Stat 8003, Homework 5

Group G: `sample( c( "David" , "Andrew", "Salam" ))`

September 30, 2014

Question 5.1. Consider a simulated dataset. Assume that the data $x_1, x_2, \cdots, x_n$ follows the following distribution:

$$x_i \sim f(x_i) = \pi_0 f_0(x_i) + \pi_1 f_1(x_i)$$

where $f_0(x_i) = 1(0 \leq x_i \leq 1)$ is the density function of the uniform and $f_1(x_i) = \beta(1 - x)^{\beta-1}$ is the density function of $Beta(1, \beta)$. The group information can be treated as a missing value and is denoted as z_i. Let $y_i = (x_i, z_i)$ be the complete data.

(a) Derive the complete likelihood function;

(b) Use the EM algorithm to derive the estimator for π_0 and β;

(c) Apply your method to the data set, estimate π_0 and β and the calculate $fdr_i = P(Z_i = 0 \mid x_i)$. (This score is called the local fdr score.)

(d) Classify x_i to the first group if $fdr_i(x_i) > 0.5$. Compare your classification with the actual group information, what is the total number of falsely classified data?

Answer:

(a) First, the *incomplete* likelihood function is given to be:

$$L(\theta\,; \mathbf{X}) = \prod_{i=1}^{n} \left(\pi_0 1 + \pi_1 \beta(1 - x_i)^{\beta-1}\right)$$

Then the *complete* likelihood function is:

$$\boxed{L(\theta\,; \mathbf{Y}) = \prod_{i=1}^{n} \left(1(Z_i = 0)\,\pi_0 + 1(Z_i = 1)\,\pi_1\,\beta(1 - x_i)^{\beta-1}\right)}$$

An alternative way of writing this likelihood is:

$$f(x_i, z_i \mid \theta) = \begin{cases} \pi_0 & \text{if } Z_i = 0 \\ \pi_1\,\beta(1 - x_i)^{\beta-1} & \text{if } Z_i = 1 \end{cases}$$

(b) To get the estimates for π_0 and β, we first find the expected value of the *log* of the *complete likelihood* function with respect to Z (the so called Q function):

$$\begin{aligned}
Q(\theta \mid \theta^t) &= \mathrm{E}\ \log(L(\theta\,; \boldsymbol{Y})) \\
&= \mathrm{E}\ \log\left(\prod_{i=1}^{n}\left(1(Z_i=0)\ \pi_0 + 1(Z_i=1)\ \pi_1\ \beta(1-x_i)^{\beta-1}\right)\right) \\
&= \mathrm{E}\ \left[\sum_{i=1}^{n}\log\left(1(Z_i=0)\ \pi_0 + 1(Z_i=1)\ \pi_1\ \beta(1-x_i)^{\beta-1}\right)\right]
\end{aligned}$$

The last expression in the brackets is either $\log(\pi_0)$ or $\log(\pi_1\ \beta(1-x_i)^{\beta-1})$, depending on the outcome of Z. So

$$\begin{aligned}
Q(\theta \mid \theta^t) &= \sum_{i=1}^{n}\left(\ \mathrm{E}\ 1(Z_i=0)\ \log(\pi_0) + \mathrm{E}\ 1(Z_i=1)\ \log(\pi_1\ \beta(1-x_i)^{\beta-1})\ \right) \\
&= \sum_{i=1}^{n}\left(\ P(Z_i=0 \mid x_i,\theta)\ \log(\pi_0) + P(Z_i=1 \mid x_i,\theta)\ \log(\pi_1\ \beta(1-x_i)^{\beta-1})\ \right)
\end{aligned}$$

Where the last equality follows because the expectation of the indicator function of a `r.v.` is simply the probability of the corresponding event.

These probabilities will be computed using Bayes rule and denoted by T^t_{ij}:

$$T^t_{ij} = P(Z_i=j \mid x_i,\theta) = \frac{P(x_i \mid Z_i=j)P(Z_i=j)}{\sum_{j=0}^{1} P(x_i \mid Z_i=j)P(Z_i=j)} \quad \text{for} \ \ j=0,1$$

Thus

$$T^t_{i0} = \frac{\pi_0}{\pi_0 + \pi_1\ \beta(1-x_i)^{\beta-1}}$$

$$T^t_{i1} = \frac{\pi_1\ \beta(1-x_i)^{\beta-1}}{\pi_0 + \pi_1\ \beta(1-x_i)^{\beta-1}}$$

Rewriting the Q function:

$$\begin{aligned}
Q(\theta \mid \theta^t) &= \sum_{i=1}^{n}\left(\ T^t_{i0}\log(\pi_0) + T^t_{i1}\log(\pi_1\ \beta(1-x_i)^{\beta-1})\ \right) \\
&= \sum_{i=1}^{n}\left(\ T^t_{i0}\log(\pi_0) + T^t_{i1}\log(1-\pi_0)\ + T^t_{i1}\log(\beta(1-x_i)^{\beta-1}\ \right)
\end{aligned}$$

Now we need to maximize this function with respect to π_0 and β......

```
.....
.....
.....
.....
```

```
invoke maximizer extrodinaire...

should have started earlier...
```

Question 5.2. (Continued from Problem 1.) It is known that the local fdr score can be written as

$$fdr_i(x_i) = \frac{\pi_0 f_0(x_i)}{f(x_i)}$$

where $f(x_i)$ is the marginal density of x_i. Assume that $\pi = 0.7$.

(a) Estimate $f(x_i)$ by using the kernel density estimation with Gaussian kernel and Silverman's h;

(b) Estimate the local fdr score;

(c) Using the same rule as in 1(d), calculate the total number of falsely classified data;

(d) Choose the bandwidth using the maximum likelihood cross validation, repeat problem (a-c), what is the total number of falsely classified data?

(e) Which method works the best in terms of having the smallest classification error?

Answer: