Machine Learning – EM Exercise

Jack 1930026143

1. Find the maximum likelihood estimate (MLE) of λ of this distribution:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

(1) $f(x) = \lambda e^{-\lambda x}$

(2) $L(\lambda) = \prod_{i-1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$

(3) $l(\lambda) = \ln L(\lambda) = \ln \lambda^n + \ln e^{-\lambda \sum_{i=1}^{n} x_i} = n \ln \lambda + - \lambda \sum_{i=1}^{n} x_i$

(4) $\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \quad \Rightarrow \quad \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$

2. MLE of normally distributed:

a.

(1) $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

(2) $L(\mu, \sigma^2) = \prod_{i-1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$

(3) $l(\mu, \sigma^2) = \ln L(\mu, \sigma^2) = \ln (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}} = \ln (2\pi\sigma^2)^{-\frac{n}{2}} + \ln e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$= -\frac{n}{2}(\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

(4) $\begin{cases} \frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \\ \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \end{cases} \Rightarrow \begin{cases} \mu = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \\ \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \end{cases}$

To conclude, $\mu_f = \frac{1}{n} \sum_{i=1}^{n} x_{i(f)} = 159.6667$, $\sigma_f^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_{i(f)} - \mu_f\right)^2 = 31.5555$

$\mu_m = 175$, $\sigma_m^2 = 16.6667$

b.

Initialization: $\pi_m^0 = \pi_f^0 = 0.5$. $\mu_m^0 = 170, \sigma_m^{2(0)} = 20$; $\mu_f^0 = 159, \sigma_f^{2(0)} = 25$.

Firstly, do the E step:

$$w_m^i = p(z^i = j) = \frac{p(x^i | z^i = j)\, p(z^i = j)}{p(x_i)}$$

$$= \frac{\pi_m^0 * \frac{1}{\sqrt{2\pi * \sigma_m^{2(0)}}} * e^{\frac{(-x_i - \mu_m^0)^2}{2*\sigma_m^{2(0)}}}}{\pi_m^0 * \frac{1}{\sqrt{2\pi * \sigma_m^{2(0)}}} * e^{\frac{(-x_i - \mu_m^0)^2}{2*\sigma_m^{2(0)}}} + \pi_f^0 * \frac{1}{\sqrt{2\pi * \sigma_f^{2(0)}}} * e^{\frac{(-x_i - \mu_f^0)^2}{2*\sigma_f^{2(0)}}}}$$

Which is based on the bayes theorem.

Thus, we can get:

w_m^1	w_m^1
0.9984	0.0016
0.9263	0.0737
0.6663	0.3337
0.9901	0.0099
0.0856	0.9144
0.0055	0.9945
0.5515	0.4485
0.2127	0.7873
0.0003	0.9997

Then we can proceed to the M step:

$$\pi_m^1 = \frac{\sum_{j=1} w_{m(j)}}{m} = 0.4930, \ \pi_f^1 = \frac{\sum_{j=1} w_{f(j)}}{m} = 0.5070$$

$$\mu_f^1 = \frac{\sum_{j=1} w_{f(j)} x_j}{\sum_{j=1} w_{f(j)}} = 158.1956, \ \sigma_f^{2(1)} = \frac{\sum_{j=1} w_{f(j)} \left(x - \mu_f^1\right)\left(x - \mu_f^1\right)^T}{\sum_{j=1} w_{f(j)}} = 32.7712$$

$$\mu_m^1 = \frac{\sum_{j=1} w_{m(j)} x_j}{\sum_{j=1} w_{m(j)}} = 171.5475, \ \sigma_m^{2(1)} = \frac{\sum_{j=1} w_{m(j)} (x - \mu_m^1)(x - \mu_m^1)^T}{\sum_{j=1} w_{m(j)}} = 38.8944.$$