Expectation Maximization Notes

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

Consider a model where: X is observed
θ 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:

$$p(x|\theta) = \sum_{z} \overbrace{p(x, z|\theta)}^{\text{complete data likelihood}}$$

we assume z is our latent variable and discrete here.

Goal, learn θ via ML

$$\theta_{ML} = \underset{\theta}{\operatorname{argmax}} \underbrace{\sum_{z} p(x, z|\theta)}$$

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

Derive a bound on the likelihood

We use Jensen's Inequality where:

$f(\cdot)$ is some convex function

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

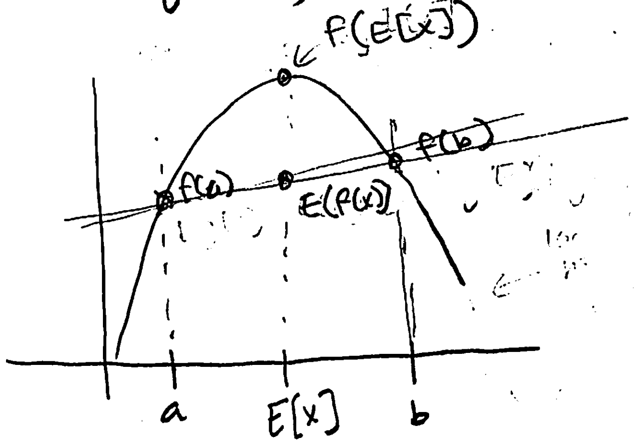
If $f(\cdot)$ is concave, the inequality is reversed:

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

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



$X = \{a, b\}$

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

↑ concave function

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

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

variational distribution → $q(z)$

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

use Jensen's inequality

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

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

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

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

$$= \underbrace{E_q[\log p(x, z|\theta)] + H(q)}$$

Key idea: We can improve these terms via EM!

E-step

1) Choose $q(z^{i+1}) = p(z|x, \theta^i)$

the variational distribution is the posterior over our latent variables z as a function of θ from the previous iteration.

We can show that:

$$\ln p(x|\theta) = \overbrace{\sum_z q(z) \ln \left\{ \frac{p(z,x|\theta)}{q(z)} \right\}}^{\text{Lower bound}} - \overbrace{\sum_z q(z) \ln \left\{ \frac{p(z|x,\theta)}{q(z)} \right\}}^{\text{KL divergence}}$$

Plugging in $\Rightarrow \ln p(z,x|\theta) = \ln p(z|x,\theta) + \ln p(x|\theta)$

$$\ln p(x|\theta) = \sum_z q(\cdot) \ln p(z,x|\theta) - \cancel{\sum_z q(\cdot) \ln q(z)} - \sum_z q(z) \ln p(z|x,\theta) + \cancel{\sum_z q(z) \ln q(z)}$$

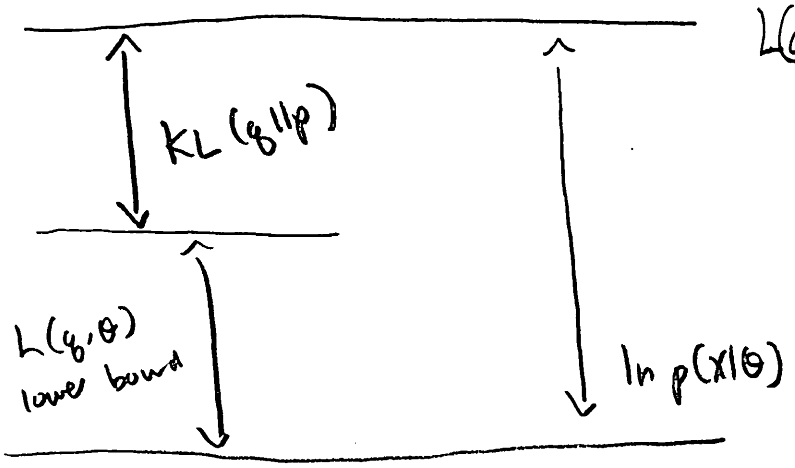
$$= \cancel{\sum_z q(\cdot) \ln p(z|x,\theta)} + \sum_z q(\cdot) \ln p(x|\theta) - \cancel{\sum_z q(z) \ln p(z|x,\theta)}$$

$$= \sum_z q(z) \ln p(x|\theta) = \ln p(x|\theta)$$

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

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

Thus in the E-step, we specify our $q(z) = p(z|x,\theta_{old})$.



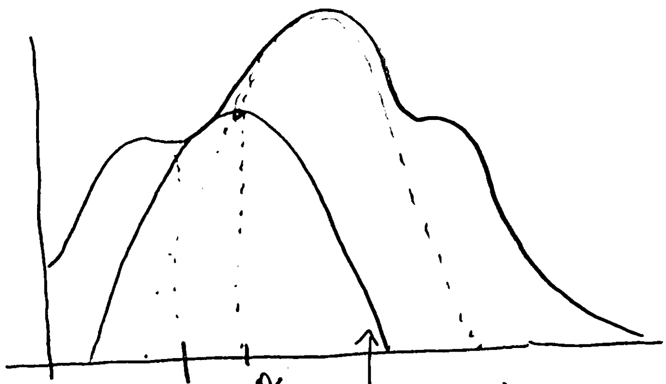
$$L(q,\theta) = \sum_z p(z|X,\theta^{old}) \ln p(X,z|\theta) - \sum_z p(z|X,\theta^{old}) \ln p(z|X,\theta^{old})$$

The M-step takes the variational distribution $q(z)$ to be fixed and the lower bound is maximized w.r.t to θ.

EM Algorithm

1. Choose some parameters for θ which we will call θ_{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 $\mathcal{L}(q,\theta)$.
3. Maximize the $\underbrace{E_q[\log p(X,Z|\theta^*)]}_{\text{expectation of the complete data log likelihood}}$ (M-step)

   a) Since the expectation is maximized w.r.t to θ_{old} the new θ^* will increase the original log-likelihood of $\ln p(X|\theta_{old})$



$\mathcal{L}(\theta,\theta^{old})$ = convex function guaranteed

E-step determines this curve

M-step maximizes this lower bound

GMM - Gaussian Mixture Model

Complete Data Likelihood

$$p(x_i, z_i | \mu, \Sigma) = \sum_K \mathbb{1}(z_i, k)\, \pi_k\, N(x_i | \mu_k, \Sigma_k)$$

$p(z_i = k) \triangleq \pi_k$

Likelihood Function

$$p(x|\mu,\Sigma) = \prod_{i=1}^{N} \sum_K \pi_k\, N(x_i | \mu_k, \Sigma_k)$$

has K^N terms

Step 1 for GMM - E-step

Set $q(z_i) = p(z_i = k \mid x_i, \mu, \Sigma) = \dfrac{p(z_i = k, x_i \mid \mu, \Sigma)}{p(x_i \mid \mu, \Sigma)}$

$$= \frac{\pi_k N(x_i \mid \mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l N(x_i \mid \mu_l, \Sigma_l)}$$

Compute the new bound

$$E_q[\log p(X, Z \mid \theta)] = \sum_{i=1}^{N} \sum_{k=1}^{K} p(z_i = k \mid x_i, \mu, \Sigma)\left[\log \pi_k + \log N(x_i \mid \mu_k, \Sigma_k)\right]$$

Step 2 for GMM - M-Step

Maximize w.r.t μ, Σ

$$\mu_k^{new} = \frac{1}{N_k} \sum_{i=1}^{n} p(z_i = k \mid x_i, \mu, \Sigma)\, x_i$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{i=1}^{n} p(z_i = k \mid x_i, \mu, \Sigma)(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N} \quad \text{where } N_k = \sum_{i=1}^{N} p(z_i = k \mid x_i, \mu, \Sigma)$$

Step 3

Check convergence
btw log likelihood
$\ln p(X \mid \theta^{i}) \to \ln p(X \mid \theta^{(i+1)}) \ldots$