CS671-Machine Learning
SPRING 2023

Prof. Dan A. Simovici

Office hours: MW 3:00 - 4:00 pm

This page is posted on www.cs.umb.edu/~dsim; on the same site you will find copies of the slides I am using in class, homework, and handouts relevant to the course. You should visit it often!

Machine Learning is a mathematical discipline that is the foundation for Data Mining and Data Analysis. I expect that you had previous courses in linear algebra and probabilities, but everything possible will be done to make the course self-contained. The main sources for this course are:

S. Shalev-Shwartz and S. Ben-David: Understanding Machine Learning, Cambridge University Press, 2016;

Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.T. Lin, Learning from Data, AMLbook.com, 2012;

M. J. Kearns and U. V. Vazirani, An Introduction to Computational Learning Theory, MIT Press, 1997;

K. P. Murphy: Machine Learning - A Probabilistic Perspective, MIT Press, 2012;

M. Anthony and N. Biggs: Computational Learning Theory, Cambridge University Press, 1997.

D. Simovici and C. Djeraba: Mathematical Tools for Data Mining, 2nd edition, Springer 2014.

In addition, I will use several classical probabilities texts. Your primary source for this course should be the slides that I will present in class and place on the web site.

We shall cover the following topics:

1. Probably Approximately Correct (PAC) Learning
2. The Vapnik-Chervonenkis Dimension
3. Support Vector Machines
4. Decision Trees
5. Boosting
6. Regression
7. Stochastic Gradient Descent
8. Kernel Methods
9. Clustering

Homework should be entirely the product of your work; you may discuss it with colleagues and I encourage you to talk to me if you have difficulties. Cheating in any form will be severely sanctioned. Learn LaTeX and use it to write your homework. Homework and class participation determine your grade.

Machine Learning is a wonderful subject and I hope that you will enjoy it!