

## CMPS 142 Syllabus, Winter '10

Here is the syllabus for the Data Mining and Machine Learning class. We will be learning about and experimenting with methods for learning from data. The text is *Machine Learning An Algorithmic Perspective* by Marsland. You might find an introduction to probability as well as these other books useful:

- *Neural Networks for Pattern Recognition* by Bishop
- *Pattern Classification* by Duda, Hart and Stork or the earlier *Pattern Classification and Scene Analysis* by Duda and Hart
- *Machine Learning* by Mitchell
- *Artificial Intelligence A Modern Approach* by Russell and Norvig.
- *The Elements of Statistical Learning* by Hastie, Tibshirani, Friedman
- *Machine Learning* by Alpaydin

I am assuming that students have some background in algorithms and probability (at least at the CMPS 101 level).

Planned Topics:

1. Introduction: What is Machine Learning? What is Data Mining? (ch 1)
2. Nearest Neighbor
3. Perceptron algorithm (ch 2)
4. Neural Networks (ch 3)
5. Decision Trees (ch 6)
6. Bayesian learning and parameter estimation (ch 8)
7. Linear Discrimination and Support vector machines (ch 5)
8. Boosting (AdaBoost) (ch 7)
9. Clustering, EM Algorithm and K-means (ch 9)
10. On-line prediction (paper)
11. Reinforcement Learning (ch 13)

Additional topics may be inserted and/or some topics may be skipped based on the interests of the class. The syllabus is aggressive, and we may not get to everything on this list.

Evaluation in the course will be based on periodic homework assignments and exercises (20%), an in-class midterm in the eighth week (50%) and a term project (30%).