The EM algorithm

The EM (Expectation Maximization) algorithm is a general technique for computing the parameters of unknown probability distributions. In some cases the results are similar to variants of k-means, and can be used for clustering data. Perhaps the most well known application is using the EM to compute the parameters of a Gaussian Mixture Model (GMM).

Gaussian density/distribution

The Gaussian density function of dimension D vectors is:

$$g(x;\mu,\Sigma) = \frac{1}{(\sqrt{2\pi})^D |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

The parameters μ, Σ are the mean and the covariance matrix and can be estimated from the data by:

$$\mu = \frac{\sum_{i=1}^m x_i}{m}, \quad \Sigma = \frac{\sum_i (x_i - u)(x_i - u)^T}{m-1}$$

Mixture of k Gaussians is the following density function:

$$f(x,\Phi) = \sum_{j=1}^k p_j g(x; u_j, \Sigma_j), \qquad \sum_{j=1}^k p_j = 1$$

Here the parameters to be estimated are where $\Phi = \{p_1, u_1, \Sigma_1, p_2, u_2, \Sigma_2, \ldots, p_k, u_k, \Sigma_k\}$.

The EM algorithm can be used to find Φ by the following iteration. Start with a guess of Φ.

Keeping Φ unchanged compute $c(i,j)$ by:

$$c(i,j) = \frac{p_j g(x_i; u_j, \Sigma_j)}{\sum_{t=1}^k p_t g(x_i; u_t, \Sigma_t)} \tag{6}$$

Here we introduce $c(i,j)$ as the likelihood that x_i belongs to class j.
Keeping $c(i,j)$ unchanged, the values of u_j, Σ_j, p_j are computed by:

$$\begin{aligned} u_j &= \frac{\displaystyle\sum_i c(i,j) x_i}{\displaystyle\sum_i c(i,j)} \\ \Sigma_j &= \frac{\sum_i c(i,j)(x_i - u_j)(x_i - u_j)^T}{\sum_i c(i,j)} \\ p_j &= \frac{1}{m} \sum_{i=1}^m c(i,j) \end{aligned} \tag{7}$$

To obtain clustering from the output of the EM:

$$c(i) = \arg\max_j c(i,j)$$