Expectation Maximisation Algorithm

Maximum likelihood from incomplete data.

We can use EM for finding a local maximum of a gaussian mixture:

$$\text{log likelihood} = \ln(p(X|\pi,\mu,\Sigma) = \sum_{n=1}^{N} \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)\right)$$

There is no closed form for calculating the local maxima so we must use an iterative approach.

$$\mu_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})x_n \Sigma_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T \pi_k = \frac{N_k}{N} \text{where } \gamma(z_{nk}) = p(z=1|x_n) \text{ and } N_k = \sum$$

To initialise EM we choose our starting values for mu, sigma and pi as well as the number of gaussians k.

We can then compute responsibilities which is the probabilty that data point n comes from mixture K

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}(\pi_j \mathcal{N}(x_j,\mu_j\Sigma_j)} \qquad = \text{Bayes theorem}$$

We then plug the new responsabilities into the previous formulae to calculate new constants for each gaussian.

General Case:

$$\ln p(X|\theta) = \ln\left\{\sum_z p(X,Z|\theta)\right\} \text{Z are hidden variables (not observed) also called latent variables\{X, Z\} is the}$$

$$\text{Let q(Z) be the distribution over the hidden variables} \ln(p(X|\theta)) = l(q,\theta) + KL(q||p) l(q,\theta) = \sum_Z q(Z) \ln\left(\frac{p(X}{q(}\right.$$

The KL function above is the KL divergence between probablity distributions p and q. Its a measure of how different those distributions are. KL divergence cannot become negative.

if p = q then KL(p,q) = 0.

Our aim to to maximise the l function above

- In the E-step, increase l by updating q
- In the M-step, increase l by updating theta

In the E-step we want to maximise the l function, we can use the decomposition of the log likelihood above to realise that if we minimise KL then l must increase. So we can set p = q.

In the M-step we want to find parameters theta that maximise the L function leaving q fixed. This will increase the log likelihood since KL = 0.

We can repeat these steps to constantly increase the log likelihood.