Brain Tissue Segmentation using EM Algorithm

Course Title: Medical Image Segmentation and Applications (MISA)
Group Member: Md Imran Hossain, Muhammad Zain Amin
Email: hossainimran.maia@gmail.com, zainamin1@outlook.com
Erasmus Mundus Joint Master in Medical Imaging Applications (MAIA), University of Girona, Spain

Objective(s):

- To understand the concept of the Gaussian Mixture Model (GMM) and Expectation and Maximization (EM) and its application for image segmentation.
- To develop the Estimation-Maximization (EM) model from the scratch.
- To perform the brain tissue segmentation from MR Images (T1 and T2 FLAIR) using the EM algorithm.
- To compute the segmentation evaluation metric e.g., dice score.

Introduction:

Image segmentation is an essential part of many computer vision systems and medical applications. It is widely used for developing Computer-Aided Diagnosis systems for detecting disease, which helps to take the early steps for clinical treatment. Brain tissue segmentation is one of the challenging tasks. Segmentation can be performed in both supervised and unsupervised manners. However, the cluster-based Expectation-Maximization (EM) combined with a Gaussian Mixture Model (GMM), an unsupervised machine learning algorithm, can perform brain tissue such as White Matter (WM), Gray Matter (GM), and Cerebrospinal Fluid (CSF) segmentation from the MRI perfectly. In this approach, every feature of the feature space has probabilities belonging to different clusters based on Gaussian Probability Distribution, which is considered as a soft assignment technique. In addition, to get a better segmentation outcome, some pre-processing techniques including bias field correction and skull stripping are required. In this lab, we are focused on developing the EM algorithm from scratch to perform the brain tissue segmentation and computing the dice score to understand the segmentation performance.

Expectation-Maximization (EM) Algorithm:

The Expectation-Maximization (EM) algorithm is a powerful iterative method for finding parameters in statistical models. It is commonly used in various fields such as machine learning, computer vision, natural language processing, and bioinformatics. The algorithm iterates between two steps (expectation and maximization) until convergence is achieved. The EM algorithm guarantees monotonic improvement in the likelihood function at each step and converges to a local maximum, though not necessarily a global maximum. Here is the general outline of the EM algorithm:

1. Initialization: *Choose initial values for the parameters (mean, covariance).*
2. E-step: *Compute the expected value of the log-likelihood function.*
3. M-step: *Maximize the expected log-likelihood with respect to the parameters.*
4. Convergence check: *Check if the parameters have converged. If not, go to step 2.*

Initialization: For initialization, we can use two different approaches such as random initialization and k-mean clustering. However, k-mean initialization is more preferable than random initialization to achieve the convergence faster. Means, Covariances, and Mixing components can be used as the initial parameter for the Gaussian Mixture Model (GMM).

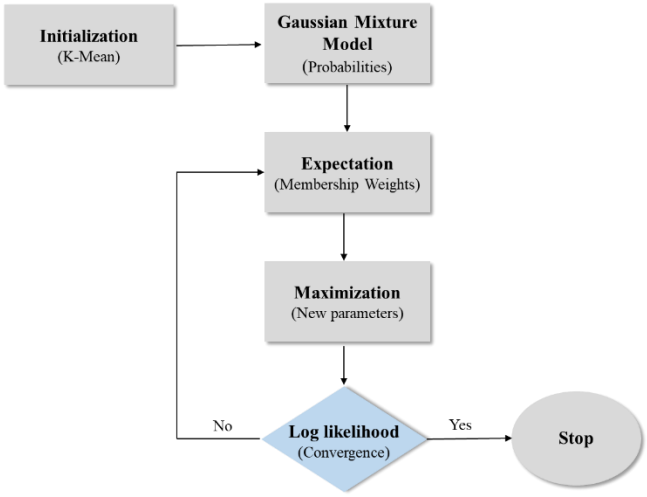

Figure 1: Flow chart of the Expectation and Maximization (EM) algorithm.

Gaussian Mixture Model: A Gaussian Mixture Model (GMM) is defined as a combination of *K* components (clusters) where each component *k* represents the gaussian density characterized by the following parameters in its multivariate version:

- Means/centroid of clusters: μ_k
- Covariance matrix of clusters**:** $\sum_k$
- Mixing components of clusters: α_k

Each component is defined by a gaussian density function by following equation (1):

$$p_k\,(x|\theta_k) = \frac{1}{(2\pi)^{d/2}\,|\sum_k|^{1/2}}\; e^{-\frac{1}{2}(x-\mu_k)^{T}\,\sum_k^{-1}\,(x-\mu_k)} \tag{1}$$

where, *d* is the dimension of data point *x* and θ_k the set of parameters of the gaussian component *k* ($\mu_k \; and \; \Sigma_k$).

The final density of the mixture model of the K gaussian components is given in equation (2), given below:

$$p(\mathrm{x}|\,\Theta) = \sum_{k=1}^{K} \alpha_k p_k\,(x\,|z_k, \theta_k) \tag{2}$$

where, z_k is the latent variable which represent the identity of the mixture component c, that is presumed to have generated the sample *x*. It is a vector of *K* elements where the c^{th} element is one and the remaining elements are zero. In other terms we can say that, it is a one-hot encoding which indicates the belonging of point *x* to cluster *k*.

The objective of the segmentation method based on clustering is to identify the set of optimal values, denoted as Θ, for three key parameters μ, Σ, and, α for each gaussian component, ensuring the best possible fits with the given data points. This is achieved through the application of Maximum Likelihood.

Gaussian mixture models extend the concept of the K-Means algorithm by not only taking into account the centers of the latent Gaussians but also by incorporating the covariance structure of the data points associated with each Gaussian component.

Expectation Step: During this phase, the existing parameters are employed to determine the belonging of each point (voxel) to one of the components. This is achieved through a Bayesian methodology utilizing the means, covariances, and prior probabilities of the Gaussian components. These factors are leveraged to compute the posterior probability or responsibilities, which signifies the likelihood that a given point belongs to a specific component. The posterior probabilities, often

referred to as membership weights, are computed for each data point and for each mixture component in accordance with equation (3).

$$w_{ik} = \frac{p_k(x_i|z_k,\theta_k).\alpha_k}{\sum_{m=1}^{k} p_m(p_m(x_i|z_m,\theta_m).\alpha_m)} \quad (3)$$

The weight w_{ik} denotes the probability that the data sample x_i belongs to the component k given the data sample itself and the current set of parameters Θ. For each point x_i , the sum of the weights across all gaussian components must sum to 1.

Maximization Step: During the maximization step, the weights computed in the E-Step, along with the data samples, are utilized to derive the updated parameters for the Gaussian components using the equations presented in (4), (5), and (6).

$$\alpha_k^{new} = \frac{N_k}{N} \quad (4)$$

$$\mu_k^{new} = \frac{1}{N_k}\sum_{i=1}^{N} w_{ik} \,.\, x_i \quad (5)$$

$$\sum_k^{new} = \frac{1}{N_k}\sum_{i=1}^{N} w_{ik}.(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T \quad (6)$$

Here, *N* represents the overall count of samples, while N_k indicates the sum of membership weights for component *k* across all data samples.

Convergence: Since EM is a iterative method, it necessitates a convergence condition to stop the iterations. One method to track the algorithm's convergence is by calculating the log-likelihood after each iteration and stopping the process when there is no significant change in this metric between two consecutive iterations. Formula 7 defines the log-likelihood.

$$\text{logl}(\theta) = \sum_{i=1}^{N}(\log \sum_{K=1}^{K} \alpha_k p_k(x_i|z_k,\theta_k)) \quad (7)$$

Implementation:

In this section, we provided a detailed explanation of our lab work, where we have developed a segmentation framework using Python that leverages Gaussian Mixture Models (GMM) and Expectation-Maximization (EM) techniques.

To execute the segmentation process, the framework requires two key inputs: the volume intended for segmentation and the specified number of gaussian mixture components to utilize. The framework supports multivariate inputs, meaning that each voxel can encompass one or more channels of information. Before applying our segmentation method to the provided brain MRI scans, we conduct a preliminary skull-stripping step to retain solely the brain tissues. This is achieved by consolidating the three tissue classes, Gray Matter (GM), White Matter (WM), and Cerebrospinal Fluid (CSF), into a unified class, utilizing the provided ground truth mask. Subsequently, this mask is used before giving the MRI scan as an input to the segmentation method to filter out the background voxels. This process ensures that the flattened input vector exclusively contains voxels belonging to one of the three brain tissues.

a) Initialization: To kickstart the GMM-EM segmentation process, the initial step involves is to initialize the set of parameters Θ. We offer various initialization choices for each of these parameters.

- *Means:* Means can be initialized in two distinct ways, either by selecting random values from the provided data samples or by applying k-means clustering to the data and using the resultant centroids as the initial means for the Expectation-Maximization (EM) process, The

size of the means matrix is determined by *K x d*, where *K* represents the number of components and d corresponds to the dimension of the data, signifying the number of features.

- *Covariance:* In the case of initializing means randomly or manually, a common covariance matrix, derived from all data points, is applied to all components. However, if the k-means algorithm has been employed, the labels assigned to each data point are utilized to compute individual covariance matrices for each cluster. These covariance matrices for each component have a size of *d x d*, reflecting the dimensionality of the data.

- *Mixing components:* The choice for mixing components can either be uninformative, where equal probabilities are assigned to all components ($\alpha = 1/N$) for each class, or they can be user-defined by providing specific priors as parameters. The mixing components or priors vector has a size of *1 x K*, corresponding to the number of components.

b) E-Step: During the expectation step, the posterior probabilities, or membership weights, are calculated using Equation (3) and are dependent on the current parameters. The data samples vector, denoted as *x*, has a dimension of *N x d*, with *N* representing the total number of voxels to be categorized, and the resulting weights will have a dimension of *N x K*. It's worth emphasizing once more that the classification is exclusively applied to voxels within the brain tissue mask.

c) M-Step: During the maximization step, we compute the updated parameters, including means, covariance matrices, and mixing components. These calculations are based on the weights acquired from the E-Step, and we have implemented the equations (5), (6), and (7) from the preceding section.

d) Convergence Criteria: To determine when to conclude the EM process, we employ two stopping criteria: Log-likelihood: As explained in the previous section, if the difference in log-likelihood between two consecutive iterations falls below a user-defined tolerance parameter, the EM procedure halts. Maximum iterations: If the algorithm doesn't converge based on the previous criterion even after a predetermined number of iterations, the EM procedure is terminated.

Experimental Results:

To investigate the performance of the developed EM algorithm, we performed a set of experiments on the provided brain image dataset which consists of 5 [T1, T2_FLAIR] images along with their corresponding ground truth. Each MR image has 48 slices. Table 1 illustrates the experimental results such as dice score, number of iterations, and time. The dice score indicates how good the segmentation is. It can be computed by the Equation (8). It is clear from the table that the developed EM model provided a good result. As a result, we got the maximum dice score of 0.9099 for CSF segmentation of the multi-modality images (T1+T2_FLAIR) 1. However, in most of the cases, the EM model performed better for the single modality image (T1).

$$Dice\ Score = \frac{2|A \cap B|}{|A| + |B|} \tag{8}$$

The total number of iteration and time both of them indicates how fast the EM model can perform the segmentation. It is shown that the in case of a single modality the developed EM model takes less number of iterations and times to perform the segmentation tasks. In some cases, the multi-modality image takes almost double the time to be segmented. Therefore, the developed EM model can perform more better for the single modality image than the multi-modality images. The segmentation results of single modality (T1) image and multi-modality (T1+T2_FLAIR) are shown in the Table (2).

Table-1: Experimental results of the developed EM model

Data	Modalities	Iteration	Time (sec)	Dice Score		
				WM	GM	CSF
01	T1	52	13.43	0.8650	**0.8782**	0.8177
	T1 + T2	67	20.70	0.8625	0.8273	**0.9099**
02	T1	38	08.95	0.8019	0.7969	0.8645
	T1 + T2	67	17.11	0.8023	0.7460	0.8048
03	T1	35	06.35	0.8633	0.7949	0.8490
	T1 + T2	66	12.76	0.8504	0.7796	0.8590
04	T1	36	04.12	0.8738	0.8231	0.8632
	T1 + T2	52	12.86	0.8739	0.8394	0.9034
05	T1	45	05.93	**0.9012**	0.8595	0.8425
	T1 + T2	44	09.14	0.8968	0.8524	0.8720

The box plot indicates the mean and standard deviation of the computed dice scores in the case of CSF, GM, and WM. The mean and standard deviation of the dice score are (0.86, 0.81, 0.86) and (0.03, 0.03, 0.03) respectively.

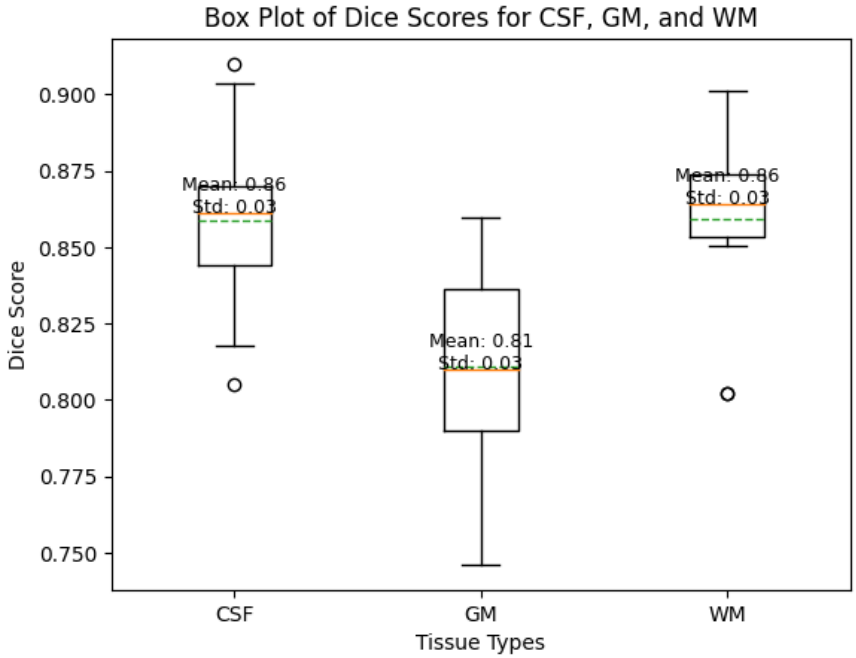

Figure 02: Mean and Standard deviation of dice scores

Table-2: Segmented (T1, T1+T2_FLAIR) images

T1	T2_FLAIR	Segmented Image (T1)	Segmented Image (T1+T2_FLAIR)	Ground Truth