CSCI3320 Homework 3 1155110208 Chung Tsz Ting

1) Let x be the result of test and x=1 be the positive testing result while x=0 be the negative testing result, y be the actual result whether the patient has the disease and y=1 is positive while y=0 is negative. The probability that the patient has the disease given the test result is positive is P(y=1|x=1).

$$P(y=1|x=1) = \frac{P(x=1|y=1)P(y=1)}{P(x=1)}$$
$$= \frac{P(x=1|y=1)P(y=1)}{P(x=1|y=1)P(y=1)+P(x=1|y=0)P(y=0)}$$
$$= \frac{0.99\cdot(\frac{1}{1000000})}{0.99\cdot\left(\frac{1}{1000000}\right)+\left(\frac{1}{1000}\right)\cdot\left(\frac{999999}{1000000}\right)}$$
$$= 0.000989$$

The probability that the patient has the disease given the positive testing result is only 0.000989 which is because of the rarity of the disease.

2) As the log odd is $log\frac{p(C_1|x)}{p(C_2|x)}$, the discriminant function is

$$g_i(x) = \log\ p(C_i|x)$$
$$= \log(\ p(x|C_i)p(C_i))$$
$$g(x) = \log\frac{p(C_1|x)}{p(C_2|x)}\ and\ choose\ \ C_1\ if\ g(x) > 0\ else\ C_2$$
$$= log\frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}$$

3) As the expected risk for an action a_i is $R(a_i|x) = \sum_{k=1}^{K}\lambda_{ik}P(C_k|x)$
$where\ x\ is\ data,$
$P(C_k|x)\ is\ the\ probability\ as\ class\ C_k\ given\ x,$
$\lambda_{ik}\ is\ the\ loss\ of\ action\ \ a_i\ when\ state\ is\ C_k.$

Thus, $R(a_1|x) = \lambda_{11}P(C_1|x) + \lambda_{12}P(C_2|x)$
$= 0\cdot P(C_1|x) + 10\cdot P(C_2|x)$
$= 10\cdot(1 - P(C_1|x))$

And, $R(a_2|x) = \lambda_{21}P(C_1|x) + \lambda_{22}P(C_2|x)$
$= 5\cdot P(C_1|x) + 0\cdot P(C_2|x)$
$= 5\cdot P(C_1|x)$

Thus, the decision is performed $a_1$if $R(a_1|x) < R(a_2|x)$ else perform a_2.

$$R(a_1|x) < R(a_2|x)$$
$$10\cdot(1 - P(C_1|x)) < 5\cdot P(C_1|x)$$
$$10 < 15\cdot P(C_1|x)$$
$$\therefore P(C_1|x) < \frac{2}{3}\ \ then\ perform\ a_1\ else\ perform\ a_2$$

4) As Support (X → Y) = P(X,Y) and Confidence (X → Y) = P(Y|X),
Support (milk → bananas) = 2/6
=1/3

Confidence (milk → bananas)= 2/4
= ½

Support (bananas → milk) = 2/6
= 1/3

Confidence (bananas → milk) = 2/2
= 1

Support (milk → chocolate) = 3/6
= 1/2

Confidence (milk → chocolate) = 3/4

Support (chocolate → milk) = 3/6
= 1/2
Confidence (chocolate → milk) = 3/5

5) As $P(x_1, x_2, \dots, x_K) = \prod_{i=1}^{K} p_i^{x_i}$ for K class in which $\sum_{i=1}^{K} p_i = 1$ and p_i stands for the probability of class I, $x_1, \dots, x_K$ be the indicator variables such that $x_i = 1$ for outcome as class i and 0 otherwise.

Thus, the log likelihood is

$$
\begin{aligned}
\mathcal{L}(\theta|X) &= \log p(X|\theta) \\
&= \log \textstyle\prod_{i=1}^{K} \prod_{t=1}^{N} p(x_i^t|\theta) \quad \mathit{for\ t\ classes\ and\ N\ numbers\ of\ sample} \\
&= \textstyle\sum_{i=1}^{K} \sum_{t=1}^{N} \log p(x^t|\theta) \\
&= \textstyle\sum_{i=1}^{K} \sum_{t=1}^{N} \log p_i^{x_i^t}
\end{aligned}
$$

As the constraint of $\sum_{i=1}^{K} p_i = 1$ also need to be satisfied, Lagrange multipliers is used, i.e. to maximize $\mathcal{L}(\theta|X)$ subject to $\sum_{i=1}^{K} p_i = 1$

$$\mathcal{L}(x_1, \dots, x_K, \lambda) = \mathcal{L}(x_1, x_2, \dots, x_K) + \lambda(1 - \sum_{i=1}^{K} p_i)$$

$$\mathcal{L}(x_1, \dots, x_K, \lambda) = \sum_{i=1}^{K} \sum_{t=1}^{N} \log p_i^{x_i^t} + \lambda(1 - \sum_{i=1}^{K} p_i)$$

$$\frac{d\mathcal{L}(x_1, \dots, x_K, \lambda)}{dp_i} = \frac{d \sum_{i=1}^{K} \sum_{t=1}^{N} x_i^t \cdot \log p_i}{dp_i} + \frac{d\lambda(1 - \sum_{i=1}^{K} p_i)}{dp_i}$$

$$\frac{d\mathcal{L}(x_1, \dots, x_K, \lambda)}{dp_i} = \sum_{t=1}^{N} \frac{x_i^t}{p_i} + \frac{d\lambda}{dp_i} - \frac{d\lambda \sum_{i=1}^{K} p_i}{dp_i}$$

$$0 = \frac{\sum_{t=1}^{N} x_i^t}{p_i} + 0 - \lambda$$

$$\lambda p_i = \sum_{t=1}^{N} x_i^t$$
$$\sum_{i=1}^{K} \lambda p_i = \sum_{i=1}^{K} \sum_{t=1}^{N} x_i^t$$
$$\lambda = \sum_{t=1}^{N} \sum_{i=1}^{K} x_i^t \quad (\because \sum_{i=1}^{K} p_i = 1)$$
$$\lambda = \sum_{t=1}^{N} 1$$
$$\lambda = N$$

$$\therefore \widehat{p_i} = \frac{\sum_{t=1}^{N} x_i^t}{\lambda}$$
$$= \frac{\sum_{t=1}^{N} x_i^t}{N}$$

6) As for $p(x){:}\, \mathcal{N}(\mu, \sigma^2)$, $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\,[-\frac{(x-\mu)^2}{2\sigma^2}]$
   The discriminant points are,

$$P(C_1|x) = P(C_2|x)$$
$$P(x|C_1)P(C_1) = P(x|C_2)P(C_2)$$
$$\log P(x|C_1) + \log P(C_1) = \log P(x|C_2) + \log P(C_2)$$
$$\log \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(x-\mu_1)^2}{2{\sigma_1}^2}\right] + \log P(C_1) = \log \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{(x-\mu_2)^2}{2{\sigma_2}^2}\right] + \log P(C_2)$$
$$-\tfrac{1}{2} \log 2\pi - \log \sigma_1 - \tfrac{(x-\mu_1)^2}{2{\sigma_1}^2} + \log P(C_1) = -\tfrac{1}{2} \log 2\pi - \log \sigma_2 - \tfrac{(x-\mu_2)^2}{2{\sigma_2}^2} + \log P(C_2)$$
$$-\log \sigma_1 - \frac{x^2 - 2x\mu_1 + {\mu_1}^2}{2{\sigma_1}^2} + \log P(C_1) = -\log \sigma_2 - \frac{x^2 - 2x\mu_2 + {\mu_2}^2}{2{\sigma_2}^2} + \log P(C_2)$$
$$\left(\frac{1}{2{\sigma_2}^2} - \frac{1}{2{\sigma_1}^2}\right)x^2 + \left(\frac{\mu_1}{{\sigma_1}^2} - \frac{\mu_2}{2{\sigma_2}^2}\right)x - \frac{\mu_1}{2{\sigma_1}^2} + \frac{\mu_2}{2{\sigma_2}^2} + \log \frac{\sigma_2}{\sigma_1} + \log \frac{P(C_1)}{P(C_2)} = 0$$

   Thus, the discriminant points are

$$x = \frac{-\left(\frac{\mu_1}{{\sigma_1}^2} - \frac{\mu_2}{2{\sigma_2}^2}\right) + \sqrt{\left(\frac{\mu_1}{{\sigma_1}^2} - \frac{\mu_2}{2{\sigma_2}^2}\right)^2 - 4\left(\frac{1}{2{\sigma_2}^2} - \frac{1}{2{\sigma_1}^2}\right)\left(\frac{\mu_2}{2{\sigma_2}^2} - \frac{\mu_1}{2{\sigma_1}^2} + \log \frac{\sigma_2}{\sigma_1} + \log \frac{P(C_1)}{P(C_2)}\right)}}{2\left(\frac{1}{2{\sigma_2}^2} - \frac{1}{2{\sigma_1}^2}\right)}$$

or

$$x = \frac{-\left(\frac{\mu_1}{{\sigma_1}^2} - \frac{\mu_2}{2{\sigma_2}^2}\right) - \sqrt{\left(\frac{\mu_1}{{\sigma_1}^2} - \frac{\mu_2}{2{\sigma_2}^2}\right)^2 - 4\left(\frac{1}{2{\sigma_2}^2} - \frac{1}{2{\sigma_1}^2}\right)\left(\frac{\mu_2}{2{\sigma_2}^2} - \frac{\mu_1}{2{\sigma_1}^2} + \log \frac{\sigma_2}{\sigma_1} + \log \frac{P(C_1)}{P(C_2)}\right)}}{2\left(\frac{1}{2{\sigma_2}^2} - \frac{1}{2{\sigma_1}^2}\right)}$$