CS578/STAT590: Introduction to Machine Learning	**Fall 2014**
Problem Set 4	
Gen Nishida	*Handed In: December 2, 2014*

Questions

1. **Boosting**

(1) The weak hypothesis used at each round, and its error

1st round $h_1(x) = \text{sign}(x_1 > 6)$, $\epsilon_1 = 0.2$ ($\alpha_1 = 0.6931$)

index	x_1	x_2	y	prediction
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	-	+

Table 1: The prediction by h_1

2nd round $h_2(x) = sign(x_2 > 9)$, $\epsilon_1 = 0.25$ ($\alpha_2 = 0.5493$)

index	x_1	x_2	y	prediction
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	-	-

Table 2: The prediction by h_1

(2) The distribution D_i over the examples for each round

The distribution D_i evolves over the rounds as shown in Table 3.

index	D_1	D_2	D_3
1	0.1	0.0625	0.125
2	0.1	0.0625	0.0417
3	0.1	0.0625	0.125
4	0.1	0.0625	0.0417
5	0.1	0.0625	0.125
6	0.1	0.0625	0.125
7	0.1	0.0625	0.0417
8	0.1	0.0625	0.0417
9	0.1	0.25	0.1666
10	0.1	0.25	0.1666

Table 3: The distribution D_i

(3) The final hypothesis after running two rounds.
$H_{\text{final}}(x) = \text{sign}\left[0.6931 \times \text{sign}(x_1 > 6) + 0.5493 \times \text{sign}(x_2 > 9)\right]$

2. **Naive Bayes**

(1) Write the joint log-likelihood of a document and labels (i.e., $\log P(D_i, y_i)$).
First, the joint probability $P(d_i, y_i)$ is

$$\left[\frac{n!}{a_i!b_i!c_i!d_i!}\alpha_1^{a_i}\beta_1^{b_i}\gamma_1^{c_i}\delta_1^{d_i} \times P(y=1)\right]^y \times \left[\frac{n!}{a_i!b_i!c_i!d_i!}\alpha_0^{a_i}\beta_0^{b_i}\gamma_0^{c_i}\delta_0^{d_i} \times P(y=0)\right]^{(1-y)}$$

Since we do not know a prior probability of $P(y)$, we just use the uniform distribution, which is $P(y) = 1/2$. Thus,

$$\frac{1}{2} \times \left[\frac{n!}{a_i!b_i!c_i!d_i!}\alpha_1^{a_i}\beta_1^{b_i}\gamma_1^{c_i}\delta_1^{d_i}\right]^y \times \left[\frac{n!}{a_i!b_i!c_i!d_i!}\alpha_0^{a_i}\beta_0^{b_i}\gamma_0^{c_i}\delta_0^{d_i}\right]^{(1-y)}$$

Then, its logarithm becomes

$$\begin{aligned}
&\log\frac{1}{2} + y\log\left[\frac{n!}{a_i!b_i!c_i!d_i!}\alpha_1^{a_i}\beta_1^{b_i}\gamma_1^{c_i}\delta_1^{d_i}\right] + (1-y)\log\left[\frac{n!}{a_i!b_i!c_i!d_i!}\alpha_0^{a_i}\beta_0^{b_i}\gamma_0^{c_i}\delta_0^{d_i}\right] \\
&= \log\frac{1}{2} + \log\frac{n!}{a_i!b_i!c_i!d_i!} + y\left[a_i\log\alpha_1 + b_i\log\beta_1 + c_i\log\gamma_1 + d_i\log\delta_1\right] \\
&\quad +(1-y)\left[a_i\log\alpha_0 + b_i\log\beta_0 + c_i\log\gamma_0 + d_i\log\delta_0\right]
\end{aligned}$$

3. **Bayesian Network**

(1) How many parameters are needed to define the network?

(2) Write the expression calculating (a) $P(A=1, D=2)$ (b) $P(A=1, D=2|C=1)$
(a) $P(A=1)P(D=2)$
(b) $\frac{P(A=1)P(C=1|A=1)P(D=2|C=1)}{P(C=1)}$

4. **Variable Elimination**