GMDL, HW#2

208727164 – אורן יקואל , 336132014 - יונתן לויטס, 318512910 – גל אלון

Problem 1:

We'll show that $argmax_{(\theta_1,\dots,\theta_k,\pi_1,\dots,\pi_k)} Q(\theta,\theta^t)$ s.t $\sum_{k=1}^{K}\pi_k = 1$.

As we saw in class we know that:

$Q(\theta,\theta^t) = \left(\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}^t \log \pi_k\right) + \left(\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}^t \log N(x_i;\mu_k,\Sigma_k)\right)$

Let's denote a LaGrange multiplier function, with the constraints:

$$L(\pi_k,\lambda) = \left(\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}^t \log \pi_k\right) + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)$$

So we'll partially derive:

$\frac{\partial L(\pi_k,\lambda)}{\partial \pi_k} = \left(\sum_{i=1}^{N}\frac{r_{i,k}^t}{\pi_k}\right) + \lambda$ and $\frac{\partial L(\pi_k,\lambda)}{\partial \lambda} = \sum_{k=1}^{K}\pi_k - 1$

So: $\left(\sum_{i=1}^{N}\frac{r_{i,k}^t}{\pi_k}\right) + \lambda = 0$

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

And: $\sum_{k=1}^{K}\pi_k - 1 = 0 \rightarrow \sum_{k=1}^{K}\pi_k = 1$

together:

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

We'll recall,

$$\boldsymbol{\pi_k} = \frac{-\sum_{i=1}^{N} r_{i,k}^t}{\lambda} = \frac{\sum_{i=1}^{N} r_{i,k}^t}{\sum_{k=1}^{K}\sum_{i=1}^{N} r_{i,k}^t} = \frac{\sum_{i=1}^{N} r_{i,k}^t}{\sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k}^t} = \frac{\sum_{i=1}^{N} r_{i,k}^t}{\sum_{i=1}^{N} 1} = \frac{\boldsymbol{\sum_{i=1}^{N} r_{i,k}^t}}{\boldsymbol{N}} \quad \blacksquare$$

Problem 2:

(i) We'll notice that if we choose $\alpha_k = 1\ \forall k$ we'll get:

$$Dir(\boldsymbol{\pi}|\alpha = 1) = \frac{\Gamma(\Sigma_{k=1}^{K} 1)}{\Gamma(1)^K} * \Pi_{k=1}^{K} 1 = \frac{\Gamma(K)}{\Gamma(1)^K}$$

Because we get an expression which doesn't involve π, and is a close expression, we can say Π is sampled from a uniform distribution.

(ii) Uniform $\boldsymbol{\pi}$: $\boldsymbol{\pi} = \left[\frac{1}{\boldsymbol{k}},\dots,\frac{1}{\boldsymbol{k}}\right] \in \mathbb{R}^k$

On the other hand , $\boldsymbol{\pi}$ that is sampled from uniform distribution is (as proven above):

$$Dir(\boldsymbol{\pi}|\alpha = 1) = \frac{\Gamma(\Sigma_{k=1}^{K} 1)}{\Gamma(1)^K} * \Pi_{k=1}^{K} 1 = \frac{\Gamma(K)}{\Gamma(1)^K}$$

Meaning, a uniform $\boldsymbol{\pi}$ is a vector where entry is $\frac{1}{\boldsymbol{k}}$ and $|\boldsymbol{\pi}| = k$, while a $\boldsymbol{\pi}$ that is sampled from uniform distribution is a vector that can receive any value with equal probability.

(iii) In the context of mixture models and estimating π , using the prior instead of a pure likelihood-based approach can prevent overfitting or misleading results. In particular, if the data set is too small or noisy, the prior can provide a useful source of information about the distribution, which can help better generalize new data.

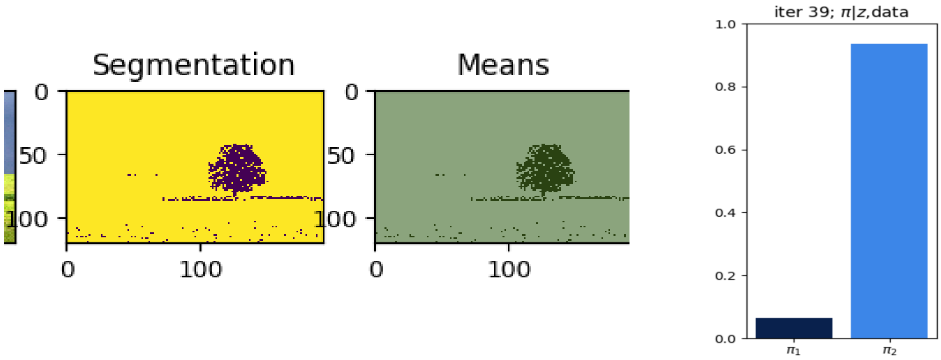
For example, if we sampled N (a relatively small number) samples from a mixture model and received data in which no sample comes from the entry with the largest weight – we'll denote as c, then a likelihood-based calculation will give $\pi_c = 0$. On the other hand, if we can assume in advance that the data is sampled from some distribution where π_c gets a relatively high value with a high probability, then the model will give a more accurate value for π_c.

Part 2:

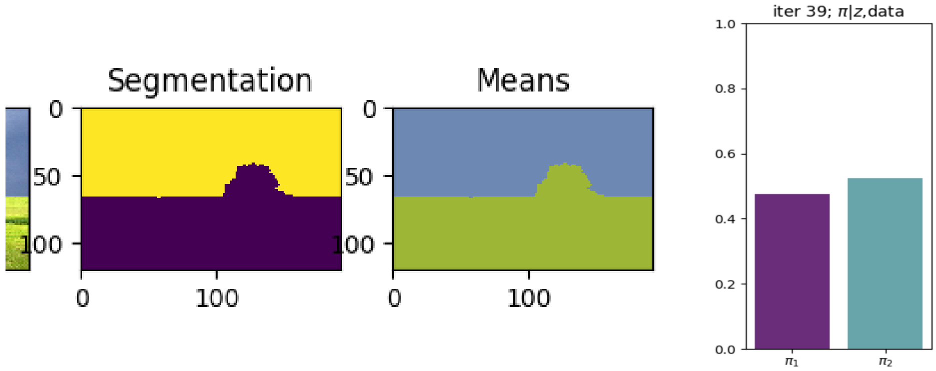
Obviously, and as we saw the K parameter affects on how many different objects the model is trying to segment.

K = 2:

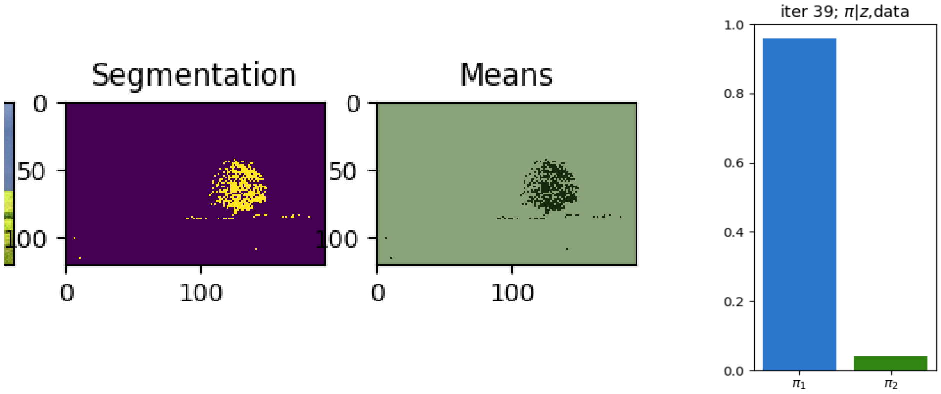
kappa = 1
nu = 1000
alphas = 100
psi = 0.2
m = 0.5



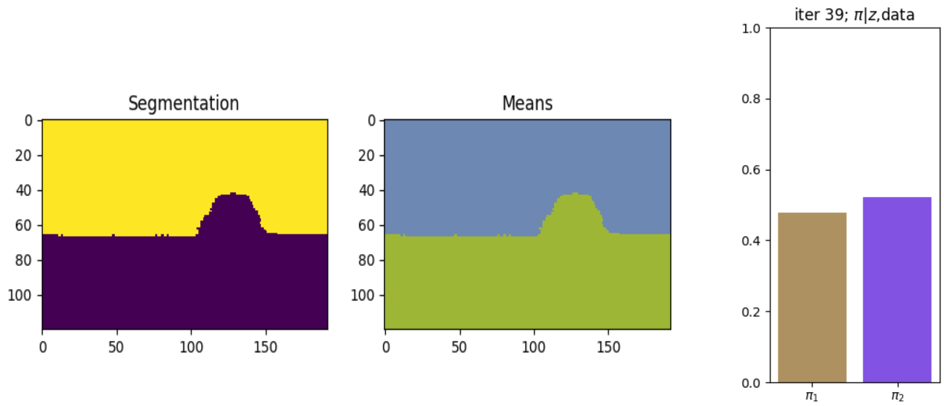
kappa = 1
nu = 100
alphas = 100
psi = 0.2
m = 0.5



kappa = 1
nu = 100
alphas = 1000
psi = 0.38
m = 0.5

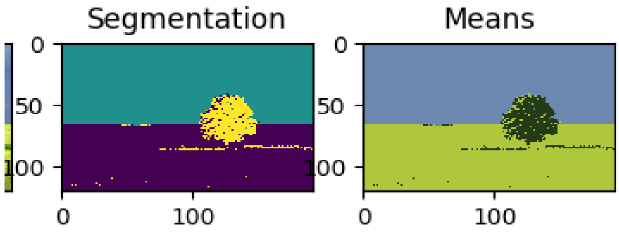


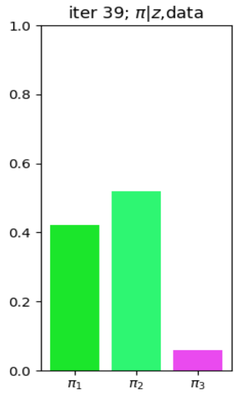
kappa = 1
nu = 1000
alphas = 100
psi = 0.2
m = 0.5



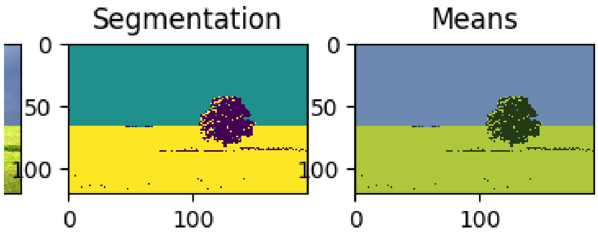
K = 3:

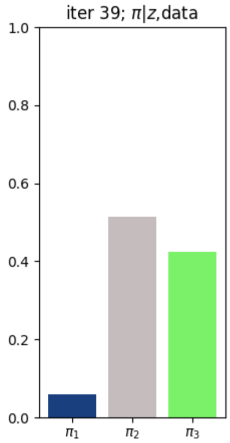
kappa = 10
nu = 10000
alphas = 100
psi = 0.1
m = 0.5



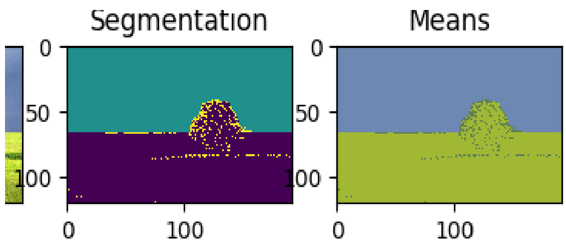


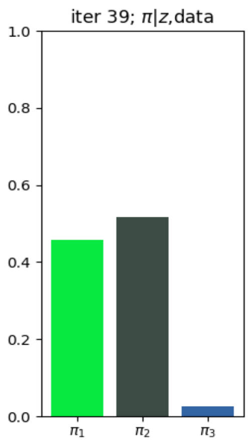
kappa = 1
nu = 10000
alphas = 100
psi = 0.1
m = 0.5



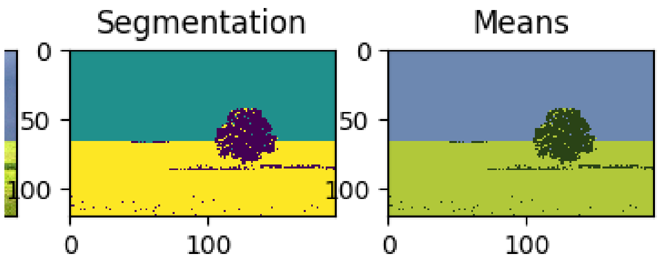


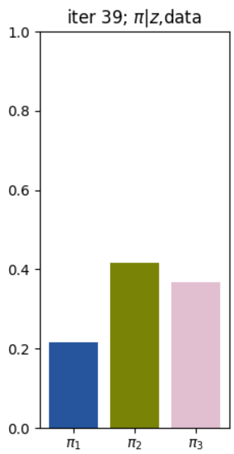
kappa = 1
nu = 10000
alphas = 100
psi = 0.1
m = 0.5





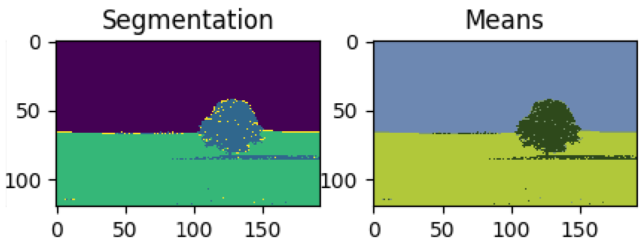
kappa = 1
nu = 10000
alphas = 10000
psi = 0.1
m = 0.5

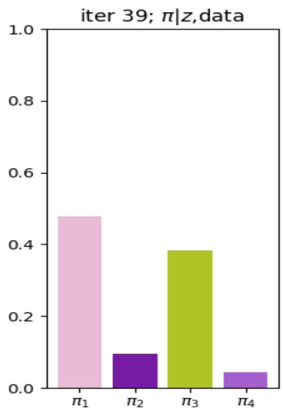




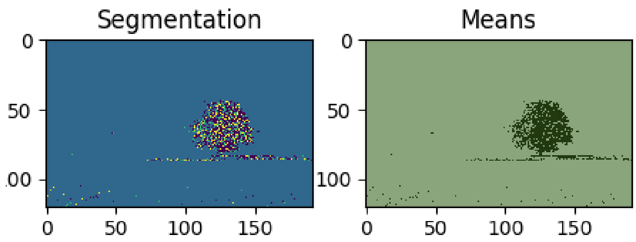
K = 4:

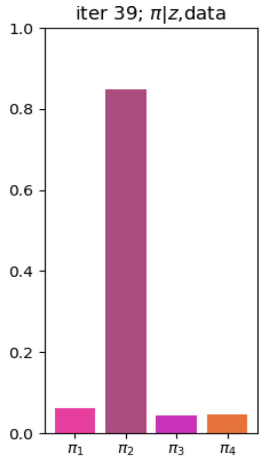
kappa = 1
nu = 100
alphas = 1000
psi = 0.2
m = 0.5



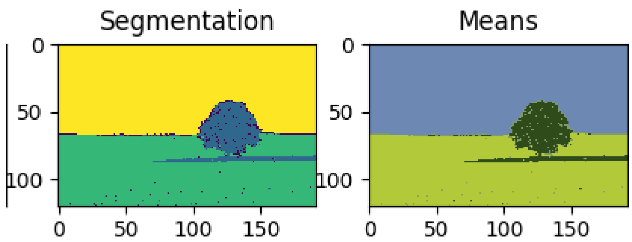


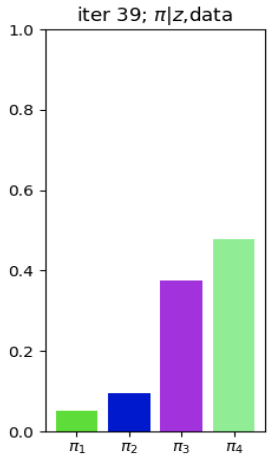
kappa = 1
nu = 1000
alphas = 1000
psi = 0.3
m = 0.5



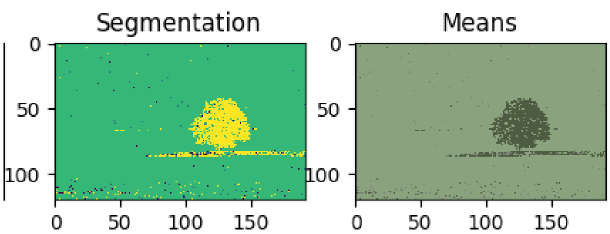


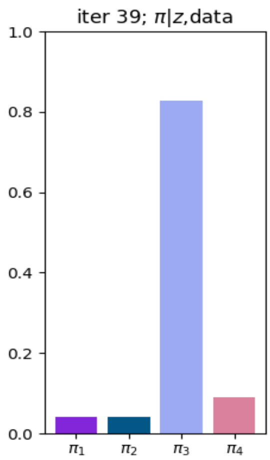
kappa = 1
nu = 5
alphas = 1000
psi = 0.2
m = 0.5





kappa = 1000
nu = 100
alphas = 1000
psi = 0.2
m = 0.5





Analysis of hyper parameters

Alphas - Dirichlet Hyperparameters, higher values of alphas result in a more peaked prior distribution, favouring fewer dominant components in the image segmentation. A higher alpha can influence the likelihood of gaussian centre to be selected and create segments bigger than they should be.

Kappa – Controls the strength of the prior belief, larger kappa is more data-driven when smaller kappa is more faithful to the prior. We can see that for a larger kappa it shifts the weight towards the larger shapes in the image.

Nu – Represents the degree of freedom in the hyperparameters, seems that a higher value of nu allows more flexibility in the segmented shapes.

Psi - Represents the prior precision matrix of the covariance matrices, a lower psi value results in more detailed segmentation.

M – Represents the prior mean vector – couldn’t find any visible affect it has.