1 Image Segmentation

1.1 Problem Statement

In this section, we focus on using probabilistic graphical models to segment an image into 2: foreground and background. Each pixel belongs to one of the segments and the assignment of each pixel can be modelled as a categorical latent variable (Z), while the pixel value is the observed variable (X). The observed variables can be modelled as a multivariate normal distribution. The observations are 3-dimensional and each pixel can either be represented in the RGB space or the Lab space.

1.2 Approach

This is a clustering problem with 2 clusters, each represented as a Gaussian. The input is an image with N pixels and each pixel is represented as a 3-dimensional random variable $\boldsymbol{x}_n$ where, $n = 1, ..., N$. $\boldsymbol{z}_n$ is a latent random variable that denotes the segment each observation belongs to and follows a categorical distribution that can be represented as:

$$p(Z) = \prod_{k=1}^{2} \pi_k^{z_k} \tag{1}$$

where the parameter $\pi = [\pi_1, \pi_2]$ represent the probability of Z taking each of the 2 states. This paramater also represents mixing coefficients of each cluster's Gaussian. The Gaussian mixture model is given by:

$$p(\boldsymbol{x}) = \sum_{k=1}^{2} \pi_k \mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \tag{2}$$

The responsibility that each segment k takes for explaining each observation $\boldsymbol{x}_n$ is given by $\gamma(\boldsymbol{Z})$ which is an N x 2 table. We aim to find the parameters, θ, that maximize the log-likelihood of the model, which is given by:

$$lnp(\boldsymbol{x}_1, ..., \boldsymbol{x}_N|\theta) = \sum_{n=1}^{N} ln \sum_{\boldsymbol{z}_n} p(\boldsymbol{x}_n, \boldsymbol{z}_n|\theta), \theta = \{\pi_1, \pi_2, \mu_1, \mu_2, \Sigma_1, \Sigma_2\} \tag{3}$$

Since we cannot obtain a closed form solution of the unknown parameters by setting the derivatives to 0, we initialize the parameters to sensible initial values and iteratively calculate new responsibilities and new parameters until the log-likelihood converges. To obtain sensible initializations, we first perform K-means clustering to estimate π_k, μ_k and Σ_k for each cluster. Then we proceeded to the expectation step where we calculated the responsibilities as:

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}\left(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{j=1}^{2} \pi_j \mathcal{N}\left(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)} \tag{4}$$

Next, in the maximization step, we re-estimate the parameters using the current responsibilities as such:

$$\boldsymbol{\mu}_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{x}_n \tag{5}$$

$$\boldsymbol{\Sigma}_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{new}\right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{new}\right)^T \tag{6}$$

$$\boldsymbol{\pi}_k^{new} = \frac{N_k}{N} \tag{7}$$

where,

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk}) \tag{8}$$

After every iteration, we calculate the log-likelihood as specified in Equation 3. The difference in the increase in log-likelihood after every iteration was computed. If the difference was less than 1×10^{-6}, the iteration was stopped and the final responsibilities were calculated. The segment that had the highest responsibility for each observation was assigned as the label.

1.3   Results and Discussion

Using EM algorithm, we successfully segmented all 4 test images into distinct foreground and background segments, with minimal noise. We experimented with using 3 features (the L, a, b channels) and 5 features (x-coordinate, y-coordinate and the L, a, b channels). We found that the results were better when 5 features were used. The input data was scaled along each feature independently using their respective means and standard deviations. We did not notice any difference in result between using the raw data and scaled data. However, the runtime was faster when scaled data was used.