LAB 3: EM-ALGORITHM FOR A SIMPLE SHAPE MODEL

Andrii Zakharchenko

January 2020

1 Algorithm derivation

1.1 Log-likelihood

To derive a log-likelihood function we will start from the fact that we are interested in estimating some parameters of the model for some dataset. In statistical language we can set up this problem as we want to maximize the probability of observing our dataset given some parameters:

$$p(\mathcal{T}|\Theta, \mathbf{u}) = p(x^1, x^2, ...x^m|\Theta, \mathbf{u}) \tag{1}$$

where $\mathcal{T}$ is the dataset and $x^l \in \mathcal{T}$.
By x_i^l we will denote the i-th pixel of l-th image. Probability from the eq. 1 can be expressed in terms of pixels:

$$p(\mathcal{T}|\Theta, \mathbf{u}) = p(x_1^1, x_2^1, ..., x_n^1, x_1^2, x_2^2, ..., x_n^1, ..., x_1^m, x_2^m, ..., x_n^m|\Theta, \mathbf{u}) \tag{2}$$

From this we can construct likelihood function as follows:

$$\ell(\Theta, \mathbf{u}) = \prod_l^m \prod_i^n p(x_i^l|u_i, \theta^l) \tag{3}$$

where $\theta^l \in \Theta$ and $u_i \in \mathbf{u}$ To get a log-likelihood we will take a natural logarithm of the eq. 3:

$$L(\Theta, \mathbf{u}) = \sum_l^m \sum_i^n \log p(x_i^l|u_i, \theta^l) \tag{4}$$

Further we will denote $p(x|u_i, \theta^l)$ and similar as $p_{u_i,\theta^l}(x)$.
Since we want to segment image in two clusters we will denote them as $s_i \in 0, 1$, where $s_i = 0$ means i-th pixel belongs to background and $s_i = 1$ means i-th pixel belongs to foreground.
From the fact that we can get $p(A)$ from $p(A, B)$ as $p(A) = \sum_B p(A, B)$ we can further decompose log-likelihood function as:

$$L(\Theta, \mathbf{u}) = \sum_l^m \sum_i^n \log \sum_{s_i} p_{u_i,\theta^l}(x_i^l, s_i) \tag{5}$$

We can also scale this function by m total number of images, but scaling should not affect $\arg\max_{\Theta,\mathbf{u}} L$.

1.2 E-step

EM algorithm optimizes lover bound of log-likelihood function. We can construct lover bound by using Jensen's inequality which for concave function $f(x)$ looks like this:

$$E[f(x)] \leq f(E[x]) \tag{6}$$

where $E[.]$ is the expected value.
By introducing latent variable $\alpha_l(s_i)$ s.t $\sum_{s_i} \alpha_l(s_i) = 1$ and $\alpha_l(s_i) \geq 0$ we can rewrite eq. 5 as:

$$L(\Theta, \mathbf{u}) = \sum_{l}^{m} \sum_{i}^{n} \log \sum_{s_i} \frac{\alpha_l(s_i)}{\alpha_l(s_i)} p_{u_i,\theta^l}(x_i^l, s_i) \tag{7}$$

note that $\sum_{s_i} \alpha_l(s_i) \frac{p_{u_i,\theta^l}(x_i^l,s_i)}{\alpha_l(s_i)}$ is the expected value of $\frac{p_{u_i,\theta^l}(x_i^l,s_i)}{\alpha_l(s_i)}$ w.r.t to s_i and we can substitute this to Jensen's inequality:

$$E[f(\frac{p_{u_i,\theta^l}(x_i^l, s_i)}{\alpha_l(s_i)})] \leq f(E[\frac{p_{u_i,\theta^l}(x_i^l, s_i)}{\alpha_l(s_i)}]) \tag{8}$$

from eq. 7 we can deduce lover bound as:

$$\begin{aligned} \log \sum_{s_i} \alpha_l(s_i) \frac{p_{u_i,\theta^l}(x_i^l, s_i)}{\alpha_l(s_i)} &\geq \sum_{s_i} \log \alpha_l(s_i) \frac{p_{u_i,\theta^l}(x_i^l, s_i)}{\alpha_l(s_i)} \\ &= \sum_{s_i} \alpha_l(s_i) \log p_{u_i,\theta^l}(x_i^l, s_i) - \alpha_l(s_i) \log \alpha_l(s_i) \end{aligned} \tag{9}$$

To deduce the E-step we recall that EM algorithm maximizes lover bound L_B, which basically means that $L_B = L$. This means in eq. 8 we will write equality sign. From the property of expected value $E[f(x)] = f(E[x])$ i.f.f $x = E[x]$(e.g. when x is constant) we can say that:

$$\frac{p_{u_i,\theta^l}(x_i^l, s_i)}{\alpha_l(s_i)} = C \implies \alpha_l(s_i) \propto p_{u_i,\theta^l}(x_i^l, s_i) \tag{10}$$

Since $\alpha_l(s_i)$ is valid probability distribution we can say that $\alpha_l(s_i) = p_{u_i,\theta^l}(s_i|x_i^l)$ since $p_{u_i,\theta^l}(s_i|x_i^l) = \frac{p_{u_i,\theta^l}(x_i^l,s_i)}{p_\theta^l(x_i^l)}$. This is how we deduce that E-step is:

$$\alpha_l(s_i) = p_{u_i,\theta^l}(s_i|x_i^l)$$

1.3 M-step

Let's say that

$$f(x_i^l, s_i) = \log p_{u_i,\theta^l}(x_i^l, s_i) = \log p_{u_i}(s_i) p_{\theta^l}(x_i^l|s_i) = \\ \log p_{u_i}(s_i) + \log p_{\theta^l}(x_i^l|s_i) \tag{11}$$

We know $p_{u_i}(s_i)$ and $p_{\theta^l}(x_i^l|s_i)$ from the task, except for the parameters u_i and θ^l. By substituting $f(x_i^l, s_i)$ to lover bound we will get:

$$\begin{aligned} L_B = &\sum_l \sum_i \sum_{s_i} \alpha_l(s_i = 1) u_i - log(1 + e^{u_i}) \\ &+ \alpha_l(s_i = 1) \log p_{\theta_1^l}(x_i^l|s_i = 1) \\ &+ \alpha_l(s_i = 0) \log p_{\theta_0^l}(x_i^l|s_i = 0) \\ &- \alpha_l(s_i) \log \alpha_l(s_i) \end{aligned} \tag{12}$$

From the above equation we can see that maximizing L_B by each u_i, θ_0^l, θ_1^l decomposes this to 3 separate tasks
Also, to show that function from point a) in the task is concave and have unique global maximum we need to take 2-nd derivative of that function and show that it's strictly negative:

$$\frac{\partial^2 f}{\partial u_i^2} = -\frac{e^{u_i}}{(1 + e^{u_i})^2} < 0 \tag{13}$$

Since 2-nd derivative is less than zero, this means that function is strictly concave, which means that it has unique global maximum.

2 Baseline models

K-means has average precision of 0.835 while Gaussian mixture model has 0.881. GMM performs better since it has more flexibility in cluster shapes. For intuitive explanation imagine, for example, 2 clusters that partially overlap. K-means will assign clusters to points just by Euclidean distance from the center of the cluster, while GMM will learn distributions of clusters and will classify points according to these distributions. So the border between these 2 clusters in area of overlap in case of k-means may look like a hard line, while for GMM it may look like a soft gradient.

3 EM algorithm for simple shape model

The average precision of the EM algorithm for the shape model is: 0.990. We see that the result is better than k-means and gmm. This is because those base methods take into account only colors of the image, while this algorithm also takes into account the shape by estimating $p(s_i)$.
Since with each iteration EM algorithms monotonically improves log-likelihood

$L(\Theta_t, \mathbf{u}_t) \leq L(\Theta_{t+1}, \mathbf{u}_{t+1})$, convergence criteria was chosen to check the difference of log-likelyhood between steps and to stop when the difference is smaller than some tolerance threshold, which would mean that EM approaches maximum of L and convergence becomes too slow.

Figure 1: "Learned shape model"

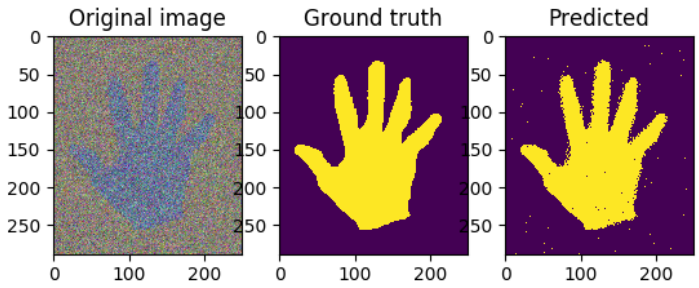

Figure 2: Segmentation of hand-0.png

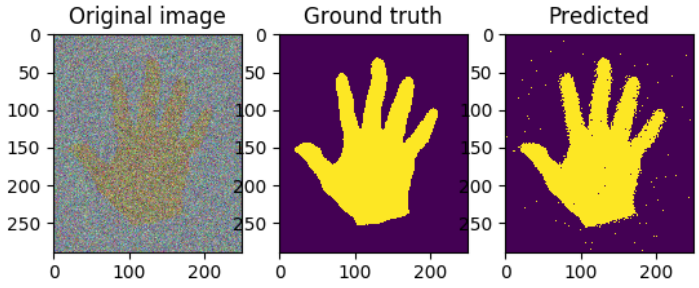

Figure 3: Segmentation of hand-17.png

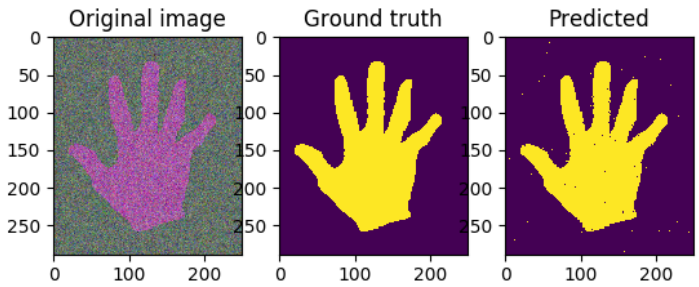

Figure 4: Segmentation of hand-49.png