CIS 732 Pattern Recognition and Machine Learning
CIS 830 Topics in Artificial Intelligence
Spring 2008
Homework 1 of 8: Problem Set (PS1)
Supervised Learning for Classification

Assigned: Fri 15 Feb 2008
Due: Fri 22 Feb 2008 (before midnight)

The purpose of this assignment is to exercise your basic understanding of intelligent agents, state space search, and game theory, and to help you apply these concepts simulate the behavior of search algorithms.

This homework assignment is worth a total of 20 points. It is required for CIS 732, CIS 830, and independent study (CIS 798) students.

Turn in hard copy or attach an electronic copy of the assignment in PDF form (converted from your word processor, or scanned) to the instructor at: **CIS732TA-L@listserv.ksu.edu** or **CIS830TA-L@listserv.ksu.edu**.

1. **(10%) Version Spaces and Inductive Bias.** Learning a particular hypothesis is sometimes referred to as "belief revision", i.e., revising $VS_{H,D}$ for consistency. Explain, in your own words, the difference between learning in this sense and optimization of *representation bias*. What is being updated in each scenario, and what is the overall objective?

2. **(10%) Tree vs. Rule Post-Pruning.** (Problem 6.3, p. 375 Han & Kamber 2^e)

   Given a decision tree, you have the option of (a) converting the decision tree to rules and then pruning the resulting rules, or (b) pruning the decision tree and then converting the pruned tree to rules. What advantage does (a) have over (b)?

3. **(10%) Complexity of Decision Tree Induction.** (Adapted from Problem 6.4, p. 375 Han & Kamber 2^e)

   It is important to calculate the worst-case computational complexity of the decision tree algorithm. Given data set D, the number of attributes n, and the number of training tuples $m = |D|$, show that the computational cost of growing a tree is in $O(nm \lg m)$.

4. **(20%) Naïve Bayes.** (Problem 6.9, pp. 375 – 376 Han & Kamber 2^e)

   Design an efficient method that performs effective naïve Bayesian classification over an *infinite* data stream (i.e., you can scan the data stream only once). If we wanted to discover the *evolution* of such classification schemes (e.g., comparing the classification scheme at this moment with earlier schemes, such as one from a week ago), what modified design would you suggest?

5. **(10%) Rule Induction using a Sequential Covering Algorithm.** Suppose you are trying to learn a rule set for a concept (Boolean classification problem) such as spam filtering. Give an example of how overfitting might occur in *Learn-One-Rule* and how post-pruning using validation data might remove conditions from the rule. You do not have to use a real data set, but describe a concrete scenario where pruning would make a difference.

6. **(20%) Winnow and Concept Learning.** Develop a way to adapt the Littlestone's Winnow algorithm for learning monotone disjunctive concepts so that you are instead learning monotone conjunctive concepts. What kind of inductive bias does this exhibit, and how does it compare to that of the disjunctions? (Can you propose an application where the bias might be useful?)

7. **(10%) Multilayer Perceptrons and Backpropagation of Error.** Consider the error surface of a multilayer perceptron trained in batch mode. What does the gradient represent? That is, when we refer to "hill climbing" in this error space, what gradient is being ascended or descended? (Hint: look at the loss function on p. 156 of the reference below.)

   Rojas, R. (1996). *Neural Networks - A Systematic Introduction*, Chapter 7, "Backpropagation", pp. 151 - 184. Berlin, Germany: Springer. Retrieved from: http://page.mi.fu-berlin.de/rojas/neural/chapter/K7.pdf

8. **(10%) Lazy vs. Eager Classification.** (Adapted from Problem 6.8, p. 375 Han & Kamber 2^e.) State two relative advantages and two relative disadvantages of *eager* classification (e.g., decision tree, Bayesian, artificial neural neural network) versus *lazy* classification (e.g., *k*-nearest neighbor, case-based reasoning).

Class participation (required). Post a question on your most unclear point on one of the following classifiers to **CIS732-L@listserv.ksu.edu**, along with a brief introduction stating your:

- name
- program (grad or undergrad) and major
- interests in machine learning and pattern recognition
- special topics you would like to see covered, if any

Classifiers discussed so far

1. Conjunctive hypotheses with don't cares
2. Decision trees
3. Naïve Bayes
4. Propositional rules (Horn clauses)
5. Perceptrons
6. Multilayer perceptrons
7. Monotone disjunctive hypotheses
8. Pure disjunctive hypotheses
9. Support vector machines
10. Nearest neighbor