Mixture Models

John Min

April 2, 2014

Given data $x_1, \dots x_n$, arguably, the simplest model for the data $\mathbf{X}$ is to use a single, unimodal probability distribution:

$$x_i \overset{iid}{\sim} p(x|\theta), i = 1, \dots, n$$

Placing a prior on the unknown θ, we can then calculate the posterior, $p(\theta|x_1, \dots x_n) \propto p(x_1, \dots x_n|\theta)p(\theta) = \prod_{i=1}^{n} p(x_i|\theta)p(\theta)$

Then, we make the predictions: $p(x_{n+1}|x_1, \dots, x_n) = \int_\theta p(x_{n+1}|\theta)p(\theta|x_1, \dots, x_n)\partial\theta$

A **mixture model** adds a layer of complexity to this.

- We have a set of K parameters, $\theta_1, \dots, \theta_K$
- Each observation picks a parameter according to a distribution: $c_i \sim \text{Discrete}(\pi)$, where π is a K-dimensional probability distribution
- Given the parameter, the value for x_i is drawn: $x_i \sim p(x|\theta_{c_i})$.

$G = \sum_{i=1}^{K} \pi_i \partial_{\theta_i} \rightarrow$ G is a probability measure.

δ_{θ_i} is a delta measure $\rightarrow \delta_{\theta_i}(\theta) = \mathbb{1}[\theta = \theta_i]$. Therefore, $G(\theta_j) = \sum_{i=1}^{K} \pi_i \delta_{\theta_i}(\theta_j) = \pi_j$ (assuming all θ_i are unique).

Dirichlet Distribution

The **Dirichlet distribution** is defined on the probability simplex. That is, if π $\sim$Dir$(\gamma_1, \dots, \gamma_K)$, then $\pi \in \Delta_K$, which means:

1. π is a K-dimensional vector
2. $\pi \geq 0$
3. $\sum_{i=1}^{K} pi_i = 1$.

Let's reparameterize using scalar α and probability vector g_0: $\alpha = \sum_{i=1}^{n} \gamma_i, g_{0_i} = \frac{\gamma_i}{\sum_{j=1}^{k} \gamma_j}$