BMIF330 (CS396) — Machine Learning for Biomedicine (Draft 1/07/09)

Spring 2009   Tuesdays and Thursdays, 4:00 pm to 5:15 pm (EBL 456)

Instructors: Subramani Mani and Hua Xu
{subramani.mani, hua.xu}@vanderbilt.edu

     This course builds on the material covered in BMIF315 by introducing several additional machine learning concepts and algorithms with a focus on biomedical decision making and discovery. Note that even though biomedical applications and examples will be discussed, the methods have broad applicability in science and engineering. The following are the core topics that we plan to cover in this course. However, certain topics may be expanded or modified based on the background of the class participants.

Pre-requisites:

For DBMI students: Methodological foundations of Biomedical Informatics (BMIF315).

For non-DBMI students: 1. A course in data structures or algorithm design and analysis.

     2. Ability to program in Matlab version 6 or later.

     3. Basic concepts of machine learning and fundamental mathematical concepts needed in machine learning at the level covered in BMIF315. CS260/360 would provide an appropriate background as a starting point. The students are encouraged to attend the BMIF315 lectures and complete the assignments for the topics not covered in CS260/360.

Grading: Based on mid-term, final, term project, assignments and class participation.

Required/Recommended Course Textbooks:
Recommended: [HTF01, HL04, MS99].
There will also be other reading assignments as appropriate.

Software:

- Matlab (base + NN toolkit + optimization toolkit + statistics toolkit)

- Various other code to be distributed or downloaded from the web

Topics

1. Natural language processing and text mining

2. Bayesian networks (Definition, properties, inference, applications)

3. Decision trees

4. Feature subset selection

5. Neural networks

6. Hidden Markov models

7. Support vector machines

8. Bayesian networks (structure and parameter learning)

9. Causal discovery using Bayesian networks

References

[HL04]   Duane C. Hanselman and Bruce L. Littlefield. *Mastering Matlab 7*. Prentice Hall, 2004.

[HTF01] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[MS99]   Christopher D. Manning and Hinrich Schutze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.