JHU Engineering for Professionals
Applied and Computational Mathematics
Data Mining: 625.740

Instructor: M. Weisman, PhD
Email: data.mining.625.740@gmail.com

Prerequisites: Multivariate calculus, linear algebra, probability and statistics, experience programming and access to a computer.

Topics Covered: Bayesian decision theory, Gaussian mixture models, supervised learning, unsupervised learning, principal components analysis, k-means and fuzzy k-means clustering, expectation-maximization, perceptrons and neural networks, support vector machines.

Required Textbook: Duda, Hart, & Stork, **Pattern Classification**, Second Edition, Wiley.

Additional References:

- **Data Mining: Practical Machine Learning Tools and Techniques**, Witten and Frank, Morgan Kaufmann.

- **Data Mining with R: Learning with Case Studies**, L. Torgo, CRC Press.

- **Elements of Statistical Learning: Data Mining, Inference, and Prediction**, Hastie, Tibshirani, and Friedman, Springer (`http://www-stat.stanford.edu/~tibs/ElemStatLearn/`).

- **Google's PageRank and Beyond: The Science of Search Engine Rankings**, Langville and Meyer, Princeton University Press.

- **Introduction to Data Mining**, Tan, Steinbach, and Kumar, Addison Wesley.

- **Introduction to Machine Learning**, Alpaydin, MIT Press.

- **Introduction to Pattern Recogntion: A MATLAB Approach**, Theodoris, et. al, Academic Press.

- **An Introduction to Statistical Learning: with Applications in R**, G. James, D. Witten, T. Hastie, and R. Tibshirani, Springer.

- **Introduction to Statistical Pattern Recognition**, Fukunaga, Academic Press.

- **Pattern Recognition**, Theodoridis and Koutroumbas, Academic Press.

Student Evaluation: The final grade for this course will be determined from a weighted average of the exams, student project, and homework.

Homework:	10%
Project:	30%
Exams:	60%

Homework Submittal: Homeworks will typically be posted at the beginning of class each week. We are requesting that homeworks by submitted electronically to `homework.625.740@gmail.com`, by the beginning of the next week's class. Homeworks will be graded

1. For completeness. Points will be scored for making an effort on each problem;

2. For correctness. Not every problem will be graded 'in depth' each week. A few (or possibly all) the problems will be selected for 'careful grading'.

You may collaborate on the homeworks but please have the writeups represent your own understanding. All sources should be referenced. Late homeworks will not be accepted.

Exams: The final and midterm exams will include a 'take-home' component involving the use of a computer. These take-home problems may be implemented in a higher-level language such as C, Java, Mathematica, MATLAB, Octave, Perl, Python, R, Weka, etc.

Academic Misconduct Policy: Students should read policies pertaining to academic misconduct at `http://ep.jhu.edu/genpolguid`.

Collaborations and discussions between students are key ingredients to success in a graduate course. You are encouraged to discuss the course material with each other as you sort through concepts that may be difficult to comprehend or controversial. However, the line between collaboration and cheating needs to be carefully delineated. Whenever you turn in work with your name on it to be evaluated, graded and included in your record it must represent an individual effort by you alone. If you include direct quotes from any source in your discussions, written assignments, the final exam, or any other submission for which you will receive a grade you must provide attribution. Students using published material without reference, or copying the work of another individual will receive a warning at the first incident. Any further incidents will result in the student receiving a zero on the assignment and the matter will be referred to the Associate Dean. Contact us if you have any questions, no matter how slight, about this policy, or if you have questions about a particular assignment.

Plagiarism: Plagiarism is defined as taking the words, ideas or thoughts of another and representing them as one's own. If you use the ideas of another, provide a complete citation in the source work; if you use the words of another, present the words in the correct quotation notation (indentation or enclosed in quotation marks, as appropriate) and include a complete citation to the source. See the course text for examples.