<u>HW3</u>
Ilana Pervoi, Pan Eyal

Problem 1:

Show that $\underset{(\theta_1,\dots,\theta_K,\pi_1,\dots,\pi_K)}{\operatorname{argmax}} Q(\theta,\theta^t)$ satisfied when $\pi_k = \frac{1}{N}\sum_i r_{i,k} \ \ \forall k \in \{1,\dots,K\}$,
subject to the constraint: $\sum_{k=1}^{K}\pi_k = 1$.

<u>Proof:</u> (Using Lagrange multiplier)
The equality constraint function can be expressed as:

$$g(\pi_1,\dots,\pi_K) = \left(\sum_{k=1}^{K}\pi_k\right) - 1 = 0$$

Let us define a new function $L(\theta,\theta^t,\lambda)$:

$$L(\theta,\theta^t,\lambda) = Q(\theta,\theta^t) + \lambda * g(\pi_1,\dots,\pi_K) = \left(\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}\log\pi_k\right) + \left(\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}\log p(\boldsymbol{x}_i;\theta_k)\right) + \lambda\left(\sum_{k=1}^{K}\pi_k\right) - \lambda$$

We will find the partial derivative of L.
To show that $\underset{(\theta_1,\dots,\theta_K,\pi_1,\dots,\pi_K)}{\operatorname{argmax}} Q(\theta,\theta^t)$ is given by Equation 1, all the θ_k params are irrelevant.

For any k:

$$\frac{\partial}{\partial\pi_k}L = \frac{\partial}{\partial\pi_k}\left(\sum_{i=1}^{N}\sum_{l=1}^{K} r_{i,l}\log\pi_l\right) + \frac{\partial}{\partial\pi_k}\left(\sum_{i=1}^{N}\sum_{l=1}^{K} r_{i,l}\log p(\boldsymbol{x}_i;\theta_l)\right) + \frac{\partial}{\partial\pi_k}\lambda\left(\sum_{l=1}^{K}\pi_l\right) - \frac{\partial}{\partial\pi_k}\lambda$$

$$= \frac{\partial}{\partial\pi_k}\left(\sum_{i=1}^{N}\sum_{l=1}^{K} r_{i,l}\log\pi_l\right) + \frac{\partial}{\partial\pi_k}\left(\sum_{i=1}^{N}\sum_{l=1}^{K} r_{i,l}\log\sum_{m=1}^{K}\pi_m\mathcal{N}(x_i;\mu_m,\Sigma_m)\right) + \frac{\partial}{\partial\pi_k}\lambda\left(\sum_{l=1}^{K}\pi_l\right) - \frac{\partial}{\partial\pi_k}\lambda$$

$$\underbrace{=}_{\forall l\neq k:\frac{\partial}{\partial\pi_k}\{non-\pi_k\ function\}=0} \frac{\partial}{\partial\pi_k}\left(\sum_{i=1}^{N} r_{i,k}\log\pi_k\right) + \frac{\partial}{\partial\pi_k}\lambda\pi_k = \frac{1}{\pi_k}\sum_{i=1}^{N} r_{i,k} + \lambda$$

And,

$$\frac{\partial}{\partial\lambda}L = \sum_{l=1}^{K}\pi_l - 1$$

Now we will solve the equality system:

$$\begin{cases} \dfrac{1}{\pi_k}\displaystyle\sum_{i=1}^{N} r_{i,k} + \lambda = 0, \quad \forall k\in\{1,\dots,K\} \\ \displaystyle\sum_{l=1}^{K}\pi_l - 1 = 0 \end{cases} \rightarrow$$

$$\begin{cases} \pi_k = -\dfrac{\sum_{i=1}^{N} r_{i,k}}{\lambda}, \quad \forall k\in\{1,\dots,K\} \\ \displaystyle\sum_{l=1}^{K}\pi_l = 1 \end{cases} \rightarrow$$

$$\sum_{l=1}^{K}\pi_l = \sum_{l=1}^{K} -\frac{\sum_{i=1}^{N} r_{i,k}}{\lambda} = \frac{1}{-\lambda}\sum_{l=1}^{K}\sum_{i=1}^{N} r_{i,k} = 1$$

$$\rightarrow \lambda = -\sum_{l=1}^{K}\sum_{i=1}^{N} r_{i,k}$$

$$\pi_k = \frac{-\sum_{i=1}^{N} r_{i,k}}{-\sum_{l=1}^{K}\sum_{i=1}^{N} r_{i,k}} = \frac{\sum_{i=1}^{N} r_{i,k}}{\sum_{l=1}^{K}\sum_{i=1}^{N} r_{i,k}} = \frac{N_k}{N}$$

As requested.

Problem 2:

<u>Solution:</u>

I. To make a uniform distribution over the space of all K-dimensional categorical distributions, we will need that $\mathrm{Dir}(\pi;\ \alpha_1, \ldots, \alpha_K) \propto \prod_{k=1}^{K} \pi_k^{\alpha_k - 1} = c$, where c is a constant.
Hence, for $\alpha_k = 1\ \forall k$ we will obtain that: $\prod_{k=1}^{K} \pi_k^{\alpha_k - 1} = \prod_{k=1}^{K} \pi_k^{0} = \prod_{k=1}^{K} 1 = 1$.

II. A uniform $\boldsymbol{\pi}$ has the same value for each of its entries, $\pi_k = \frac{1}{K}$.
A sampled $\boldsymbol{\pi}$ from a uniform distribution is not necessarily distributed uniformly itself, it means the probability of obtaining any $\boldsymbol{\pi}$ from the distribution is the same.

III. A prior usage might result in a noise reduction of the data from a prior knowledge information we already have on the problem. When choosing the $\alpha' s$ parameters, in a sense, we are defining how we expect the samples to be sampled.

Computer Problem 1:

We will experiment on $K = \{2,3,4\}$:

<u>For $K = 2$:</u>

$$\Psi = \begin{bmatrix} 0.1^5 & 0 & 0 \\ 0 & 0.1^5 & 0 \\ 0 & 0 & 0.1^5 \end{bmatrix}, m = [0.1, 0.1, 0.1], \kappa = 1, \nu = 10, \alpha' s = [1, 1]$$

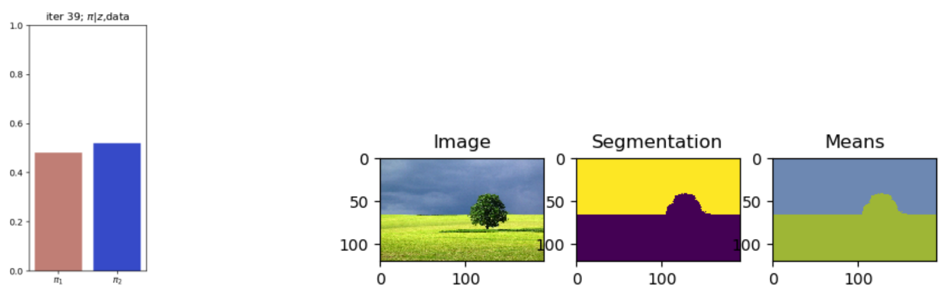


Figure 1

$$\Psi = \begin{bmatrix} 0.1^2 & 0 & 0 \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.1^2 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 1, \nu = 1000, \alpha' s = [100, 100]$$

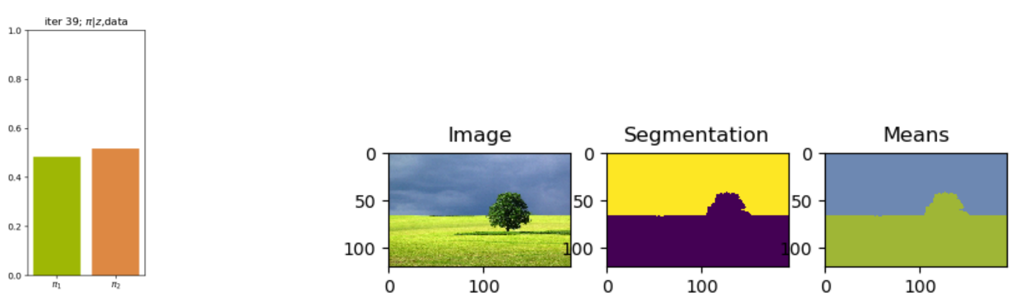


Figure 2

$$\Psi = \begin{bmatrix} 0.1^2 & 0 & 0 \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.1^2 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 1000, \nu = 1000, \alpha's = [100, 100]$$

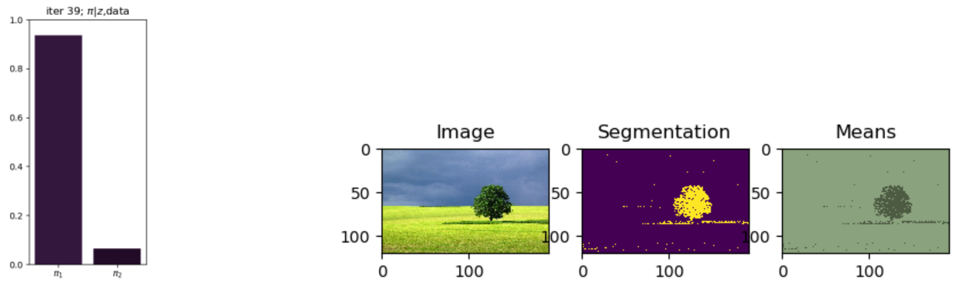


Figure 3

$$\Psi = \begin{bmatrix} 0.1^2 & 0 & 0 \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.1^2 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 1, \nu = 3, \alpha's = [100000, 1]$$

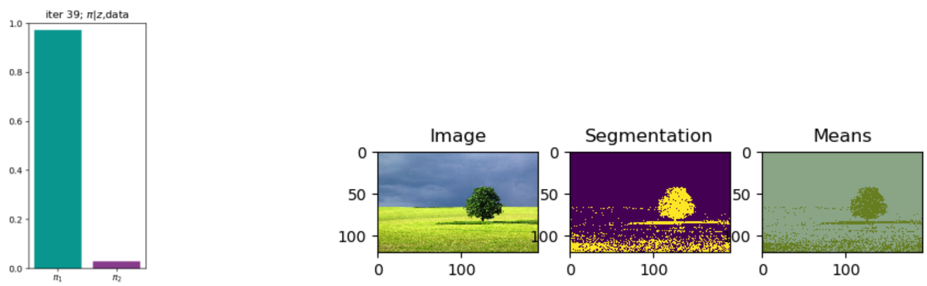


Figure 4

<u>For $K = 3$:</u>

$$\Psi = \begin{bmatrix} 0.1^5 & 0 & 0 \\ 0 & 0.1^5 & 0 \\ 0 & 0 & 0.1^5 \end{bmatrix}, m = [0.1, 0.1, 0.1], \kappa = 1, \nu = 10, \alpha's = [1, 1, 1]$$

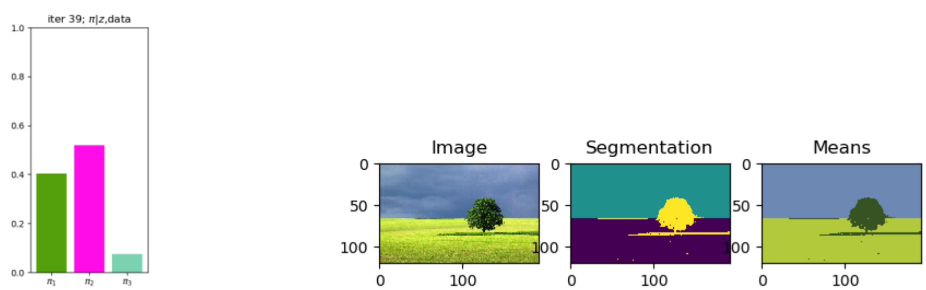


Figure 5

$$\Psi = \begin{bmatrix} 0.1^1 & 0 & 0 \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.1^5 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 1, \nu = 1000, \alpha's = [100, 100, 100]$$

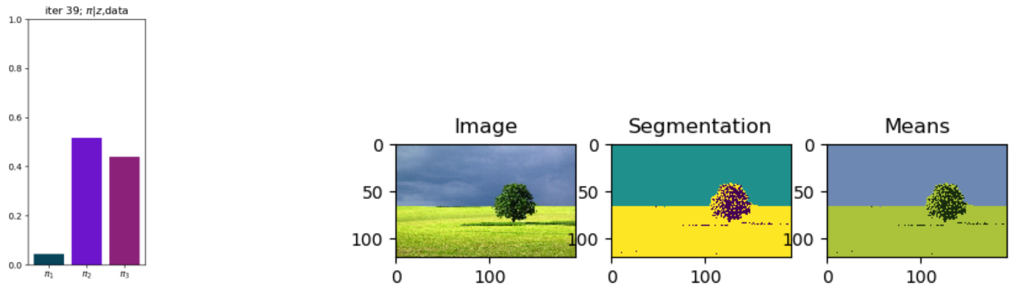


Figure 6

$$\Psi = \begin{bmatrix} 0.1^2 & 0 & 0 \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.1^2 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 100000, \nu = 1000, \alpha's = [100, 100, 100]$$

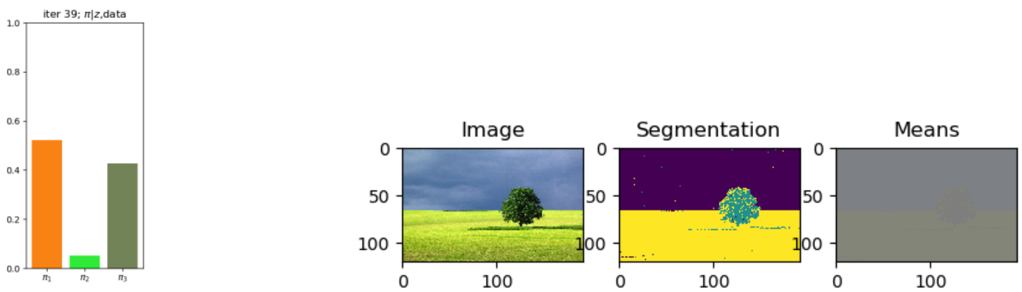


Figure 7

$$\Psi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 1, \nu = 1000, \alpha's = [100, 100, 100]$$

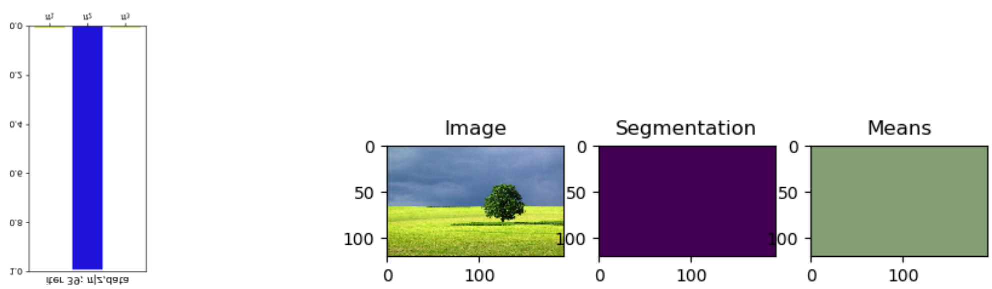


Figure 8

For $K = 4$:

$$\Psi = \begin{bmatrix} 0.1^5 & 0 & 0 \\ 0 & 0.1^5 & 0 \\ 0 & 0 & 0.1^5 \end{bmatrix}, m = [0.1, 0.1, 0.1], \kappa = 1, \nu = 10, \alpha's = [1, 1, 1]$$

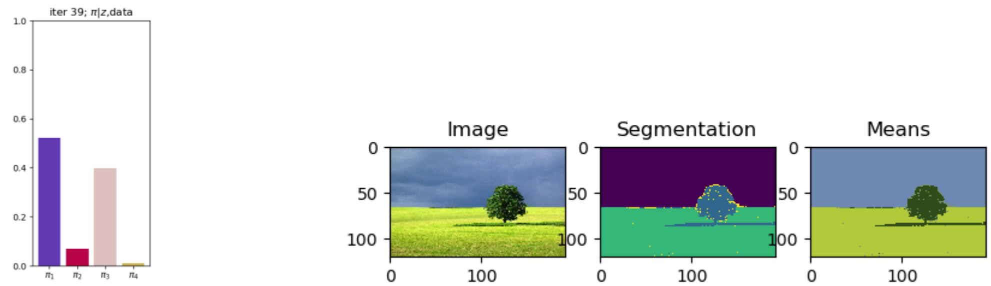


Figure 9

$$\Psi = \begin{bmatrix} 0.1^1 & 0 & 0 \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.1^5 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 1, \nu = 1000, \alpha's = [100, 100, 100]$$

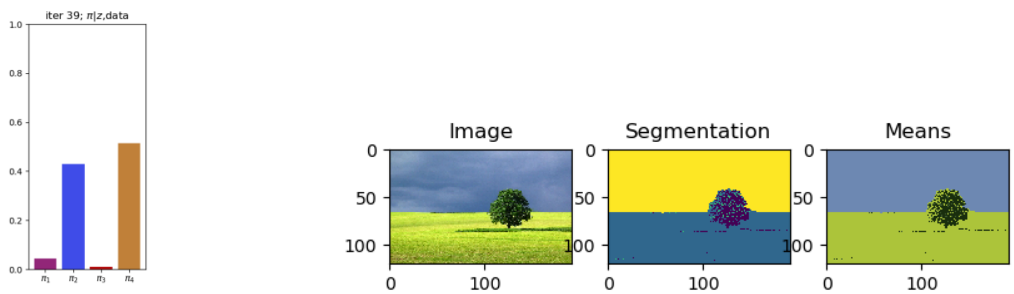


Figure 10

$$\Psi = \begin{bmatrix} 0.1^2 & 0 & 0 \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.1^2 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 100000, \nu = 1000, \alpha's = [100, 100, 100]$$

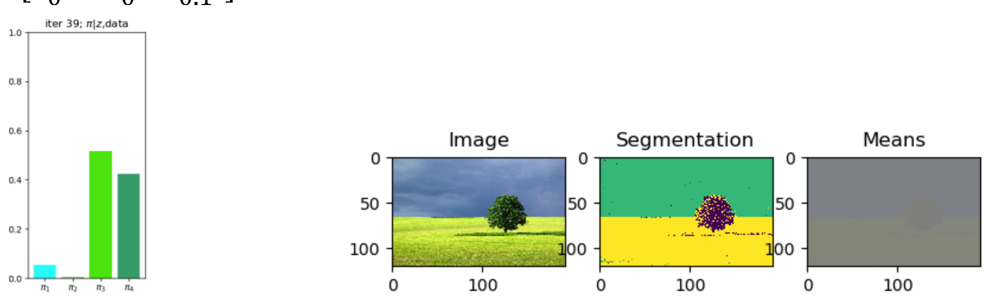
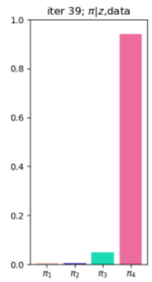


Figure 11

$$\Psi = \begin{bmatrix} 0.1^1 & 0 & 0 \\ 0 & 0.1^1 & 0 \\ 0 & 0 & 0.1^1 \end{bmatrix}, m = [0.5, 0.5, 0.5], \kappa = 1, \nu = 1000, \alpha's = [100, 100, 100]$$



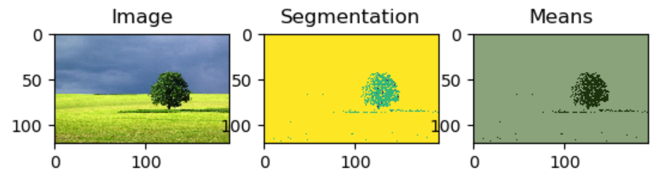


Figure 12

Speculations:

- For Ψ, we have seen that large Ψ entries will cause losing range of colors, as just one gaussian will be dominant that will be picked most of the time.
(Can be seen in figures 8 and 12)
- For κ, we know that the covariance matrix equals to $\frac{1}{\kappa}\Sigma$, therefore, for large κ the covariance will be very small and thus, the image will be noisier as the pixels of the image are not 'effects' each other. (Can be seen in figures 7 and 11 where the means picture gets very grey)
- For $\alpha's$, (as the prior variable) we have seen that unbalancing it will cause a tendency towards one of the gaussians over the other. (Can be seen in figure 4 where the bright grass merged with the skies and only the dark green in the image clustered together)