Extrinsic Gaussian Processes for Regression and Classification on Manifolds

김성민

서울대학교
통계학과, 베이즈통계 연구실

2022. 02. 03

CONTENTS

CONTENTS

Material

- Lin et al. (2019) "Extrinsic Gaussian processes for regression and classification on manifolds." *Bayesian Analysis* 14.3 : 887-906.

CONTENTS

Motivation

- Focus on regression and classification on *known manifolds*.
- How to set a *positive semi-definite kernel*?

CONTENTS

Regression on manifolds

- Let M be a smooth manifold where the predictors lie.
- Given data (x_i, y_i) with $x_i \in M$ and $y_i \in \mathbb{R}$

$$y_i = F(x_i) + \epsilon_i,$$

where $F : M \to \mathbb{R}$ is the regression function on M.

Regression on manifolds

- Let $\Pi(F)$ be a prior distribution for F.
- Gaussian process can be viewed as a probability distribution on the space of functions.
- A GP is prior for regression function F.

- Consider the following model:

$$y_i \sim Bernoulli(\pi_i), \quad \pi_i = \Phi(F(x_i)), \quad F(\cdot) \sim GP(0, K_{ext}),$$

where Φ is the standard normal cdf.

CONTENTS

GP on Manifolds

- A stochastic process $w(x)$ indexed by $x \in M$ is a GP on M.
- $w(x)$ is a GP with mean function $\mu(x)$ and covariance kernel $K(\cdot, \cdot)$ if for any $x_1, \ldots, x_n \in M$,

$$(w(x_1), \ldots, w(x_n)) \sim N((\mu(x_1), \ldots, \mu(x_n)), \Sigma),$$
$$\text{where } \Sigma_{ij} = cov(w(x_i), w(x_j)) = K(x_i, x_j).$$

Kernel on Manifolds

- $K : M \times M \to \mathbb{R}$ is a *positive semi-definite kernel* on M.
- For any points $x_1, \dots, x_n$ on M and real numbers $a_1, \dots, a_n$,

$$\sum_i \sum_j a_i a_j K(x_i, x_j) \geq 0.$$

- The fundamental difficulty in imposing a GP prior on a manifold stems from the highly challenging task of constructing a valid covariance kernel $K(\cdot, \cdot)$.

Construct valid covariance kernel

- Let $J : M \to \mathbb{R}^D$ be an embedding of M into some higher dimensional Euclidean space $\mathbb{R}^D$ $(D \geq dim\, M)$.
- Denote the image of embedding as $\tilde{M} = J(M)$.
- Given a positive semi-definite kernel $\tilde{K}$ on $\mathbb{R}^D$, we can define a positive semi-definite kernel on M by

$$K_{ext}(x_1, x_2) = \tilde{K}(J(x_1), J(x_2))$$

- We call the Gaussian process with the $K_{ext}(\cdot, \cdot)$ defined above an extrinsic Gaussian process(eGP).

- Let $||\cdot||$ be the Euclidean norm.
- We define the extrinsic distance on the Manifold M as

$$\rho(x_1, x_2) = ||J(x_1) - J(x_2)||$$

- Squared exponential kernel

$$K_{ext}(x_1, x_2) = \alpha exp(-\beta \rho^2(x_1, x_2))$$

- Matérn covariance kernel

$$K_{ext}(x_1, x_2) = \sigma^2 \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{\sqrt{2\nu}\rho(x_1, x_2)}{\kappa} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu}\rho(x_1, x_2)}{\kappa} \right)$$

Choice of Embedding J

- The embedding J is not unique.
- It is desirable to have an embedding that preserves as much geometry as possible.
- An *equivariant embedding* is one type of embedding that preserves a substantial amount of geometry.

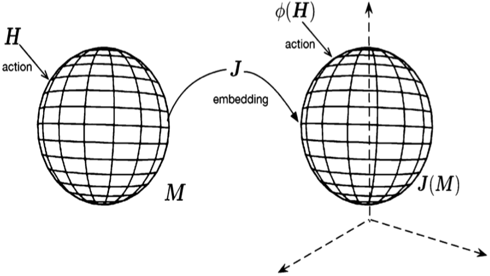

Figure 1: An simple illustration of equivariant embeddings.

CONTENTS

Posterior distribution

- With a valid covariance kernel on M, one can specify an eGP as a prior $\Pi(F)$.

$$y_i = F(x_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

- U is a measurable set in the product space $\mathcal{M} \times (0, \infty)$ with $\mathcal{M}$ denoting the space of all $M \to \mathbb{R}$ regression functions.

$$\Pi(U|(x_1, y_1), \ldots, (x_n, y_n)) = \frac{\int_U \Pi_{i=1}^n N(y_i; F(x_i), \sigma^2)\pi(\sigma^2)\Pi(dF)}{\int \Pi_{i=1}^n N(y_i; F(x_i), \sigma^2)\pi(\sigma^2)\Pi(dF)}$$

CONTENTS

Spheres

$$\mathcal{S}^d = \{x \in \mathbb{R}^{d+1} : ||x|| = 1\}$$

- To construct a valid extrinsic covariance kernel on $\mathcal{S}^d$, note that $\mathcal{S}^d$ is a sub-manifold of $\mathbb{R}^{d+1}$
- Consider the extrinsic squared exponenetial kernel on $\mathcal{S}^d$,

$$K_{ext}(x, x') = \alpha exp(-\beta||J(x) - J(x')||^2) = \alpha exp(-\beta||x - x'||^2)$$

- Compare with intrinsic kernel which is a valid covariance kernel on a sphere,

$$K_{int}(x, x') = \alpha exp(-\beta d(x, x')) = \alpha exp(-2\beta \arcsin\left(\frac{1}{2}||x - x'||\right))$$

Positive definite matrices

- Consider 3×3 positive definite matrices, and we denote it as $SPD(3)$.
- Embed it into the space $Sym(3)$ by log-map,

$$log : SPD(3) \to Sym(3).$$

- Given $A_1, A_2 \in SPD(3)$, their extrinsic distance under the log embedding is given by

$$\rho(A_1, A_2) = ||log(A_1) - log(A_2)||_F$$

CONTENTS

Mean square differentiabillity

- Let $x \in M$ and $v \in T_x M$. Choose a smooth path $\gamma : (-\epsilon, \epsilon) \to M$ such that $\gamma(0) = x$ and $\gamma'(0) = v$. The stochastic process w is **MS differentiable** at x with repect to v if, as $a \to 0$, the random variable $\dfrac{w(\gamma(a)) - w(x)}{a}$ converges to some limit $D_v w$ in mean squares, i.e.

$$\mathbb{E}\left[\left(\frac{w(\gamma(a)) - w(x)}{a} - D_v w\right)^2\right] \to 0.$$

Posterior contraction rates of eGPs

- Let F_0 be the true regression function
- We say the eGP posterior contracts to F_0 at a rate of ϵ_n if

$$\Pi(U_{\epsilon_n}(F_0)^C | (x_1, y_1), \ldots, (x_n, y_n)) \to 0, a.s. \, P_{F_0}^n,$$

where $U_{\epsilon_n}(F_0) = \{F : d_M(F, F_0) > R\epsilon_n\}$, as $n \to \infty$ for some large constant R and distance d_M.

References I

Lin, L., Mu, N., Cheung, P. and Dunson, D. (2019). Extrinsic gaussian processes for regression and classification on manifolds, *Bayesian Analysis* **14**(3): 887–906.