

Maximum Likelihood - Mixture of Gaussians

Bernt Schiele, TU Darmstadt

November 17, 2005

In the following I derive the standard equation for the Maximum Likelihood estimation for a mixture of Gaussians. I will concentrate on the mean of a single Gaussian. The other estimates (for the variance and the mixing coefficients) can be derived in a similar way.

In the class we had the following equations for a mixture of Gaussians:

$$p(x) = \sum_{j=1}^M p(x|j)P(j) \quad (1)$$

with $p(x|j)$ a single Gaussian and $P(j)$ the mixing coefficients.

In order to derive the maximum likelihood estimate let's make the involved parameters explicit. Those are μ_j and σ_j^2 for each Gaussian and the mixing coefficients themselves: $P(j) = \alpha_j$. The parameter vector θ therefore contains $3 \times M$ parameters:

$$\theta = (\mu_1, \mu_2, \dots, \mu_M, \sigma_1^2, \sigma_2^2, \dots, \sigma_M^2, \alpha_1, \alpha_2, \dots, \alpha_M) \quad (2)$$

In the following I want to make the dependency of the likelihood of a data point x_n from this parameter vector θ more explicit. I therefore introduce the following notations:

$$p(x_n|j, \theta) = p(x_n|\mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_n - \mu_j)^2}{2\sigma_j^2}\right) \quad (3)$$

$$p(x_n|\theta) = \sum_{j=1}^M p(x_n|j, \theta)P(j|\theta) \quad (4)$$

$$= \sum_{j=1}^M p(x_n|\mu_j, \sigma_j^2)\alpha_j \quad (5)$$

Let's also write out the a posterior probability of a mixture component j given a particular data point x_n since we will need it later in the derivation. This probability is $P(j|x_n, \theta)$ – again making the dependency on the parameter vector θ explicit:

$$P(j|x_n, \theta) = \frac{p(x_n|j, \theta)P(j|\theta)}{p(x_n|\theta)} \quad (6)$$

$$= \frac{p(x_n|\mu_j, \sigma_j^2)\alpha_j}{p(x_n|\theta)} \quad (7)$$

The standard log-likelihood for the training set $X = \{x_1, x_2, \dots, x_N\}$ is then defined as:

$$E = -\ln L(\theta) = -\sum_{n=1}^N \ln p(x_n|\theta) \quad (8)$$

In order to find the ML (Maximum Likelihood) estimate for example for the mean μ_j we compute the following partial derivative:

$$\frac{\partial}{\partial \mu_j} E = -\frac{\partial}{\partial \mu_j} \sum_{n=1}^N \ln p(x_n|\theta) \quad (9)$$

$$= -\sum_{n=1}^N \frac{1}{p(x_n|\theta)} \frac{\partial}{\partial \mu_j} p(x_n|\theta) \quad (10)$$

$$= -\sum_{n=1}^N \frac{1}{p(x_n|\theta)} \frac{\partial}{\partial \mu_j} \sum_{i=1}^M p(x_n|\mu_i, \sigma_i^2) \alpha_i \quad (11)$$

$$= -\sum_{n=1}^N \frac{1}{p(x_n|\theta)} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_n - \mu_j)^2}{2\sigma_j^2}\right) \frac{-2(x_n - \mu_j)}{2\sigma_j^2} (-1) \alpha_j \quad (12)$$

$$= -\sum_{n=1}^N \frac{1}{p(x_n|\theta)} p(x_n|\mu_j, \sigma_j^2) \frac{(x_n - \mu_j)}{\sigma_j^2} \alpha_j \quad (13)$$

$$= -\sum_{n=1}^N P(j|x_n, \theta) \frac{(x_n - \mu_j)}{\sigma_j^2} \quad (14)$$

Here I used (3) to get from equation (11) to (12) and calculated the respective partial derivative. Also note, that I used equation (7) to get from equation (12) to (13).

To find the *maximum* log-likelihood we set the derivative equal zero:

$$0 \stackrel{!}{=} \frac{\partial}{\partial \mu_j} E \quad (15)$$

$$0 = -\sum_{n=1}^N P(j|x_n, \theta) \frac{(x_n - \mu_j)}{\sigma_j^2} \quad (16)$$

$$0 = -\sum_{n=1}^N P(j|x_n, \theta) x_n + \mu_j \sum_{n=1}^N P(j|x_n, \theta) \quad (17)$$

$$\mu_j = \frac{\sum_{n=1}^N P(j|x_n, \theta) x_n}{\sum_{n=1}^N P(j|x_n, \theta)} \quad (18)$$

Similarly we can derive the Maximum Likelihood estimates for the parameters σ_j^2 and α_j . Those are actually derived explicitly also in the tutorial of Bilmes [?]