The Expectation Maximation (EM) Algorithm

Peter Johansen

April 4, 2001

Given a set of N data $\{x_i\}$, $0 < i \leq N$ from R^n. We want to represent the data by a distribution $p(x)$ which is a *mixture* of k Gaussians with mean μ_j, and with *mixing parameters* $P(j)$. We can think of $P(j)$ as the prior probability that a new datum drawn from the distribution belongs to class c_j.

$$p(x) = \sum_j p(x|j)P(j) \tag{1}$$

The j'th Gaussian has the expression

$$p(x|j) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\{-\frac{(x-\mu_j)^t(x-\mu_j)}{2\sigma^2}\} \tag{2}$$

Thus, the covariance matrix is σ^2 multiplied by the identity matrix.

We want to maximize the likelihood that exactly the given x_i are being drawn. We maximize by varying the center position μ_j. We assume x_i are independent. Hence, the probability to draw exactly those is

$$\prod_x \sum_j p(x|j)P(j)$$

Let us maximize the logarithm, this will give the same result, but let us first show the Gaussian differentiated with respect to its mean value.

$$\begin{aligned}
\frac{dp(x|j)}{d\mu_j} &= \frac{d}{d\mu_j}\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\{-\frac{(x-\mu_j)^t(x-\mu_j)}{2\sigma^2}\} && (3)\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\{-\frac{(x-\mu_j)^t(x-\mu_j)}{2\sigma^2}\}\frac{x-\mu_j}{\sigma^2} && (4)\\
&= p(x|j)\,\frac{x-\mu_j}{\sigma^2} && (5)
\end{aligned}$$

To maximize

$$\sum_x \log(p(x)) \quad = \quad \sum_x \log(\sum_j p(x|j)P(j)) \tag{6}$$

we differentiate with respect to μ_j and set equal to zero.

$$\begin{aligned}
0 &= \frac{d}{d\,\mu_j} \sum_x \log(p(x)) && (7)\\
&= \sum_x \frac{1}{p(x)} \frac{d}{d\,\mu_j} p(x|j)P(j) && (8)\\
&= \sum_x \frac{1}{p(x)} p(x|j)\, \frac{x-\mu_j}{\sigma^2} P(j) && (9)\\
&= \sum_x P(j|x)\, \frac{x-\mu_j}{\sigma^2} && (10)
\end{aligned}$$

where

$$P(j|x) \quad = \quad \frac{p(x|j)P(j)}{p(x)} \qquad\qquad (\text{Bayes}'\,\text{theorem})$$

is the posterior probability of class c_j.

We get this expression for μ_j that maximizes the likelihood of the feature vectors x_i.

$$\mu_j \quad = \quad \frac{\sum_x P(j|x)x}{\sum_x P(j|x)} \tag{11}$$

Notice that σ is absent. The position of the centers are independent of its value.

The EM Algorithm

At entry to the algorithm we know x_i, $i = 1, \ldots, N$, prior probabilities $P(j)$, $j = 1, \ldots, k$ for the clusters, and center positions μ_j, $j = 1, \ldots, k$. The distribution $p(x|j) = N(\mu_j, \sigma^2)$. The algorithm repeats the steps below.

1. Compute the probability density distribution by summing the Gaussians for the k clusters weighted by their estimated probabilities.

$$p(x) = \sum_j p(x|j)P(j) \tag{12}$$

2. Compute for each datum x_i the probability that it belongs to cluster j.

$$P(j|x) = \frac{p(x|j)P(j)}{p(x)}$$

3. (Estimation step) Compute the estimated probability for class c_j.

$$P(j) = \frac{1}{N}\sum_x P(j|x)$$

4. (Maximation step) Compute the new center of each class thereby maximizing the likelihood

$$\mu_j = \frac{\sum_x P(j|x)x}{\sum_x P(j|x)} \tag{13}$$

The result of the algorithm is the final set of centers $\{\mu_j\}$, $0 < j \leq k$ and probabilities $P(j)$, $j = 1, \ldots, k$ for the clusters.

Relation to k-means algorithm

The algorithm specializes to the k-means algorithm when the standard variation goes to zero. This is seen by considering the proportion between $P(i|x)$ and $P(j|x)$ as σ goes to 0.

$$\begin{aligned}
\frac{P(i|x)}{P(j|x)} &= \frac{\frac{p(x|i)P(i)}{p(x)}}{\frac{p(x|j)P(j)}{p(x)}} \\
&= \frac{p(x|i)P(i)}{p(x|j)P(j)} \\
&= \frac{P(i)}{P(j)} \frac{\exp\{-\frac{(x-\mu_i)^t(x-\mu_i)}{2\sigma^2}\}}{\exp\{-\frac{(x-\mu_j)^t(x-\mu_j)}{2\sigma^2}\}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{P(i)}{P(j)} \frac{\exp\{-\frac{d_i^2}{2\sigma^2}\}}{\exp\{-\frac{d_j^2}{2\sigma^2}\}} \\
&= \frac{P(i)}{P(j)} \exp\{\frac{d_j^2 - d_i^2}{2\sigma^2}\}
\end{aligned}
$$

If for instance $d_j < d_i$ we note that the expression converges to 0 as σ goes to 0. The consequence is that a feature vector will belong to the class with the closest center.

References

[1] Christopher M. Bishop. *Neural Nets for Pattern Recognition.* Clarendon Press, Oxford, 1995.