February 11, 2020

Abstract

1

2 Introduction

This report presents the development of a Bayesian Non parametric library containing methods for density estimation and clustering. In a Bayesian Nonparametric setting we focused on the Dirichlet process (DP) and its extensions, one of the most widely used priors due to its flexibility and computational ease.

3 Dirichlet Process

Formal definition : Let $M > 0$ and G_0 be a probability measure defined on S. A DP with parameters (M, G_0) is a random probability measure G defined on S which assigns probability $G(B)$ to every (measurable) set B such that for each (measurable) finite partition $B_1, ..., B_k$ of S, the joint distribution of the vector $(G(B_1), ..., G(B_k))$ is the Dirichlet distribution with parameters

$$(MG_0(B_1), ..., MG_0(B_k)). \tag{1}$$

The parameter M is called the precision or total mass parameter, G_0 is the centering measure, and the product MG_0 is referred to as the base measure of the DP.

The basic DP model has the form:

$$\begin{aligned} y_i|G &\overset{\text{iid}}{\sim} G, \quad i = 1, \dots, n \\ G &\sim DP(MG_0) \end{aligned}$$

A key property is that the DP is conjugate with respect to i.i.d sampling so that the posterior base distribution is a weighted average of the prior base distribution G_0 and the empirical distribution of the data, with the weighting controlled by M :

$$G|\mathbf{y} \sim DP(MG_0 + \sum_{i=1}^{n} \delta_{y_i}). \tag{2}$$

And the marginal distribution will be the result of the product of the conditionals:

$$p(y_i|y_1, ..., y_{i-1}) = \frac{1}{M+i-1} \sum_{h=1}^{n-1} \delta_{y_h}(y_i) + \frac{M}{M+i-1} G_0(y_i). \tag{3}$$

An important property of the DP is the discrete nature of G. As a discrete random probability measure we can always write G as a weighted sum of point masses. A useful property based on the discrete nature of the process is his stick-breaking representation, i.e. G can be written as:

$$G(\cdot) = \sum_{k=1}^{+\infty} w_k \delta_{m_k}(\cdot) \tag{4}$$

with $m_k \stackrel{\text{iid}}{\sim} G_0$ and the random weights constructed as $w_k = v_k \prod_{l<k}(1-v_l)$ where v_k are independent Be(1,M)random variables.

In many applications in which we are interested in a continuous density estimation this discreteness can represents a limit. It's common choice to use a Dirichlet Process Mixture (DPM) model where the DP random measure is the mixing measure for the parameters of a parametric continuous kernel function.

4 Dirichlet Process Mixture Model

Extending the DP by convolving G with a kernel F, the model will have the form:

$$
\begin{aligned}
y_i|G \sim F_G(y) &= \int F(y,\vartheta)\, G(\mathrm{d}\vartheta), \quad i=1,\dots,n \\
G &\sim DP(MG_0)
\end{aligned}
$$

An equivalent hierarchical model is:

$$
\begin{aligned}
y_i|\vartheta_i &\stackrel{\perp\!\!\!\perp}{\sim} F(\cdot,\vartheta_i), \quad i=1,\dots,n \\
\vartheta_i|G &\stackrel{\text{iid}}{\sim} G, \quad i=1,\dots,n \\
G &\sim DP(MG_0)
\end{aligned}
$$

where the *latent variables* ϑ_i are introduced, one per unit. Since G is discrete, we know that two independent draws ϑ_i and ϑ_j from G can be equal with positive probability. In this way the DPM model induces a probability model on clusters and an object of interest starting from this model is the partitioning induced by the clustering as well as the density estimation.

Considering n data units, each ϑ_i will have one of the k unique values phi_j. An estimation of the number of the unique values is $M\log(n) \ll n$. Calling c_i the *allocation* parameters to the clusters such that $c_i = j$ if $\vartheta_i = \phi_j$ the model can be thought as the limit as K goes to infinity of finite mixture model with K components:

$$
\begin{aligned}
(Y_i|\phi, c_i) &\sim F(\cdot,\phi_{c_i}) \\
(c_i|\mathbf{p}) &\sim \sum_{k=1}^{K} p_k\delta_k(\cdot) \\
\phi_c &\sim G_0 \\
\mathbf{p} &\sim \mathrm{Dir}(M/K,\dots,M/K)
\end{aligned}
$$

where $(p_1, ..., p_K)$ represent the mixing proportions for the classes and each theta is defined by the latent class c and the corresponding parameters ϕ_c.

4.1 Normal Normal-InverseGamma Model

5 Methods

5.1 Neal2

5.2 Neal8

5.3 Blocked Gibbs