CM224 HW 6 Solution

Ashish Kumar Singh (UID:105479019)

November 28, 2021

Problem 1. Forward Backward Algorithm

Sub-Problem 1a. What is $F_h(3,1)$

Answer 1a. $F_h(3,1) = 0.079996$

Solution 1a. Given, haplotype $h = 0110101$, we see that $h_2 = 1$, therefore, $\xi(h_2, e_{2i}) = h_2 * e_{2i} + (1 - h_2) * (1 - e_{2i}) = e_{2i}$

$$F_h(3,1) = \sum_{i=1}^{3} F_h(2,i)\xi(h_2, e_{2i})\delta_{i1}$$

$$F_h(3,1) = \sum_{i=1}^{3} F_h(2,i)e_{2i}\delta_{i1}$$

$$F_h(3,1) = F_h(2,1)e_{21}\delta_{11} + F_h(2,2)e_{22}\delta_{21} + F_h(2,3)e_{23}\delta_{31}$$

$$F_h(3,1) = 0.235 * (1/3) * 0.85 + 0.211 * (1/4) * 0.15 + 0.22 * (1/2) * 0.05$$

$$F_h(3,1) = 0.0665833 + 0.0079125 + 0.0055$$

$$F_h(3,1) = 0.079996$$

Sub-Problem 1b. What is $B_h(3,3)$

Answer 1b. $B_h(3,3) = 0.1280375$

Solution 1b. Given, haplotype $h = 0110101$, we see that $h_4 = 0$, therefore, $\xi(h_4, e_{4i}) = h_4 * e_{4i} + (1 - h_4) * (1 - e_{4i}) = (1 - e_{4i})$

$$B_h(3,3) = \sum_{i=1}^{3} B_h(4,i)\xi(h_4, e_{4i})\delta_{3i}$$

$$B_h(3,3) = \sum_{i=1}^{3} B_h(4,i)(1 - e_{4i})\delta_{3i}$$

$$B_h(3,3) = B_h(4,1)(1-e_{41})\delta_{31} + B_h(4,2)(1-e_{42})\delta_{32} + B_h(4,3)(1-e_{43})\delta_{33}$$
$$B_h(3,3) = 0.13 * (3/4) * 0.05 + 0.182 * (7/8) * 0.05 + 0.16 * (4/5) * 0.90$$
$$B_h(3,3) = 0.004875 + 0.0079625 + 0.1152$$
$$B_h(3,3) = 0.1280375$$

Sub-Problem 1c. Probability of each state at $s = 3$. What is the most likely state?

Answer 1c. iii) State 3
$P(z_{h_3} = 1) = 0.001403$, $P(z_{h_3} = 2) = 0.002589$ and $P(z_{h_3} = 3) = 0.006926$

Solution 1c. Similar to above calculations,

$$F_h(3,2) = F_h(2,1)e_{21}\delta_{12} + F_h(2,2)e_{22}\delta_{22} + F_h(2,3)e_{23}\delta_{32}$$
$$F_h(3,2) = 0.235 * (1/3) * 0.1 + 0.211 * (1/4) * 0.75 + 0.22 * (1/2) * 0.05$$
$$F_h(3,2) = 0.0078333 + 0.0395625 + 0.0055$$
$$F_h(3,2) = 0.052896$$

$$F_h(3,3) = F_h(2,1)e_{21}\delta_{13} + F_h(2,2)e_{22}\delta_{23} + F_h(2,3)e_{23}\delta_{33}$$
$$F_h(3,3) = 0.235 * (1/3) * 0.05 + 0.211 * (1/4) * 0.1 + 0.22 * (1/2) * 0.9$$
$$F_h(3,3) = 0.0039167 + 0.005275 + 0.099$$
$$F_h(3,3) = 0.108192$$

$$B_h(3,1) = B_h(4,1)(1-e_{41})\delta_{11} + B_h(4,2)(1-e_{42})\delta_{12} + B_h(4,3)(1-e_{43})\delta_{13}$$
$$B_h(3,1) = 0.13 * (3/4) * 0.85 + 0.182 * (7/8) * 0.1 + 0.16 * (4/5) * 0.05$$
$$B_h(3,1) = 0.082875 + 0.015925 + 0.0064$$
$$B_h(3,1) = 0.1052$$

$$B_h(3,2) = B_h(4,1)(1-e_{41})\delta_{21} + B_h(4,2)(1-e_{42})\delta_{22} + B_h(4,3)(1-e_{43})\delta_{23}$$
$$B_h(3,2) = 0.13 * (3/4) * 0.15 + 0.182 * (7/8) * 0.75 + 0.16 * (4/5) * 0.10$$
$$B_h(3,2) = 0.014625 + 0.1194375 + 0.0128$$
$$B_h(3,2) = 0.1468625$$

Given, haplotype $h = 0110101$, we see that $h_3 = 1$, therefore,
$\xi(h_3, e_{3i}) = h_3 * e_{3i} + (1-h_3) * (1-e_{3i}) = e_{3i}$
Probability for state 1 at $s = 3$;

$$P(z_{h_3} = 1) = F_h(3,1)B_h(3,1)\xi(h_3, e_{31})$$
$$P(z_{h_3} = 1) = F_h(3,1)B_h(3,1)e_{31}$$

$$P(z_{h_3} = 1) = 0.079996 * 0.1052 * (1/6)$$
$$P(z_{h_3} = 1) = 0.001403$$

Probability for state 2 at $s = 3$;

$$P(z_{h_3} = 2) = F_h(3,2)B_h(3,2)\xi(h_3, e_{32})$$
$$P(z_{h_3} = 2) = F_h(3,2)B_h(3,2)e_{32}$$
$$P(z_{h_3} = 2) = 0.052896 * 0.1468625 * (1/3)$$
$$P(z_{h_3} = 2) = 0.002589$$

Probability for state 3 at $s = 3$;

$$P(z_{h_3} = 3) = F_h(3,3)B_h(3,3)\xi(h_3, e_{33})$$
$$P(z_{h_3} = 3) = F_h(3,3)B_h(3,3)e_{33}$$
$$P(z_{h_3} = 3) = 0.108192 * 0.1280375 * (1/2)$$
$$P(z_{h_3} = 3) = 0.006926$$

We see that, $P(z_{h_3} = 1) = 0.001403$, $P(z_{h_3} = 2) = 0.002589$ and $P(z_{h_3} = 3) = 0.006926$. Thus the most likely state at $s = 3$ is state 3.

Problem 2. Gaussian Mixture Model, find π_1, π_2

Answer 2. $\pi_1 = 0.45956$ and $\pi_2 = 0.54044$, code is attached

Problem 3. Gaussian Mixture Model with admixed, find π_1^{t+1}

Answer 3. $\pi_1^{t+1} = 0.861955$, code is attached

Solution 3.

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

where, $p_1 = \pi_1^2$, $p_2 = \pi_2^2$ and $p_3 = 2\pi_1\pi_2$, Maximization step for probability update will be to maximize below using lagrange multiplier,

$$\sum_{i=1}^{n}(a_{i1}log(\pi_1^2) + a_{i2}log(\pi_2^2) + a_{i3}log(2\pi_1\pi_2)) - \lambda(\pi_1 + \pi_2 - 1)$$

Differentiating wrt to λ, π_1, π_2 and setting to zero we get following,

$$\lambda = 2\sum_{i=1}^{n}(a_{i1} + a_{i2} + a_{i3}) = 2n$$

$$\pi_1^{t+1} = \frac{\sum_{i=1}^{n}(2a_{i1} + a_{i3})}{2n}$$

$$\pi_2^{t+1} = \frac{\sum_{i=1}^{n}(2a_{i2} + a_{i3})}{2n}$$

Solving for the given data, we get $\pi_1^{t+1} = 0.861955$