CS725: Assignment 1

20 Marks, Due on August 22^{nd} Midnight

1. One important part of machine learning is decision theory. Given

   - a choice of various actions, a and
   - given that the world may be in one of many states $\mathbf{x} \in \Re^k$ such that which one occurs may be influenced by the choice of action
   - a probability distribution $\Pr(\mathbf{x}|a)$
   - a utility function $U(\mathbf{x}, a)$ which specifies the payoff you receive when the world is in state $\mathbf{x}$ and you chose action a.

   the task of decision theory is to select the action that maximizes the expected utility

   $$E[U|a] = \int d^k\mathbf{x} U(\mathbf{x}, a) \Pr(\mathbf{x}|a)$$

   (a) In the context of expected utilities and decision tree, read up the *Substitution Axiom* and the *Archemdian Axiom*[1] and present their implications as summary to `http://www.cscs.umich.edu/~spage/teaching_files/modeling_lectures/MODEL5/M21trees.pdf`.
   (**5 Marks**)

   (b) One of the challenges of decision theory is figuring out exactly what a good utility function is. The utility of money, for example, is notoriously nonlinear for most people. In fact, the behaviour of many people cannot be captured by a coherent utility function, as illustrated by the Allais paradox.

   First consider these choices:

   i. \$20 million guaranteed.
   ii.
      - 89% chance of \$20 million;
      - 10% chance of \$50 million;

[1] `http://en.wikipedia.org/wiki/Utility`

- 1% chance of nothing.

Now consider these choices:

iii.
- 89% chance of nothing;
- 11% chance of \$20 million.

iv.
- 90% chance of nothing;
- 10% chance of \$50 million.

Many people prefer (i) to (ii), and, at the same time, (iv) to (iii). Prove that these preferences are inconsistent with any utility function $U(\mathbf{x})$ for money.

(*Hint:* Read `http://en.wikipedia.org/wiki/Allais_paradox`)

(**5 Marks**)

2. Explain atleast 2 possible implementations of decision tree learning under different assumptions on the features (you can ignore the pruning phase) and analyse the complexity of each. You can use the algorithm in Figure 2 from the paper entitled *The Time Complexity of Decision Tree Induction*[2] for reference as one possible implementation. For each implementation that you describe, state both the time complexity and the space complexity.

   Two possible assumptions (there could be other scenarios)

   (a) no numeric attributes, only discrete valued attributes

   (b) only numeric attributes

   State your assumptions clearly.

   (**5 Marks**)

3. Suppose we have three categories with $P(c_1) = \frac{1}{2}$, $P(c_2) = P(c_3) = \frac{1}{4}$ and the distributions are as follows:

   (a) $p(x|c_1) = \mathcal{N}(0, 1)$

   (b) $p(x|c_2) = \mathcal{N}(.5, 1)$

   (c) $p(x|c_3) = \mathcal{N}(1, 1)$

   and that we have a random sample of the following four points (and sampled in the corresponding order): $x_1 = 0.6$, $x_2 = 0.1$, $x_3 = 0.9$ and $x_4 = 1.1$.

   (a) Calculate the probability that the sequence x_1, x_2, x_3, x_4 actually came from the class sequence c_1, c_3, c_3, c_2.

   (b) Find the class sequence having the maximum probability of having generated x_1, x_2, x_3, x_4.

   (**5 Marks**)

[2] `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.2193&rep=rep1&type=pdf`