Homework 1

Bohan Niu

November 9, 2018

1

Suppose that she choose the first door at the beginning. Then we define three situation after her choice:

A: The first one contains prize.

B: The second one contains prize.

C: She chose the first door and then the host opens the third door, which contains no prize.

Before the host opens the door, the probability of each door containing the prize is the same, which equals to $\frac{1}{3}$.Under this situation, if she will not switch, the posterior probability that she wins is $P(A|C)$, using Bayes rule, we get:

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)}$$

The prior probabilities of C in this case equals to the probability the first door has prize then host opens the third plus the probability the first door does not have prize then host opens the third. That is:

$$P(C) = P(A)P(C|A) + P(\overline{A})P(C|\overline{A}) = \frac{1}{3} * \frac{1}{2} + \frac{2}{3} * \frac{1}{2} = \frac{1}{2}$$

Therefore, we can calculate the $P(A|C)$:

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)} = \frac{\frac{1}{2} * \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

On the other hand, the the posterior probability that she changes the door and wins $P(B|C)$ equals to:

$$P(B|C) = 1 - P(A|C) = 1 = \frac{1}{3} = \frac{2}{3}$$

Due to the fact that $P(B|C) > P(A|C)$, I suggest her to swich doors.

2

In this problem, $X_i \sim$ Multinomial(π), i.i.d. for $i = 1, ..., N$. From Bayes rule, we know that:

$$\begin{aligned} p(\pi|X_1, ..., X_k) &\propto p(X_1, ..., X_k|\pi)p(\pi) \\ &\propto \frac{N!}{X_1!X_2!...X_k!} \prod_{i=1}^{k} \pi_i^{X_i} p(\pi) \\ &\propto \prod_{i=1}^{k} \pi_i^{X_i} p(\pi) \end{aligned}$$

Then, due to the equation we get, we assume that $p(\pi)$ follows the Dirichlet distribution, which looks very similar to the equation:

$$\begin{aligned} p(\pi|X_1, ..., X_k) &\propto \prod_{i=1}^{k} \pi_i^{X_i} p(\pi) \\ &\propto \prod_{i=1}^{k} \pi_i^{X_i} \frac{1}{B(\alpha)} \prod_{i=1}^{k} \pi_i^{\alpha_i - 1} \\ &\propto \prod_{i=1}^{k} \pi_i^{X_i + \alpha_i - 1} \end{aligned}$$

$\frac{N!}{X_1!X_2!...X_k!}$ and $\frac{1}{B(\alpha)}$ are ignored for they have nothing to do with $p(\pi)$. It is clear that the posterior distribution of π is also follows the Dirichlet distribution, that is, $p(\pi|X_1, ..., X_k) \sim Dir(X + \alpha)$.The most obvious feature about the parameters of this posterior distribution is that it is the original parameter combined with X.

3

a. By Bayes rule, we can get:

$$\begin{aligned} p(\lambda|x_1, ..., x_n) &\propto p(x_1, ...x_n|\lambda)p(\lambda) \\ &\propto \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \\ &\propto \prod_{i=1}^{N} (e^{-\lambda}\lambda^{x_i}) \lambda^{a-1} e^{-b\lambda} \\ &\propto e^{-N\lambda} \lambda^{\Sigma_i x_i} \lambda^{a-1} e^{-b\lambda} \\ &\propto \lambda^{\Sigma_i x_i + a - 1} e^{-(b+N)\lambda} \end{aligned}$$

Which is similar to Gamma distribution. Therefore, the the posterior of λ follows the Gamma distribution, that is $\lambda \sim \Gamma(\Sigma_i x_i + a, N + b)$

b.It is reasonable to assume that the new observation $x^* \sim P(\lambda)$ i.i.d. and $x^* \in \mathbb{N}$. Therefore:

$$\begin{aligned}
p(x^*|x_1, ..., x_n) &= \int_0^\infty p(x^*|\lambda)p(\lambda|x_1, ..., x_n)d\lambda \\
&= \int_0^\infty p(x^*|\lambda)\frac{(N+b)^{\Sigma_i x_i + a}}{\Gamma(\Sigma_i x_i + a)}\lambda^{\Sigma_i x_i + a - 1}e^{-(b+N)\lambda}d\lambda \\
&= \frac{(N+b)^{\Sigma_i x_i + a}}{\Gamma(\Sigma_i x_i + a)}\int_0^\infty \frac{e^{-\lambda}\lambda x^*}{x^*!}\lambda^{\Sigma_i x_i + a - 1}e^{-(b+N)\lambda}d\lambda \\
&= \frac{(N+b)^{\Sigma_i x_i + a}}{\Gamma(\Sigma_i x_i + a)x^*!}\int_0^\infty \lambda^{\Sigma_i x_i + x^* + a - 1}e^{-(b+N+1)\lambda}d\lambda
\end{aligned}$$

To make $\int_0^\infty \lambda^{\Sigma_i x_i + a - 1}e^{-(b+N)\lambda}d\lambda$ a complete the Gamma distribution, we first multiply $\frac{(N+b+1)^{\Sigma_i x_i + x^* + a}}{\Gamma(\Sigma_i x_i + x^* + a)}$, then divide the same one on the left of $\int$. That is:

$$\begin{aligned}
p(x^*|x_1, ..., x_n) &= \frac{(N+b)^{\Sigma_i x_i + a}}{\Gamma(\Sigma_i x_i + a)x^*!}\frac{\Gamma(\Sigma_i x_i + x^* + a)}{(N+b+1)^{\Sigma_i x_i + x^* + a - 1}}\cdot \\
&\int_0^\infty \frac{(N+b+1)^{\Sigma_i x_i + x^* + a}}{\Gamma(\Sigma_i x_i + x^* + a)}\lambda^{\Sigma_i x_i + x^* + a - 1}e^{-(b+N+1)\lambda}d\lambda \\
&= \frac{(N+b)^{\Sigma_i x_i + a}}{\Gamma(\Sigma_i x_i + a)x^*!}\frac{\Gamma(\Sigma_i x_i + x^* + a)}{(N+b+1)^{\Sigma_i x_i + x^* + a}} \\
&= \frac{\Gamma(\Sigma_i x_i + x^* + a)}{\Gamma(\Sigma_i x_i + a)\Gamma(x^* + 1)}\frac{(N+b)^{\Sigma_i x_i + a}}{(N+b+1)^{\Sigma_i x_i + x^* + a}}
\end{aligned}$$

4

b. confusion matrix

	spam	non-spam
spam	176	6
non-spam	71	209

From the confusion matrix, we can learn that the program does well in finding spam. The performance of distinguishing non-spam is a little weak, but is still OK.

c. We choose email No.26, No.188, No. 208. The the predictive probabilities of them are (1.0, 1.2979686149e-83), (2.28477999601e-08, 0.999999977152), (1.25719243598e-07, 0.999999874281). For each pair, the first value is the predictive probability when y=0, the second one is when y = 1.

d.The three most ambiguous predictions is email No.240, No.431 and No.432. Their predictive probabilities are (0.347852183361, 0.652147816639), (0.257858818686, 0.742141181314), (0.701556878927, 0.298443121073), respectively.