Estimation of a density living near a manifold

Clément Berenfeld (Dauphine), Paul Rosa (Oxford), Judith Rousseau

September 21, 2022

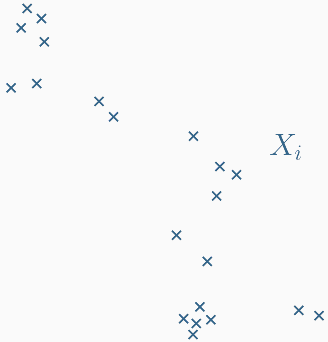
Outline

1. Motivation : low dimensional latent structure
2. Manifold driven anisotropic densities
3. Prior
4. theory
5. Simulations

Motivation : low dimensional latent structure

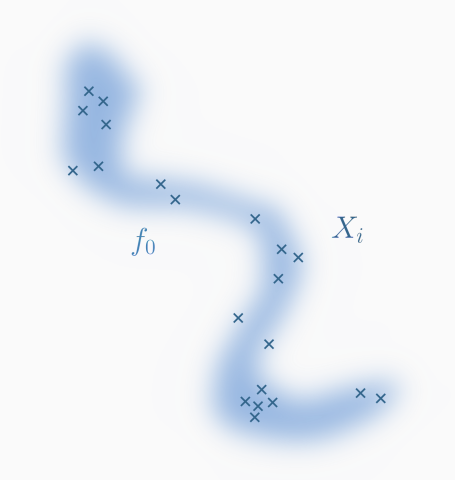
low dimensional latent structures in many examples

- the data lie near an unknown submanifold of $\mathbb{R}^D$.



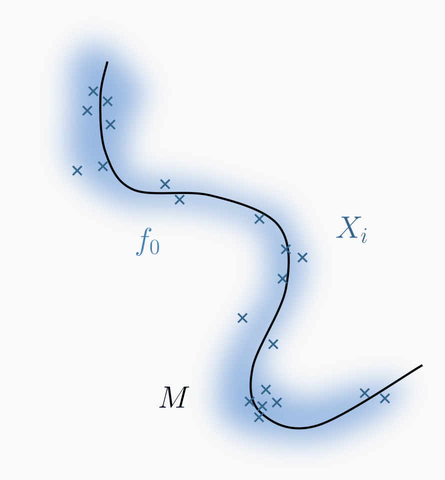
low dimensional latent structures in many examples

- the data lie near an unknown submanifold of $\mathbb{R}^D$.



low dimensional latent structures in many examples

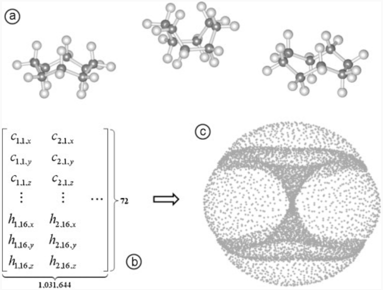
- the data lie near an unknown submanifold of $\mathbb{R}^D$.



Motivation

This is a relevant hypothesis when:

- There is a physical, underlying submanifold;



Cyclooctane ($\mathrm{C_8H_{16}}$) conformations [martin2010topology]

Motivation

This is a relevant hypothesis when:

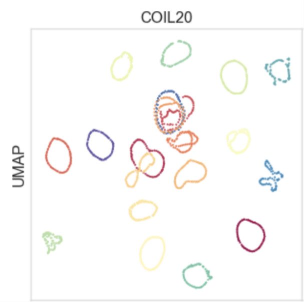
- There is a physical, underlying submanifold;
- The data is high-dimensional, but a few implicit parameters are suspected to drive the dataset.

Cat images from the COIL-20 dataset [**?**].

Motivation

This is a relevant hypothesis when:

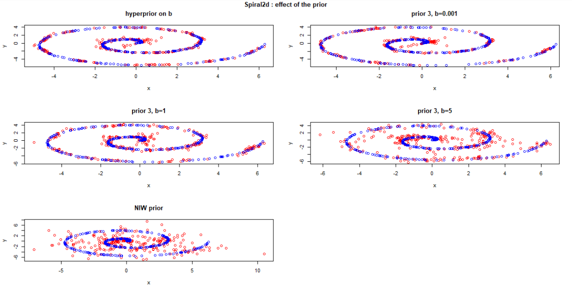
- There is a physical, underlying submanifold;
- The data is high-dimensional, but a few implicit parameters are suspected to drive the dataset.



Low-dimensional representation of the COIL-20 dataset with UMAP [mcinnes2018umap].

Motivation : low dimensional (latent) structure- Toy example

- high dimensional datasets $X_i \in \mathbb{R}^D$: underlying low dimensional structure



Aim : estimate the density of X_i

Dimension reduction

- Lots of methods to do dimension reduction
  - Linear : data belongs to a low dimensional subspace
  - Non linear :
    - data belongs to a low dimensional manifold (Ozakin et al. 09, Berenfeld et al. 21, Divol 21, Tang 22)
    - latent low dimensional manifold : e.g. $X_i = Y_i + E_i,\ Y_i \in M$

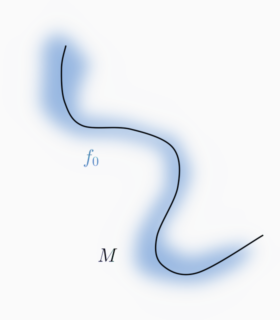
Manifold driven anisotropic densities

Mathematical formalisation

- Data $X_i \in \mathbb{R}^D \overset{iid}{\sim} f$ but

$$X_i \in M^\delta = \{x, d(x,M) \leq \delta\}, \quad M = d - \text{dim. manifold}, \quad d < D$$

- M is unknown, δ is small
- M has reach $\geq \tau > 0$ and $\delta < \tau$.

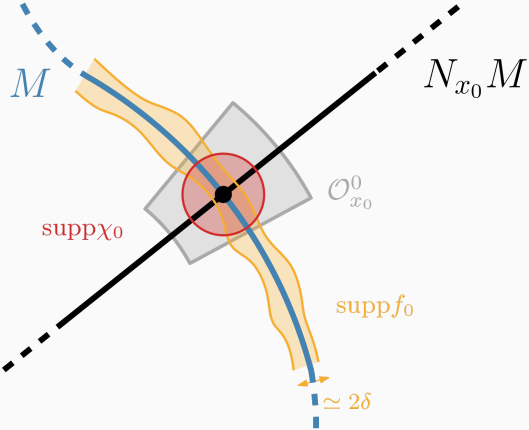


Mathematical formalisation

- Data $X_i \in \mathbb{R}^D \overset{iid}{\sim} f$ but

$$X_i \in M^\delta = \{x, d(x, M) \leq \delta\}, \quad M = d - \text{dim. manifold}, \quad d < D$$

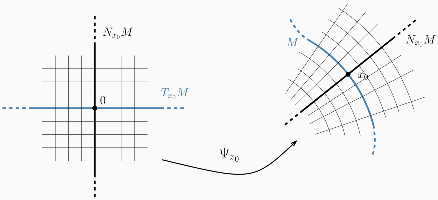
- M is unknown, δ is small
- M has reach $\geq \tau > 0$ and $\delta < \tau$.



Mathematical formalisation II: smoothness definition

- local parametrization :

$$\bar{\Psi}_{x_0} : (v,\eta) \in T_{x_0}M \times N_{x_0}M \to \underbrace{\Psi_{x_0}(v)}_{\in M} + \underbrace{N_{x_0}(v,\eta)}_{\in N_{\Psi(x_0)(v)}M},\ N_{x_0}(v,\delta\eta) = \delta N_{x_0}(v,\eta)$$



$$\bar{f}_{x_0,\delta} : \underbrace{B(0,\tau)}_{\subset T_{x_0}M} \times \underbrace{B(0,1)}_{\subset N_{x_0}M} \to \mathbb{R}^+, \quad \bar{f}_{x_0,\delta}(v,\eta) = f(\underbrace{\bar{\Psi}_{x_0}(v,\delta\eta)}_{\in M^\delta})$$

$$\bar{f}_{x_0,\delta} \in \mathcal{H}(\underline{\beta}, L), \quad \underline{\beta} = (\underbrace{\beta_0,\cdots,\beta_0}_{d}, \underbrace{\beta_\perp,\cdots,\beta_\perp}_{D-d})$$

Examples of manifold-driven Hölder densities

$$X_i = Y_i + \delta\epsilon_i, \quad Y_i \sim f_* \text{ on } M, \quad \epsilon_i \in B(0,1), \quad f_* \in \mathcal{H}_M(\beta_0, L)$$

M is $\beta_M + 1$ Hölder : $\Psi_{x_0} \in \mathcal{H}(\beta_M, C)$.

- Orthogonal noise (eg Wasserman et al. , Nigoyi et al.) : $\epsilon_i | Y_i \sim K(\epsilon)\mu_{N_{Y_i}M}(d\epsilon)$ $K \in \mathcal{H}(\beta_\perp, C)$
- Isotropic noise (eg Wu& Dunson) : $\epsilon_i \sim K(\epsilon)$ ind of Y_i, $K \in \mathcal{H}(\beta_\perp, C)$; $\beta_0 \leq \beta_\perp$.

Examples of manifold-driven Hölder densities

$$X_i = Y_i + \delta\epsilon_i, \quad Y_i \sim f_* \text{ on } M, \quad \epsilon_i \in B(0,1), \quad f_* \in \mathcal{H}_M(\beta_0, L)$$

M is $\beta_M + 1$ Hölder : $\Psi_{x_0} \in \mathcal{H}(\beta_M, C)$.

- Orthogonal noise (eg Wasserman et al. , Nigoyi et al.) : $\epsilon_i | Y_i \sim K(\epsilon)\mu_{N_{Y_i}M}(d\epsilon)$ $K \in \mathcal{H}(\beta_\perp, C)$
- Isotropic noise (eg Wu& Dunson) : $\epsilon_i \sim K(\epsilon)$ ind of Y_i, $K \in \mathcal{H}(\beta_\perp, C)$; $\beta_0 \leq \beta_\perp$.

$$\beta_M - 1 \geq \beta_\perp$$

Prior

Prior on f: Location scale Gaussian mixtures

$$f_P(x) = \int \varphi_\Sigma(x-\mu)dP(\mu,\Sigma)$$

- Choice of P: Crucial
  - if $P \sim DP(A, G_0)$, $G_0 = \mathcal{N} \times IW$ + standard density on $\mathbb{R}^D$
    $\rightarrow$ suboptimal rates (De Blasi et al. , Naulet & R)
    If density near manifold
    empirically bad behaviour (Mukhopadhyay& Dunson)
  - If Location mixture : $P(d\mu, d\Sigma) = \delta_{(\Sigma)}P_1(d\mu)$: Good rates for homogeneous smoothness but not adapted here
  - Hybrid location scale mixture: Good rates for homogeneous smoothness and flexible (Ghosal& VdV 07, Naulet & R)

2 types of NP mixtures

- Partial location - scale mixtures $\Sigma = O^T \Lambda O$, $\Lambda = diag(\lambda_1, \cdots, \lambda_D)$.

$$f_{P,\Lambda}(x) = \int \varphi_{O^T\Lambda O}(x-\mu)dP(\mu, O), \quad P = \sum_{k=1}^{K} p_k \delta_{\mu_k, O_k}\ K \in \mathbb{N} \cup \{\infty\}$$

$\lambda_j \overset{iid}{\sim} G_1$

- Hybrid location scale mixture $P = \sum_{k=1}^{\infty} p_k \delta_{\mu_k, O_k, \Lambda_k}$

$$f_P(x) = \int \varphi_{O^T\Lambda O}(x-\mu)dP(\mu, O, \Lambda)$$

$$[P(d\mu, dO, d\Lambda)|Q] \sim DP(A_1 G_0(d\mu, dO) \times Q^{\otimes D}(d\lambda)),$$

$$Q \sim DP(A_2 G_1(d\Lambda))$$

theory

Theoretical results: posterior contraction rates

$$X_i \overset{iid}{\sim} f^o \quad \text{true density}$$

- For the 2 types of Mixtures with G_1: $\sqrt{\lambda_i} \sim IG(b_1, b_2)$
- if for all $x_0 \in M$ $f_0 \in \mathcal{H}(\underline{\beta}, L)$ $+$

$$f_0(x) \lesssim e^{-c\|x\|^\tau}, \quad \int \left(\frac{L(\bar{\Psi}_{x_0,\delta}(v,\eta))}{f^o_{x_0,\delta}} \right)^m f^o_{x_0,\delta}(v,\eta) dv d\eta < \infty$$

Then

$$\Pi\left(\|f_0 - f_P\|_1 > M\epsilon_n (\log n)^q | X^n\right) = o_{P_0}(1)$$

with

$$\epsilon_n \lesssim n^{-\frac{\beta_0}{2\beta_0 + d + (D-d)\beta_0/\beta_\perp}} \vee \left(\frac{\delta^{-\frac{d+(D-d)\beta_0/\beta_\perp}{1-\beta_0/\beta_\perp}}}{n} \right)$$

- Case $\beta_\perp = +\infty$: $\epsilon_n \lesssim n^{-\frac{\beta_0}{2\beta_0+d}} \vee \frac{1}{n\delta^d}$

Comments

- We recover the same rates as for *linear* anisotropy (homogeneous)
- Adaptive estimation
- Location - scale mixtures : best to use the hybrid version

How does it work ?

- Key step find P_0 st $KL(F_o, f_{P_0}) \lesssim \epsilon_n^2$:

$$K_\Sigma f(x) = \int_{\mathbb{R}^D} \varphi_{\Sigma(\mu)}(x-\mu) f(\mu) d\mu, \quad \Sigma(\mu) = O_\mu^T \Lambda O_\mu$$

$$\Lambda = \begin{pmatrix} \sigma^{\alpha_0} Id_d & 0 \\ 0 & \delta \sigma^{\alpha_\perp} Id_{D-d} \end{pmatrix}, \quad \alpha_\perp = \frac{\beta_\perp}{\beta}, \quad \alpha_0 = \frac{\beta_0}{\beta}$$

O_μ matrix of $z \to (Pr_{T_\mu} z, Pr_{N_\mu} z)$ Then $\sigma \to 0$

$$K_\Sigma f^o = f^o + o(1), \quad \text{and} \quad \exists f_1 : K_\Sigma f_1 = f^o + O(\sigma^\beta)$$

Simulations

Simulations

- Partially location scale mixture

$$f_{P,\Lambda}(x) = \int_{\mu,O} \varphi_{O^T \Lambda O}(x-\mu) dP(\mu, O)$$

- $P \sim DP(A\mathcal{N}(\cdot|\mu_0, \Sigma_0) \otimes p_{ML})$
- $\lambda_j \overset{ind}{\sim} IG(1, b_j)$

Simulations

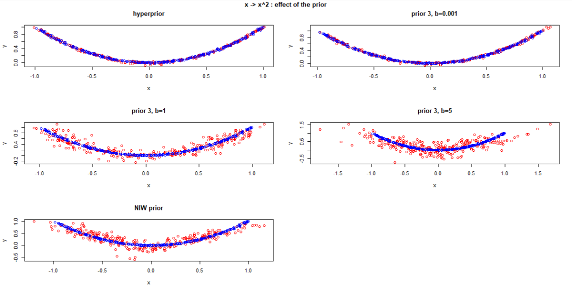
- Partially location scale mixture

$$f_{P,\Lambda}(x) = \int_{\mu,O} \varphi_{O^T \Lambda O}(x-\mu) dP(\mu, O)$$

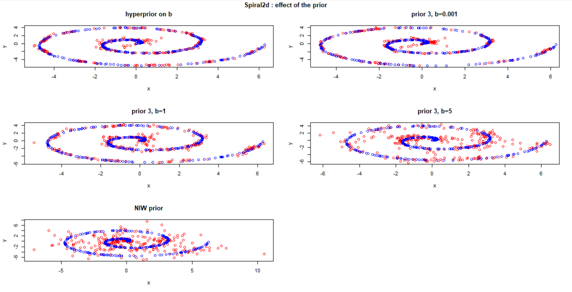
  - $P \sim DP(A\mathcal{N}(\cdot|\mu_0, \Sigma_0) \otimes p_{ML})$
  - $\lambda_j \overset{ind}{\sim} IG(1, b_j)$

- Fixed $b_1 = \cdots = b_D$ or Hierarchical $b_j \overset{iid}{\sim} Exp(1)$

Parabole

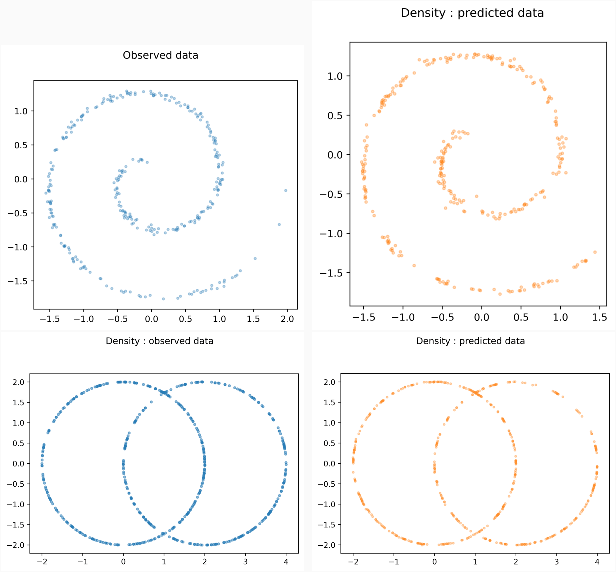

x -> x^2 : effect of the prior
hyperprior
prior 3, b=0.001
prior 3, b=1
prior 3, b=5
NIW prior
x
y

spirale

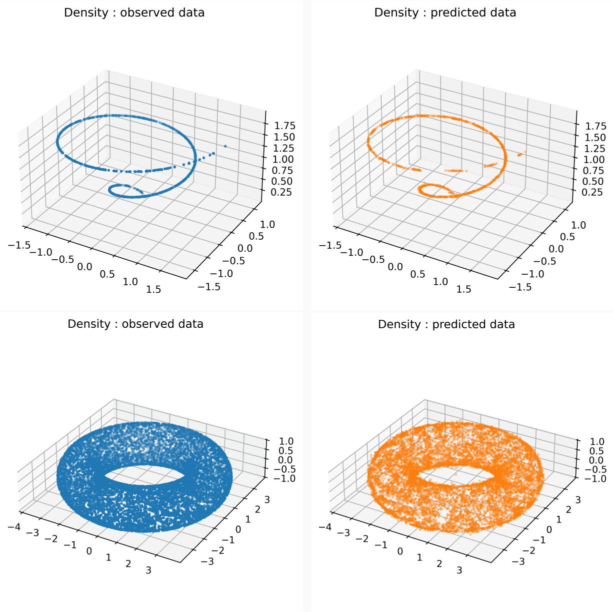


Spiral2d : effect of the prior
hyperprior on b
prior 3, b=0.001
prior 3, b=1
prior 3, b=5
NIW prior
x
y

Simulations



Simulations



Open questions

- Scalable algorithm in dim D large
- Implementation of the hybrid
- Is the rate optimal if δ very small ?

Thank you

European Research Council
Established by the European Commission