Assignment 1

Sankalp Parashar and Utkarsh Ranjan

Spring 2022

Contents

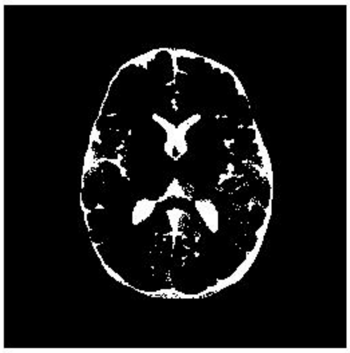
Question 1

Used Parameters

q: 1.6
Neighbourhood Mask(w) -

Initial Estimates

Initial Membership Estimates :
For getting initial membership estimates I first used K-means estimation to get the class means as it provides a good primitive choice of means to start with. Then I updated the values of initial membership by choosing membership values as 0 or 1 in a binary manner based on which class mean is closest to the pixel value. In this manner I obtained the following three initial estimates -

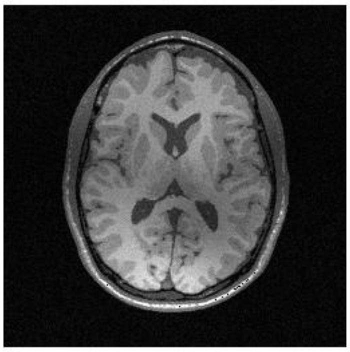


Initial Mean Estimates :

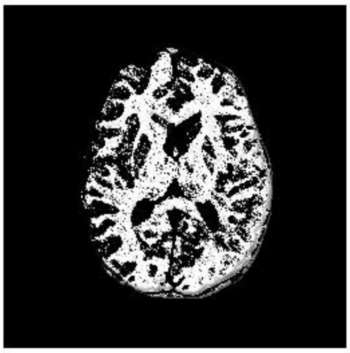
As described before I have use K-means algorithm as it gives a good starting point by estimating membership in a binary manner while ignoring the bias present in the image. The values that I got for the initial mean were - 0.6342, 0.2241 and 0.4531 respectively in the same order as the membership estimates given in the previous part.

Final Results

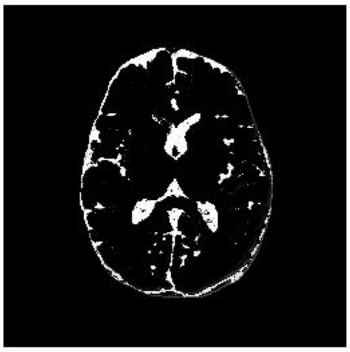
Objective Function :
Values of the objective function with with Iteration number is as follows -
Objective function at Iteration 1 is 0.024770
Objective function at Iteration 2 is 0.019060
Objective function at Iteration 3 is 0.013530
Objective function at Iteration 4 is 0.011551
Objective function at Iteration 5 is 0.009852
Objective function at Iteration 6 is 0.007110
Objective function at Iteration 7 is 0.006694
Objective function at Iteration 8 is 0.006672
Objective function at Iteration 9 is 0.006649
Objective function at Iteration 10 is 0.006513
Objective function at Iteration 11 is 0.006392
Objective function at Iteration 12 is 0.006374
Objective function at Iteration 13 is 0.006386
Images :
Corrupted Image:



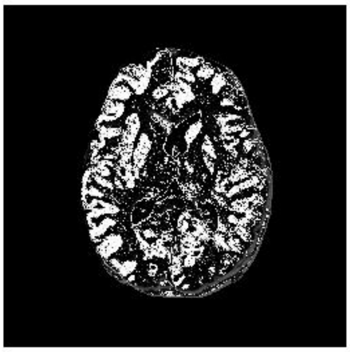
Membership Estimate 1:



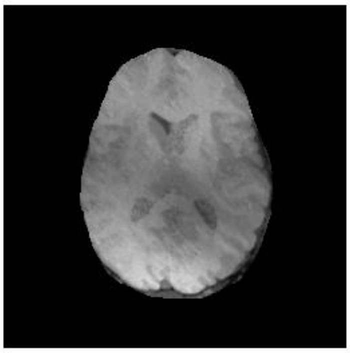
Membership Estimate 2:



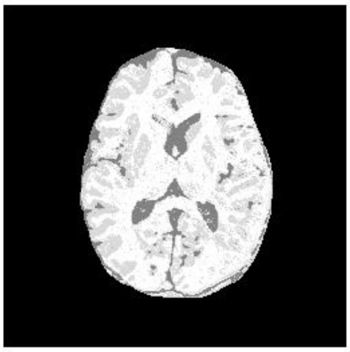
Membership Estimate 3:



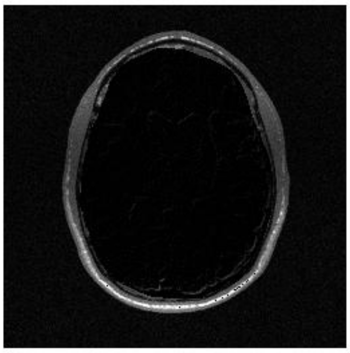
Bias Field Estimate:



Bias Removed Image:



Residual:



Class Means :
The optimal class means are 1.0298, 0.5102 and 0.8570 respectively in same order as the membership estimates given above.

Unique Solution

The formulation discussed in class could possibly give different solutions based on different initial parameters(q and w), different inititalizations of membership or mean and different stopping conditions. However when we fix these parameters for FCM formulation the solution will be unique because no internal randomization is involved in FCM algorithm. A possible way of making it unique is fixing these parameters to certain appropriate values that give good results. In my implementation I have used q = 1.6, w as a Gaussian filter of size 4x4, for initializing the memberships I have used k-means which could lead to some randomization, however, if instead we perform initialization by dividing the range of intensities in the image into three equal divisions and then use this initialization it will also get rid of any randomization in the initialization of membership and mean(means can be set as the middle of these divisions), finally the stopping condition is such that it Iterates until either the objective function starts increasing or a certain maximum number of Iterations is reached. I have implemented such an initialization which makes sure that there is no randomization(instead of kmeans) in q1_unique.m.

Question 2

Used Parameters

$\beta = 0.35$

Initial Estimates

Label Image x

For the label image x. The initial estimate was done based on the range of pixel values.
All pixels were subtracted with the least pixel value in the image. So Every pixel of the brain had a value relative to this minimum pixel. Out of these relative pixel values, those in range [0,max/3] were labelled 1. Those within range [max/3,2*max/3] were assigned 2. While those within [2*max/3,max] were assigned 3. Here max is the maximum relative pixel value.

The Gaussian parameters θ

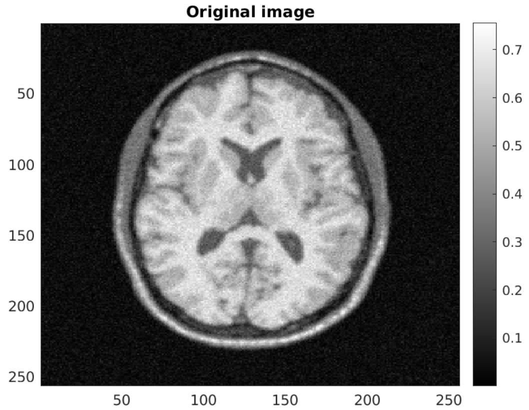
For the Gaussian Parameters of each label class we took the ML estimate of the data labels we estimated using the algorithm above. So mean was simply the average of pixels values for respective label. While Sigma of the underlying gaussian distribution was the standard deviation of the respective pixel values.

Final Results

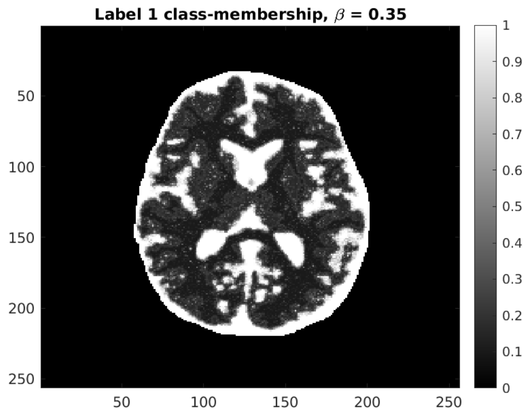
Log posterior probability for the labels :

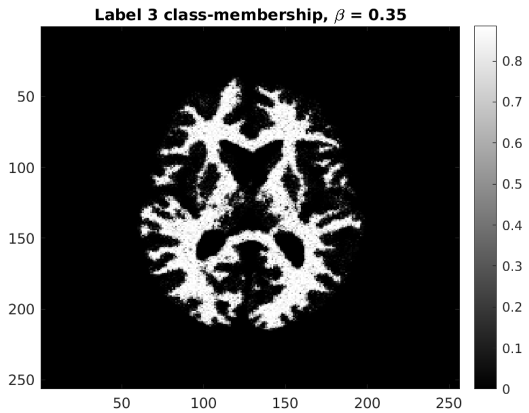
Iter 1 : Before ICM : 2.6803, After ICM :2.8922
Iter 2 : Before ICM : 3.7904, After ICM :4.0153
Iter 3 : Before ICM : 4.4595, After ICM :4.7639
Iter 4 : Before ICM : 5.0659, After ICM :5.2636
Iter 5 : Before ICM : 5.5556, After ICM :5.7108
Iter 6 : Before ICM : 6.0085, After ICM :6.1493
Iter 7 : Before ICM : 6.4524, After ICM :6.5757
Iter 8 : Before ICM : 6.8672, After ICM :6.9787
Iter 9 : Before ICM : 7.2197, After ICM :7.3251
Iter 10 : Before ICM : 7.4861, After ICM :7.5697
Iter 11 : Before ICM : 7.6550, After ICM :7.7338
Iter 12 : Before ICM : 7.7729, After ICM :7.8329
Iter 13 : Before ICM : 7.8495, After ICM :7.8930
Iter 14 : Before ICM : 7.9006, After ICM :7.9312
Iter 15 : Before ICM : 7.9349, After ICM :7.9591
Iter 16 : Before ICM : 7.9612, After ICM :7.9776
Iter 17 : Before ICM : 7.9790, After ICM :7.9922
Iter 18 : Before ICM : 7.9931, After ICM :8.0010
Iter 19 : Before ICM : 8.0016, After ICM :8.0084
Iter 20 : Before ICM : 8.0088, After ICM :8.0121
Iter 21 : Before ICM : 8.0122, After ICM :8.0154
Iter 22 : Before ICM : 8.0157, After ICM :8.0189
Iter 23 : Before ICM : 8.0192, After ICM :8.0220
Iter 24 : Before ICM : 8.0221, After ICM :8.0251
Iter 25 : Before ICM : 8.0252, After ICM :8.0277

Images :
Corrupted Image:

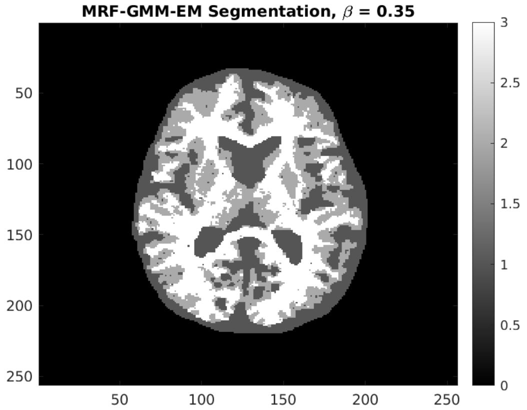


Class Membership Estimate for chosen β:

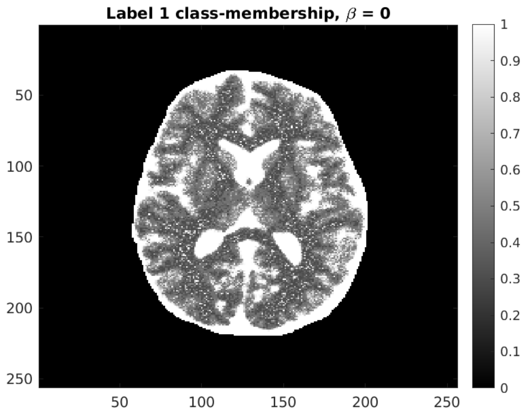


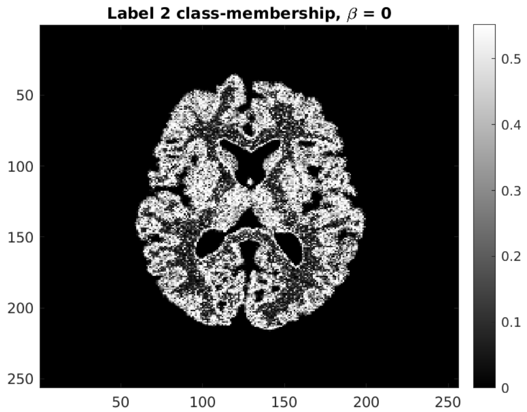


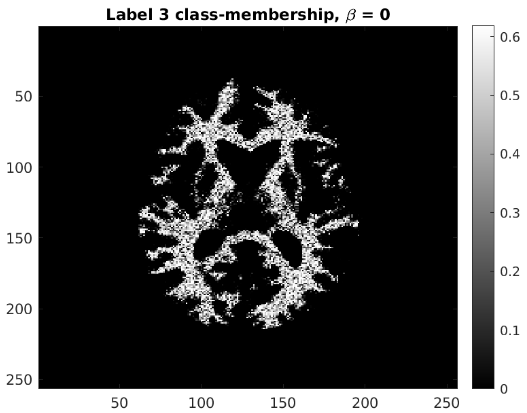
Label Image Estimate for choosen β:



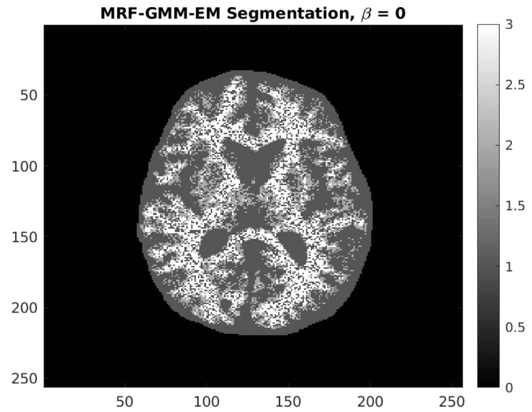
Class Membership Estimate for $\beta = 0$:







Label Image Estimate for $\beta = 0$:



Class Means :
The optimal estimates for class means are 0.4264, 0.5443. 0.6365 respectively in same order as the membership estimates given above for the chosen .

Question 3

How the E step changes

When we have the prior information on the parameter θ in the form of a probability distribution $P(\theta)$, Instead of using EM optimization to perform ML estimation we can perform MAP estimation. So our problem is now to find:

$$max_\theta \; P(y|\theta)P(\theta) \;=\; max_\theta \int_x P(y,x|\theta)P(\theta)\; dx$$

So the EM optimization now maximizes the log posterior instead of log likelihood.
Log posterior can be written as: $F(q,\theta) - KL(q||p(x|y,\theta)) + logP(\theta)$
where $F(q,\theta) = logP(y|\theta) - KL(p(x|y,\theta_i)||p(x|y,\theta))$ remains the same as the case when log-likelihood function was being maximized. This function is also equal to the log-posterior at $\theta = \theta_i$ and is always less than the log-posterior at every $\theta \neq \theta_i$. Hence, The prior term does not influence the E step of the EM algorithm, which proceeds exactly as before

How the M step changes

- Now we have choose θ to maximize $F(q,\theta) + P(\theta)$
- We rewrite $F(q,\theta)$ as we did in ML estimation as:- $F(q,\theta) = E_{q(.)}[logP(x,y|\theta)] + H(q)$, where $H(q)$ = entropy of $q(.)$ = NOT function of θ
- So our problem reduces to maximize $Q(\theta;\theta_i) + log(P(\theta))$ where $Q(\theta;\theta_i) = E_{q(.)}[logP(x,y|\theta)]$

Prior Models

A suitable selection for priors in case of the associated parameters(mean vectors, Covariance matrices and weights) are their respective conjugate priors. Using the conjugate prior provides the added benefit that the posterior belongs to the same family of distribution as the prior distribution. The conjugate prior for mean vectors will be multivariate Gaussians, the prior of covariance matrix is inverse-Wishart distribution and the conjugate prior of weights(which have multinomial distribution) is Dirichlet distribution.

Prior of mean vectors - multivariate Gaussian

$$P(\mu_k; m_k, \Sigma_{\mu_k}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_{\mu_k}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mu_k - m_k)^T \Sigma_{\mu_k}^{-1} (\mu_k - m_k)}$$

where Σ_{μ_k} is the covariance matrix in prior of μ_k, m_k is the mean of prior of μ_k and d is the no. of dimensions in μ_k.

Prior of Covariance matrix - inverse - Wishart

$$P(\Sigma_k; \nu_k, \Psi_k) = \frac{|\Psi_k|^{\nu_k/2}}{2^{\nu_k d/2} \Gamma_d\left(\frac{\nu_k}{2}\right)} |\Sigma_k|^{-(\nu_k + d + 1)/2} e^{-\frac{1}{2} tr(\Psi_k \Sigma_k^{-1})}$$

where Σ_k is a $d \times d$ matrix, ν_k is degrees of freedom of prior distribution of Σ_k, Ψ_k is a $d \times d$ scale matrix, tr is the trace function and Γ_d is the multivariate gamma function.

Prior of Weights - Dirichlet Distribution

$$P(\omega; \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} \omega_i^{\alpha_i - 1}$$

where $B(\alpha) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}$

where K is the no. of Gaussians in the GMM model and $\alpha = (\alpha_1, \ldots, \alpha_K)$

Spiral

M Step for given Prior

Date

As seen in previous part the M step changes so that

$$Q'(\theta;\theta_i) = Q(\theta;\theta_i) + \log(P(\theta))$$

Where $Q(\theta;\theta_i) = E_{q(\cdot)}[\log(P(x,y|\theta))]$

Hence update for μ_k changes as

$$\frac{\partial(Q'(\theta;\theta_i))}{\partial\mu_k} = \sum_{n=1}^{N}\frac{\partial}{\partial\mu_k}\left(\gamma_{nk}\left(-0.5\log|\Sigma_k| - 0.5(y_n-\mu_k)\Sigma_k^{-1}(y_n-\mu_k)\right)\right.$$

$$+ \frac{\partial}{\partial\mu_k}\log P(\mu_k;\nu_k,\Sigma_{\mu_k})$$

$$= \left(\sum_{n=1}^{N}\gamma_{nk}\Sigma_k^{-1} + \Sigma_{\mu_k}^{-1}\right)\mu_k = \sum_{n=1}^{N}\gamma_{nk}\Sigma_k^{-1}y_n + \Sigma_{\mu_k}^{-1}\nu_k$$

$$\boxed{\mu_k = \frac{\Sigma_k^{-1}\sum_{n=1}^{N}\gamma_{nk}y_n + \Sigma_{\mu_k}^{-1}\nu_k}{\Sigma_k^{-1}\sum_{n=1}^{N}\gamma_{nk} + \Sigma_{\mu_k}^{-1}}}$$

Update for Covariance matrix changes as –

Consider multivariate Gaussian

$$\prod_{n,k} G(y_n;\mu_k,\Sigma_k)^{\gamma_{nk}}$$

$$= \prod_{n,k}\frac{1}{(\Sigma_k)^{\frac{\gamma_{nk}}{2}}}\prod_{n,k} e^{-0.5((y_n-\mu_k)^T C_k^{-1}(y_n-\mu_k))\gamma_{nk}}$$

$$= \prod_{k}\frac{1}{(\Sigma_k)^{\frac{\gamma_k}{2}}}\prod_{n,k} e^{-0.5\,\mathrm{tr}(C_k^{-1}A)}$$ (Using Cholesky distribution)

Here A is $\sum_n \gamma_{n,k}(y_n-\mu_k)^T(y_n-\mu_k)$

~~Using Cholesky~~.

We know estimate of Σ_k for this form is

$\Sigma_k = \frac{A}{\gamma_k}$ according to slides.

Spiral

Date

On ~~state~~ introducing Inverse-Wishart prior which has the following form -

$$\frac{\Psi}{\cancel{\quad}} \frac{1}{(|\Sigma_k|)^{\gamma_k/2}} e^{-0.5\,tr(\Sigma_k^{-1}\Psi)}$$

Final estimate of C_k is

$$\boxed{C_k = \frac{A+\Psi}{\gamma_k+\beta_k}}$$

Finally, for updating weights -

$$\frac{\partial\left(Q'(\theta;\theta_i)\right)}{\partial\omega_k} = \frac{\partial}{\partial\omega_k}\sum_{n=1}^{N}\gamma_{nk}(\log\omega_k) + \frac{\partial}{\partial\omega_k}P(\omega_k|\alpha_k)$$

$$= \sum_{n=1}^{N}\frac{\gamma_{nk}}{\omega_k} + \frac{(\alpha_k-1)}{\omega_k}$$

$$L(\omega_k) = Q(\theta;\theta_i) + \lambda(\Sigma\omega_k - 1)$$

$$\Rightarrow \frac{d\{L(\omega_k)\}}{d\omega_k} = \sum_{n=1}^{N}\frac{\gamma_{nk}}{\omega_k} + \frac{\alpha_k-1}{\omega_k} + \lambda = 0$$

$$\Rightarrow \omega_k = -\frac{\left(\sum_{n=1}^{N}\gamma_{nk}+\alpha_k-1\right)}{\lambda}$$

Also $\Sigma\omega_k = 1$

$$\therefore \omega_k = \frac{\left(\sum_{n=1}^{N}\gamma_{nk}+\alpha_k-1\right)}{\sum_{k=1}^{K}\left(\sum_{n=1}^{N}\gamma_{nk}+\alpha_k-1\right)}$$

Spiral