Computer Science 6100/4100: Assignment 2

Due: November 6, 2008

Note: Problems 1 and 2 may be done in pairs. Use the same format for the writeup for Problems 1 and 2 as you did for Assignment 1 (the `sig-alternate` class), and adhere to a 3 page limit for the combined writeups for the two problems. Again, if your writeup does not meet the formatting guidelines, I will not read or grade it.

1 Problem 1 (30 points, teams of 2 allowed)

Compare the performance of ID3 (without pruning), bagged ID3 (without pruning), and boosted decision stumps on the "Heart Disease" dataset from the UCI repository (decision stumps are decision trees restricted to test and split on only one attribute). For this assignment you should split on the information gain as discussed in class. Some specifics about the data:

1. Find it at:
   `ftp://ftp.ics.uci.edu/pub/machine-learning-databases/heart-disease/`
2. Read the ".names" file.
3. Use only the processed Cleveland data.
4. Convert it into a two-class problem attempting to differentiate between presence (1,2,3,4) or absence (0) of heart disease as mentioned in point 4 of the names file.
5. Note that there are a number of attributes that can be treated as continuous.

You have complete freedom in terms of the implementation. You may use existing implementations or write your own (be sure to appropriately document and cite anything you use or write). Use ten-fold cross-validation to estimate performance. You may want to repeat it a few times to get a good estimate. Writing up your design decisions and results clearly and well is absolutely critical. Specifically, be sure to address the following questions:

1. What are the training and test errors for each classifier?
2. How are the training and test errors affected by the number of rounds of boosting or bagging that are carried out?
3. Does the training error decay to 0 as more and more rounds of boosting are carried out? If not, think carefully about the assumptions of the theorem. Which of these assumptions is being violated? Provide empirical evidence for your answer.

2 Problem 2 (20 points, teams of 2 allowed)

A weighted coin with probability p of coming up heads is flipped repeatedly. You have some whole dollar amount of money that is less than a hundred dollars. At each time, you can choose to bet any whole dollar amount that is less than or equal to the amount you possess. If the coin comes up heads you double the amount you bet, and if it comes up tails you lose the entire amount you bet. Your goal is to accumulate \$100 – after that you retire. So there are two eventual outcomes – accumulate \$100 or go bust (end up with \$0).

The state is your capital $s \in \{1, 2, \ldots, 99\}$. The actions are the amounts you may choose to bet. In state s, $a \in \{1, 2, \ldots, s\}$. If you ever reach \$100 in capital, you receive a reward of 1, and the reward received is 0 in every other state transition.

What does the value function for a state represent? (Please, make sure you know the answer to this before you proceed!)

Implement value iteration and solve for the optimal policy for $p = 0.25$ and $p = 0.55$. Present your results for each case as 2 graphs, one showing the final value estimates as a function of the capital, and one showing the policy (how much you should bet) as a function of the capital. Write down your interpretation of the forms of the optimal policies in the two cases.

3 Problem 3 (15 points, no teams!)

Consider the variance update for the Kalman filter described in class for a one-dimensional random walk:

$$\sigma_{t+1}^2 = \frac{(\sigma_t^2 + \sigma_x^2)\sigma_z^2}{\sigma_t^2 + \sigma_x^2 + \sigma_z^2}$$

σ_t^2 converges to a fixed point as $t \to \infty$. Give an expression for the value of this fixed point as a function of σ_x^2 and σ_z^2.

Analytically derive approximations for the fixed point σ^2 as $\sigma_x^2 \to 0$ and $\sigma_z^2 \to 0$ (for simplicity, you may hold the other one fixed at 1 in deriving these approximations). Interpret your results: effectively, explain:

1. The meaning of these two cases in terms of the underlying process.
2. How the behavior of the fixed point differs for the two cases and why this makes sense in terms of (1).