CS446: Machine Learning	**Spring 2017**
Problem Set 6	
Bangqi Wang	*Handed In: April 20, 2017*

1. Answer to problem 1

   a. $y = sign()\vec{w}^T x + \theta$
   $\vec{w} = \{\vec{w_0}, \vec{w_1}, \vec{w_2}, \vec{w_3}, \vec{w_4}, \vec{w_5}, \vec{w_6}, \vec{w_7}, \vec{w_8}\}$
   For $f_{TH(4,9)}$, $\vec{w} = \{1, 1, 1, 1, 1, 1, 1, 1, 1\}$, and $\theta = -3.9$. $f_{TH(4,9)}$ has a linear decision surface over the 9 dimensional boolean cube.

   b.

$$\begin{aligned}
h &= argmax_{y\in\{0,1\}} P(y) = \prod_{i=0}^{n} P(x_i|y) = \prod_{i=0}^{8} P(x_i|y) && (1)\\
P(x_i|y) &= \frac{P(x_i)P(y|x_i)}{P(y)} && (2)\\
P(y=0) &= \binom{9}{0}0.5^9 + \binom{9}{1}0.5^9 + \binom{9}{2}0.5^9 + \binom{9}{3}0.5^9 = \frac{65}{256} && (3)\\
P(y=1) &= 1 - P(y=0) = \frac{191}{256} && (4)\\
P(x=0|y=0) &= \frac{\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \binom{8}{3}}{\binom{9}{0} + \binom{9}{1} + \binom{9}{2} + \binom{9}{3}} = \frac{93}{130} && (5)\\
P(x=0|y=1) &= \frac{\binom{8}{4} + \binom{8}{5} + \binom{8}{6} + \binom{8}{7} + \binom{8}{8}}{0.5^9 - \binom{9}{0} + \binom{9}{1} + \binom{9}{2} + \binom{9}{3}} = \frac{163}{382} && (6)\\
P(x=1|y=0) &= 1 - P(x=0|y=0) = \frac{37}{130} && (7)\\
P(x=1|y=1) &= 1 - P(x=0|y=1) = \frac{219}{382} && (8)
\end{aligned}$$

$$y = argmax_y(P(y=0)\prod_{i=0}^{8} P(x_i|y=0, P(y=1)\prod_{i=0}^{8} P(x_i|y=1)$$

   c. Assume $\vec{w} = \{1, 1, 1, 0, 0, 0, 0, 0, 0\}$, which $f_{TH(4,9)} = 0$.

$$\begin{aligned}
h(y=0) &= \frac{65}{256} \times (\frac{37}{130})^3 \times 93130^6 = 7.8e-4 && (9)\\
h(y=1) &= \frac{191}{256} \times (\frac{219}{382})^3 \times 163382^6 = 8.5e-4 && (10)
\end{aligned}$$

   The true label is 0, but the predicts is 1 becasue $h(y=1) > h(y=0)$. The prediction is wrong, so the hypothesis does not represent this function.

d. The naive bayes assumption is not satisfied by $f_{TH(4,9)}$, because the label from function depends on the the number of activate features. Given the label, the value of each dimension is conditionally dependent to others.

2. Answer to problem 2

a. Prior probabilities and parameter values:

$\Pr(Y = A) = \frac{3}{7}$	$\Pr(Y = B) = \frac{4}{7}$
$\lambda_1^A = 2$	$\lambda_1^B = 4$
$\lambda_2^A = 5$	$\lambda_2^B = 3$

Table 1: Parameters for Poisson naïve Bayes

b.

$$P(X_1 = 2|Y = A) = \frac{e^{-2}2^2}{2!} = 2e^{-2} \tag{11}$$

$$P(X_2 = 3|Y = A) = \frac{e^{-5}5^3}{3!} = \frac{125}{6}e^{-5} \tag{12}$$

$$P(X_1 = 2|Y = B) = \frac{e^{-4}4^2}{2!} = 8e^{-4} \tag{13}$$

$$P(X_2 = 3|Y = B) = \frac{e^{-3}3^3}{3!} = \frac{9}{2}e^{-3} \tag{14}$$

$$\frac{P(X_1 = 2, X_2 = 3|Y = A)}{P(X_1 = 2, X_2 = 3|Y = B)} = \frac{2e^{-2} \times \frac{125}{6}e^{-5}}{8e^{-4} \times \frac{9}{2}e^{-3}} = \frac{125}{108} \tag{15}$$

c. Given $X_1 = x_1$ and $X_2 = x_2$, $Y = A$ iff $\frac{P(Y=A|X_1=2,X_2=3)}{P(Y=B|X_1=2,X_2=3)} \geq 1$

$$\frac{P(Y = A|X_1 = 2, X_2 = 3)}{P(Y = B|X_1 = 2, X_2 = 3)} = \frac{P(X_1 = 2, X_2 = 3|Y = A) \cdot P(Y = A)}{P(X_1 = 2, X_2 = 3|Y = B) \cdot P(Y = B)} \tag{16}$$

$$= \frac{e^{-\lambda_1^A}(-\lambda_1^A)^{x_1}e^{-\lambda_2^A}(-\lambda_2^A)^{x_2} \cdot P(Y = A)}{e^{-\lambda_1^B}(-\lambda_1^B)^{x_1}e^{-\lambda_2^B}(-\lambda_2^B)^{x_2} \cdot P(Y = B)} \tag{17}$$

$$\geq 1 \tag{18}$$

Taking *log*:

$$h(x_1, x_2) = \lambda_1^B + \lambda_2^B - \lambda_1^A - \lambda_2^A + log\frac{P(Y = A)}{P(Y = B)} + x_1 log\frac{\lambda_1^A}{\lambda_1^B} + x_2 log\frac{\lambda_2^A}{\lambda_2^B} \geq 0 \tag{19}$$

We predict $Y = A$, iff $h(x_1, x_2) \geq 0$.

d.

$$\begin{aligned}
h(x_1,x_2) &= \lambda_1^B+\lambda_2^B-\lambda_1^A-\lambda_2^A+log\frac{P(Y=A)}{P(Y=B)}+x_1log\frac{\lambda_1^A}{\lambda_1^B}+x_2log\frac{\lambda_2^A}{\lambda_2^B} && (20)\\
&= 4+3-2-5+log\frac{\frac{3}{7}}{\frac{4}{7}}+2log\frac{2}{4}+3log\frac{5}{3} && (21)\\
&= log\frac{3}{4}+2log\frac{1}{2}+3log53 && (22)\\
&= -0.1415<0 && (23)
\end{aligned}$$

The classifier predicts Y as B, since $h(x_1,x_2)<0$.

3. Answer to problem 3

a. The order of the words.

b.

$$\begin{aligned}
P(D_i|y_i) &= P(D_i|Y=y_i)\cdot P(Y=y_i) && (24)\\
&= (P(D_i|Y=1)\cdot P(Y=1))^{y_i}(P(D_i|Y=0)\cdot P(Y=0))^{1-y_i} && (25)\\
&= (\theta\frac{n!}{a_i!b_i!c_i!}\alpha_1^{a_i}\beta_1^{b_i}\gamma_1^{c_i})^{y_i}((1-\theta)\frac{n!}{a_i!b_i!c_i!}\alpha_0^{a_i}\beta_0^{b_i}\gamma_0^{c_i})^{1-y_i} && (26)
\end{aligned}$$

Taking log:

$$\begin{aligned}
L = & \sum_i log(P(D_i|y_i)) && (27)\\
= & \sum_i y_i(log\theta+log\frac{n!}{a_i!b_i!c_i!}+a_ilog\alpha_1+b_ilog\beta_1+c_ilog\gamma_1) && (28)\\
& +(1-y_i)(log(1-\theta)+log\frac{n!}{a_i!b_i!c_i!}+a_ilog\alpha_0+b_ilog\beta_0+c_ilog\gamma_0) && (29)
\end{aligned}$$

c.

$$\begin{aligned}
L' = & \sum_i y_i(log\theta+log\frac{n!}{a_i!b_i!c_i!}+a_ilog\alpha_1+b_ilog\beta_1+c_ilog\gamma_1) && (30)\\
& +(1-y_i)(log(1-\theta)+log\frac{n!}{a_i!b_i!c_i!}+a_ilog\alpha_0+b_ilog\beta_0+c_ilog\gamma_0) && (31)\\
& -\lambda_1(\alpha_1+\beta_1+\gamma_1-1)-\lambda_0(\alpha_0+\beta_0+\gamma_0-1) && (32)
\end{aligned}$$

$$\frac{\partial L'}{\partial \alpha_1} = \sum_i y_i \frac{a_i}{\alpha_1} - \lambda_1 = 0 \tag{33}$$

$$\frac{\partial L'}{\partial \beta_1} = \sum_i y_i \frac{b_i}{\beta_1} - \lambda_1 = 0 \tag{34}$$

$$\frac{\partial L'}{\partial \gamma_1} = \sum_i y_i \frac{c_i}{\gamma_1} - \lambda_1 = 0 \tag{35}$$

$$\tag{36}$$

$$\Rightarrow \alpha_1 = \frac{1}{\lambda} \sum_i y_i a_i \tag{37}$$

$$\Rightarrow \beta_1 = \frac{1}{\lambda} \sum_i y_i b_i \tag{38}$$

$$\Rightarrow \gamma_1 = \frac{1}{\lambda} \sum_i y_i c_i \tag{39}$$

Since $\alpha_1 + \beta_1 + \gamma_1 = \frac{1}{\lambda} \sum_i (a_i y_i + b_i y_i + c_i y_i) = 1$, we know that $\lambda = n \sum_i y_i$.

$$\Rightarrow \qquad \alpha_1 = \frac{\sum_i y_i a_i}{n \sum_i y_i}, \ \beta_1 = \frac{\sum_i y_i b_i}{n \sum_i y_i}, \ \gamma_1 = \frac{\sum_i y_i c_i}{n \sum_i y_i} \tag{40}$$

$$\Rightarrow \qquad \alpha_0 = \frac{\sum_i (1-y_i) a_i}{n \sum_i (1-y_i)}, \ \beta_0 = \frac{\sum_i (1-y_i) b_i}{n \sum_i (1-y_i)}, \ \gamma_0 = \frac{\sum_i (1-y_i) c_i}{n \sum_i (1-y_i)} \tag{41}$$

4. Answer to problem 4

$$P = \qquad (p^2)^k (1-p^2)^(n-k) \tag{42}$$

$$\Rightarrow logP = \qquad 2klogp + (n-k)log(1-p^2) \tag{43}$$

P is max, when $\frac{\partial P}{\partial p} = 0$.

$$\frac{\partial P}{\partial p} \qquad = \frac{2k}{p} + \frac{-2p(n-k)}{1-p^2} = 0 \tag{44}$$

$$\Rightarrow \qquad 2k = 2np^2 \tag{45}$$

$$\Rightarrow \qquad p = \sqrt{\frac{k}{n}} \tag{46}$$

In this sequence, $k = 4$ and $n = 10$. So, $p = \sqrt{\frac{4}{10}} = \sqrt{\frac{2}{5}}$.