Sample complexity and effective dimension for regression on manifolds

Andrew D. McRae

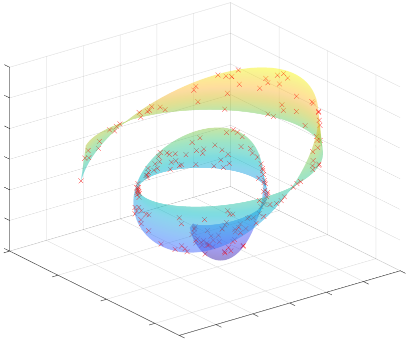
Georgia Tech School of Electrical and Computer Engineering
admcrae@gatech.edu
Joint work with Mark Davenport and Justin Romberg

Bernoulli-IMS One World Symposium

August 24–28, 2020

Learning function on manifold domain

- Common statistics and machine learning model
- m-dimensional (Riemannian) manifold domain $\mathcal{M}$ embedded in $\mathbf{R}^d$ ($d \gg m$)
- Traditional statistics: need $n \gtrsim C^d$ samples to estimate function
- Can we get complexity that scales only with m?

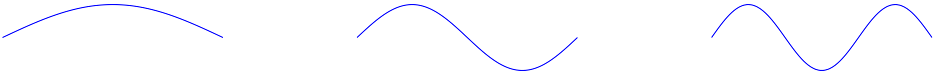


Manifold function spaces

- Analysis tool: spectral decomposition of manifold Laplacian
  - Equivalent to $-\sum_i \frac{\partial^2}{\partial x_i^2}$ in $\mathbf{R}^m$.

$$\Delta_{\mathcal{M}} f = \sum_{\ell=0}^{\infty} \omega_\ell^2 \langle f, v_\ell \rangle_{L_2} v_\ell$$

- ω_ℓ frequency associated with mode v_ℓ
  - $\{v_\ell\}$ orthonormal basis for $L_2(\mathcal{M})$
- Example: Fourier series on interval/circle
- Weyl law: $\#\{\ell : \omega_\ell \leq \Omega\} \sim c_m \, \mathrm{vol}(\mathcal{M}) \, \Omega^m$ as $\Omega \to \infty$
  - Dimension of space of Ω-bandlimited functions



Main result 1: nonasymptotic complexity

Theorem

If $\mathcal{M}$ has curvature bounded by κ, and $\Omega \gtrsim \sqrt{m^3\kappa}$, then

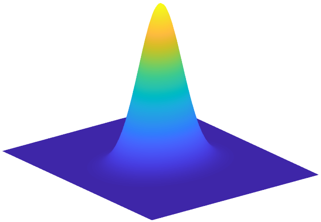
$$\#\{\ell\colon \omega_\ell \leq \Omega\} \leq C_m \operatorname{vol}(\mathcal{M})\Omega^m =: p(\Omega).$$

- First bound with explicit constants
- Hard bound on function space complexity on manifold

Kernel regression estimates

- Consider *heat kernel* $k_t(x,y) = \sum_\ell e^{-\omega_\ell^2 t/2} v_\ell(x) v_\ell(y)$
  - Analogous to Gaussian RBF $(2\pi t)^{-m/2} e^{-\|x-y\|^2/2t}$
  - Also closely approximated by RBF for small t
- Heat kernel reproducing kernel Hilbert space (RKHS) $\mathcal{H}_t$ has norm

$$\|f\|_{\mathcal{H}_t}^2 = \sum_{\ell=0}^{\infty} e^{\omega_\ell^2 t/2} \langle f, v_\ell \rangle_{L_2}^2$$



$k_t(x,y)$ on a section of the sphere

Main result 2: kernel regression error bounds

- Sample $y_i = f^*(x_i) + \xi_i$, $i \in \{1, \ldots, n\}$
- x_i's i.i.d. uniformly on $\mathcal{H}$; ξ_i's i.i.d. subexponential with variance σ^2
- Consider regularized estimate

$$\hat{f} = \arg\min_{f \in \mathcal{H}_t} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \alpha \|f\|_{\mathcal{H}_t}^2$$

Theorem

If $\Omega \gtrsim \sqrt{m^3 \kappa}$, $n \gtrsim p(\Omega) \log p(\Omega)$, *and* $\alpha \approx e^{-\Omega^2 t/2}$, *then*

$$\|\hat{f} - f^*\|_{L_2} \lesssim e^{-\Omega^2 t/4} \|f^*\|_{\mathcal{H}_t} + \sqrt{\frac{p(\Omega)}{n}} \sigma.$$

Discussion

Our bound:

$$\|\hat{f} - f^*\|_{L_2} \lesssim e^{-\Omega^2 t/4} \|f^*\|_{\mathcal{H}_t} + \sqrt{\frac{p(\Omega)}{n}}\sigma \text{ if } n \gtrsim p(\Omega) \log p(\Omega)$$

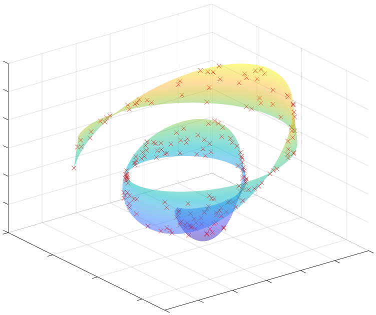
- Optimal choice of Ω gives "almost parametric" rate

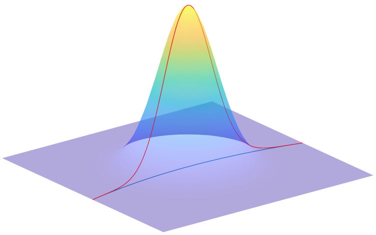
$$\|\hat{f} - f^*\|_{L_2} \lesssim \sqrt{\frac{\log^{m/2} n}{n}}$$

- For Ω such that $e^{-\Omega^2 t/4}$ is very small, $p(\Omega)$ is *effective dimension* of $\mathcal{H}_t$
- Remember $p(\Omega) \approx \Omega^m$

Summary and next steps

- Wanted to show how sample complexity scales with the manifold dimension
- Result 1: non-asymptotic $O(\Omega^m)$ bound on function space complexity on manifolds
- Result 2: effective-dimension based learning theory result for kernel methods
- Future work: can we get similar results with manifold-agnostic methods?
  - For example, Gaussian RBF in Euclidean space as approximation





Wrap-up

- Preprint: `https://arxiv.org/abs/2006.07642`