

## Problem 1

(1)

For each datapoint  $y_i$ , we introduce a latent variable  $z_i$ , which indicates the normal distribution  $y_i$  is from, *i.e.*  $z_i = 0 \iff y_i$  is from  $N(\mu_1, \sigma_1^2)$  and vice versa. The EM algorithm is derived below.

In E-step, we first derive  $p(z_n = k|x_n, \theta)$  using Bayesian formula where  $\theta$  denotes  $\mu_i, \sigma_i^2, w$ .

$$\begin{aligned} p(z_n = k|y_n, \theta) &= \frac{p(z_n = k, y_i|\theta)}{\sum_k p(z_n = k, y_i|\theta)} \\ &= \frac{w_k N(y_i|\mu_k, \sigma_k^2)}{\sum_k w_k N(y_i|\mu_k, \sigma_k^2)} \\ &= \gamma_{n,k} \end{aligned} \quad (1)$$

where  $\gamma_{n,k}$  can be seen as a soft label of latent variable  $z_n$ . The log likelihood function is then

$$\begin{aligned} l(\theta) &= \sum_n \sum_k p(z_k|x_n, \theta) \log p(z_k, x_n|\theta) \\ &= \sum_n \sum_k \gamma_{n,k} (\log w_k - \frac{1}{2} 2\pi - \frac{1}{2} \log \sigma_k^2) - \frac{(y_n - \mu_k)^2}{2\sigma_k^2} \end{aligned} \quad (2)$$

In M-step, we take the derivative of  $l(\theta)$ , and set it to zero w.r.t.  $w_k, \mu_k$  and  $\sigma_k$ .

a) max  $l(\theta)$  w.r.t.  $w_k$ , using Lagrange multiplier since  $\sum w_k = 1$

$$\begin{aligned} \sum_n \sum_k \gamma_{n,k} \frac{1}{w_k} - \lambda &= 0, \quad \forall k \\ w_k &\propto \sum_n \gamma_{n,k} \\ w_k &= \frac{\sum_n \gamma_{n,k}}{\sum_k \sum_n \gamma_{n,k}} \end{aligned} \quad (3)$$

b) max  $l(\theta)$  w.r.t.  $\mu_k$

$$\begin{aligned} \sum_n \gamma_{n,k} \frac{(y_n - \mu_k)}{\sigma_k^2} &= 0 \\ \mu_k &= \frac{\sum_n \gamma_{n,k} y_n}{\sum_n \gamma_{n,k}} \end{aligned} \quad (4)$$

c) max  $l(\theta)$  w.r.t.  $\sigma_k^2$

$$\begin{aligned} \sum_n \gamma_{n,k} \left( -\frac{1}{2\sigma_k^2} + \frac{(y_n - \mu_k)^2}{2(\sigma_k^2)^2} \right) &= 0 \\ \sigma_k^2 &= \frac{\sum_n \gamma_{n,k} (y_n - \mu_k)^2}{\sum_n \gamma_{n,k}} \end{aligned} \quad (5)$$

The EM algorithm is shown below

---

**Algorithm 1:** EM for GMM

---

```

Result:  $\mu, \sigma^2, w$ 
begin
  initialize  $\mu, \sigma^2, w$ 
  while not converge do
    calculate  $\gamma_{n,k}$ 
    update  $\mu, \sigma^2, w$ 
  end
end

```

---

(2)

EM algorithm does not always converge on this dataset. In my implementation, four modes are found.

**Problem 2**

(1)

The likelihood function for  $\theta$  is

$$L(\theta) = \prod_n e^{-(x_n\theta + r_n)} \frac{(x_n\theta + r_n)^{y_n}}{y_n!} \quad (6)$$

(2)

Similar to the E-step in Problem1, we first calculate  $E(z_{j1}|x_j, \theta)$ . Notice that  $z_{j1}|y_j \sim \text{Binominal}(y_j, \frac{x_j\theta}{x_j\theta + r_j})$ . No need to calculate  $E(z_{j2}|x_j, \theta)$  since it's irrelevant to  $\theta$  in log likelihood function.

$$E(z_{j1}|x_j, \theta) = y_j \frac{x_j\theta}{x_j\theta + r_j} \quad (7)$$

In M-step, we derive its log likelihood function first

$$\begin{aligned}
 l(\theta) &= \sum_j -x_k + z_{j1} \log x_j\theta - \log z_{j1}! - r_j + z_{j2} \log r_j - \log z_{j2}! \\
 \frac{dl(\theta)}{d\theta} &= \sum_j -x_j + z_{j1} \frac{1}{\theta}
 \end{aligned} \quad (8)$$

Substitute  $z_{j1}$  by  $E(z_{j1}|x_j, \theta)$  and set the derivative to zero, we have

$$\theta' = \frac{\theta}{\sum_j x_j} \sum_j \frac{y_j x_j}{x_j\theta + r_j} \quad (9)$$

which can be seen as a fix point iteration of true likelihood function.

---

**Algorithm 2:** EM for sum of Poission

---

```

Result:  $\theta$ 
begin
  initialize  $\theta$ 
  while not converge do
    calculate  $E(z_{j1}|x_j, \theta)$ 
    update  $\theta$ 
  end
end

```

---

(3)

The MLE is 5.606063396561341, which yields 1.78e-15 in true likelihood function.

(4)

By taking second derivative of log likelihood function in (1), we have

$$-\frac{\partial^2 l(\theta)}{\partial \theta^2} = \sum_n y_n \frac{x_n^2}{(x_n \theta + r_n)^2} = 2.423 \quad (10)$$

, and the complete information being

$$-\frac{\partial^2 q(\theta)}{\partial \theta^2} = \sum_n y_n \frac{x_n}{(x_n \theta + r_n) \theta} = 2.588 \quad (11)$$

, so the fraction of missing information is 0.064

### Problem 3

(1)

$$l(\pi) = 100 \log \pi_{11} + 50 \log \pi_{12} + 75 \log \pi_{21} + 75 \log \pi_{22} \quad (12)$$

The assumption under such a log likelihood includes

- missing data do not count for likelihood, although they are partially observed

(2)

We derive the likelihood function for all data first

$$L(\pi) = \prod_n \pi_{11}^{y_{n1}=1, y_{n2}=1} \pi_{12}^{y_{n1}=1, y_{n2}=2} \pi_{21}^{y_{n1}=2, y_{n2}=1} \pi_{22}^{y_{n1}=2, y_{n2}=2} \quad (13)$$

E-step: For the observed 300 cases, all  $y_{n1}, y_{n2}$  are observed thus fixed, while for missing data, one of them need to be estimated, e.g. for the 28 datapoints with  $Y_1$  missing, the estimation is simply from Bayesian formula

$$p(y_{n1} = 1, y_{n2} = 1) = \frac{\pi_{11}}{\pi_{11} + \pi_{12}}, \quad p(y_{n1} = 1, y_{n2} = 2) = \frac{\pi_{12}}{\pi_{11} + \pi_{12}} \quad (14)$$

We do this for all missing data.

In M-step, we take the derivative of log likelihood function and set it to zero w.r.t.  $\pi_{ij}$

$$\begin{aligned} \sum_n p(y_{n1} = i, y_{n2} = j) / \pi_{ij} - \lambda &= 0 \\ \pi_{ij} &\propto \sum_n p(y_{n1} = i, y_{n2} = j) \\ \pi_{ij} &= \sum_n p(y_{n1} = i, y_{n2} = j) / \sum_{i,j} \sum_n p(y_{n1} = i, y_{n2} = j) \end{aligned} \quad (15)$$

(3)

The MLE of  $\pi$  from EM algorithm is [0.27912927 0.17190756 0.24112718 0.30783599]

(4)

The odds ratio from complete data is  $100 * 75 / (75 * 50) = 2.0$ , while that from MLE is 2.07, not equal.