles to processing including managerial obstacles, problems of data consistency, problems of data accuracy, data-reduction methods,

and Big-data distributed processing methods. Topics on machine learning include concept learning, decision trees, Bayesian models, linear models, neural networks,

case-based reasoning, genetic algorithms, sequence learning, and assessment techniques. There are some variations in the topics from quarter to quarter.

Instructor:

Prof. Arijit Das, adas@nps.edu, contact via Teams for 1-1 questions.

Textbook and notes:

Textbook: I. Witten, E. Frank, M. Pal, and C. Hall, Data Mining: Practical Machine Learning Tools and Techniques, fourth edition, Morgan Kaufmann, 2016.

The video lectures from the authors of the book will be used in the learning of the concepts for this entire class. There are some good videos explaining Weka and data mining at www.youtube.com/user/WekaMOOC. For the class video lectures in SAKAI click on the "i" next to the name of the video file, and you will get the password.

Getting the book from O'Reilly:

To access O'Reilly you need to go through the Navy MWR Digital Library in order to get a free O'Reilly account; go to www.navymwrdigitallibrary.org.

To get started, log in to the digital library with a CAC (or register for a DS Logon account). Once you've logged into the library, find the O'Reilly section

and use the 'Visit' button to register for an O'Reilly Learning account.

A shortcut is <https://www.oreilly.com/self-registration/dod-mwr-libraries-navy/>.

After your O'Reilly account has been established, you can access it from outside the Digital Library at <https://learning.oreilly.com> or downloading the O'Reilly app.

Lecture:

Tu/TH: 1700-1900 PST.

Join ZoomGov Meeting

<https://nps-edu.zoomgov.com/j/16124674856>

Meeting ID: 161 2467 4856

Passcode: B!g2jLPcQX

Quizzes:

Open book & open notes. The quiz will be an assignment that you can do over a few days. It will be independent effort and no collaboration.

Software:

As a data scientist, you need to be familiar with a variety of tools. In particular, you will need to run these open-source tools, which will require installing them on a personal computer (Windows O/S).

WEKA (https://waikato.github.io/weka-wiki/downloading_weka/): This is a simple machine-learning tool that does most of the work for you, as it is intended for people that are not programming experts. The only major challenge is in creating input in a format that it will accept.

SQL Developer and the Oracle XE 11G Database:

SQL Developer is a GUI (graphical user interface) to the Oracle XE Database (both running on your laptop).

I will provide instruction for both in a SAKAI folder.

Once connected you can load CSV files into Database tables, use SQL queries to clean up the data and save it back into a CSV file.

Excel: You are free to use Excel for some data cleanup and analysis also.

You are best to use a Windows machine for this class as the Oracle XE Database runs on Windows.

Grading:

Assignments will be graded one/zero and expect about 12 mini assignments. In November we will have our first quiz of the class we will have 2-3 quizzes. Sample quizzes and starting material will be provided in advance. Your best preparation will be to do the assignments as the quizzes will be extensions of the assignments.

All assignments have to be completed and returned with a check for a minimum of B else grade will be less than a B (B-, C+, C depending on # of assignments completed).

If students get an A in both quizzes, then their final grade will be an A, else it will be A-, B+ or a B.

I tend to grade easy so you can get a high score but depending on how you compare with others your grade can be anything from a B to an A.

Suggested readings to enhance your understanding (book mentioned above)

Chapters: 1, 2 and 3.

Sections: 4.1, 4.2, 4.3, 4.4, 4.5, 4.7, 4.8, 5.1, 5.2, 5.3, 5.8, 6.1, 6.2 and 8.1.

Section 7.2: introduction, The Maximum Margin Hyperplane, Nonlinear Class Boundaries, and Multilayer Perceptron including Backpropagation.

The videos at www.youtube.com/user/WekaMOOC also provide learning about Weka Machine Learning concepts.

Course Outline

- Part 1: Introduction and data setup
- Part 2: Data manipulation (wrangling)
- Part 3: Introduction to machine learning
- Part 4: Logical learning
- Part 5: Decision graphs
- Part 6: Bayesian models
- Part 7: Neural networks
- Part 8: Case-based reasoning
- Part 9: Sequence learning
- Part 10: Evolutionary learning (optional)
- Part 11: Further directions and conclusions (optional)