General EM Problem

Given

- Observed data $\{\mathbf{x}_d\}_{d \in D}$
- Unobserved data $\{\mathbf{z}_d\}_{d \in D}$ where $\mathbf{z}_d = \langle z_{d1}, \ldots, z_{dM} \rangle$
- Parameterized probability distribution $p(D|h)$ where
  - $D = \{d\}$ is the complete data where $d = \langle \mathbf{x}_d, \mathbf{z}_d \rangle$
  - h comprises the parameters

Determine ML hypothesis h' that (locally) maximizes $\mathbb{E}[\ln p(D|h')]$

General EM Algorithm

Define function $Q(h'|h) = \mathbb{E}[\ln p(D|h')|h, \{\mathbf{x}_d\}_{d \in D}]$ given current parameters h and observed data $\{\mathbf{x}_d\}_{d \in D}$ to estimate the latent variables $\{\mathbf{z}_d\}_{d \in D}$

EM Algorithm. Pick random initial h. Then, iterate

- E Step. Calculate $Q(h'|h)$ using current hypothesis h and observed data $\{\mathbf{x}_d\}_{d \in D}$ to estimate the latent variables $\{\mathbf{z}_d\}_{d \in D}$:
  $Q(h'|h) \leftarrow \mathbb{E}[\ln p(D|h')|h, \{\mathbf{x}_d\}_{d \in D}]$

- M Step. Replace hypothesis h by the hypothesis h' that maximizes this Q function: $h \leftarrow \mathrm{argmax}_{h'}\, Q(h'|h)$

Applying EM to Estimate M Means

Given

- Instances from X generated by mixture of M Gaussians with the same known variance σ^2
- Unknown means $\langle \mu_1, \ldots, \mu_M \rangle$ of the M Gaussians
- Don't know which instance x_d is generated by which Gaussian

Determine maximum likelihood (ML) estimates of $\langle \mu_1, \ldots, \mu_M \rangle$

Consider full description of each instance as $d = \langle x_d, z_{d1}, \ldots, z_{dM} \rangle$ where

- z_{dm} is unobservable and is of value 1 if m-th Gaussian is selected to generate x_d, and 0 otherwise
- x_d is observable

Applying EM to Estimate M Means

$$\begin{aligned}
p(d|h') &= p(x_d, z_{d1}, \ldots, z_{dM}|h') \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\sum_{m=1}^{M} z_{dm}(x_d - \mu'_m)^2\right) \\
\ln p(D|h') &= \ln \prod_{d \in D} p(d|h') = \sum_{d \in D} \ln p(d|h') \\
&= \sum_{d \in D}\left(\ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{m=1}^{M} z_{dm}(x_d - \mu'_m)^2\right) \\
Q(h'|h) &= \mathbb{E}[\ln p(D|h')] \\
&= \mathbb{E}\left[\sum_{d \in D}\left(\ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{m=1}^{M} z_{dm}(x_d - \mu'_m)^2\right)\right] \\
&= \sum_{d \in D}\left(\ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{m=1}^{M} \mathbb{E}[z_{dm}](x_d - \mu'_m)^2\right) \\
\text{where } \mathbb{E}[z_{dm}] &= \frac{\exp(-\frac{1}{2\sigma^2}(x_d - \mu_m)^2)}{\sum_{\ell=1}^{M} \exp(-\frac{1}{2\sigma^2}(x_d - \mu_\ell)^2)}
\end{aligned}$$

Applying EM to Estimate *M* Means

$$\arg\max_{h'} Q(h'|h)$$

$$= \arg\max_{h'} \sum_{d \in D} \left(\ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2} \sum_{m=1}^{M} \mathbb{E}[z_{dm}](x_d - \mu'_m)^2 \right)$$

$$= \arg\max_{h'} \sum_{d \in D} - \sum_{m=1}^{M} \mathbb{E}[z_{dm}](x_d - \mu'_m)^2$$

$$= \arg\min_{h'} \sum_{d \in D} \sum_{m=1}^{M} \mathbb{E}[z_{dm}](x_d - \mu'_m)^2$$

$$\mu'_m \leftarrow \frac{\sum_{d \in D} \mathbb{E}[z_{dm}]\; x_d}{\sum_{d \in D} \mathbb{E}[z_{dm}]}$$

Hold on…

Let $X_D = \{\mathbf{x}_d\}_{d \in D}$ and $Z_D = \{\mathbf{z}_d\}_{d \in D}$. Shouldn't h' be selected to maximize log-likelihood of observed data $\ln p(X_D|h')$ (and marginalize out the unobserved variables Z_D) instead of $Q(h'|h)$?

Proposition. Log-likelihood of observed data monotonically increases with an increasing number of EM iterations.

Proof.

$$
\begin{aligned}
p(X_D|h') &= p(X_D, Z_D|h')/P(Z_D|h', X_D) \\
\ln p(X_D|h') &= \ln p(X_D, Z_D|h') - \ln P(Z_D|h', X_D) \\
\sum_{Z_D} P(Z_D|h, X_D) \ln p(X_D|h') &= \sum_{Z_D} P(Z_D|h, X_D) \ln p(X_D, Z_D|h') - \sum_{Z_D} P(Z_D|h, X_D) \ln P(Z_D|h', X_D) \\
\ln p(X_D|h') &= \underbrace{\sum_{Z_D} P(Z_D|h, X_D) \ln p(X_D, Z_D|h')}_{=Q(h'|h)} \underbrace{- \sum_{Z_D} P(Z_D|h, X_D) \ln P(Z_D|h', X_D)}_{=R(h'|h)} \\
&= Q(h'|h) + R(h'|h) \qquad (1) \\
\ln p(X_D|h) &= Q(h|h) + R(h|h) \qquad \text{by subst. } h' = h \text{ in (1)} \qquad (2) \\
\ln p(X_D|h') - \ln p(X_D|h) &= Q(h'|h) - Q(h|h) + R(h'|h) - R(h|h) \qquad \text{using (1)} - \text{(2)} \\
\ln p(X_D|h') - \ln p(X_D|h) &\geq Q(h'|h) - Q(h|h) \qquad \text{by Gibbs' inequality: } R(h'|h) \geq R(h|h)
\end{aligned}
$$