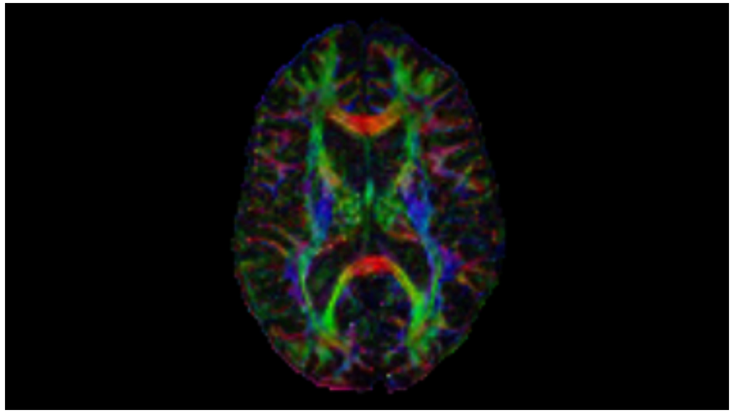
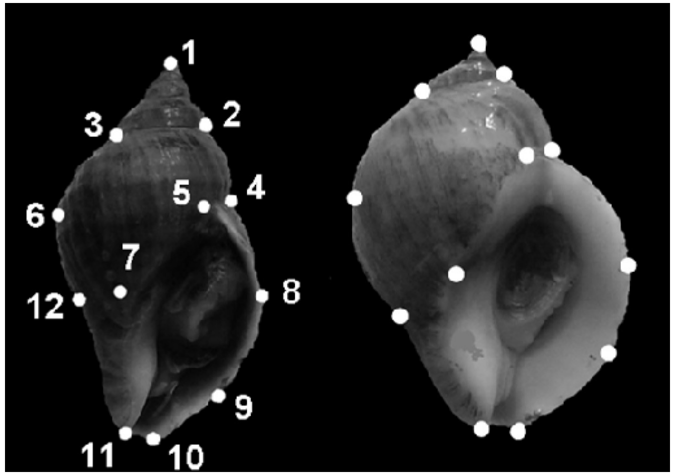
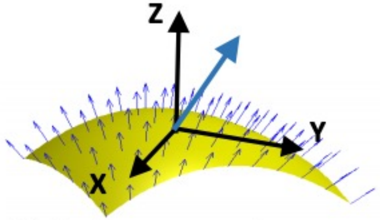
Extrinsic Gaussian Processes for Regression and Classification on Manifolds

Author: Lizhen Lin, Niu Mu, Pokman Cheung, and David Dunson

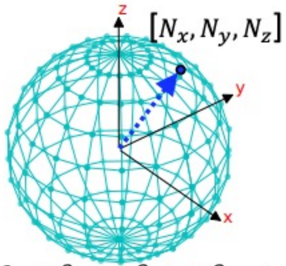
Motivations







(b) Surface normal vector



S^2: $N_x^2 + N_y^2 + N_z^2 = 1$

- DTI
- positive definite matrix

- Machine vision and medical diagnostics
- landmark-based shape space

- Surface
- Curves
- Spheres

Motivations

- Predictors not in Euclidean space but manifolds

Model

- Classification and regression with predictors on known manifolds

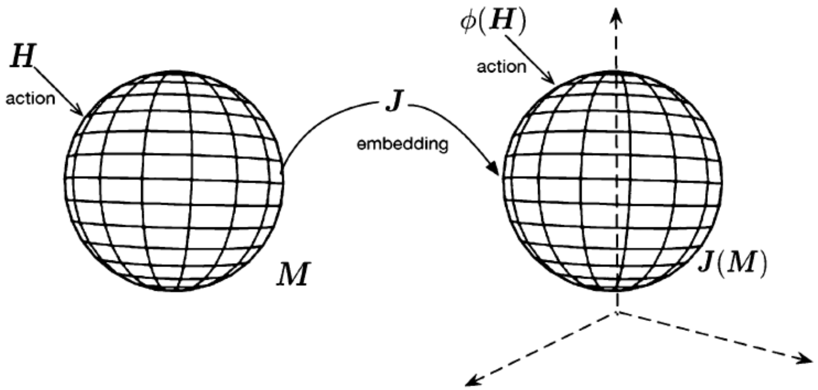
$$y_i = F(x_i) + \epsilon_i, \quad x_i \in M, y_i \in \mathbb{R}, F : M \to \mathbb{R}$$

- Binary/categorical response: classification
- Inference on F , prior $\Pi(F)$, posterior $\Pi(F|D)$
- GP: simple representation, tractability, flexibility for modeling; theoretical properties.
- Euclidean space --> manifolds

GP in manifolds

- $\omega(x)$ is a GP with mean function $\mu(x)$ and covariance kernel $K(\cdot, \cdot)$ for $x_1, \ldots, x_n \in M$
- Challenge: valid $K(\cdot, \cdot)$
- Idea (extrinsic GP):
  - Embed manifolds into Euclidean spaces $J : M \to \mathbb{R}^D, (D \geq dim(M))$
  - Extrinsic kernels on the image manifold $\tilde{M} : K_{ext}(x_1, x_2) = \tilde{K}(J(x_1), J(x_2))$

Embedding $J: M \to \mathbb{R}^D, (D \geq dim(M))$



- Smooth map: differential at each point is an injective map
  - Injective map: from the tangent space of M at x to the tangent space of $\mathbb{R}^D$ at $J(x)$
- Homeomorphism between M and $\tilde{M}$
- Equivariant embedding: preserves a substantial amount of geometry

$$J(hp) = \phi(h)J(p), \quad h \in M, p \in M, \phi : H \to GL(D, \mathbb{R})$$

- Not unique

Posterior distribution

$$\Pi(U \mid (x_1, y_1), \ldots, (x_n, y_n)) = \frac{\int_U \prod_{i=1}^n N(y_i; F(x_i), \sigma^2)\pi_{\sigma^2}\Pi(dF)}{\int \prod_{i=1}^n N(y_i; F(x_i), \sigma^2)\pi_{\sigma^2}\Pi(dF)}, \quad (7)$$

where 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.

For classification, a latent process $\omega(x)$ and link function L

$$F(x) = L(\omega(x))$$

Examples: Spheres

- Inclusion map $J : S^d \to \mathbb{R}^{d+1}$
- Equivariant w.r.t $H = SO(d+1)$ (symmetry)
- Kernel: $K_{\text{ext}}(x, x') = \alpha \exp\left(-\beta \|J(x) - J(x')\|^2\right) = \alpha \exp\left(-\beta \|x - x'\|^2\right).$

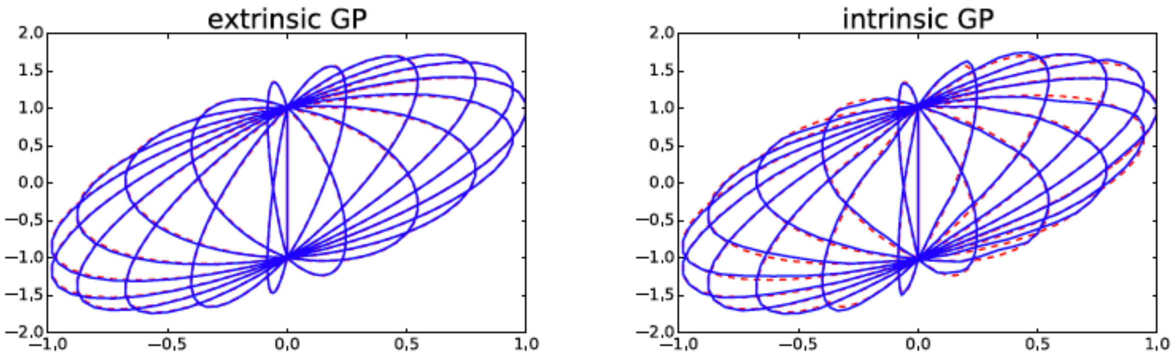


Figure 2: GP predictive results using spherical exponential kernel vs eGP with an extrinsic kernel. Truth is shown in red dashed lines and posterior mean estimates in blue.

Examples: Landmark-based shape space Σ_2^k

- Planar shapes Σ_2^k
- Veronese-Whitney embedding: $J : \Sigma_2^k \to S(k, \mathbb{C})$
- Equivariant w.r.t $H = SU(k) = \{A \in GL(k, \mathbb{C}), AA^* = I, \det(A) = I\},$
- Kernel: $K_{ext}(x_1, x_2) = \alpha \exp(-\beta\rho^2(x_1, x_2)),$

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

Examples: positive definite matrices

- Log-map embedding: $log : SPD(3) \rightarrow Sym(3)$
- Equivariant w.r.t $GL(3, \mathbb{R})$
- kernel $K_{ext}(x_1, x_2) = \alpha \exp(-\beta\rho^2(x_1, x_2)),$

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

Examples: Stiefel manifolds $V_k(\mathbb{R}^m)$ & Grassmann manifolds $Gr_k(\mathbb{R}^m)$

Grassmann manifolds

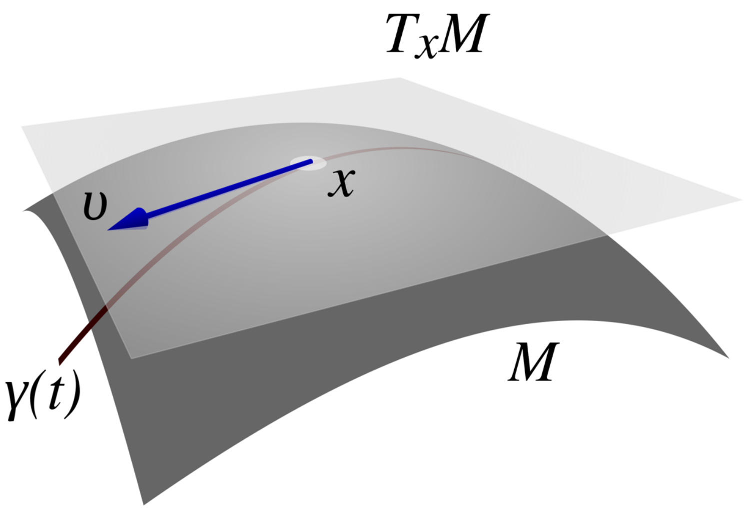
- Embedding: $J : Gr_k(\mathbb{R}^m) \to \mathbb{R}^{m^2}$ $\quad J(\sigma(X)) = XX'$ $\quad \sigma(X) = X \cdot O(k)$
- Equivariant w.r.t. $H = O(m)$
- Kernel: $K_{ext}(x_1, x_2) = \alpha \exp(-\beta \rho^2(x_1, x_2)),$

$$\rho(\sigma(X_1), \sigma(X_2)) = \|X_1X_1' - X_2X_2'\|, \qquad \sigma(X) = X \cdot O(k)$$

Stiefel manifolds

- Inclusion map: $J : V_k(\mathbb{R}^m) \to \mathbb{R}^{m \times k}$
- Equivariant embedding

Properties of eGPs



- Mean square differentiability of stochastic process $\omega(x)$
  - $$\mathbb{E}\left[\left(\frac{w(\gamma(a)) - w(x)}{a} - D_v w\right)^2\right] \to 0.$$
  - Proposition 2: differentiable μ , K is of a class C^2 --> ω MS differentiable at x
  - Proposition 3: MS derivative $D_V\omega$ has mean func. $D_V\mu$, Cov func. $D_{V^{(1)}}D_{V^{(2)}}K$
  - Corollary 1: $\mu \in C^n, K \in C^{2n}$ n-times MS differentiable.
  - Important for interpolation and prediction
  - E.g. pull $\omega \in \mathbb{R}^D$ back to $J^*\omega \in M$ with $(J^*w)(x) = w(J(x)), \quad \text{for } x \in M.$

- Posterior contraction rates adapt to dim of underlying M

$$\epsilon_n = n^{-s/(2s+d)}(\log n)^{d+1}$$

Discussions

- Other manifolds? Space beyond manifolds?
- Unknown manifolds?
- Other criterion for embedding?
- Comparison between extrinsic and intrinsic?
- Manifold-valued response? (Lin et al., 2017)