Derivation of Semi-Supervised Learning using Expectation Maximization and Gaussian Mixture Model

Sheng Lundquist

1 Notation

$$
\begin{aligned}
N_u = &\quad \text{Number of unsupervised samples} && (1)\\
N_s = &\quad \text{Number of supervised samples} && (2)\\
N = &\quad N_u + N_s && (3)\\
K = &\quad \text{Number of clusters} && (4)\\
x = &\quad \{x_{1_s}, x_{2_s}, ...x_{N_s}, x_{1_u}, x_{2_u}, ...x_{N_u}\} && (5)\\
s = &\quad \{s_1, s_2, ...s_K\} && (6)\\
&\quad \text{where each s is a one-hot vector for the correct class} && (7)\\
r = &\quad \{r_1, r_2, ...r_K\} && (8)\\
\Theta = &\quad \{\pi, \mu, \Sigma\} && (9)\\
\pi = &\quad \{\pi_1, \pi_2, ...\pi_K\} && (10)\\
\mu = &\quad \{\mu_1, \mu_2, ...\mu_K\} && (11)\\
\Sigma = &\quad \{\Sigma_1, \Sigma_2, ...\Sigma_K\} && (12)\\
S_k = &\quad \sum_i^{N_s} s_{ik} && (13)\\
R_k = &\quad \sum_i^{N_s} r_{ik} && (14)
\end{aligned}
$$

2 Objective

$$L(\Theta) = \sum_{i}^{N_s} \ln f(x_i, y_i) + \sum_{i}^{N_u} \ln f(x_i) \tag{15}$$

$$\ln f(x_i, y_i) = \sum_{k}^{K} s_k \ln \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{16}$$

$$\ln f(x_i) = \sum_{k}^{K} r_k \ln \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{17}$$

3 E step

$$P(y_i = k, x|\Theta) = \pi_k \mathcal{N}(x, \mu_k, \Sigma_k) \tag{18}$$

$$P(y_i = k|x, \Theta) = r_k = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_j^K \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)} \tag{19}$$

4 M step

4.1 M step for π

$$\begin{aligned}
\arg\max_{\pi} L(\Theta) = &\quad \sum_{i}^{N_s}\sum_{k}^{K} s_{ik}\ln(\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)) + && (20)\\
&\quad \sum_{i}^{N_u}\sum_{k}^{K} r_{ik}\ln(\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)) && (21)\\
= &\quad \sum_{i}^{N_s}(\sum_{k}^{K} s_{ik}\ln\pi_k + \sum_{k}^{K} s_{ik}\ln\mathcal{N}(x_i|\mu_k,\Sigma_k)) + && (22)\\
&\quad \sum_{i}^{N_u}(\sum_{k}^{K} r_{ik}\ln\pi_k + \sum_{k}^{K} r_{ik}\ln\mathcal{N}(x_i|\mu_k,\Sigma_k)) && (23)\\
= &\quad \sum_{k}^{K} S_k\ln\pi_k + \sum_{k}^{K} R_k\ln\pi_k + C && (24)\\
= &\quad (N)(\sum_{k}^{K}\frac{S_k}{N}\ln\pi_k + \sum_{k}^{K}\frac{R_k}{N}\ln\pi_k) && (25)\\
= &\quad (N)\sum_{k}^{K}(\frac{S_k+R_k}{N})\ln\pi_k && (26)\\
\leq &\quad (N)\sum_{k}^{K}(\frac{S_k+R_k}{N})\ln\frac{S_k+R_k}{N} && (27)\\
\pi_k = &\quad \frac{S_k+R_k}{N} && (28)
\end{aligned}$$

4.2 M step for μ

$$\begin{aligned}
\frac{dL}{d\mu_k} = & \sum_i^{N_s} s_{ik} \frac{d}{d\mu_k}(\ln \mathcal{N}(x_i|\mu_k, \Sigma_k)) + & (29)\\
& \sum_i^{N_u} r_{ik} \frac{d}{d\mu_k}(\ln \mathcal{N}(x_i|\mu_k, \Sigma_k)) & (30)\\
\ln \mathcal{N}(x|\mu, \Sigma) = & -\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) & (31)\\
\frac{d}{d\mu} = & \Sigma^{-1}(x-\mu) & (32)\\
\frac{dL}{d\mu_k} = & \sum_i^{N_s} s_{ik}\Sigma_k^{-1}(x_i-\mu_k) + \sum_i^{N_u} r_{ik}\Sigma_k^{-1}(x_i-\mu_k) & (33)\\
0 = & \sum_i^{N_s} s_{ik}\Sigma_k^{-1}(x_i-\mu_k) + \sum_i^{N_u} r_{ik}\Sigma_k^{-1}(x_i-\mu_k) & (34)\\
0 = & \sum_i^{N_s} s_{ik}(x_i-\mu_k) + \sum_i^{N_u} r_{ik}(x_i-\mu_k) & (35)\\
0 = & \sum_i^{N_s} s_{ik}(x_i-\mu_k) + \sum_i^{N_u} r_{ik}(x_i-\mu_k) & (36)\\
0 = & \sum_i^{N_s} s_{ik}x_i - \sum_i^{N_s} s_{ik}\mu_k + \sum_i^{N_u} r_{ik}x_i - \sum_i^{N_u} r_{ik}\mu_k & (37)\\
\mu_k(S_k + R_k) = & \sum_i^{N_s} s_{ik}x_i + \sum_i^{N_u} r_{ik}x_i & (38)\\
\mu_k = & \frac{\sum_i^{N_s} s_{ik}x_i + \sum_i^{N_u} r_{ik}x_i}{S_k + R_k} & (39)
\end{aligned}$$

4.3 M step for Σ

$$\frac{dL}{d\Sigma_k^{-1}} = \frac{\Sigma_k}{2} S_k - \frac{1}{2}\sum_i^{N_s} s_{ik}(x_i - \mu_k)(x_i - \mu_k)^T + \tag{40}$$

$$\frac{\Sigma_k}{2} R_k - \frac{1}{2}\sum_i^{N_u} r_{ik}(x_i - \mu_k)(x_i - \mu_k)^T \tag{41}$$

$$0 = \frac{\Sigma_k}{2} S_k - \frac{1}{2}\sum_i^{N_s} s_{ik}(x_i - \mu_k)(x_i - \mu_k)^T + \tag{42}$$

$$\frac{\Sigma_k}{2} R_k - \frac{1}{2}\sum_i^{N_u} r_{ik}(x_i - \mu_k)(x_i - \mu_k)^T \tag{43}$$

$$\frac{\Sigma_k}{2}(S_k + R_k) = \frac{1}{2}(\sum_i^{N_s} s_{ik}(x_i - \mu_k)(x_i - \mu_k)^T + \sum_i^{N_u} r_{ik}(x_i - \mu_k)(x_i - \mu_k)^T) \tag{44}$$

$$\Sigma_k = \frac{\sum_i^{N_s} s_{ik}(x_i - \mu_k)(x_i - \mu_k)^T + \sum_i^{N_u} r_{ik}(x_i - \mu_k)(x_i - \mu_k)^T}{S_k + R_k} \tag{45}$$