CMPS 242 Midterm review topics, Fall ’03

The Midterm will be closed book, but each student may have one three-by-five card of handwritten notes. Here are the topics we have covered in the class. Also review the homework problems.

1. What is Machine learning

2. Concept learning, version/hypothesis space, FIND-S algorithm

3. On-line learning
   - conjunctions
   - halving algorithm, randomized halving algorithm
   - standard optimal algorithm
   - weighted majority
   - winnow, standard optimal algorithm

4. Perceptron algorithm and convergence theorem

5. Neural Networks
   - logistic function
   - networks of sigmoid units
   - gradient descent and backpropagation algorithm

6. Decision trees
   - Information as measure of impurity
   - greedy construction of trees using information gain
   - overfitting and how to fix it

7. AdaBoost algorithm

8. proof that AdaBoost’s training error decreases exponentially

9. AdaBoost is (somewhat) resistant to overfitting

10. Bayesian learning
    - Review of probability basics
    - Bayes rule and the sock paradox
    - maximum likelihood
    - priors, maximum a posteriori and mean a posteriori estimates and predictions
    - Losses, risk, and bayes optimal predictions
• decision boundaries and error probabilities
• fitting gaussians to data - maximum likelihood and unbiased parameter estimates
• prediction based on bits through multiple channels
• Naive Bayes algorithm - its assumptions and weaknesses
• Estimating probabilities, the Laplace and m-estimates
• Regression: least squares, and maximum likelihood
• MDL learning

11. EM Algorithm

12. Bayesian networks and how they represent (conditional) independence

13. Non-parametric (instance based) learning
   • nearest neighbor and voronoi diagrams
   • importance of distance function, problems with irrelevant and/or unscaled features
   • nearest neighbor not optimal and curse of dimensionality
   • instance based (parzon window) density estimation

14. PAC learning (just started)
   • basic definitions
   • simple examples (singletons, initial segments, conjunctions, polygons)
   • Consistent learners with finite hypothesis spaces