Manifold-valued Dirichlet Processes

Hyunwoo J. Kim, Jia Xu, Baba C. Vemuri, Vikas Singh

Reviewed by Zhao Song

April 22, 2016
Multivariate general linear model (MGLM)

- Given covariates $x_i \in \mathbb{R}^d$ and response variables $y_i \in \mathbb{R}^{d'}$,

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_d x_{id} + \epsilon$$

- Non-linear mapping

$$y_i = \beta_i^0 + \beta_i^1 x_{i1} + \ldots + \beta_i^d x_{id} + \epsilon$$

- A mixture of an infinite number of linear models.
- Overfitting.
- The Dirichlet processes (DP) mixture model

$$(x_i, y_i) | \theta_i \sim F(\theta_i), \theta_i | G \sim G, G \sim DP(G_0, \nu)$$
DP-MGLM

A DP mixture of multivariate general linear models [Hannah et al. 2011]

• Covariates $X$ modelled by a mixture of normal distributions.
• Responses $Y$ modelled by MGLMs conditioned on the covariates.
• The full model:

$$y_i | x_i, \theta_{i,y} \sim \mathcal{N}(\hat{y}_i, \sigma^2_{i,y}), \quad \hat{y}_i = MGLM(x_i, \theta_{i,y})$$

$$x_i | \theta_{i,x} \sim \mathcal{N}(\mu_{i,x}, \sigma^2_{i,x}), \quad \theta_{i,x} = (\mu_{i,x}, \sigma^2_{i,x})$$

$$\theta_i | G \sim G, \quad G \sim DP(G_0, \nu), \quad \theta_i = (\theta_{i,x}, \theta_{i,y})$$

$Y$ is manifold-valued?
Basic notations

- The exponential map: Mapping from the tangent space to the manifold.
  \[ \text{Exp}(y_i, \cdot) : T_{y_i}M \rightarrow M \]

- The logarithm map: Mapping from the manifold to the tangent space.
  \[ \text{Log}(y_i, \cdot) : M \rightarrow T_{y_i}M \]

- \( B \in \mathcal{M} \) is an anchor point.

- \( \{v^j\}_{j=1}^{d} \in T_B \mathcal{M} \) represent tangent vectors.
The model for MGLM on Riemannian manifolds [Kim et al. 2014]:

\[ y = \text{Exp}\left(\text{Exp}\left(B, \sum_{j=1}^{d} v^j x^j\right), \epsilon\right) \]
The model for MGLM on Riemannian manifolds:

\[
y = \text{Exp}(\text{Exp}(B, \sum_{j=1}^{d} v^j x^j), \epsilon)
\]
Geometry of SPD(n)

Consider the DP-MGLM on the space of SPD(n), $n \times n$ symmetric positive definite matrices.

- Tangent space: $\text{Sym}(n)$, $n \times n$ symmetric matrices.
- Geodesic distance:
  \[
  d(B, Y)^2 = \text{tr}(\log^2(B^{-1/2}Y B^{-1/2}))
  \]
- The exponential map:
  \[
  \text{Exp}(B, V) = B^{1/2} \exp(B^{-1/2}V B^{-1/2}) B^{1/2}
  \]
- The logarithm map:
  \[
  \text{Log}(B, V) = B^{1/2} \log(B^{-1/2}Y B^{-1/2}) B^{1/2}
  \]
DP-MGLM on Riemannian manifolds

- Joint distribution in one cluster:

\[ y_i | x_i, \theta_{i,y} \sim N_{\text{SPD}}(\hat{y}_i, \sigma^2_y), \quad \text{where} \quad \hat{y}_i = \text{Exp}(B_i, V_i x_i) \]

\[ x_i | \theta_{i,x} \sim N(\mu_{i,x}, \sigma^2_{i,x}), \quad \text{where} \quad \theta_{i,x} = (\mu_{i,x}, \sigma^2_{i,x}) \]

where the generalized normal distribution is represented as

\[ N_{\text{SPD}}(y; \mu_y, \sigma^2_y) = \frac{1}{Z(\mu_y, \sigma_y)} \exp \left( - \frac{d(y, \mu_y)^2}{2\sigma^2} \right) \]
The base distribution $G_0$ over $\theta = \{\mu_x, \sigma_x^2, B, V\}$:

- $\mu_x | \mu_0, \sigma_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$
- $\log(\sigma_x^2) | M_\sigma, \Sigma_\sigma \sim \mathcal{N}(M_\sigma, \Sigma_\sigma^2)$
- $B | \mu_B, \sigma_B^2 \sim \mathcal{N}_{SPD}(B | \mu_B, \sigma_B^2)$
- $V | \mu_V, B \sim \mathcal{N}_{Sym}(V | \mu_V, B)$

where

$$
\mathcal{N}_{Sym}(V | \mu_V, B) = \frac{1}{Z} \exp \left( -\frac{1}{2} \text{tr} \left[ ((V - \mu_V)B^{-1})^2 \right] \right)
$$
Posterior Sampling

- For $\theta_x = (\mu_x, \sigma_x^2)$: Slice sampling [Neal 2000].
- For $\theta_y = (B, V)$: Hamiltonian Monte Carlo (HMC)

\[
H(q, p) = U(q) + K(p) \\
U(q) = -\log[\pi(q) L(D|q)] \\
K(p) = p^T M^{-1} p / 2
\]

- Transition from $(q, p)$ to $(q^*, p^*)$ with probability
  \[
  \min\{1, \exp (H(q, p) - H(q^*, p^*))\}
  \]
Hamiltonian function for DP-MGLM

- The Hamiltonian function:

\[ H(B, V, \dot{B}, \dot{V}) = U(B, V) + K(\dot{B}, \dot{V}) \]

- The potential function:

\[ U(B, V) := \frac{1}{\sigma^2} E(B, V) - \log f_{\text{SPD}}(B) - \log f_{\text{Sym}}(V) \]

where \( E(B, V) := \frac{1}{2} \sum_i d(y_i, \hat{y}_i)^2 \).

- The kinetic energy:

\[ K(\dot{B}, \dot{V}) := \frac{1}{2} \| \dot{B} \|_B + \frac{1}{2} \sum_{j=1}^d \| \dot{V}^j \|_B \]

with \( \langle U, V \rangle = \text{tr}(B^{-1/2}UB^{-1}VB^{-1/2}) \) for \( U, V \in T_B\mathcal{M} \).
Algorithm 1: HMC algorithm for DP-MGLM on Riemannian manifolds

1: Input: \((B_{\text{cur}}, V_{\text{cur}}) \in \mathcal{M} \times T_B \mathcal{M}^n\), Leapfrog parameters \(e \in \mathbb{R}_{++}, L \in \mathbb{Z}_{++}\)
2: Output: \((B_{\text{next}}, V_{\text{next}}) \in \mathcal{M} \times T_B \mathcal{M}^n\)
3: Sample \((\dot{B}_{\text{cur}}, \dot{V}_{\text{cur}}) \in T_B \mathcal{M} \times T_B \mathcal{M}^n\) from independent normal distribution w.r.t. Riemannian metric.
4: Initialize \((B, V, \dot{B}, \dot{V}) \leftarrow (B_{\text{cur}}, V_{\text{cur}}, \dot{B}_{\text{cur}}, \dot{V}_{\text{cur}})\)
5: \(\dot{B} \leftarrow \dot{B} - \frac{e}{2} \nabla_B U(B, V)\) and \(\dot{V} \leftarrow \dot{V} - \frac{e}{2} \nabla_V U(B, V)\)
6: \(\text{for } i \in \{1, \cdots, L\} \text{ do}\)
7: \(B' \leftarrow B, B \leftarrow \text{Exp}(B, e\dot{B}), V \leftarrow V + e\dot{V}\)
8: \((V', \dot{B}', \dot{V}') \leftarrow (\Gamma_{B' \rightarrow B} V, \Gamma_{B' \rightarrow B} \dot{B}, \Gamma_{B' \rightarrow B} \dot{V})\)
\hspace{1cm} /* Parallel transport */
9: \(\text{if } i != L \text{ then}\)
10: \(\dot{B} \leftarrow \dot{B} - e \nabla_B U(B, V)\) and \(\dot{V} \leftarrow \dot{V} - e \nabla_V U(B, V)\)
11: \(\text{end if}\)
12: \(\text{end for}\)
13: \(\dot{B} \leftarrow \dot{B} - \frac{e}{2} \nabla_B U(B, V)\) and \(\dot{V} \leftarrow \dot{V} - \frac{e}{2} \nabla_V U(B, V)\)
14: Accept \((B, V)\) with probability
\[\min[1, \exp(H(\dot{B}_{\text{cur}}, \dot{V}_{\text{cur}}, B_{\text{cur}}, V_{\text{cur}}) - H(\dot{B}, \dot{V}, B, V))]\]
Experiments

- DP mixtures of MGLM on SPD
Experiments (cont.)

- Estimating Models for Spatially-based Covariates
Experiments (cont.)

- How do facial landmark appearances evolve with age?
