CS578/STAT590: Introduction Machine Learning **Fall 2014**

Problem Set 4

Handed Out: Dec 3 *Due: Dec 18*

Questions

1. **Boosting**
   In this problem you will have to use the AdaBoost algorithm to learn a function, mapping examples in $\mathbb{R}^2$ to a boolean value. The space of weak learner considered by the algorithm consists of hypotheses of the form: $x_i > A$ where A is an Integer, and $i = \{1, 2\}$. Run the AdaBoost algorithm for two rounds using the data appearing in the table below, at each round AdaBoost chooses the weak learner that minimizes the error (ϵ). Your answer should consist of :
   (1) The weak hypothesis used at each round, and its error
   (2) The distribution D_i over the examples for each round
   (3) The final hypothesis after running two rounds.

index	x_1	x_2	y
1	1	10	-
2	4	4	-
3	8	7	+
4	5	6	-
5	3	16	-
6	7	7	+
7	10	14	+
8	4	2	-
9	4	10	+
10	8	8	-

2. **Naïve Bayes**
   You are given a collection of m documents written in a language consisting of four symbols a,b,c,d, each document is n symbols long. Each document is associated with a binary label ("good" or "bad" document). We can represent the documents using a multinomial distribution. Given a document, we denote by a_i, b_i, c_i, d_i, the count of each one of the symbols in the document, and define the distribution:

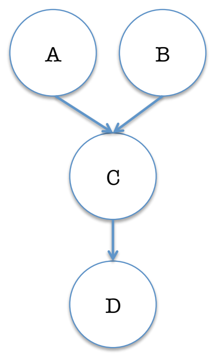
   P(D_i |y=1) = $\frac{n!}{a_i!b_i!c_i!}\alpha_1^{a_i}\beta_1^{b_i}\gamma_1^{c_i}\delta_1^{d_i}$
   where α_1 ($\beta_1, \gamma_1, \delta_1$) is the probability that a symbol a (and respectively, b,c,d) appears in a document labeled as "good". Similarly we define

   P(D_i |y=0) = $\frac{n!}{a_i!b_i!c_i!}\alpha_0^{a_i}\beta_0^{b_i}\gamma_0^{c_i}\delta_0^{d_i}$

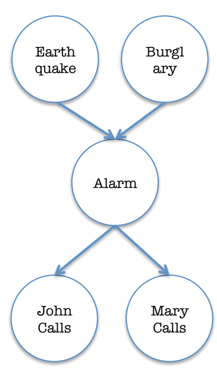
where α_0 $(\beta_0, \gamma_0, \delta_0)$ is the probability that a symbol a (and respectively, b,c,d) appears in a document labeled as "bad"[1].

(1) Write the joint log-likelihood of a document and labels (i.e., log $\mathrm{P}(D_i, y_i)$).

3. **Bayesian Network**
You are given a Bayesian network, defined over four variables. The variable A is binary, and the rest of the variables take 3 values.
(1) How many parameters are needed to define the network?
(2) Write the expression calculating **(a)** P(A=1, D=2) **(b)** P(A=1,D=2| C=1)



4. **Variable Elimination**
Given the "Burglar Alarm" network we saw in class, compute P(Mary calls).
(1) Write the expression calculating P(Mary calls), according to the network below. How many operations are needed to compute it?
(2) Run the variable elimination algorithm. How many operations are needed now? (write down each step)



[1]Note that $\alpha_i + \beta_i + \gamma_i + \delta_i$=1