CM224 HW 5 Solution

Ashish Kumar Singh (UID:105479019)

November 12, 2021

Problem 1. MAF estimation, $J \in \{\{0,1\},1,0,2,\{0,1,2\}\}$, $P_t = 0.43$, find P_{t+1}

Answer 1. $P_{t+1} = 0.44614$

Solution 1. Estimating expected number of 1's for each observation,
$J_1 = \{0,1\} \Rightarrow \dfrac{2p(1-p)}{(1-p)^2 + 2p(1-p)} = \dfrac{2*0.43*0.57}{0.57*0.57+2*0.43*0.57} = 0.6013986$
$J_2 = 1 \Rightarrow 1$
$J_3 = 0 \Rightarrow 0$
$J_4 = 2 \Rightarrow 2$
$J_5 = \{0,1,2\} \Rightarrow \dfrac{2p(1-p)+2p^2}{(1-p)^2+2p(1-p)+p^2} = \dfrac{2*0.43*0.57+2*0.43*0.43}{0.57*0.57+2*0.43*0.57+0.43*0.43} =$ 0.86

$$P_{t+1} = \frac{0.6013986 + 1 + 0 + 2 + 0.86}{2*5} = 0.44614$$

Problem 2. Q-function

Answer 2.

$$Q((\pi,\lambda)|(\pi^t,\lambda^t)) = c + \sum_{i=1}^{n}\sum_{j=0}^{2} a_{ij}[log(\pi_j) + x_i log(\lambda_j) - \lambda_j]$$

where c is a constant, $a_{ij} = \dfrac{\pi_j^t f_j^t(x_i)}{\sum_{m=0}^{2}\pi_m^t f_m^t(x_i)}$ and $f_j(x) = \dfrac{\lambda_j^x e^{-\lambda_j}}{x!}$

Solution 2. For Poisson distributed variable, the pdf of samples originated from the j^{th} clusters is,

$$f_j(x) = \frac{\lambda_j^x e^{-\lambda_j}}{x!}$$

Let $x_1, ..., x_n$ be the count of bacteria in n individuals. To use the EM algorithm, we will define the latent variables as the assignment of each sample to a cluster. Specifically, for

each sample i we have a variable $Z_i \in \{0,1,2\}$ which denotes the cluster of i.
In the t^{th} iteration of the EM, lets assume that the parameters are given by $\{\pi^t, \lambda^t\}$. Given that π_j is the cluster probability, let a_{ij} be the probability that i^{th} sample is assigned to j^{th} cluster, then

$$a_{ij} = P(Z_i = j|(\pi^t, \lambda^t)) = \frac{\pi_j^t f_j^t(x_i)}{\sum_{m=0}^{2} \pi_m^t f_m^t(x_i)}$$

Now, we can write the Q-function as below,

$$Q((\pi, \lambda)|(\pi^t, \lambda^t)) = E[logP(Z|(\pi, \lambda))]$$

$$Q((\pi, \lambda)|(\pi^t, \lambda^t)) = \sum_{i=1}^{n} E[logP(Z_i|(\pi, \lambda))]$$

$$Q((\pi, \lambda)|(\pi^t, \lambda^t)) = \sum_{i=1}^{n} \sum_{j=0}^{2} a_{ij}[log(\pi_j) + log(f_j(x_i))]$$

$$Q((\pi, \lambda)|(\pi^t, \lambda^t)) = \sum_{i=1}^{n} \sum_{j=0}^{2} a_{ij}[log(\pi_j) + x_i log(\lambda_j) - \lambda_j - log(x_i!)]$$

$log(x_i!)$ is a constant that we can remove from Q-function,

$$Q((\pi, \lambda)|(\pi^t, \lambda^t)) = c + \sum_{i=1}^{n} \sum_{j=0}^{2} a_{ij}[log(\pi_j) + x_i log(\lambda_j) - \lambda_j]$$

where c is a constant, $a_{ij} = \frac{\pi_j^t f_j^t(x_i)}{\sum_{m=0}^{2} \pi_m^t f_m^t(x_i)}$ and $f_j(x) = \frac{\lambda_j^x e^{-\lambda_j}}{x!}$
The Q-function can be differentiated to find update for π and λ and soft assignments can be done by updating a_{ij}.