Expectation Maximization Notes

A general technique for finding ML solutions for latent variable models.

Consider a model where $x$ is observed

$\theta$ are parameters.

The likelihood would be $p(x | \theta)$, but imagine this is very difficult to compute, (i.e., not in any exponential family).

We write the marginal likelihood as:

$$p(x | \theta) = \sum_z p(x, z | \theta)$$

In this model $z$ is our latent variable, and discrete here.

Goal, learn $\theta$ via ML

$$\Theta_{ML} = \arg\max_{\theta} \sum_z p(x, z | \theta)$$

completely intractable since this is a summation over exponential $z$ of terms.

Derive a bound on the likelihood

We use Jensen's inequality where $f(\cdot)$ is some convex function

$$f(E[x]) \leq E[f(x)]$$

If $f(\cdot)$ is concave, the inequality is reversed:

$$f(E[x]) \geq E[f(x)]$$

Let $f(x)$ be the $\log(x)$ where $x$ is some probability value then:

$$\log E(x) \geq E[\log(x)]$$
Jensen's inequality continued

\[ \log E[x] \geq E[\log(x)] \]

\[ \Rightarrow \text{ concave function} \]

\[ \log p(x|\theta) = \log \sum_z \frac{p(x, z|\theta)}{q(z)} \]

\[ = \log \sum_z q(z) \frac{p(x, z|\theta)}{q(z)} \]

\[ \Rightarrow q(z) \]

\[ = \log E_q \left[ \frac{p(x, z|\theta)}{q(z)} \right] \]

\[ \text{use Jensen's inequality} \]

\[ \log p(x|\theta) = \log E_q \left[ \frac{p(x, z|\theta)}{q(z)} \right] \]

\[ \log E_q \left[ \frac{p(x, z|\theta)}{q(z)} \right] \geq E_q \left[ \log \frac{p(x, z|\theta)}{q(z)} \right] \]

\[ = E_q \left[ \log p(x, z|\theta) - \log q(z) \right] \]

\[ = E_q \left[ \log p(x, z|\theta) \right] - E \left[ \log q(z) \right] \]

\[ = E_q \left[ \log p(x, z|\theta) \right] + H(q) \]

\text{Key idea: we can improve these terms via EM!}

\text{E-step}

\text{D) Choose} \quad q(z^{(i+1)}) = p(z|x, \theta^i)
We can show that: \[ \ln p(x \mid \theta) = \sum_z g(z) \ln \left( \frac{p(z \mid x \mid \theta)}{q(z)} \right) - \sum_z g(z) \ln \left( \frac{p(z \mid x \mid \theta)}{q(z)} \right) \]

Plugging in \( \ln p(z, x \mid \theta) = \ln p(z \mid x, \theta) + \ln p(x \mid \theta) \)

\[ \ln p(x \mid \theta) = \sum_z g(z) \ln p(z \mid x \mid \theta) - \sum_z g(z) \ln q(z) - \sum_z g(z) \ln p(z \mid x, \theta) + \sum_z g(z) \ln q(z) \]

\[ = \sum_z g(z) \ln p(z \mid x, \theta) + \sum_z g(z) \ln p(x \mid \theta) - \sum_z g(z) \ln p(z \mid x, \theta) \]

\[ = \sum_z g(z) \ln p(x \mid \theta) = \ln p(x \mid \theta) \]

It turns out that increasing this lower bound results in minimizing the KL divergence between our true posterior \( p(z \mid x, \theta) \) and our variational distribution \( q(z) \).

When \( q(z) = p(z \mid x, \theta) \), the KL divergence vanishes and we reach the maximum of that bound.

Thus in the E-step, we specify our \( q(z) = p(z \mid x, \theta) \).

\[ L(q, \theta) = \sum_z p(z \mid x, \theta^{old}) \ln p(x, z \mid \theta) - \sum_z p(z \mid x, \theta^{old}) \ln p(z \mid x, \theta^{DM}) \]

The M-step takes the variational distribution \( q(z) \) to be fixed and the lower bound is maximized w.r.t to \( \theta \).
**EM Algorithm**

1. Choose some parameters for $\theta$ which we will call $\theta_{old}$

2. Evaluate $p(z|x, \theta_{old})$ (E-step)
   
   a) This results in removing the KL divergence term leaving you with the lower bound $L(\theta, \theta_{old})$.

3. Maximize the $E_{\theta} \left[ \log p(x, z | \theta) \right]$ (M-step)

   expectation of the complete data log likelihood

4. Since the expectation is maximized w.r.t to $\theta_{old}$, the new $\theta$ will increase the original log-likelihood of $\ln p(x | \theta_{old})$

---

**GMM - Gaussian Mixture Model**

**Complete Data Likelihood**

$$p(x_i, z_i | \mu, \Sigma) = \sum_{k} \Pi(z_i = k) \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)$$

**Likelihood Function**

$$p(x | \mu, \Sigma) = \prod_{L=1}^{N} \sum_{K} \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)$$

\( \leftarrow \text{has } K^N \text{ terms} \)
Step 1 for GMM - E-step

Set \( q(z_i) = \frac{p(z_i = k | x_i, \mu, \Sigma)}{p(x_i | \mu, \Sigma)} = \frac{\tau_i N(x_i | \mu_k, \Sigma_k)}{\sum_k \tau_i N(x_i | \mu_k, \Sigma_k)} \)

Compute the new bound

\[
E_q \left[ \log p(x, z | \theta) \right] = \sum_{i=1}^{N} \sum_{k=1}^{K} p(z_i = k | x_i, \mu, \Sigma) \left[ \log \tau_k + \log N(x_i | \mu_k, \Sigma_k) \right]
\]

Step 2 for GMM - M-Step

Maximise w.r.t. \( \mu, \Sigma \)

\[
\mu_k^{\text{new}} = \frac{1}{N_k} \sum_{i=1}^{N} p(z_i = k | x_i, \mu, \Sigma) x_i
\]

\[
\Sigma_k^{\text{new}} = \frac{1}{N_k} \sum_{i=1}^{N} p(z_i = k | x_i, \mu, \Sigma) (x_i - \mu_k^{\text{new}})(x_i - \mu_k^{\text{new}})^T
\]

\[
\tau_k^{\text{new}} = \frac{N_k}{N} \quad \text{where} \quad N_k = \sum_{i=1}^{N} p(z_i = k | x_i, \mu, \Sigma)
\]

Step 3

Check convergence between log likelihood

\[
\ln p(x | \theta) \to \ln p(x | \theta^{(t+1)})
\]