CPSC 633-600 Homework 2, part I/II (Total 50 points)
Decision Tree Learning

See course web page for the **due date**.
Use **elearning.tamu.edu** to submit your assignments, or submit a hard copy.

Instructor: Yoonsuck Choe

February 23, 2013

1 Entropy

Given a random variable X that can take on values $\{\oplus, \ominus\}$, the entropy is defined as:

$$E(X) = -\sum_{x \in \{\oplus, \ominus\}} P(X = x) \log_2 P(X = x).$$

Since $P(X = \oplus) + P(X = \ominus) = 1$, $E(X)$ can be rewritten as a function of $P(X = \oplus)$: Letting $p_\oplus = P(X = \oplus)$:

$$E(X) = f(p_\oplus) = -p_\oplus \log_2 p_\oplus - (1 - p_\oplus) \log_2 (1 - p_\oplus).$$

Figure 1 shows how $f(p_\oplus)$ behaves as $p_\oplus$ changes.

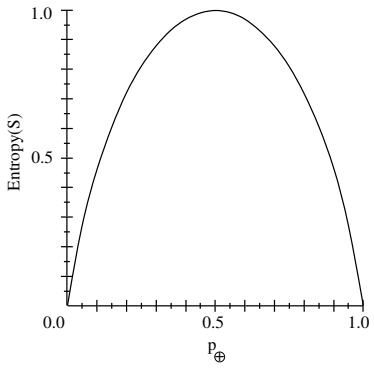

Figure 1: **Entropy.**

Problem 1 (Written: 10 pts): Extend the above analysis to a random variable Y that can take on values $\{\alpha, \beta, \gamma\}$. Given $p_\alpha = P(Y = \alpha)$, etc.,

1. Derive $E(Y)$ as a function of p_α and p_β:

   $$E(Y) = f(p_\alpha, p_\beta) = ...$$

   Note: $p_\alpha + p_\beta + p_\gamma = 1.0$.

2. For which values of p_α and p_β does $E(Y)$ become maximal (no need to derive it exactly from $f(p_\alpha, p_\beta)$–consider when it is maximal in the 2-value case)?

3. Explain why (you don't need to provide a formal proof).

Problem 2 (Program: 20 pts): Write a short program to calculate $f(p_\alpha, p_\beta)$ derived above, and obtain the $E(Y) = f(p_\alpha, p_\beta)$ values for all combinations of $p_\alpha, p_\beta \in \{0.0, 0.01, 0.02, ..., 0.99, 1.0\}$, and plot in 3D (Octave: use `surf`; Matlab: use `surf`; or draw by hand). You have to be careful because:

- log(0) will throw an error, so you have to check for the occurrence of (0 * log(0)) and make that 0 before log(0) gets evaluated. Alternatively, you can start with a value close to 0: 0.001, 0.01, 0.02, ..., 0.99, 1.0.
- Also, you have to plot for the (p_α, p_β) that sums up to less than or equal to 1.0.

Note that you can use `meshgrid` to get the grid: `[pa,pb] = meshgrid((0.001:0.01:1));`

Problem 3 (Written: 5 pts): Based on the insight from above, when you have a random variable X that can take on four different discrete values (say, $\{a, b, c, d\}$), then (1) what should $P(X = a)$ etc. be so that the entropy of X is maximized? (2) Also, what is the value of the maximum entropy? (Note: You don't need to provide a formal proof.)

2 Decision Tree Learning (ID3)

Problem 4 (Written: 15 pts): (1) Calculate the entropy of the following training set. (2) Calculate the information gain for each of the three attributes. (3) Which one is the best attribute to test first?

Instance#	Hardware	Storage	OS	Decision (Buy?)
1	PC	SSD	Linux	Y
2	PC	Hard Drive	Linux	Y
3	PC	Hard Drive	Linux	N
4	Mac	SSD	Chrome	N
5	Mac	Hard Drive	MacOS	Y
6	PC	SSD	MacOS	Y
7	PC	SSD	Chrome	N
8	PC	Hard Drive	MacOS	N
9	PC	Hard Drive	Windows	Y
10	Mac	SSD	Windows	N
11	Mac	SSD	Linux	Y
12	Mac	Hard Drive	Windows	N

NOTE: Total points for this assignment is 50. Part II of this assignment will be given later, worth 50 points (as part of homework 2), on the topic of reinforcement learning.