On robust GP regression

Masha Naslidnyk [1]

[1]University College London

Hello!

- PhD candidate at UCL since November 2021, supervised by F-X Briol.
- previously Amazon, 2015-2021.
- this is upcoming work with `permutation`{Motonobu Kanagawa, Toni Karvonen, Maren Mahsereci}.

Setting

- Regression problem: $\{(x_1, f(x_1)), \ldots, (x_N, f(x_N))\}$
- Fit a GP $f_{\mathrm{GP}} \sim \mathcal{GP}(m, k_\theta)$: choose kernel k_θ, $\theta \in \Theta$.
- Standard: maximum likelihood (ML).
- Problem: can respond poorly to model misspecification!

Replace ML with.. leave-one-out CV

$$\hat{\theta}_{\mathrm{ML}} \in \operatorname*{argmin}_{\theta \in \Theta} \left[-\sum_{n=1}^{N} \log p(f(x_n) \mid x_n, \theta) \right],$$

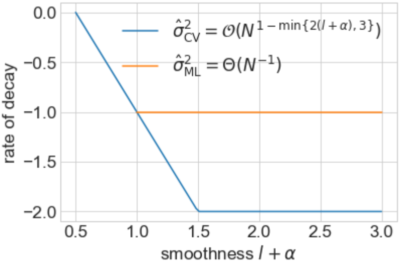
$$\hat{\theta}_{\mathrm{CV}} \in \operatorname*{argmin}_{\theta \in \Theta} \left[-\sum_{n=1}^{N} \log p(f(x_n) \mid x_n, \mathrm{x}_{\setminus n}, f(\mathrm{x}_{\setminus n}), \theta) \right].$$

- existing empirical results: leave-one-out cross-validation better in misspecified settings than ML.

[1] F. Bachoc. (2013). Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification.

Results for Brownian motion

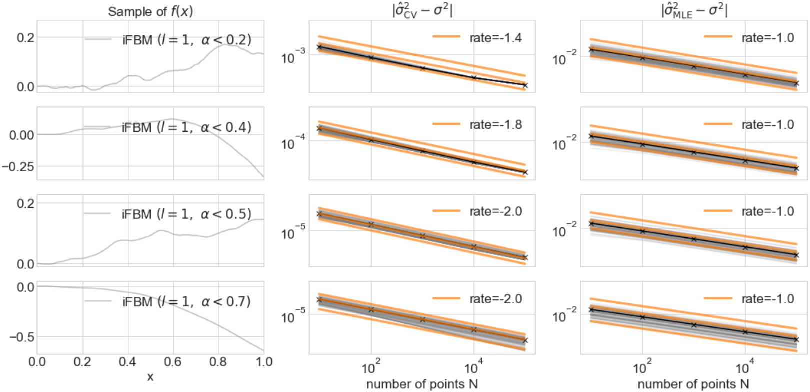
- $k_\sigma(x, x') = \sigma^2 \min(x, x')$. What happens to σ as $N \to \infty$?
- $f \in C^{l,\alpha}([0, T])$: f has continuous derivatives up to order l, and $l'th$ derivative is α-Hoelder continuous: there is a $C > 0$ such that $|f^{(l)}(x) - f^{(l)}(x')| \leq C|x - x'|^\alpha$
- distance between any two neighbouring points x_i, x_{i+1} is $\Theta(1/N)$.
- Then..



[2] T. Karvonen, G. Wynne, F. Tronarp, C. J. Oates, and S. Särkkä. (2020). Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions.

Experiments

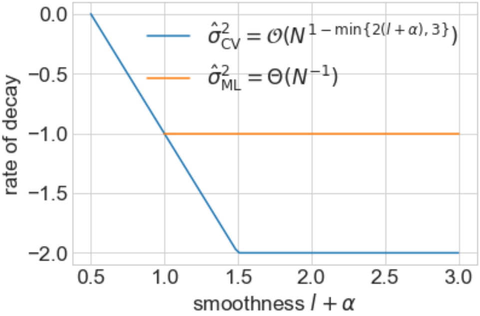
- Is there a corresponding lower bound? Looks like it!



Implications for uncertainty quantification

- Likely more reliable uncertainty quantification!

$$R_{\mathrm{CV}}(x,N) = \frac{|f(x) - m_N(x)|}{\hat{\sigma}_{\mathrm{CV}}\sqrt{k_N(x)}} \quad \text{VS} \quad R_{\mathrm{ML}}(x,N) = \frac{|f(x) - m_N(x)|}{\hat{\sigma}_{\mathrm{ML}}\sqrt{k_N(x)}}.$$

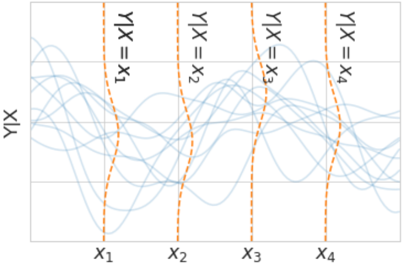


Current work: Replace ML with.. MMD?

- tangential line of work.
- Maximum mean discrepancy:

$$\mathrm{MMD}_k(\mathbb{P}, \mathbb{Q}) = \left\| \int_{\mathcal{X}} k(x, \cdot)\mathbb{P}(\mathrm{d}x) - \int_{\mathcal{X}} k(x, \cdot)\mathbb{Q}(\mathrm{d}x) \right\|_{\mathcal{H}_k}. \quad (1)$$

- how to extend to conditional models—i.e multiple distributions?



[3] F.-X. Briol, A. Barp, A. B. Duncan, and M. Girolami. (2019). Statistical inference for generative models with maximum mean discrepancy.

Conditional KMEs

- A conditional kernel mean embedding $\mu_{Y|X}$ is defined as a $\mathcal{H}_{\mathcal{Y}}$-valued, X-measurable random variable:

$$\mu_{Y|X=\cdot} = \int k_{\mathcal{Y}}(y, \cdot)\mathbb{P}_{Y|X=\cdot}(\mathrm{d}y). \tag{2}$$

- introduce $k_{\mathcal{X}}$, perform vector-valued regression to get an estimate for $\mu_{Y|X=x}$ for an arbitrary x.
- Maximum *conditional* mean discrepancy (MCMD) is a function of x, the distance between embeddings of distributions of $Y|X = x$ and $Y'|X' = x$ in $\mathcal{H}_{\mathcal{Y}}$.
- we consider eMCMD, expected value of MCMD over measure $\nu(x)$.
- ... to be continued!

[4] J. Park and K. Muandet. (2020). A measure-theoretic approach to kernel conditional mean embeddings.

Thank you!

Reach out to: masha.naslidnyk.21@ucl.ac.uk