CS446: Machine Learning **Fall 2014**

Problem Set 6

Chen Zhang *Handed In: 11/18/2014*

1 Problem 1

1.1 Part A

The function $f_{TH(3,7)}$ can be stated as:

$$\vec{w}^T = [\frac{1}{3}, \frac{1}{3}, \dots \frac{1}{3}]_{(7\times 1)}, \quad \theta = -1$$

$$f_{TH(3,7)} = \begin{cases} 1, & \vec{w}^T\vec{x_i} + \theta \geq 0 \\ 0, & \vec{w}^T\vec{x_i} + \theta < 0 \end{cases}$$

Obviously this is a linear function, thus $f_{TH(3,7)}$ has a linear decision surface over the 7 dimensional boolean cube.

1.2 Part B

Since this function can be expressed as a linear function and x_i and $f_{TH(3,7)}$ have only two values, the classifier could be expressed as

$$\theta = \log\frac{P(v_j = 1)}{P(v_j = 0)} + \sum_1^7 \log\frac{1-p_i}{1-q_i}$$
$$w_i = \log\frac{p_i}{1-p_i} - \log\frac{q_i}{1-q_i}$$

We can calculate that

$$P(v_j = 0) = (\frac{1}{2})^7 + (\frac{1}{2})^7 \times 7 + (\frac{1}{2})^7 \times C_7^2 = \frac{29}{128}$$
$$P(v_j = 1) = 1 - P(v_j = 0) = \frac{99}{128}$$
$$p_i = Pr(x_i = 1|v = 1) = \frac{Pr(xi = 1, v = 1)}{P(v = 1)} = \frac{57}{99}$$
$$q_i = Pr(x_i = 1|v = 0) = \frac{Pr(xi = 1, v = 0)}{P(v = 0)} = \frac{7}{29}$$

Plug in the expression for θ and w_i, we have

$$\theta = 7 \times \log \frac{\frac{42}{99}}{\frac{22}{29}} + \log \frac{\frac{99}{128}}{\frac{29}{128}} = 7 \times (-0.8384) + 1.77 = -4.1$$

$$w_i = \log \frac{p_i}{1-p_i} - \log \frac{q_i}{1-q_i} = \log \frac{\frac{57}{99}}{\frac{42}{99}} - \log \frac{\frac{7}{29}}{\frac{22}{29}} = 0.44 - (-1.652) = 2.1$$

1.3 Part C

Suppose that we have a vector

$$\vec{X} = [1,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0]$$

Plug in the classifier we have

$$\vec{w}^T \vec{x_i} + \theta = 0.1 > 0$$

Thus based on the classifier, $f_{TH(\vec{X})} = 1$. But obviously this is wrong. Thus the final hypothesis does not represent this function.

1.4 Part D

It's not satisfied. The Naive Bayes states that $(x_i|A)$ and $(x_j|A)$ are independent. Thus that

$$Pr(x_i, x_j|v) = Pr(x_i|v)Pr(x_j|v)$$

We know that

$$Pr(x_1 = 1, x_2 = 1|v = 0) = (\frac{1}{2})^7 = \frac{1}{128}$$
$$Pr(x_1 = 1|v = 0) \times Pr(x_2 = 1|v = 0) = (\frac{7}{29})^2 = \frac{49}{841}$$

These two values are not equal. Thus we can see that the Naive Bayes assumption is not satisfied.

2 Problem 2

We need to solve

$$argmax \quad Pr(D) = argmax \quad \prod Pr(D_i, y_i)$$

Call $\eta = Pr(y_i = 1)$, it follows

$$Pr(D) = \prod Pr(D_i, y_i) = \prod [\eta \frac{n!}{a_i!b_i!c_i!} \alpha_1^{a_i} \beta_1^{b_i} \gamma_1^{c_i}]^{y_i} [(1-\eta) \frac{n!}{a_i!b_i!c_i!} \alpha_0^{a_i} \beta_0^{b_i} \gamma_0^{c_i}]^{1-y_i}$$

Taking the logrithmic, it follows

$$\log Pr(D) = \sum \log Pr(D_i, y_i) = \sum y_i[\log \eta + C + a_i \log \alpha_1 + b_i \log \beta_1 + c_i \log \gamma_1] + \\ (1 - y_i)[\log (1 - \eta) + C' + a_i \log \alpha_0 + b_i \log \beta_0 + c_i \log \gamma_0]$$

Using Lagrange Multiplier to put in the constraints, we have

$$\log Pr(D) = \sum \log Pr(D_i, y_i) = \sum y_i[\log \eta + C + a_i \log \alpha_1 + b_i \log \beta_1 + c_i \log \gamma_1] + \\ (1 - y_i)[\log (1 - \eta) + C' + a_i \log \alpha_0 + b_i \log \beta_0 + c_i \log \gamma_0] \\ + \xi_1(\alpha_1 + \beta_1 + \gamma_1 - 1) + \xi_0(\alpha_0 + \beta_0 + \gamma_0 - 1) + \xi_2(a_i + b_i + c_i - n)$$

We take derivative with respect to α_1, β_1,γ_1,α_0, β_0,γ_0,ξ_0,ξ_1,ξ_2 and we get the following equation.

$$\begin{aligned}
\frac{d \log Pr(D)}{\alpha_1} &= \frac{1}{\alpha_1} \sum y_i a_i + \xi_1 = 0 \quad (1) \\
\frac{d \log Pr(D)}{\beta_1} &= \frac{1}{\beta_1} \sum y_i b_i + \xi_1 = 0 \quad (2) \\
\frac{d \log Pr(D)}{\gamma_1} &= \frac{1}{\gamma_1} \sum y_i c_i + \xi_1 = 0 \quad (3) \\
\frac{d \log Pr(D)}{\xi_1} &= \alpha_1 + \beta_1 + \gamma_1 - 1 = 0 \quad (4) \\
\frac{d \log Pr(D)}{\alpha_0} &= \frac{1}{\alpha_0} \sum (1 - y_i) a_i + \xi_1 = 0 \quad (5) \\
\frac{d \log Pr(D)}{\beta_0} &= \frac{1}{\beta_0} \sum (1 - y_i) b_i + \xi_1 = 0 \quad (6) \\
\frac{d \log Pr(D)}{\gamma_0} &= \frac{1}{\gamma_0} \sum (1 - y_i) c_i + \xi_1 = 0 \quad (7) \\
\frac{d \log Pr(D)}{\xi_0} &= \alpha_0 + \beta_0 + \gamma_0 - 1 = 0 \quad (8) \\
\frac{d \log Pr(D)}{\xi_2} &= a_i + b_i + c_i - n = 0 \quad (9)
\end{aligned}$$

Using (1)(2)(3)(4), we have

$$(\frac{\sum y_i b_i}{\sum y_i a_i} + \frac{\sum y_i c_i}{\sum y_i a_i} + \frac{\sum y_i a_i}{\sum y_i a_i})\alpha_1 = 1$$

Using (9), we have

$$(\frac{\sum y_i b_i}{\sum y_i a_i} + \frac{\sum y_i c_i}{\sum y_i a_i} + \frac{\sum y_i a_i}{\sum y_i a_i})\alpha_1 = \frac{\sum (a_i + b_i + c_i) y_i}{\sum y_i a_i}\alpha_1 = \frac{n \sum y_i}{\sum y_i a_i}\alpha_1 = 1$$

It leads to

$$\alpha_1 = \frac{\sum y_i a_i}{n \sum y_i}$$

Using the same method, we have

$$\left(\frac{\sum y_i b_i}{\sum y_i b_i} + \frac{\sum y_i c_i}{\sum y_i b_i} + \frac{\sum y_i a_i}{\sum y_i b_i}\right)\beta_1 = \frac{\sum (a_i + b_i + c_i) y_i}{\sum y_i b_i}\beta_1 = \frac{n \sum y_i}{\sum y_i b_i}\beta_1 = 1$$

and

$$\left(\frac{\sum y_i b_i}{\sum y_i c_i} + \frac{\sum y_i c_i}{\sum y_i c_i} + \frac{\sum y_i a_i}{\sum y_i c_i}\right)\gamma_1 = \frac{\sum (a_i + b_i + c_i) y_i}{\sum y_i c_i}\gamma_1 = \frac{n \sum y_i}{\sum y_i c_i}\gamma_1 = 1$$

For the result, we have

$$\beta_1 = \frac{\sum y_i b_i}{n \sum y_i}$$
$$\gamma_1 = \frac{\sum y_i c_i}{n \sum y_i}$$

Using (5)(6)(7)(8)(9), we could have

$$\left(\frac{\sum (1 - y_i) b_i}{\sum (1 - y_i) a_i} + \frac{\sum (1 - y_i) c_i}{\sum (1 - y_i) a_i} + \frac{\sum (1 - y_i) a_i}{\sum (1 - y_i) a_i}\right)\alpha_0 = 1$$

Using (9), we have

$$\left(\frac{\sum (1 - y_i) b_i}{\sum (1 - y_i) a_i} + \frac{\sum (1 - y_i) c_i}{\sum (1 - y_i) a_i} + \frac{\sum (1 - y_i) a_i}{\sum (1 - y_i) a_i}\right)\alpha_0 = \frac{\sum (a_i + b_i + c_i)(1 - y_i)}{\sum (1 - y_i) a_i}\alpha_0 = \frac{n \sum (1 - y_i)}{\sum (1 - y_i) a_i}\alpha_0 = 1$$

It leads to

$$\alpha_0 = \frac{\sum (1 - y_i) a_i}{n \sum (1 - y_i)}$$

Using the same method, we have

$$\left(\frac{\sum (1 - y_i) b_i}{\sum (1 - y_i) b_i} + \frac{\sum (1 - y_i) c_i}{\sum (1 - y_i) b_i} + \frac{\sum (1 - y_i) a_i}{\sum (1 - y_i) b_i}\right)\beta_0 = \frac{\sum (a_i + b_i + c_i)(1 - y_i)}{\sum (1 - y_i) b_i}\beta_0 = \frac{n \sum (1 - y_i)}{\sum (1 - y_i) b_i}\beta_0 = 1$$

and

$$\left(\frac{\sum (1 - y_i) b_i}{\sum (1 - y_i) c_i} + \frac{\sum (1 - y_i) c_i}{\sum (1 - y_i) c_i} + \frac{\sum (1 - y_i) a_i}{\sum (1 - y_i) c_i}\right)\gamma_0 = \frac{\sum (a_i + b_i + c_i)(1 - y_i)}{\sum (1 - y_i) c_i}\gamma_0 = \frac{n \sum (1 - y_i)}{\sum (1 - y_i) c_i}\gamma_0 = 1$$

For the result, we have

$$\beta_0 = \frac{\sum (1 - y_i) b_i}{n \sum (1 - y_i)}$$
$$\gamma_0 = \frac{\sum (1 - y_i) c_i}{n \sum (1 - y_i)}$$

3 Problem 3

3.1 Part A

It's easy to see that $Pr(Y = A) = \frac{3}{7}$ and $Pr(Y = B) = \frac{4}{7}$. For the parameters, we need to solve

$$argmax \quad Pr(D) = argmax \quad \prod Pr(D_i)$$

Using the Bayes rule, it becomes

$$argmax \quad \prod Pr(D_i) = argmax \quad \prod Pr(D_i|Y)Pr(Y)$$

Plug in the values, we could have

$$Pr(D) = C \times \exp[-3\lambda_1^A](\lambda_1^A)^6 \exp[-3\lambda_2^A](\lambda_2^A)^{15} \exp[-4\lambda_1^B](\lambda_1^B)^{16} \exp[-4\lambda_2^B](\lambda_2^B)^{12}$$

where $C = \frac{3}{7}\frac{4}{7}\frac{1}{0!4!2!3!8!4!6!3!2!5!2!5!1!4!}$. By taking derivatives and setting to zeros, it follows

$$\frac{dPr(D)}{d\lambda_1^A} = \frac{dPr(D)}{d\lambda_2^A} = \frac{dPr(D)}{d\lambda_1^B} = \frac{dPr(D)}{d\lambda_2^B} = 0$$

Solving the equation, it leads to $\lambda_1^A = 2, \lambda_2^A = 5, \lambda_1^B = 4, \lambda_2^B = 3$.

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

Table 1: Parameters for Poisson naïve Bayes

3.2 Part B

$$\begin{aligned}\frac{Pr(X_1 = 2, X_2 = 3|Y = A)}{Pr(X_1 = 2, X_2 = 3|Y = B)} &= \frac{Pr(X_1 = 2|Y = A)Pr(X_2 = 3|Y = A)}{Pr(X_1 = 2|Y = B)Pr(X_2 = 3|Y = B)} \\ &= \frac{\exp[-\lambda_1^A](\lambda_1^A)^2 \exp[-\lambda_2^A](\lambda_2^A)^3}{\exp[-\lambda_1^B](\lambda_1^B)^2 \exp[-\lambda_2^B](\lambda_2^B)^3}.\end{aligned}$$

By solving, we have

$$\begin{aligned}\frac{Pr(X_1 = 2, X_2 = 3|Y = A)}{Pr(X_1 = 2, X_2 = 3|Y = B)} &= \exp[-\lambda_1^A - \lambda_2^A + \lambda_1^B + \lambda_2^B](\frac{\lambda_1^A}{\lambda_1^B})^2(\frac{\lambda_2^A}{\lambda_2^B})^3 \\ &= \frac{125}{108}\end{aligned}$$

3.3 Part C

Using Naive Bayes, it could be stated as

$$argmax \quad Pr(X_1, X_2|Y) = argmax \quad Pr(X_1|Y)Pr(X_2|Y)$$

Thus we can determine $Y = A$ if

$$\frac{Pr(X_1, X_2|Y = A)Pr(Y = A)}{Pr(X_1, X_2|Y = B)Pr(Y = B)} = \frac{Pr(X_1|Y = A)Pr(X_2|Y = A)Pr(Y = A)}{Pr(X_1|Y = B)Pr(X_2|Y = B)Pr(Y = B)} > 1$$

Plugging in the values, we have $Y = A$ if

$$\exp[-\lambda_1^A - \lambda_2^A + \lambda_1^B + \lambda_2^B](\frac{\lambda_1^A}{\lambda_1^B})^2(\frac{\lambda_2^A}{\lambda_2^B})^3\frac{3}{4} > 1$$

and $Y = B$ otherwise.

Plugging in the value of $X_1 = 3, X_2 = 3$, we have

$$\exp[-\lambda_1^A - \lambda_2^A + \lambda_1^B + \lambda_2^B](\frac{\lambda_1^A}{\lambda_1^B})^2(\frac{\lambda_2^A}{\lambda_2^B})^3\frac{3}{4} = \frac{125}{144} < 1$$

Thus the label is $Y = B$. We could see the role of the prior played in this case.

4 Problem 4

We could calculate that the probability of having an H is $Pr(H) = p^2$ and the probability of having a T is $Pr(T) = 1 - p^2$.

Thus with the Bernoulli model, the probability of having TTHTHHTHTT is

$$Pr = (1 - p^2)^6 p^8$$

. The MLE can be calculated as

$$\frac{dPr}{dp} = 8(1 - p^2)^6 p^7 - 12(1 - p^2)^5 p^9 = 0$$

The most likely value of p can be solved as

$$p = \frac{\sqrt{10}}{5} = 0.632$$