CMPS 142/242 Midterm Topics (Fall 2011)

The midterm will be in class Monday, November 21. The exam will be closed-book, although
students may have one 3x5 card of handwritten notes (both sides). The reading assignments and
lecture slides have been posted on the class web page, and reviewing the homework problems is
recommended. An old sample exam has been posted. Here are this quarter’s topics:

1. What is Machine Learning
2. Supervised batch learning, examples
3. Regression
4. Unsupervised learning
5. Concept Learning: hypothesis class, domain, target, inductive bias, version space
6. Noise and causes (label errors, attribute errors, features or hypothesis class may
not fit phenomena exactly)
7. Model Selection, Generalization, and Overfitting – Using test sets and cross
   validation to estimate generalization error
8. Feature Selection/Creation
9. Basic probability (sample space, events, random variables, independence,
   conditional probability, Bayes rule, sum rule)
10. Estimating probabilities (e.g. the bias of a coin flip): maximum likelihood, priors,
    maximum a’ posteriori, mean a’ posteriori, predictive distribution, and Laplacian
    estimates of probabilities
11. Hypotheses as models generating the data, the maximum likelihood hypothesis,
    the maximum a’ posteriori hypothesis, and the predictive posterior distribution
12. Discriminative and Generative Models
13. Decision Theory with asymmetric losses, Bayes’ risk, and Bayes’ optimal
    predictions
15. Prediction by using Gaussians to model class-conditional distributions
16. Instance based learning: Nearest Neighbor, curse of dimensionality, kNN, density
    estimation, instance based regression
17. Decision Trees: Greedy construction of trees, information gain criterion, applying
    a split criterion (impurity function, e.g. information gain) to select tests at nodes,
    over-fitting and pruning
18. Naïve Bayes and its independence assumption, naïve Bayes for text classification
19. Linear Regression: least squares as maximum likelihood, bias-variance
    decomposition, regularization, Bayesian linear regression
20. Linear Classification: add-a-dimension trick, Fisher’s Linear Discriminant (LDA),
21. Perceptron algorithm and convergence,
22. Logistic regression and softmax function
23. Feed forward artificial neural networks, backprop algorithm (gradient descent)
24. Support vector machines and margin maximization, Lagrangian, dual problem,
    and kernel functions
25. Clustering: hierarchical vs partitional methods, K-means
26. Mixture of Gaussians: learn with EM, use for clustering and density estimation
27. A very little bit on Hidden Markov Models