GMDL HW2

Yair Gross, Neta Elmaliach, Sharon Hendy

25/05/2023

Mixture Models

Problem 1

$$\underset{\theta_1,\ldots,\theta_k,\pi_1,\ldots,\pi_k}{\text{argmax}} Q(\theta,\theta^t) = \underset{\theta_1,\ldots,\theta_k,\pi_1,\ldots,\pi_k}{\text{argmax}} (\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}^t \log(\pi_k) + (\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}^t \log p(x_i;\theta_k^t))$$

According to Lagrange multipliers we would like to maximize a new function:

$$\underset{\theta_1,\ldots,\theta_k,\pi_1,\ldots,\pi_k}{\text{argmax}} Q(\theta,\theta^t) = \underset{\theta_1,\ldots,\theta_k,\pi_1,\ldots,\pi_k}{\text{argmax}} (\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}^t \log(\pi_k) + (\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}^t \log p(x_i;\theta^t)) + \lambda[\sum_{k=1}^{K} \pi_k - 1]$$

Derivative according to $\pi_{\mathbf{k}}$:

$$\nabla_{\pi_{\mathbf{k}}} = \sum_{i=1}^{N} r_{i,k}^t \frac{N}{\pi_k} + \lambda$$

$$\sum_{i=1}^{N} r_{i,k}^t \frac{N}{\pi_k} + \lambda = 0$$

$$\frac{N}{\pi_k}\sum_{i=1}^{N} r_{i,k}^t + \lambda = 0$$

$$\pi_k = \frac{-1}{\lambda}(\sum_{i=1}^{N} r_{i,k}^t * N)$$

Using the Lagrange multipliers constraint:

$$\sum_{k=1}^{K}(\frac{-1}{\lambda}(\sum_{i=1}^{N} r_{i,k}^t * N)) = 1$$

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

$$\frac{-1}{\lambda} * N^2 = 1$$

$$\lambda = -N^2$$

subtitute λ in $\pi_{\mathbf{k}}$:

$$\pi_k = \frac{-1}{\lambda}(\sum_{i=1}^{N} r_{i,k}^t * N) = \frac{-1}{-N^2}(\sum_{i=1}^{N} r_{i,k}^t * N) = \frac{1}{N}(\sum_{i=1}^{N} r_{i,k}^t)$$

Problem 2

Part 1

For $1 \leq i \leq K$, take $a_i = 1$.

$$\pi \sim Dir(\pi; a_1, ..., a_K) \propto \prod_{k=1}^{K} {\pi_k}^{a_k - 1} = \prod_{k=1}^{K} {\pi_k}^{0} = 1 \propto constant$$

When the distribution is propotional to constant it means that it is uniform distribution.

Part 2

In a uniform π, there is an equal probability for every option. On the other hand, if π is sampled from a uniform distribution, it is not necessary that the probability for every option is equal. Instead, there is an equal probability for each distribution from the distribution space from which we sampled.

Part 3

The motivation to use the prior is that if we had early knowledge on the distribution in our problem, we would like to give this knowledge weight when choosing the distribution.

Computer Exercise 1

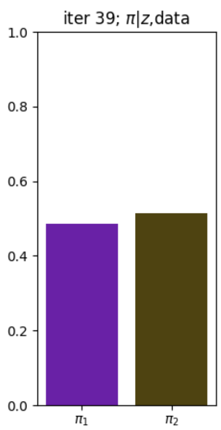
After running the image segmenting code using different configurations, we came to the following conclusions:

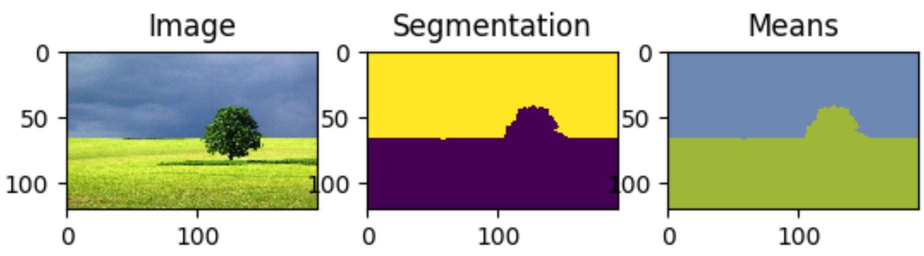
- ψ and ν are the parameters which have the most influence on distribution, and therefore the results.
- ψ, the scale matrix, affects the shape and prior knowledge of the covariance matrix. A larger value of ψ represents "higher certainty", meaning that we will see a specific structure – for example, more concentration around the mean or assigning adjacent pixels to the same cluster.
  On the other hand, as the value of ψ decreases, we expect to see higher variability and a less structured result.
- ν represents the degrees of freedom of the NIW distribution.
  Larger values of ν symbol a higher probability to receive covariance matrices which are closer to the mean covariance matrix. As a result, we expect to see more smoother and structured segmentation. However, values of ν can lead to overfitting and a less accurate ability to capture details.
  Smaller values of ν lead to a higher probability to receive larger covariance matrices, which can lead to more diverse and less structured segmentations.

We can see the effect of the different values of the hyperparameters in the NIW distribution, in the following results for k = 2,3,4 :

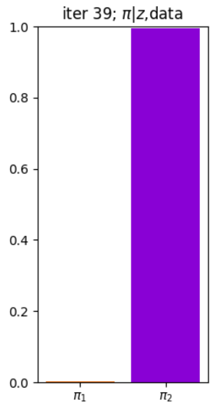
K = 2:

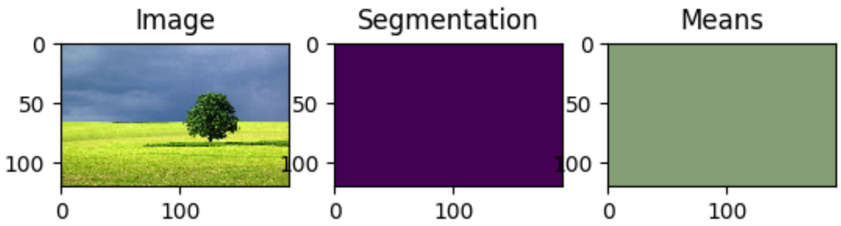
Initial values:



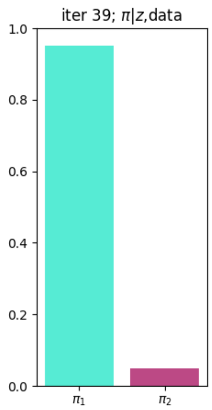


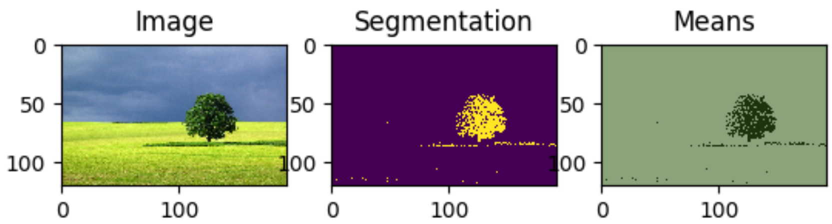
$\boldsymbol{\psi = (0.1)^{0.1}, \nu = 1000}$:



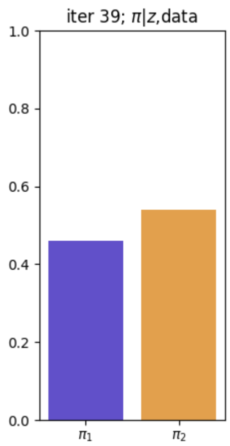


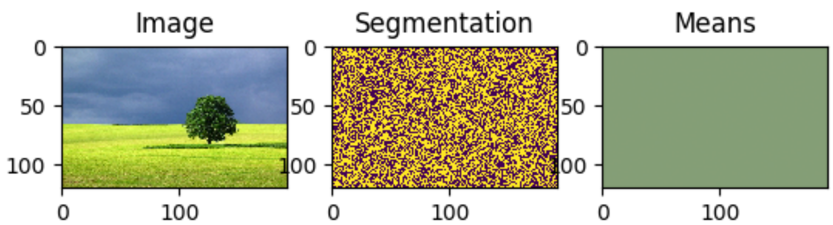
$\boldsymbol{\psi = 0.1, \nu = 1000}$:





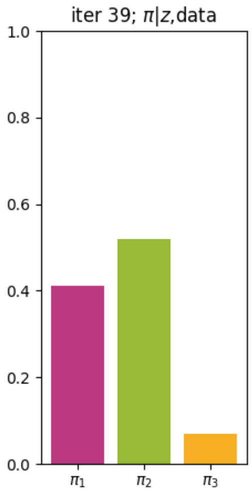
$\boldsymbol{\psi = (0.1)^{0.1}, \nu = 1000000}$:

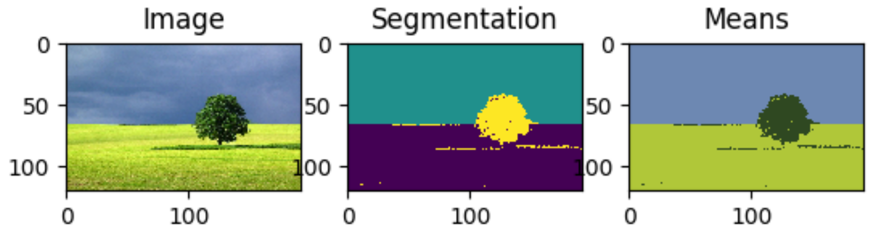




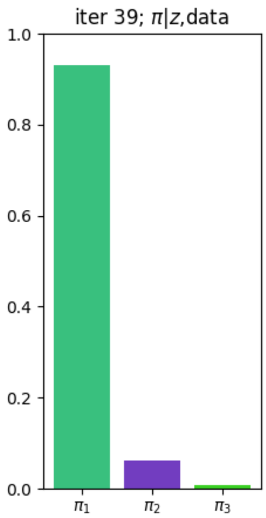
K = 3:

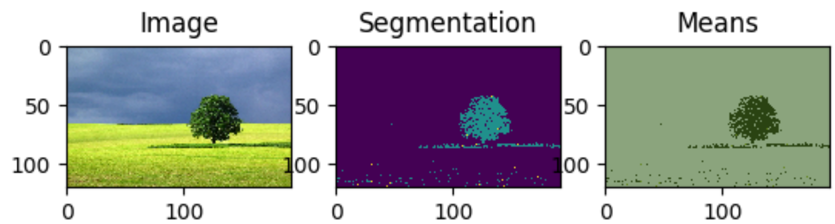
Initial values:



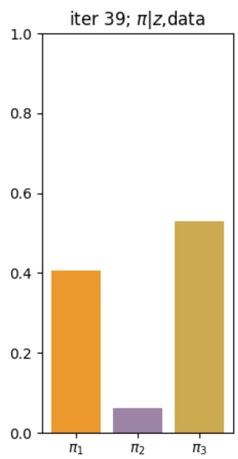


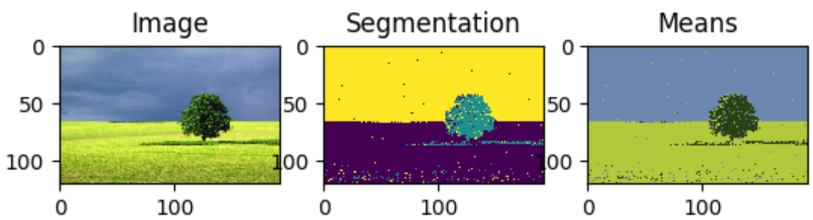
$\boldsymbol{\psi = (0.1)^{1.4}, \nu = 1000}$:



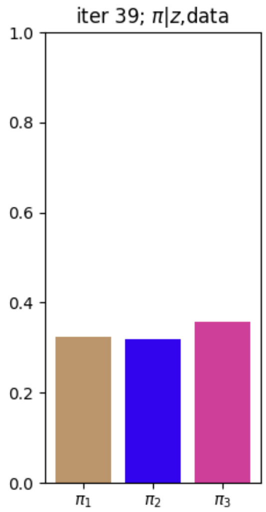


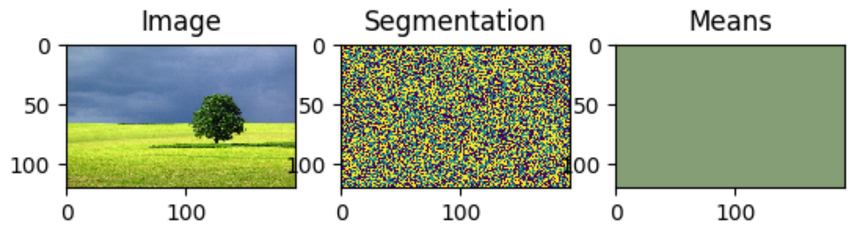
$\boldsymbol{\psi = (0.1)^{1.5}, \nu = 1000000}$:





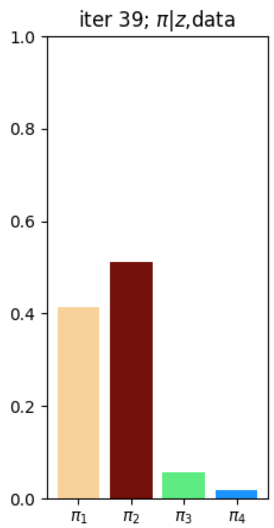
$\boldsymbol{\psi = (0.1)^{0.8}, \nu = 1000000}$**:**

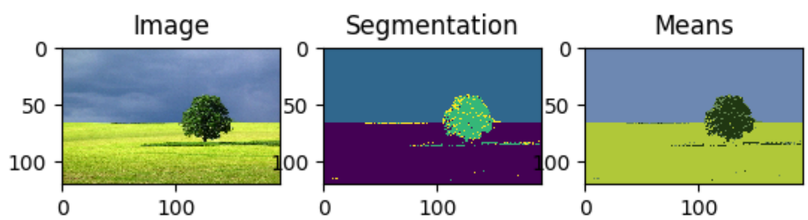




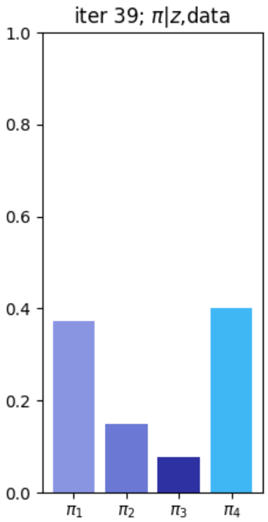
K = 4:

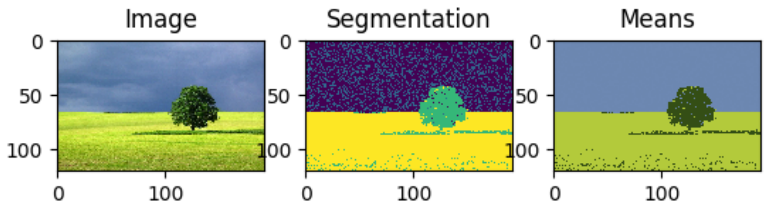
***Initial values*:**



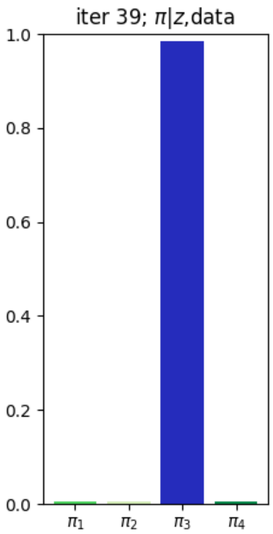


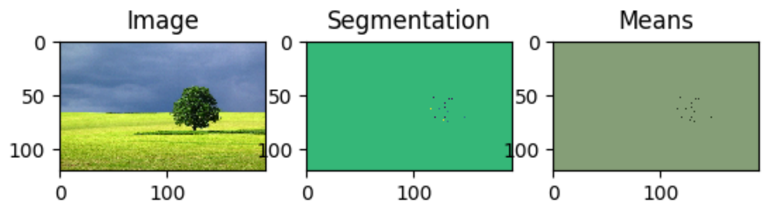
$\psi = (0.1)^2, \nu = 1000000$:





$\psi = (0.1)^{0.7}, \nu = 1000$:





$\psi = (0.1)^{0.7}, \nu = 1000000$: