Finite Mixture Model

Introduction

Imagine we randomly choose a distribution to sample from. Let k denote which distribution we choose.

- Distribution 1
  - Parameter Ω_1
  - Probability of choosing distribution 1 is π_1
  - $x|k = 1, \Omega_1 \ \sim \Pr(x|k = 1, \Omega_1)$
- Distribution 2
  - Parameter Ω_2
  - Probability of choosing distribution 2 is π_2
  - $x|k = 2, \Omega_2 \ \sim \Pr(x|k = 2, \Omega_2)$

$\vdots$

- Distribution C
  - Parameter Ω_C
  - Probability of choosing distribution C is π_C
  - $x|k = C, \Omega_C \ \sim \Pr(x|k = C, \Omega_C)$

In general we have data $\mathbf{x} = \{x_1, x_2, ...x_n\}$. Suppose we know which distribution each sample was sampled from $\mathbf{k} = \{k_1, k_2, ...k_n\}$. We know the parameters of each distribution $\boldsymbol{\Omega} = \{\Omega_1, \Omega_2, ...\Omega_C\}$. Finally we know the probability of choosing each distribution $\boldsymbol{\pi} = \{\pi_1, \pi_2, ...\pi_C\}$. We can represent each sample as:

$$x_i|k_i, \boldsymbol{\Omega} \ \sim \Pr(x_i|k_i, \boldsymbol{\Omega})$$

Bayes Theorm

$$\Pr(\mathbf{k}, \boldsymbol{\Omega}|\mathbf{x}) \propto \Pr(\mathbf{x}|\mathbf{k}, \boldsymbol{\Omega})\Pr(\mathbf{k}, \boldsymbol{\Omega}, \boldsymbol{\pi})$$

Prior

We often times assume independence between the some of the parameters

$$\Pr(\mathbf{k}, \boldsymbol{\Omega}, \boldsymbol{\pi}) = \Pr(\mathbf{k}|\boldsymbol{\pi})\Pr(\boldsymbol{\pi})\Pr(\boldsymbol{\Omega})$$

$$\begin{aligned}\Pr(\mathbf{k}, \boldsymbol{\Omega}, \boldsymbol{\pi}) &= \Pr(\mathbf{k}, \boldsymbol{\pi})\Pr(\boldsymbol{\Omega}) \\ &\propto \Pr(\mathbf{k}|\boldsymbol{\pi})\Pr(\boldsymbol{\pi})\Pr(\boldsymbol{\Omega})\end{aligned}$$

Prior $\mathbf{Pr}(\boldsymbol{\pi})$

We typically sample $\boldsymbol{\pi}$ from a Dirichlet Distribution. If so we have the following:

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{\alpha}) \mid \boldsymbol{\alpha} \in [0, \infty)^C$$

using the PDF of the Dirichlet Distribution we can get:

$$\Pr(\boldsymbol{\pi}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{j=1}^{C} \pi_j^{\alpha_j - 1}$$

The normalization constant is found using this B function. In this $\Gamma(x)$ denotes the Gamma function.

$$B(\boldsymbol{\alpha}) = \frac{\prod_{j=1}^{C} \Gamma(\alpha_j)}{\Gamma(\sum_{j=1}^{C} \alpha_j)}$$

Prior $\mathbf{Pr}(\mathbf{k}|\boldsymbol{\pi})$

for a single sample k_i the probability is just the associated π

$$\Pr(k_i|\boldsymbol{\pi}) = \pi_{k_i}$$

We can represent in a diffrent way were we product over all the π

$$\Pr(k_i|\boldsymbol{\pi}) = \prod_{j=1}^{C} \pi_j^{\delta_{k_i j}}$$

$\delta_{k_i j}$ Is the Kronecker delta function $\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$. Lets see what this looks like for when $k_i = 2$

$$\begin{aligned} \Pr(k_i = 2|\boldsymbol{\pi}) &= \prod_{j=1}^{C} \pi_j^{\delta_{2j}} \\ &= \pi_1^{\delta_{21}} \pi_2^{\delta_{22}} \pi_2^{\delta_{23}} \dots \pi_C^{\delta_{2C}} \\ &= \pi_1^0 \pi_2^1 \pi_2^0 \dots \pi_C^0 \\ &= \pi_2 \end{aligned}$$

We can see that this product is another way to represent the math. Going back to finding $\Pr(\mathbf{k}|\boldsymbol{\pi})$ we can assume that k_i are independent of each other

$$\begin{aligned} \Pr(\mathbf{k}|\boldsymbol{\pi}) &= \prod_{i=1}^{n} \Pr(k_i|\boldsymbol{\pi}) \\ &= \prod_{i=1}^{n} \prod_{j=1}^{C} \pi_j^{\delta_{k_i j}} \\ &= \prod_{i=1}^{n} \pi_j^{n_j} \;\Bigg|\; n_j \text{ is number of samples in category j} \end{aligned}$$

Prior $\mathbf{Pr}(\boldsymbol{\Omega})$

if we assume independent parameters:

$$\Pr(\boldsymbol{\Omega}) = \prod_{j=1}^{C} \Pr(\Omega_j)$$

Likelihood

We can look at a single sample and do a smilar thing that we did for $\Pr(\mathbf{k}|\boldsymbol{\pi})$

$$\Pr(x_i|k_i, \boldsymbol{\Omega}) = \prod_{j=1}^{C} \Pr(x_i|k_i = j, \Omega_j)^{\delta_{k_i j}}$$

Looking at all samples we get

$$\begin{aligned}
\Pr(\mathbf{x}|\mathbf{k}, \boldsymbol{\Omega}) &= \prod_{i=1}^{n} \Pr(x_i|k_i, \boldsymbol{\Omega}) \\
&= \prod_{i=1}^{n} \prod_{j=1}^{C} \Pr(x_i|k_i = j, \Omega_j)^{\delta_{k_i j}} \\
&= \prod_{j=1}^{C} \prod_{i \big| k_i = j} \Pr(x_i|k_i = j, \Omega_j)
\end{aligned}$$