Variational Inference for Bayesian Density Regression.

Eric Chuu
Department of Statistics, Texas AM University
ericchuu@tamu.edu

1. Introduction

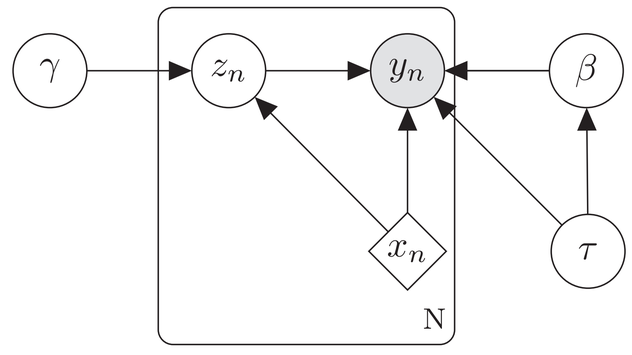
In the Bayesian density regression problem, we observe data $(y_n, x_n), n = 1, \ldots, N$, and the goal is the estimate the conditional density of $y \mid x$. A common approach for doing this is to model the density using a mixture of Gaussians. We extend this idea by allowing the covariates to enter the weights through a logit link function.

$$f(y \mid x) = \sum_{k}^{K} \pi_k(x) \cdot \mathcal{N}\left(y \mid \mu_k(x), \tau_k^{-1}\right)$$

where $\mu_k(x) = x^\intercal \beta_k$ and $\pi_k(x) \propto \exp(x^\intercal \gamma_k)$. While this increases the flexibility of the model, it also increases the computational complexity. In order to perform fast inference on the model parameters, we adopt a variational approach to obtain an approximating distribution to the true posterior. This setup, however, requires an approximation of the softmax function to achieve closed form updates for the coordinate ascent algorithm. We demonstrate the algorithm on both synthetic and real datasets, and we consider extensions of the algorithm via variable selection.

2. Notation

- Data: $\mathbf{y} = \{y_{1:N}\}, \mathbf{X} = \{x_{1:N}\} \subseteq \mathbb{R}^D$
- Coefficient Vector (Gaussian): $\boldsymbol{\beta} = \{\beta_{1:K}\}$
- Precision (Gaussian): $\boldsymbol{\tau} = \{\tau_{1:K}\}$
- Coefficient Vector (weights): $\boldsymbol{\gamma} = \{\gamma_{1:K}\}$
- Cluster Indicator: $\mathbf{Z} = \{z_{1:N}\} \subseteq \mathbb{R}^K$



3. Model Setup

Introducing the cluster indicators, we can simplify the conditional density. In addition, we consider conjugate priors to ease computation.

$$p\left(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{Z}\right) = \prod_{n}\prod_{k} \mathcal{N}\left(y_n \mid x_n^\intercal \beta_k, \tau_k^{-1}\right)^{z_{nk}}$$

$$p\left(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\gamma}\right) = \prod_{n}\prod_{k} \left[\frac{e^{x_n^\intercal \gamma_k}}{\sum_{j=1}^{K} e^{x_n^\intercal \gamma_j}}\right]^{z_{nk}}$$

$$p\left(\boldsymbol{\gamma}\right) = \prod_{k} \mathcal{N}\left(\gamma_k \mid 0, \mathrm{I}_D\right)$$

$$p\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right) = \prod_{k} p\left(\beta_k \mid \tau_k\right) p\left(\tau_k\right)$$

$$p\left(\beta_k \mid \tau_k\right) = \mathcal{N}\left(\beta_k \mid m_0, \left(\tau_k \Lambda_0\right)^{-1}\right)$$

$$p\left(\tau_k\right) = \mathrm{Gamma}\left(\tau_k \mid a_0, b_0\right)$$

6. Variable Selection

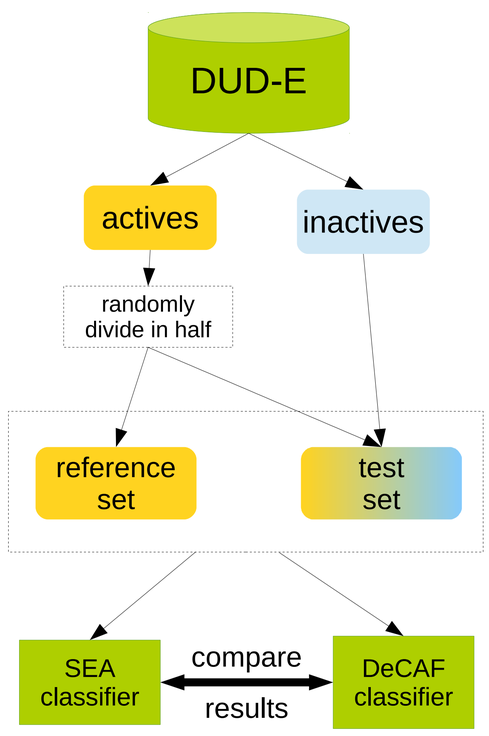
3. Variational Approximation

To measure similarity of two molecules or to combine them into one model, DeCAF first finds their **maximum common substructure (MCS)**. To provide fast, but accurate method for solving MCS problem, we combined Generic Match Algorithm (GMA) [**?**] with backtracking algorithm proposed by Yiqun Cao [**?**].
Here we present comparison of molecules with similar and with different structures. DeCAF scores and **Tanimoto coefficient (Tc)** values are shown in red and black, respectively.

4. Application to Bimodal Conditional Densities

DeCAF is a versatile tool with many possible applications. It allows to compare two molecules or more complex models created from sets of ligands. Our method can be used to align multiple ligands and find crucial pharmacophoric features in a set of active compounds. Pharmacophore models can help in database screening for molecules with desired properties. DeCAF is also suitable for comparing entire sets of ligands, e.g. to analyse properties of proteins in drug repositioning process. DeCAF is a versatile tool with many possible applications. It allows to compare two molecules or more complex models created from sets of ligands. Our method can be used to align multiple ligands and find crucial pharmacophoric features in a set of active compounds. Pharmacophore models can help in database screening for molecules with desired properties. DeCAF is also suitable for comparing entire sets of ligands, e.g. to analyse properties of proteins in drug repositioning process.

5. Application to Speedflow Data



We examine the results of this algorithm applied to both synthetic datasets and a real dataset involving the distribution of speed for different traffic flows on one of California's freeways. We examine the results of this algorithm applied to both synthetic datasets and a real dataset involving the distribution of speed for different traffic flows on one of California's freeways. We examine the results of this algorithm applied to both synthetic datasets and a real dataset involving the distribution of speed for different traffic flows on one of California's freeways. We examine the results of this algorithm applied to both synthetic datasets and a real dataset involving the distribution of speed for different traffic flows on one of California's freeways.