

732A90 Computational Statistics

Lecture 6

Derivation of EM algorithm for Slide 15

KRZYSZTOF BARTOSZEK

7 XII 2020

E-step: derive

$$Q(\theta, \theta^k) = \mathbb{E} [\log \text{lik}(\theta | Y, Z) | \theta^k, Y] = \mathbb{E} \left[-\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^n (y_i - \mu)^2 \right) \middle| \theta^k, Y \right] \stackrel{*}{=} .$$

Our observed data is $Y = \{y_1, \dots, y_r\}$ and the latent variables (unobserved data) is $Z = \{y_{r+1}, \dots, y_n\}$. Hence, continuing we have

$$\begin{aligned} & \stackrel{*}{=} -\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathbb{E} \left[\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2 \middle| \theta^k, Y \right] \\ &= -\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathbb{E} \left[\sum_{i=1}^r y_i^2 + \sum_{i=r+1}^n y_i^2 - 2\mu \sum_{i=1}^r y_i - 2\mu \sum_{i=r+1}^n y_i + n\mu^2 \middle| \theta^k, Y \right] \\ &= -\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^r y_i^2 + \sum_{i=r+1}^n \mathbb{E} [y_i^2 | \theta^k] - 2\mu \sum_{i=1}^r y_i - 2\mu \sum_{i=r+1}^n \mathbb{E} [y_i | \theta^k] + n\mu^2 \\ &= -\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^r y_i^2 + \sum_{i=r+1}^n (\mu_k^2 + \sigma_k^2) - 2\mu \sum_{i=1}^r y_i - 2\mu \sum_{i=r+1}^n \mu_k + n\mu^2 \\ &= -\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^r (y_i - \mu)^2 + (n-r)(\mu - \mu_k)^2 + (n-r)\sigma_k^2 \\ &= \{\text{setting the unobserved } y_i \equiv \mu_k\} \\ &= -\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 + \frac{n-r}{2\sigma^2} \sigma_k^2 = Q(\theta, \theta^k). \end{aligned}$$

M-step: taking the derivatives with respect to the parameters μ and σ^2

$$\begin{aligned} Q(\theta, \theta^k)'_{\mu} &= \frac{2}{2\sigma^2} \sum_{i=1}^n (y_i - \mu), \\ Q(\theta, \theta^k)'_{\sigma^2} &= -\frac{n}{2} \frac{1}{2\pi\sigma^2} \cdot 2\pi - \left(\sum_{i=1}^n (y_i - \mu)^2 \right) \frac{-1}{2} \frac{1}{(\sigma^2)^2} - \frac{1}{2} (n-r) \sigma_k^2 (-1) \frac{1}{(\sigma^2)^2}. \end{aligned}$$

Setting the first equation to 0 we obtain $0 = \sum_{i=1}^n (y_i - \mu)$, giving $n\mu = \sum y_i$ and in turn

$$\mu_{k+1} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Turning to the second equation

$$\begin{aligned} 0 &= -\frac{n}{2} \frac{1}{2\pi\sigma^2} \cdot 2\pi - \left(\sum_{i=1}^n (y_i - \mu)^2 \right) \frac{-1}{2} \frac{1}{(\sigma^2)^2} - \frac{1}{2}(n-r)\sigma_k^2(-1) \frac{1}{(\sigma^2)^2} \\ 0 &= -\frac{n}{2} + \frac{1}{(\sigma^2)^2} \left(\sum_{i=1}^n (y_i - \mu)^2 \right) + \frac{\sigma_k^2}{(\sigma^2)^2} (n-r) \\ n\sigma^2 &= \left(\sum_{i=1}^n (y_i - \mu)^2 \right) + (n-r)\sigma_k^2 \\ \sigma^2 &= \frac{1}{n} \left(\sum_{i=1}^n (y_i - \mu)^2 \right) + \frac{1}{n}(n-r)\sigma_k^2 \\ &= \frac{1}{n} \left(\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2 + (n-r)\sigma_k^2 \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^r y_i^2 + \sum_{i=r+1}^n \mu_k^2 - 2\mu \sum_{i=1}^r y_i - 2\mu \sum_{i=r+1}^n \mu_k + n\mu^2 + (n-r)\sigma_k^2 \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^r y_i^2 + (n-r)\mu_k^2 - 2\mu \sum_{i=1}^r y_i - 2(n-r)\mu\mu_k + n\mu^2 + (n-r)\sigma_k^2 \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^r y_i^2 + (n-r)(\mu_k^2 + \sigma_k^2) \right) + \mu^2 - \frac{2}{n}\mu \sum_{i=1}^n y_i. \end{aligned}$$

As we now have the estimate at the $(k+1)$ st step of μ as μ_{k+1} , then

$$\mu_{k+1}^2 - \frac{2}{n}\mu_{k+1} \sum_{i=1}^n y_i = \mu_{k+1}^2 - 2\mu_{k+1} \frac{1}{n} \sum_{i=1}^n y_i = \mu_{k+1}^2 - 2\mu_{k+1}^2 = -\mu_{k+1}^2$$

and we obtain

$$\sigma_{k+1}^2 = \frac{1}{n} \left(\sum_{i=1}^r y_i^2 + (n-r)(\mu_k^2 + \sigma_k^2) \right) - \mu_{k+1}^2.$$

The next step could then be to e.g. consider the second derivative and show that this is actually a maximum.