COMP3620 Tutorial 4: Learning

1. You are a physician. You think it is quite likely (say 60% probability) that one of your patients has strep throat, but you aren't sure. You take some swabs from the throat and send them to a lab for testing. The test is (like nearly all lab tests) not perfect. If the patient has strep throat, then 70% of the time the lab says YES. But 30% of the time it says NO. If, however, the patient does not have strep throat, then 90% of the time the lab says NO. But 10% of the time it says YES. You send five successive swabs to the lab, from the same patient. You get back these results, in order: YES, NO, YES, NO, YES.

   From that, which of the following can you conclude and why?

   (a) The results are worthless.

   (b) It is likely that the patient does not have strep throat.

   (c) It is more likely than not that the patient does have strep throat.

2. Let $X = \{0,1\}^n$ and H a set of functions from X to $\{0,1\}$. The class H is said to be PAC (probably approximately correct) learnable if there exists an algorithm L satisfying the following: given any $\epsilon, \delta \in (0,1)$, there is an integer $m(\epsilon, \delta)$ such that for all $m \geq m(\epsilon, \delta)$, for any $t \in H$ and any probability distribution μ on X, with probability at least $1 - \delta$, given a sample of size m drawn independently according to μ and labelled with t, the error of the hypothesis $h \in H$ output by L with respect to t and μ defined by

   $$er_\mu(h,t) = \mu\{x \in X : h(x) \neq t(x)\}$$

   is less than ϵ. The class H is said to be efficiently PAC learnable if, in addition to the above, L runs in time polynomial in $m, 1/\epsilon, 1/\delta, n$.

   Let k-$DL(n)$ be the class of all decision lists over $\{0,1\}^n$ where each test in an internal node is a conjunction of at most k tests on the input attributes. For example,

$$f(x_1, \dots, x_{10}) = \textit{if } x_2 \wedge \neg x_4 \wedge x_3 \textit{ then } 1 \textit{ else if } x_1 \wedge \neg x_2 \textit{ then } 1 \textit{ else } 0$$

is a member of 3-$DL(10)$.

Show that the class k-$DL(n)$ is efficiently PAC learnable.

3. Learn and discuss decision-tree pruning methods. The χ^2 pruning method described in the textbook is a good starting place.

4. Consider an ensemble learning algorithm that uses simple majority voting among M learned hypotheses. Suppose that each hypothesis has error ϵ and that the errors made by each hypothesis are independent of the others'. Give a formula for the error of the ensemble algorithm in terms of M and ϵ, and evaluate it for the cases where $M = 5, 10$, and 20 and $\epsilon = 0.1, 0.2$, and 0.4.

5. Write out in full the computation shown on slide 10 of the lecture on Bayesian networks.

6. Prove the CTW Theorem given on slide 18 of the third lecture. Look up arithmetic coding and show how it can be used in conjunction with the CTW algorithm to achieve excellent data compression.