

(http://www.mimicry.com/succeed.html.php?)

Succeed in OMSCS (/)

helping omscs students find information

Sign in



CS 7641 Machine Learning

Menu

Preparation Advice

CS 7641 Machine Learning is not an impossible course.

But it is a hard course.

Preparing in advance is a good idea, since from the beginning you will need to review (learn) a lot of information before you can start working on the first assignment.

CS 7641's Syllabus is very similar to this one (http://www.cc.gatech.edu/~isbell/classes/2009/cs7641_spring/) (except that there's no group project for the OMSCS version). Assignment 1 covers lessons 1-6 from the "Supervised Learning" section of the course, so in a short window of time you need to: watch the lectures, work on the assigned readings, pick two datasets (and clean/preprocess them), learn a ML framework (Weka/Java/Python/R/Matlab/etc), run experiments many times, write a 12-page paper.

Many people feel overwhelmed due to all this work, and end up submitting a weak assignment. Because of that, a recommended preparation would be:

- 1) **Theory:** Watch lectures 1-6 in advance. Try to understand them. Take notes. If you find many concepts too high level and would like an introduction, watch other videos like Andrew Ng's (a very popular choice). Also you need to know in advance: Multivariate Calculus, Linear Algebra, Statistics and Probability. If you took those classes in undergrad, you should be fine. If not, a MOOC on those topics could help.
- 2) **Choose an ML framework:** Python, R, Matlab, Java, it's up to you. But it's recommended to use a language that you are already proficient in. If you don't do that you will dedicate (waste) time to learn the language, while you could be using that precious time running experiments.

The Packt books: Machine Learning with R (<https://www.packtpub.com/big-data-and-business-intelligence/machine-learning-r>) and Python Machine Learning (<https://www.packtpub.com/big-data-and-business-intelligence/python-machine-learning>) are very recommended. They explain not only ML APIs and libraries, but also relevant ML concepts (theory). It's important that you find a way to automate the execution of experiments with different parameters (the caret library in R, scikit-learn in python, etc).

(TO-DO, information about WEKA, Matlab, and other frameworks/libraries).

- 3) **Pick your datasets:** For the assignments you need to choose two datasets. Unless you have already worked extensively on ML and want to use this class to do something fancy, it's better to keep things simple. Choose datasets from the UCI Repository (<http://archive.ics.uci.edu/ml/datasets.html>), it's better if you choose classification datasets. It's not a requirement, but again, if you are a newbie it's better not to overcomplicate things (gigantic datasets, dirty datasets, etc). There's no hard rule, that's why many people "waste" time in this step. Once you have your "candidate" datasets, apply what you learned in the step #2 above, and run a few supervised learning algos over them and "see what happens".

Congratulations! At this point you should already have a head start for the course. Have fun.

Resources

Fall 2015 course schedule with the list of readings is available here (<http://omscs.wikidot.com/local--files/courses:cs7641/CS7641-Fall-2015-Schedule.pdf>).

The required textbook for the course is Machine Learning by Tom Mitchell, 1997 (<http://www.cs.cmu.edu/~tom/mlbook.html>).

Software suggestions for Assignments (from preceding semesters' reviews):

- Assignment 1 - Weka (<http://www.cs.waikato.ac.nz/ml/weka/>) (many also used Python and R)
- Assignment 2 - ABAGAIL (<https://github.com/pushkar/ABAGAIL>) (This has a lot of starter code to help you)
- Assignment 3 - Scikit Learn (<http://scikit-learn.org/stable/>) (Weka has ICA missing)
- Assignment 4 - BURLAP (<http://burlap.cs.brown.edu/>) (Python's or R's mdptoolbox can also be used)

Machine Learning with R:

Notes on R (<https://docs.google.com/document/d/1ceUoFEpr3UpDIR4rpYQ3RgKyNs-bR0DF3xkyYB2Ojrs/edit>), by Brent Wagenseller

knitr (<http://yihui.name/knitr/>): Elegant, flexible and fast dynamic report generation with R

caret (<http://topepo.github.io/caret/index.html>): Set of functions that attempt to streamline the process for creating predictive models.

Learning Ensembles with R (<http://machinelearningmastery.com/machine-learning-ensembles-with-r/>)

Machine Learning in R for Beginners (<https://www.datacamp.com/community/tutorials/machine-learning-in-r#gs.pdhCt5A>) (DataCamp tutorial)

Machine Learning with Python

Mimicry (<https://github.com/mjs2600/mimicry>)

Using ABAGAIL and Jython: <https://www.youtube.com/watch?v=oFvQsArCSXo> (<https://www.youtube.com/watch?v=oFvQsArCSXo>)

Fellow Student - github repo: shared machine learning algos for learning purposes

Scikit-learn (<http://scikit-learn.org/stable/>) - A common, easy to use Python machine learning library

Additional readings and tutorials:

Slides for Tom Mitchell Machine Learning Book (<http://www.cs.cmu.edu/~tom/mlbook-chapter-slides.html>)

Tom Mitchell's Machine Learning new chapters. (<http://www.cs.cmu.edu/~tom/NewChapters.html>)

Tom Mitchell has posted old hws and exam material for his past classes:

1. <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-3/www/ml.html> (<http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-3/www/ml.html>) (1998)
2. <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/mlc/> (<http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/mlc/>) (2003)

The Open Source Data Science Masters (<http://datasciencemasters.org/>)

Deep Learning: MIT Press (HTML version free) (<http://www.deeplearningbook.org/>) - Ian Goodfellow (Google), Yoshua Bengio and Aaron Courville (both University of Montreal)

Journey into Machine Learning at Georgia Tech OMSCS: Tips and Considerations (<https://www.whenthereisdata.com/2016/01/journey-into-machine-learning-at-georgia-tech-omscs/>)

A student-created cheatsheet with a list of key topics (https://github.com/eternalmothra/ml_cheat_sheet/blob/master/ml_cheat_sheet.pdf)

Help (<http://www.wikidot.com/doc>) | Terms of Service (<http://www.wikidot.com/legal:terms-of-service>) | Privacy (<http://www.wikidot.com/legal:privacy-policy>) | Report a bug | Flag as objectionable

Powered by Wikidot.com (<http://www.wikidot.com>)

Unless otherwise stated, the content of this page is licensed under Creative Commons Attribution-ShareAlike 3.0 License (<http://creativecommons.org/licenses/by-sa/3.0/>)