The Expectation-Maximisation (EM) algorithm

Adahlia Lux

School of Computer Science
University of Birmingham

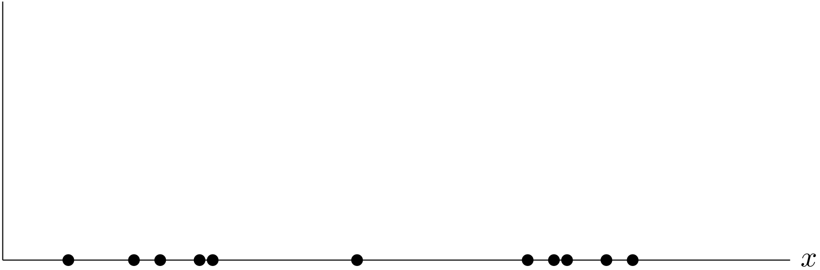
19 November 2015

What is EM?

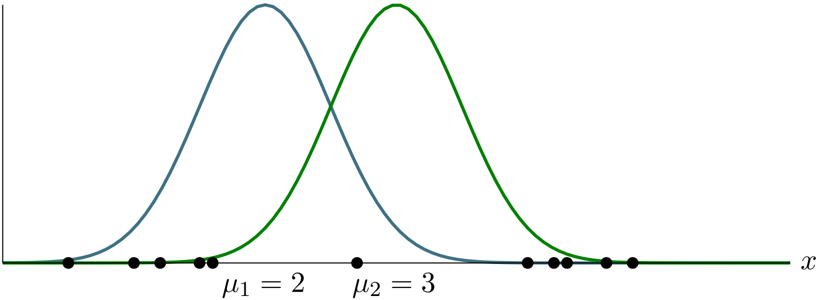
- Iterative way to find maximum likelihood (ML) or maximum a posteriori (MAP) estimates of model parameters θ.
- Useful when the model has *latent variables*.

Example: Mixture of 1-D Gaussians

- Fit k Gaussians to these data? (e.g. $k = 2$)



1. Guess the means μ_1 and μ_2. We are told $\sigma = 0.5$.



2. (E-step) For each point x, calculate posterior for each distribution i.

$$\omega_{ij} \equiv P(j \mid i) = \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-(x_j-\mu_i)^2/2\sigma^2} P(i)}{\sum_{k=1}^{K} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_j-\mu_k)^2/2\sigma^2} P(k)} \tag{1}$$

x_j	**0.5**	**1.0**	**1.2**	**1.5**	**1.6**	**2.7**	**4.0**	**4.2**	**4.3**	**4.6**	**4.8**
ω_{1j}	1	.998	.995	.982	.973	.310	.002	.001	.001	0	0
ω_{2j}	0	.002	.005	.018	.027	.690	.998	.999	.999	1	1

3. (M-step) Re-estimate the means.

$$\mu_i{}' = \frac{\sum_{j=0}^{N} \omega_{ij} x_j}{\sum_{j=0}^{N} \omega_{ij}} \tag{2}$$

$$\mu_1{}' = \frac{0.5*1+1*.998+...+4.8*0}{1+.998+...+0} = \mathbf{1.250}$$

$$\mu_2{}' = \frac{0.5*0+1*.002+...+4.8*1}{0+.002+...+1} = \mathbf{4.152}$$

- Could also estimate variances in the same way.

3. (M-step) Can also re-estimate priors.

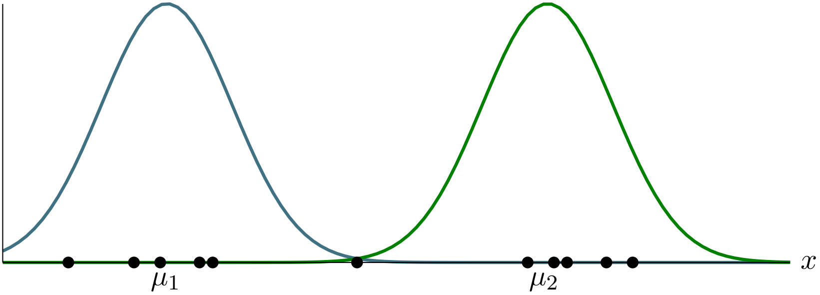
$$P(i)' = \frac{\sum_{j=1}^{N} \omega_{ij}}{N} \tag{3}$$

$$P(i=1)' = \frac{1+.998+...+0}{11} = \mathbf{.478}$$

$$P(i=2)' = \frac{0+.002+...+1}{11} = \mathbf{.522}$$

4. Repeat E-step and M-step until convergence.
   - Check for convergence: $Q(\theta' \mid \theta) < \epsilon$.

- After one iteration:



Deriving EM: lower bound 1

- We want to maximise likelihood of model parameters θ.

$$\begin{aligned}\mathcal{L}(\theta) &\equiv \ln P(x \mid \theta) \\ &= \ln \left[\sum_z P(x \mid z, \theta) P(z \mid \theta) \right]\end{aligned}$$

$$= \ln \left[\sum_z P(z \mid x, \theta') \frac{P(x \mid z, \theta) P(z \mid \theta)}{P(z \mid x, \theta')} \right] - \ln P(x \mid \theta') + \mathcal{L}(\theta')$$

Deriving EM: lower bound 2

- Use Jensen's inequality to get lower bound on $\mathcal{L}(\theta)$:

$$\geq \sum_z P(z \mid x, \theta') \ln \left[\frac{P(x \mid z, \theta) P(z \mid \theta)}{P(z \mid x, \theta') P(x \mid \theta')} \right] + \mathcal{L}(\theta')$$

$$\geq \sum_z P(z \mid x, \theta') \ln \left[\frac{P(x, z \mid \theta)}{P(z, x \mid \theta')} \right] + \mathcal{L}(\theta')$$

Deriving EM: E-step and M-step

$$
\begin{aligned}
\theta'' &= \arg\max_{\theta} \sum_{z} P(z \mid x, \theta') \ln \left[\frac{P(x, z \mid \theta)}{P(z, x \mid \theta')} \right] \\
&= \arg\max_{\theta} \sum_{z} P(z \mid x, \theta') \left[\ln P(x, z \mid \theta) \right] \\
&= \arg\max_{\theta} \sum_{z} P(z \mid x, \theta') \left[\ln P(x \mid z, \theta) P(z \mid \theta) \right]
\end{aligned}
$$

Deriving EM: E-step and M-step

$$\theta'' = \arg\max_{\theta} \sum_{z} P(z \mid x, \theta') \ln \left[\frac{P(x, z \mid \theta)}{P(z, x \mid \theta')} \right]$$

$$= \arg\max_{\theta} \sum_{z} P(z \mid x, \theta') \left[\ln P(x, z \mid \theta) \right]$$

$$= \underbrace{\arg\max_{\theta}}_{\text{M-step}} \underbrace{\sum_{z} P(z \mid x, \theta') \left[\ln P(x \mid z, \theta) P(z \mid \theta) \right]}_{\text{E-step}}$$