Lecture Summary 14

In this lecture we will look at mixture models. These are currently among the most popular models being used today. We start with the simplest of all;

$$k(x|\theta) = w\,N(x|\mu_1, \sigma_1^2) + (1-w)\,N(x|\mu_2, \sigma_2^2),$$

where $\theta = (w, \mu_1, \sigma_1, \mu_2, \sigma_2)$. So w is the weight for the two normal components. To sample from this we would take a variable from the first normal component with probability w or from the second normal component with probability $1-w$.

As it stands, the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} \left[w\,N(x_i|\mu_1, \sigma_1^2) + (1-w)\,N(x_i|\mu_2, \sigma_2^2) \right]$$

and this is far from easy to work with directly. The standard approach here is to introduce latent variables $d_i \in \{1,2\}$ which tell us the component, 1 or 2, each x_i comes from.

To use these variables we consider the joint model

$$k(x, d|\theta) = w_d\,N(x|\mu_d, \sigma_d^2)$$

where $w_1 = w$ and $w_2 = 1-w$. We can use this joint model since the marginal for y is correct; i.e.

$$\sum_{d=1}^{2} k(x, d|\theta) = k(x|\theta)$$

and hence

$$\sum_{d_1=1}^{2} \cdots \sum_{d_n=1}^{2} f(\theta, d_1, \ldots, d_n|\text{data}) = f(\theta|\text{data}).$$

Given priors $f(w)$, $f(\mu_1)$, $f(\mu_2)$, $f(\sigma_1)$ and $f(\sigma_2)$ we can then use a Gibbs sampling framework to sample approximately from the posterior; we also need to include the sampling of the (d_i) within this Gibbs sampler. The full conditionals need to be extracted from

$$f(\theta|\text{data}) \propto f(\theta) \prod_{i=1}^{n} w_{d_i}\,N(x_i|\mu_{d_i}, \sigma_{d_i}^2).$$

We will look at how to get these in class and also how we can use priors which lead to simple full conditional densities for each variable.