Monotonicity and Double Descent in Uncertainty Estimation with Gaussian Processes

Liam Hodgkinson, Chris van der Heide, Fred Roosta & Michael W. Mahoney

Gaussian Processes (GP)

A stochastic process $f \sim \mathcal{GP}(m, k)$ on $\mathbb{R}^d$ such that $(f(x_1), \ldots, f(x_n))$ is multivariate normal with ***mean*** $\{m(x_i)\}_{i=1}^n$ and ***covariance matrix*** $(k(x_i, x_j))_{i,j=1}^n$.

Regression

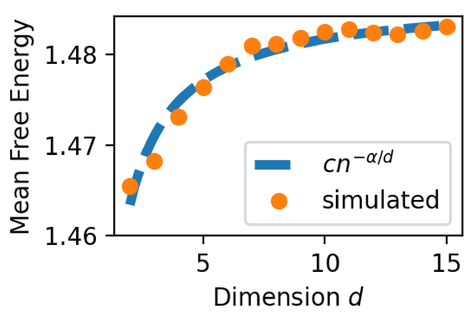
Data:	input-output pairs $(X_i, Y_i)_{i=1}^n$
Model:	find a predictor $y = f(x)$
Likelihood:	output is normal w/ mean $f(X_i)$ and variance σ_n^2
Prior:	a Gaussian process $f \sim \mathcal{GP}(0, \sigma_f^2 k)$

Measures of Model Error (accounting for uncertainty)

(A) Posterior Predictive Loss $\mathbb{E}_{f\sim\rho} L(f)$

(B) (Mean) Bayes Free Energy $\mathcal{F}_n = -\frac{1}{n} \log \underbrace{\mathbb{E}_{f\sim\pi}}_{\text{prior}} \prod_{i=1}^n \underbrace{p(Y_i|f, X_i)}_{\text{likelihood}}$

Common Rule of Thumb



(I) ***Curse of dimensionality:*** Free energy penalises extra covariates

(II) ***Free energy*** $\sim$ ***posterior loss*** (Fong & Holmes, 2019)

Figure 1: Free energy of GP with $n = 200$, RBF kernel, $\sigma_n^2 = \sigma_f^2 = 1$.

Acknowledgements

MM would like to acknowledge the IARPA (contract W911NF20C0035), NSF, and ONR for providing partial support of this work. This research was also partially supported by the Australian Research Council through an Industrial Transformation Training Centre for Information Resilience (IC200100022) and the Australian Centre of Excellence for Mathematical and Statistical Frontiers (CE140100049).

Theorem 1: Tuning Negates Curse of Dimensionality

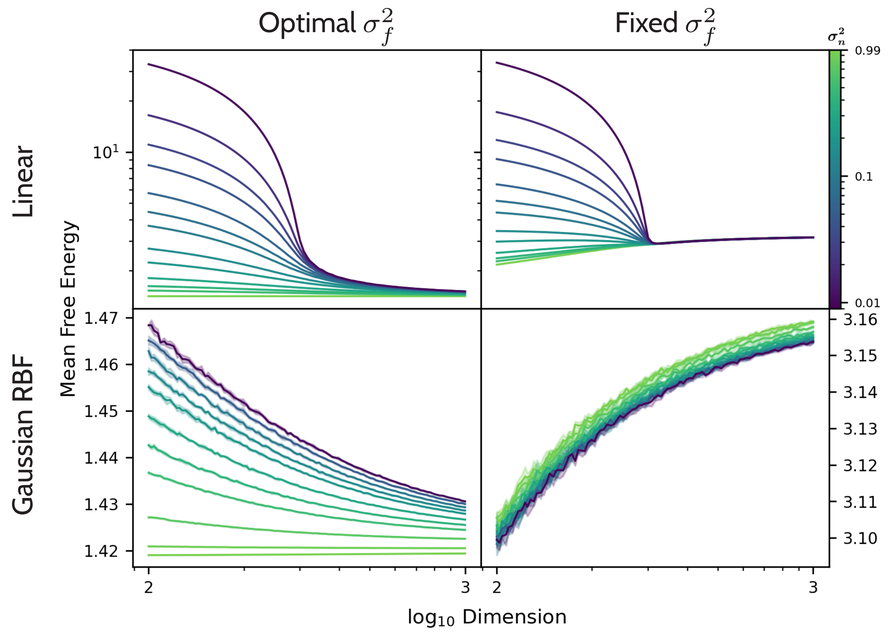


Figure 2: Free energies on synthetic Gaussian data for linear (top) and Gaussian RBF (bottom) kernels with optimal σ_f^2 (left, error decreases) and $\sigma_f^2 = 0.01$ (right, error increases).

Theorem 2: Measures Behave Differently

Bayes free energy: monotone decay. ***Posterior loss***: exhibits double-descent.

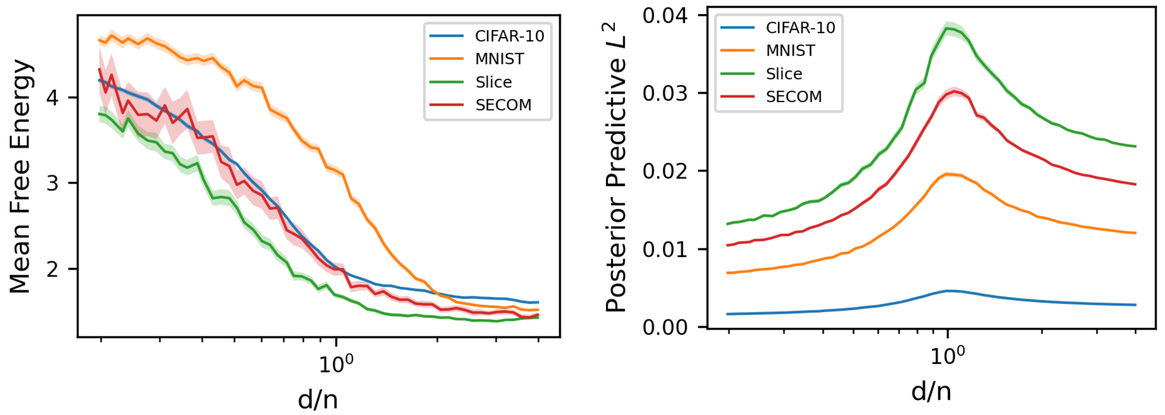


Figure 3: Mean free energies (left) and L^2 losses (right) for multiple datasets linear GP, optimal σ_f^2

The University of Melbourne · Berkeley University of California · ICSI · ICML International Conference On Machine Learning