Manifold Regression via Brownian Motion

Xu Wang and Gilad Lerman

Department of Mathematics
University of Minnesota

August 19, 2015

Kernel regression

Nadaraya-Watson kernel estimator

Given observations $\{(t_i, x_i)\}_{i=1}^n \in \mathbb{R} \times \mathbb{R}$, learn the function $f_0 : t \in \mathbb{R} \to x \in \mathbb{R}$. The Nadaraya-Watson estimator is

$$
\begin{aligned}
\hat{f}_0(t) &= \frac{\sum_{i=1}^n K_h(t - t_i) x_i}{\sum_{i=1}^n K_h(t - t_i)} \\
&= \frac{\sum_{i=1}^n K_h(t - t_i) f_0(t_i)}{\sum_{i=1}^n K_h(t - t_i)} + \frac{\sum_{i=1}^n K_h(t - t_i)(x_i - f_0(t_i))}{\sum_{i=1}^n K_h(t - t_i)} \\
&\approx (I - h\Delta_{p^2(t)}) f_0 + (error) \\
&= f_0(t)
\end{aligned}
$$

when K_h is the Gaussian kernel.

Problem: One can not add x_i if they lie on a general manifold.

Bayes theorem:

$$\mathbb{P}(func|data) = \frac{\mathbb{P}(data|func) \cdot \mathbb{P}(func)}{\mathbb{P}(data)}$$

Terminology: Prior distribution : $\mathbb{P}(func)$
Posterior distribution : $\mathbb{P}(func|data)$
MAP estimator: $\hat{f} = \mathrm{argmax}_{f \in C([0,1],M)} \mathbb{P}(f|\{(t_i, x_i)\}_{i=1}^n)$

Question: Consistency of this estimator requires the convergence of $\hat{f}$ to f_0.

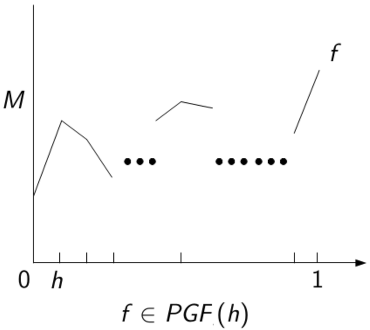
Given a compact manifold M (e.g., $\mathbb{S}^D$, $\mathrm{SO}(3)$, space of shapes) and observations $\{(t_i, x_i)\}_{i=1}^n \in [0,1] \times M$, learn the function $f_0 : [0,1] \to M$.

$$x|t \sim p_{\sigma^2}(f_0(t), x).$$

Here $p_{\sigma^2}(f_0(t), x)$ denotes the heat kernel on M, determining how x_i deviates from $f_0(t_i)$

When $M = \mathbb{R}^D$:

$$p_{\sigma^2}(f_0(t), \mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^D} \exp\left(\frac{-\|\mathbf{x} - f_0(t)\|^2}{2\sigma^2}\right),$$



The class of piecewise geodesic functions $PGF(h) \subset C([0,1], M)$

The prior distribution $\Pi_{c,h}$ with density:

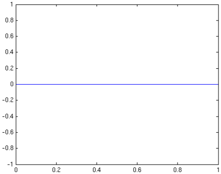
$$\pi_{c,h}(f) = \frac{1}{\mu(M)} \prod_{k=1}^{1/h} p_{ch}(f(kh-h), f(kh)).$$

Bayesian framework

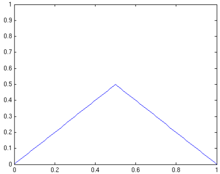
Prior distribution (discrete BM)

Let $c = 1, h = 0.5,\ f \in PGF(h) \subset C([0,1], \mathbb{S}^1)$ is determined by 3 values $f(0), f(0.5), f(1) \in \mathbb{S}^1 \equiv [0, 2\pi]/\{0, 2\pi\}$.

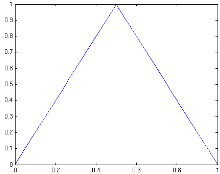
	$f(0)$	$f(0.5)$	$f(1)$	$\pi_{c,h}(f)$
f_1	0	0	0	0.051
f_2	0	0.5	0	0.031
f_3	0	1	0	0.007



f_1



f_2



f_3

The continuous BM prior on $\mathcal{P} \equiv C([0,1], M)$ is the limit of the discrete BM prior. It is the unique probability measure Π such that for any $n \in \mathbb{N}$, $0 = t_0 < \ldots < t_n = 1$, and open subsets $U_0, U_1, \ldots, U_n \in M$, the following identify is satisfied

$$\Pi(f \in \mathcal{P} \mid f(t_0) \in U_0, \ldots, f(t_n) \in U_n) = \int_{U_0} \frac{d\mu(x_0)}{\mu(M)} \int_{U_1 \times \ldots \times U_n} \prod_{i=1}^{n} p_{c(t_i - t_{i-1})}(x_i, x_{i-1}) d\mu(x_i).$$

Given observations $\{(t_i, x_i)\}_{i=1}^{n}$ i.i.d drawn from

$$x|t \sim p_{\sigma^2}(f_0(t), x),\ t \sim p(t)$$

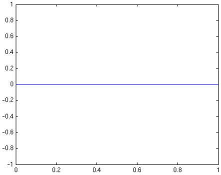
the posterior distribution of Π has the density function

$$\begin{aligned}\Pi(f \in A|\{(t_i, x_i)\}_{i=1}^{n}) &\propto \int_{f \in A} \prod_{i=1}^{n} p(t_i, x_i|f) d\Pi(f) \\ &= \int_{f \in A} \prod_{i=1}^{n} p_{\sigma^2}(f(t_i), x_i) p(t_i) d\Pi(f).\end{aligned}$$

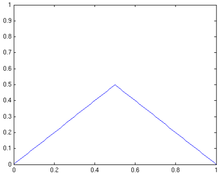
Bayesian framework
Posterior distribution

If $\sigma^2 = 0.3$ and we observe $(0.5, 0.6) \in [0,1] \in \mathbb{S}^1$, then

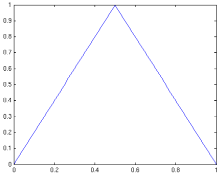
	$f(0)$	$f(0.5)$	$f(1)$	$\pi_{c,h}(f)$	$\pi_{c,h}(f \mid f(0.5) = 0.6)$
f_1	0	0	0	0.051	0.020
f_2	0	0.5	0	0.031	0.022
f_3	0	1	0	0.007	0.004



f_1



f_2



f_3

DBM algorithm

Algorithm 1 Bayesian Manifold regression via discrete BM

Input: Observations $\{(t_i, x_i)\}_{i=1}^n \subset [0,1] \times M$, grid spacing h ($m = 1/h + 1$), scaling constant c, variance σ^2, temperature T for Simulated Annealing (SA).

Output: A function in $PGF(h)$ determined by $\{(t_j^*, x_j^*)\}_{j=1}^m$.

Steps:

- $t_j^* = (j-1)h$, $x_j^{(0)}$ i.i.d. uniformly sampled from M and $f^{(0)} \in PGF(h)$ determined by $\{(t_j^*, x_j^{(0)})\}_{j=1}^m$
- Energy function $E(f) = \pi_{c,h}(f) \prod_{i=1}^n p_{\sigma^2}(f(t_i), y_i) p(t_i)$
- $f^* = SA(E, f^{(0)}, T)$

return The function f^* in $PGF(h)$ determined by $\{(t_j^*, x_j^*)\}_{j=1}^m$.

CBM algorithm

Algorithm 2 Bayesian Manifold regression via continuous BM

Input: Observations $\{(t_i, x_i)\}_{i=1}^n \subset [0,1] \times M$, $\{t_j^*\}_{j=1}^m$, scaling constant c, variance σ^2, temperature T for Simulated Annealing (SA).

Output: Predictions $\{(t_j^*, x_j^*)\}_{j=1}^m$.

Steps:

- Let $\{t_j'\}_{i=1}^{n+m}$ be sorted array of $\{t_i\}_{i=1}^n \cup \{t_j^*\}_{j=1}^m$, $x_j^{(0)}$ i.i.d. uniformly sampled from M and $f^{(0)}$ such that $f^{(0)}(t_j') = x_j^{(0)}$
- $\tilde{\pi}_c(f) = \prod_{i=1}^{n+m} p_{c(t_i'-t_{i-1}')}(f(t_i'), f(t_{i-1}'))$
- Energy function $E(f) = \tilde{\pi}_c(f) \prod_{i=1}^n p_{\sigma^2}(f(t_i), y_i) p(t_i)$
- $f^* = SA(E, f^{(0)}, T)$

return $\{(t_j^*, x_j^*)\}_{j=1}^m$.

The metric topology is defined by

$$d_q(f_1, f_2) = \left(\int_{t \in [0,1]} \mathrm{dist}_M(f_1(t), f_2(t))^q p(t) dt \right)^{1/q}$$

The weak topology

$$N_\epsilon(f_0) = \left\{ f \in \mathcal{P} : \left| \int_{[0,1] \times M} p_g p_f dt d\mu(x) - \int_{[0,1] \times M} p_g p_{f_0} dt d\mu(x) \right| \leq \epsilon, \forall g \in \mathcal{P} \right\}$$

Here $p_f = p_{\sigma^2}(f(t), x) p(t)$.

Theorem

If M is a compact Riemannian manifold and if the true underlying function $f_0 \in \mathcal{P}$ of the regression model is Lipschitz continuous, then the posterior distribution $\Pi(\cdot|\{(t_i, x_i)\}_{i=1}^n)$ is weakly consistent. In other words, for any $\epsilon > 0$,

$$\Pi(N_\epsilon(f_0)|\{(t_i, x_i)\}_{i=1}^n) \longrightarrow 1$$

almost surely w.r.t. the true probability measure P_0^n as $n \to \infty$.

Theorem

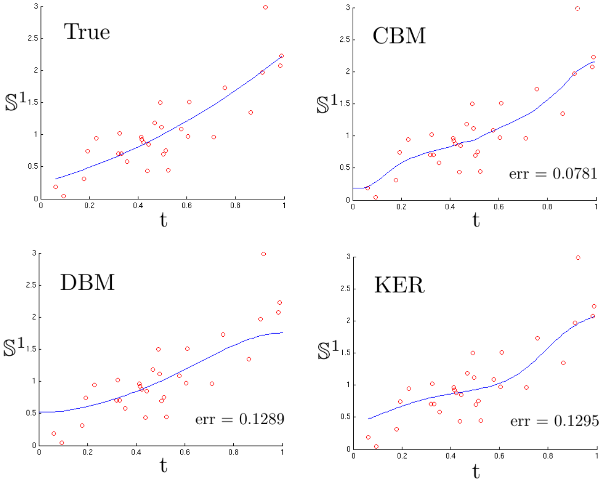
Assume $p(t)$ *is strictly positive on* $[0,1]$*, the true function* $f_0 \in \mathcal{P}$ *is Lipschitz Assume an arbitrarily fixed* $0 < \epsilon < 1/4$ *and for* $n \in \mathbb{N}$*, let* $b_n = n^{-1/2+2\epsilon}$ *be the sidelength of the set* $PGF(b_n)$ *and let* $\{\Pi_n\}_{n\in\mathbb{N}}$ *denote the sequence of discretized BM priors on* $PGF(b_n)$*. Then there exists an absolute constant* A_0 *and a fixed constant* C_0 *depending only on the positive minimum value of* $p(t)$ *on* $[0,1]$*, the volume of* M *and the Riemannian metric of* M *such that* $\Pi_n(\cdot|\{(t_i,x_i)\}_{i=1}^n)$ *contracts to* f_0 *according to the rate* $\epsilon_n = \sqrt{b_n/C_0} = O(n^{-1/4+\epsilon})$*. More precisely,*

$$\Pi_n(f : d_q(f,f_0) \geq A_0\epsilon_n|\{(t_i,x_i)\}_{i=1}^n) \to 0$$

in P_0^n*-probability as* $n \to \infty$*.*

Experiment

Comparison with kernel regression



Experiment

The hyperparameter c

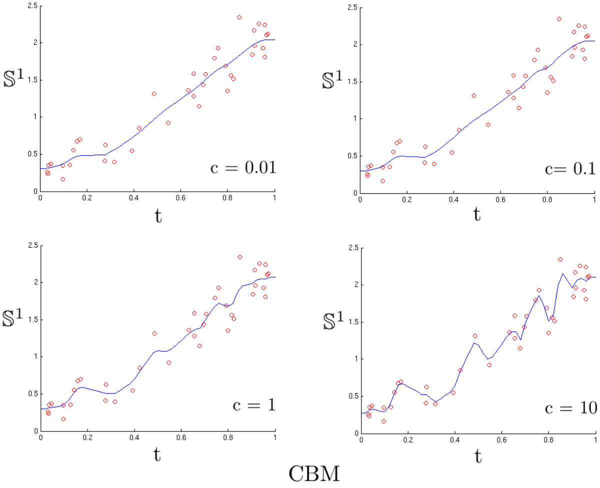


Figure : Continuous BM.

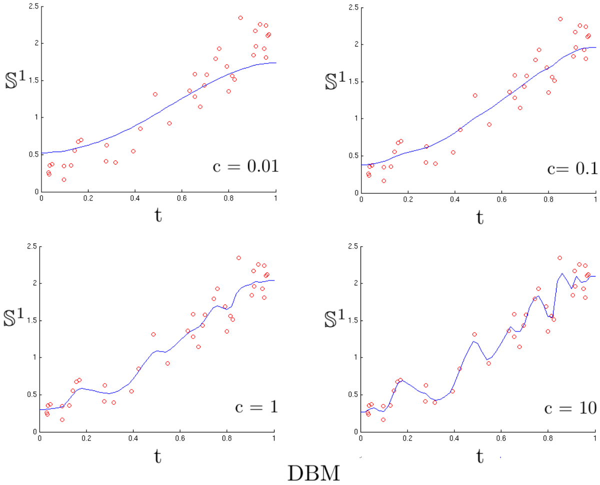


Figure : Discretized BM.

Experiment

The sidelength parameter h

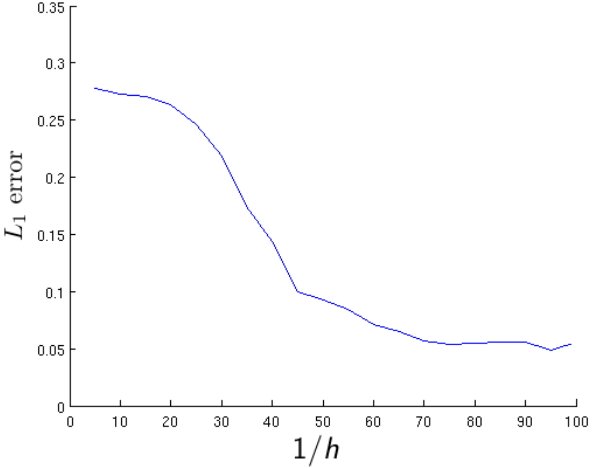


Figure : L_1 error of DBM for different sidelength h.

Thank You !