4.5

(a) $$l(\theta) = \log p(x, z|\theta)$$
$$= \sum_i \log p(x_i|z_i,\theta)\cdot p(z_i|\theta)$$
$$= \sum_i \log \prod_j \left(\pi_j^{z_{ij}}\left(\frac{1}{x_i!} e^{-\lambda_j} \lambda_j^{x_i}\right)^{z_{ij}}\right)$$
$$= \sum_i \sum_j z_{ij}\left(\log \pi_j - \log x_i! - \lambda_j + x_i \log \lambda_j\right)$$

E-step: $$\hat{z}_{ij} = p(z_i=j \mid x, \hat{\theta}^{old})$$
$$= \frac{p(x_i \mid z_i=j, \hat{\theta}^{old})\cdot p(z_i=j \mid \hat{\theta}^{old})}{\sum_j p(x_i \mid z_i=j, \hat{\theta}^{old})\cdot p(z_i=j \mid \hat{\theta}^{old})}$$
$$= \frac{\frac{1}{x_i!}\cdot e^{-\hat{\lambda}_j}\cdot \hat{\lambda}_j^{x_i}\cdot \pi_j}{\sum_k \pi_k \cdot \frac{1}{x_i!}\cdot e^{-\hat{\lambda}_k}\cdot \hat{\lambda}_k^{x_i}}$$

M-step $$\frac{\partial}{\partial \lambda_j} l(\theta) = \sum_i \hat{z}_{ij}\left(-1 + \frac{x_i}{\lambda_j}\right) = 0$$
$$\Rightarrow \quad \hat{\lambda}_j = \frac{\sum_i \hat{z}_{ij}\cdot x_i}{\sum_i \hat{z}_{ij}}$$

PS2.1 ML for a Poisson $\hat{\lambda} = \frac{\sum_i x_i}{n}$

We denote $\hat{n}_j = \sum_i \hat{z}_{ij}$ is the number of points assigned to class j

then $\sum_i \hat{z}_{ij}\cdot x_i$ is the sum of x_i value for those points

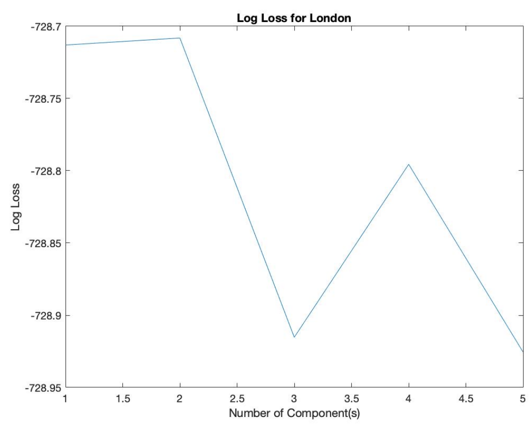
$\hat{\lambda}_j = \frac{\sum_i \hat{z}_{ij}\cdot x_i}{\hat{n}_j}$ is similar to an MLE for a Poisson with class j

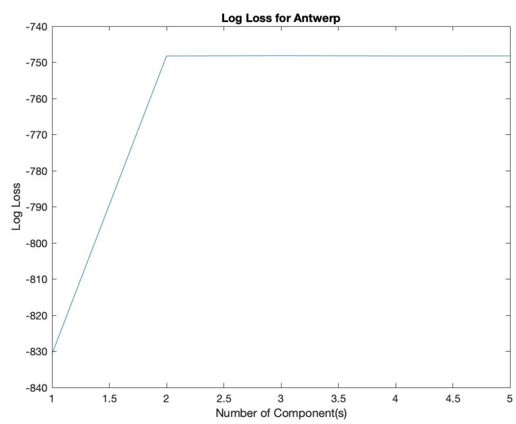
$$f = l(\theta) + \lambda\left(\sum_j \pi_j - 1\right)$$
$$\frac{\partial}{\partial \pi_j} f = 0 \Rightarrow \sum_i \hat{z}_{ij}\cdot \frac{1}{\pi_j} + \lambda = 0$$
$$\frac{\partial}{\partial \lambda} f = 0 \Rightarrow \sum_j \pi_j = 1 \qquad \sum_i \hat{z}_{ij} + \lambda \pi_j = 0$$
$$-\lambda \cdot \sum_j \pi_j = \sum_i \sum_j \hat{z}_{ij} = n$$
$$\lambda = -\frac{n}{\sum_j \pi_j} = -n$$

$$\Rightarrow \quad \hat{\pi}_j = \frac{\sum_i \hat{z}_{ij}}{n}$$

$\hat{\pi}_j$ denotes how many weights we assign to class j according to the number of samples belonging to it

(b) $$\text{log-likelihood} = \log p(x|\hat{\theta})$$
$$= \sum_i \log p(x_i|\hat{\theta})$$
$$= \sum_i \log \sum_j \hat{\pi}_j \cdot \text{Poisson}(x_i|\hat{\lambda}_j)$$





From these two figures plotted in MATLAB, we see that log likelihood of London stays around -728 for k=1 to k=5, whereas Antwerp increases significantly from k=1 to k=2, and then stay around -747.

lambda_Antwerp

5x5 double

	1	2	3	4	5
1	0.8958	0	0	0	0
2	2.3059	0.2623	0	0	0
3	1.0881	0.1392	2.4459	0	0
4	0.1202	0.5533	1.5377	2.6478	0
5	2.5882	1.6660	0.1522	0.1722	0.4533

lambda_london

5x5 double

	1	2	3	4	5
1	0.9288	0	0	0	0
2	0.9953	0.8651	0	0	0
3	1.1139	1.1220	0.6995	0	0
4	0.8226	0.7502	1.0724	1.1687	0
5	1.1456	0.8189	1.1185	1.1084	0.6273

From the lambda, we see that London tends to have a random attack, whereas in Antwerp, there are some squares with higher probability of being attack.