10-701/15-781: Homework 1 Solutions

Eric Xing, Tom Mitchell, Aarti Singh
Carnegie Mellon University
Updated on January 12, 2010

1 Decision Trees and Information Theory [35 pt, Ni Lao]

1. **The 9 ball problem [10 pt]** Divide the balls into three 3-ball groups. Compare the first two groups. If one of them is heavier, then that group has the heavier ball, or else the last group has it. Then compare two balls from the group which has the heavier ball. If one of them is heavier, then the special ball is found, or else the third ball is heavier.

   At each trial, the scale has three possible outcomes: tilt left, tilt right, or equal. We can treat the outcome of a serial of trials as a 3-nary codeword. The amount of information of picking out 1 ball from 9 balls is $H(X) = \log_3 9 = 2$ bit. Therefore, the minimum expected number of trials is 2, and our strategy has reached this minimum.

2. **Information gain and optimal encoding [10 pt]** From $I(Y_t; X|Y_1...Y_{t-1}) = 1$, we have $H(Y_t|Y_1...Y_{t-1}) \geq I(Y_t; X|Y_1...Y_{t-1}) = 1$, therefore $p(Y_t = k|Y_1...Y_{t-1}) = 1/K$.

   Since the mapping from X to Y_t are deterministic, we have $p(X = i) = p(Y_1...Y_{l_i}) = K^{-l_i}$. Therefore, the expected number of tests is optimal.

3. **The ID3 algorithm [5 pt]** Yes it will. In the first step, comparing any two non-overlapping 3-ball groups has a information gain of $H(X) - H(X|Y_1) = 2 - 1 = 1$ which is the maximum a 3-outcome test can achieve. In the second step, comparing any two balls from the selected 3-ball group also has the maximum information gain of $H(X|Y_1) - H(X|Y_2, Y_1) = 1 - 0 = 1$.

4. **The number guessing problem [5 pt]** Identifying a 4 digit number need $\log_{10}(10000) = 4$ bit of information (assuming even prior $p_i = 1/10000$). Each of Alice's answer has $5(5+1)/2 = 15$ outcomes, therefore carrying at most $\log_{10}(15) = 1.176$ bit of information. Bob need at least $4/1.176 = 3.4$ guesses on average.

5. **[Extra 5 pt]** This bound is easy to achieve in reality, because people usually have non even prior over the 10000 numbers, therefore reduces the amount of information needed to identify a number.