Assignment 1.

E-step: Deduce that the E-step reads as simple as

$$\alpha_\ell(s_i) = p_{u_i,\theta^\ell}(s_i \mid x_i^\ell), \tag{5}$$

where $\mathbf{u}, \boldsymbol{\theta}$ denote the current estimate of the parameters.

$$\sum_{i\in D} \sum_{s_i\in\{0,1\}} \alpha_\ell(s_i) \log p_{u_i,\theta^\ell}(x_i^\ell, s_i) - \alpha_\ell(s_i) \log \alpha_\ell(s_i) \to \max_{\alpha_\ell(s_i)}$$

s.t.

$$\alpha_\ell(s_i) > 0$$

$$\sum_{i\in D} \sum_{s_i\in\{0,1\}} \alpha_\ell(s_i) = 1$$

$$L = \sum_{i\in D} \sum_{s_i\in\{0,1\}} \alpha_\ell(s_i) \log p_{u_i,\theta^\ell}(x_i^\ell, s_i) - \alpha_\ell(s_i) \log \alpha_\ell(s_i) - \lambda\left(\sum_{i\in D} \sum_{s_i\in\{0,1\}} \alpha_\ell(s_i) - 1\right)$$

$$\frac{\partial L(\alpha_\ell, \lambda)}{\partial \alpha_\ell(s_i)} = \log p_{u_i,\theta^\ell}(x_i^\ell, s_i) - \lambda - \log \alpha_\ell(s_i) - 1 = 0$$

$$\log p_{u_i,\theta^\ell}(x_i^\ell, s_i) \cdot e^{-\lambda-1} = \alpha_\ell(s_i)$$

$$\sum_{i\in D} \sum_{s_i\in\{0,1\}} \alpha_\ell(s_i) = 1 \Rightarrow 1 = \sum_{i\in D} \sum_{s_i\in\{0,1\}} \log p_{u_i,\theta^\ell}(x_i^\ell, s_i) \cdot e^{-\lambda-1}$$

$$= e^{-\lambda-1} \sum_{i\in D} \sum_{s_i\in\{0,1\}} \log p_{u_i,\theta^\ell}(x_i^\ell, s_i) = e^{-\lambda-1} p_{u_i,\theta^\ell}(x_i^\ell)$$

$$\alpha_\ell(s_i) = \frac{p_{u_i,\theta^\ell}(x_i^\ell, s_i)}{p_{u_i,\theta^\ell}(x_i^\ell)} = p_{u_i,\theta^\ell}(s_i \mid x_i^\ell)$$

Deduce that the log-likelihood $L_m(\mathbf{u}, \boldsymbol{\theta})$ of the training data $\mathcal{T}^m$ decomposes into a sum over images and pixels

$$L_m(\mathbf{u}, \boldsymbol{\theta}) = \frac{1}{m} \sum_{\ell=1}^{m} \sum_{i\in D} \log \sum_{s_i\in\{0,1\}} p_{u_i,\theta^\ell}(x_i^\ell, s_i) \tag{2}$$

$$p(x_i \mid u_i) = p(x_i \mid \theta_0^\ell) \cdot p_{u_i}(\theta_0^\ell) + p(x_i \mid \theta_1^\ell) \cdot p_{u_i}(\theta_1^\ell)$$

$$= p_{u_i}(x_i, \theta_0^\ell) + p_{u_i}(x_i, \theta_1^\ell) = \sum_{s_i \in \{0,1\}} p_{u_i}(x_i, \theta_i^\ell) = \sum_{s_i \in \{0,1\}} p_{u_i, \theta^\ell}(x_i, s_i)$$

$$L_m(u, \theta) = \frac{1}{m} \sum_{\ell=1}^{m} \prod_{i \in D} p(x_i \mid u_i) = \frac{1}{m} \sum_{\ell=1}^{m} \sum_{i \in D} \log \sum_{s_i \in \{0,1\}} p_{u_i, \theta^\ell}(x_i^\ell, s_i)$$

(a) Show that the maximisation w.r.t. $\mathbf{u}$ further decomposes into independent tasks for each pixel $i \in D$

$$\frac{1}{m} \sum_{\ell=1}^{m} \alpha_\ell(s_i = 1)\, u_i - \log(1 + e^{u_i}) \to \max_{u_i} \tag{6}$$

Show that the function is concave and has a unique global maximum.

First derivative

$$\frac{1}{m} \sum_{\ell=1}^{m} \alpha_\ell(s_i = 1) - \frac{1}{1 + e^{u_i}} \cdot e^{u_i}$$

Second derivative

$$-\frac{e^{u_i}}{(1 + e^{u_i})^2} < 0$$

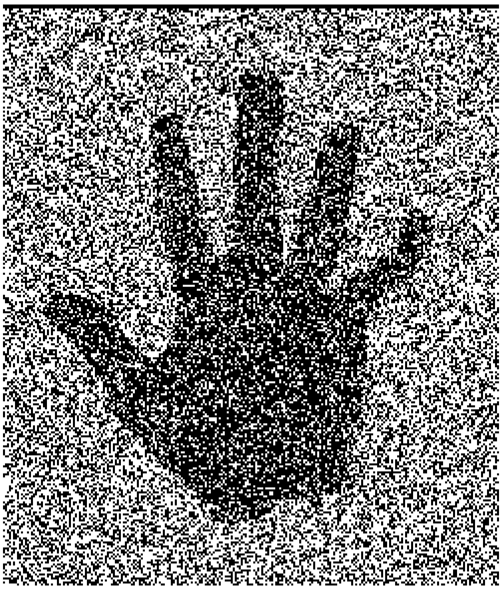
This function is always less than zero, so function is concave

Assignment 2.

Error depends on the type of picture. For example, it is very hard to distinguish the background from the foreground in the picture 47, so the error is very high. For picture 2 it is vice versa.

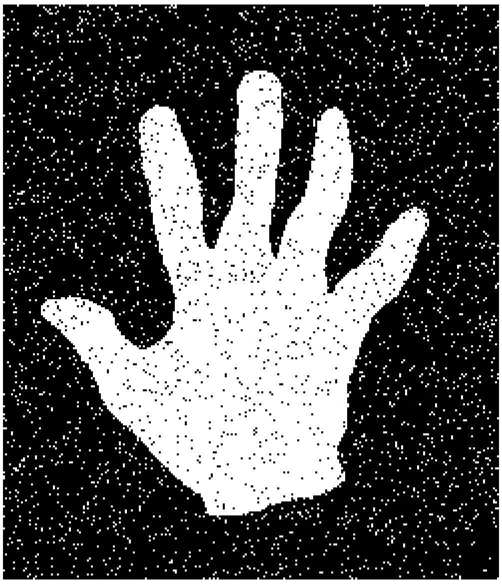
(1) k-means algorithm

Picture 47



Error = 0.729412

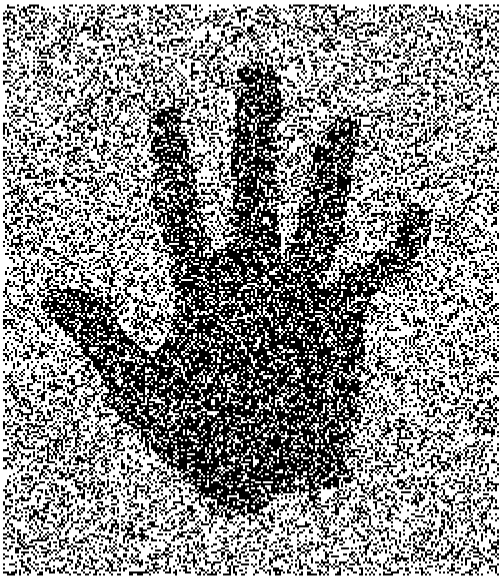
Picture 2



Error = 0.094118

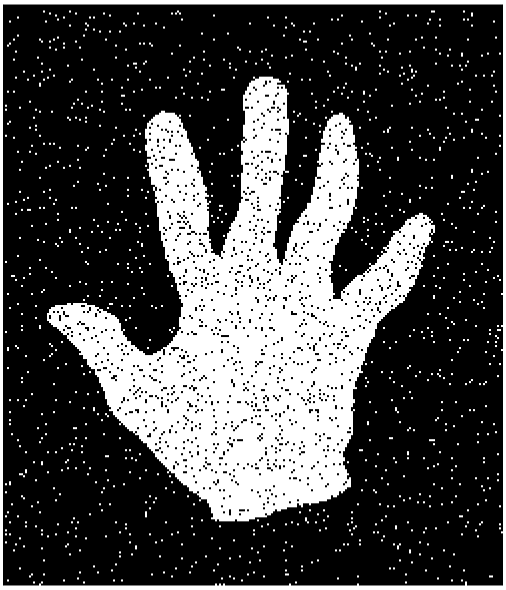
(2) Mixture of Gaussians

Picture 47



Error = 0.650980

Picture 2



Error = 0.070588

Mixture of Gaussians shows better results than k-means algorithm because Gaussian model select points more flexible than k-means. Gaussian model uses range of standard deviation for clustering points.