Expectation Maximization Algorithm

I. Number of clusters = number of Gaussian distributions K

A Gaussian distribution is given as

$$\mathcal{N}(\vec{x}\,|\,\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{1}{2\sigma}(x-\mu)^2}$$

Important is the so-called expectation value of the log-likelihood function

$$Q(\theta, \theta^{(t-1)}) = \mathbb{E}\left[\underbrace{\sum_i \log p(\vec{x}_i, \vec{z}_i | \vec{\theta})}_{\text{log-likelihood}}\right]$$

iteration number

expectation value

Here $\vec{\Theta}$ are the parameters, which we want to estimate ($\vec{\mu}$ and $\vec{\alpha}$), $\vec{x}_i$ the data, and $\vec{z}_i$ refers to unknown data.

For a Gaussian distribution this quantity Q is rewritten as

$$Q(\vec{\Theta}, \vec{\Theta}^{(t-1)}) =$$

probability that point belongs to cluster k

$$= \sum_i \sum_k r_{ik} \log[q_k] + \sum_i \sum_k r_{ik} \log[p(\vec{x}_i | \vec{\Theta}_k)]$$

with

$$r_{ik} = p(z_i = k | \vec{x}_i, \vec{\Theta}^{(t-1)})$$

II. E – step:

$$r_{ik} = \frac{q_k \, p(\vec{x}_i | \vec{\theta}^{(t-1)})}{\sum_{k'} q_{k'} \, p(\vec{x}_i | \vec{\theta}_{k'}^{(t-1)})}$$

III. M – step:

Look at part of Q, which depends on $\vec{\mu}_k$ and $\vec{\alpha}_k$:

$$\sum_k \sum_i r_{ik} \log\left[p(\vec{x}_i | \vec{\theta}_k) \right] =$$

$$= -\frac{1}{2} \sum_i r_{ik} \left[\log |\vec{\Sigma}_k| + (\vec{x}_i - \vec{\mu}_k)^T \vec{\Sigma}_k^{-1} (\vec{x}_i - \vec{\mu}_k) \right]$$

$$\frac{1}{N} \sum (\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T$$

basically the maximum likelihood estimation of a multivariate Gaussian

This gives the update

$$\vec{\mu}_k = \frac{\sum_i r_{ik} \vec{x}_i}{\sum_i r_{ik}}$$

$$\vec{\Sigma}_k = \frac{\sum_i r_{ik} (\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T}{\sum_i r_{ik}}$$

$$= \frac{\sum r_{ik} \vec{x}_i \vec{x}_i^T}{\sum r_{ik}} - \vec{\mu}_k \vec{\mu}_k^T$$