Learning Gaussian Mixture Parameters for the MATISSE algorithm

Igor Ulitsky Ron Shamir

December 7, 2006

We describe here how the Gaussian mixture parameters $\Theta = (\mu_m, \sigma_m, \mu_n, \sigma_n, p_m)$ can be learned using the EM algorithm in our probabilistic model, given the values of $P(R_i)$ for every gene. The EM procedure closely resembles the standard mixture-density parameter estimation problem addressed in detail in [?]. Thus we shall address only the differences between the standard procedure with two component densities optimization and our problem, introduced by the use of $P(R_i)$ priors.

Define:

$$\alpha_{1ij} = p_m P(R_i)P(R_j)$$

$$\alpha_{2ij} = (1 - p_m P(R_i)P(R_j))$$

$$\mu_1 = \mu_m$$

$$\sigma_1 = \sigma_m$$

$$\mu_2 = \mu_n$$

$$\sigma_2 = \sigma_n$$

We will denote by $P_N(S_{ij}, \mu_\ell, \sigma_\ell)$ the density of S_{ij} in the normal distribution $N(\mu_\ell, \sigma_\ell)$:

$$P_N(S_{ij}, \mu_\ell, \sigma_\ell) = \frac{1}{\sqrt{2\pi}\sigma_\ell} \exp(\frac{-(S_{ij} - \mu_\ell)^2}{2\sigma_\ell{}^2})$$

For every pair of genes (i, j), the probability of observing the similarity S_{ij} is given by:

$$P(S_{ij}|\Theta) = \sum_{\ell=1}^{2} \alpha_{\ell ij} P_N(S_{ij}, \mu_\ell, \sigma_\ell)$$

.

Define x_{ij} as the indicator of the pair (i, j) being mates. Given the unobserved data items $X = \{x_{ij}\}_{(i,j)\in N\times N}$, the complete-data log-likelihood is:

$$\log(L(\Theta|S, X)) = \log(P(S, X|\Theta)) = \sum_{i=1}^{N}\sum_{j=1}^{N} \log(P(S_{ij}|x_{ij})P(x_{ij}))$$

The algorithm starts with some initial parameter guess $\Theta^g = (\mu_m^g, \sigma_m^g, \mu_n^g, \sigma_n^g, p_m^g)$. The mixing parameters $\alpha_{\ell ij}$ can be thought of as prior probabilities of each mixture components, distinct for every pair (i, j). Therefore, Using Bayes's rule we can compute:

$$P(x_{ij} = \ell|S_{ij}, \Theta^g) = \frac{\alpha_{\ell ij} P_N(S_{ij}, \mu_\ell^g, \sigma_\ell^g)}{P(S_{ij}|\Theta^g)}$$

and

$$P(X|S, \Theta^g) = \prod_{i=1}^{N} P(x_{ij}, \Theta^g)$$

Following the same equation manipulations as in [?] we can write the following term for $Q(\Theta, \Theta^g)$:

$$Q(\Theta, \Theta^g) = \sum_{\ell=1}^{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \log(\alpha_{\ell ij})p(\ell|S_{ij}, \Theta^g) + \log(P_N(S_{ij}, \mu_\ell^g, \sigma_\ell^g))$$

This expression can be maximized by maximizing the term containing $\alpha_{\ell ij}$ and the term containing μ_ℓ and σ_ℓ independently, as they are not related.

As in [?], we introcude the Lagrange multiplier with the constraint that $\sum_i \sum_j \sum_\ell \alpha_{\ell ij} = 1$, and solve the following:

$$\frac{\partial}{\partial \alpha_{\ell ij}} [\sum_\ell \sum_i \sum_j \log(\alpha_{\ell ij})p(\ell|S_{ij}, \Theta^g) + \lambda(\sum_i \sum_j \sum_\ell \alpha_{\ell ij} = 1)] = 0$$

Summing both sides over ℓ, we get that $\lambda = -\sum_{i=1}^{N}\sum_{j=1}^{N} P(R_i)P(R_j)$ and find an update formula for p_m:

$$p_m = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} P(\ell|S_{ij}, \Theta^g)}{\sum_{i=1}^{N}\sum_{j=1}^{N} P(R_i)P(R_j)}$$

.

The equations for update of the μ and σ parameters are the same as in [?]:

$$\mu_\ell^{new} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} S_{ij} P(\ell|S_{ij}, \Theta^g)}{\sum_{i=1}^{N}\sum_{j=1}^{N} P(\ell|S_{ij}, \Theta^g)}$$

$$\sigma_\ell^{new} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} (S_{ij} - \mu_\ell^{new})^2 P(\ell|S_{ij}, \Theta^g)}{\sum_{i=1}^{N}\sum_{j=1}^{N} P(\ell|S_{ij}, \Theta^g)}$$