Learning from Functions with Semi-Parametric Kernel Ridge

O. Taylan Turan*, David M.J. Tax*, and Marco Loog*

Pattern Recognition Laboratory, Delft University of Technology, Van Mourik Broekmanweg 6, Delft 2628 XE, The Netherlands

o.t.turan@tudelft.nl github.com/taylanot

1. Introduction

- Meta-learning: leverages similar learning problems (tasks) for a specific similar data-scarce target learning problem (task).
- AIM : Come up with a meta-learning method that is interpretable with a convex loss to optimize!

2. Semi-Parametric Kernel Ridge

Assuming $\tilde{f} = f + h$ where $h \in span\{\psi_p\}$ and as$\{\psi_p\}_{p=1}^{M}$ are real-valued functions and if the loss is taken as $\mathcal{L} = \sum_i^N (\tilde{f}(\mathbf{x}_i) - y_i)^2$. Then,

$$\hat{\tilde{f}} = \min_{\tilde{f} \in \mathcal{H}} \sum_i^{N_a} (\tilde{f}(\mathbf{x}_i) - y_i)^2 + g(||f||_{\mathcal{H}}) \tag{3}$$

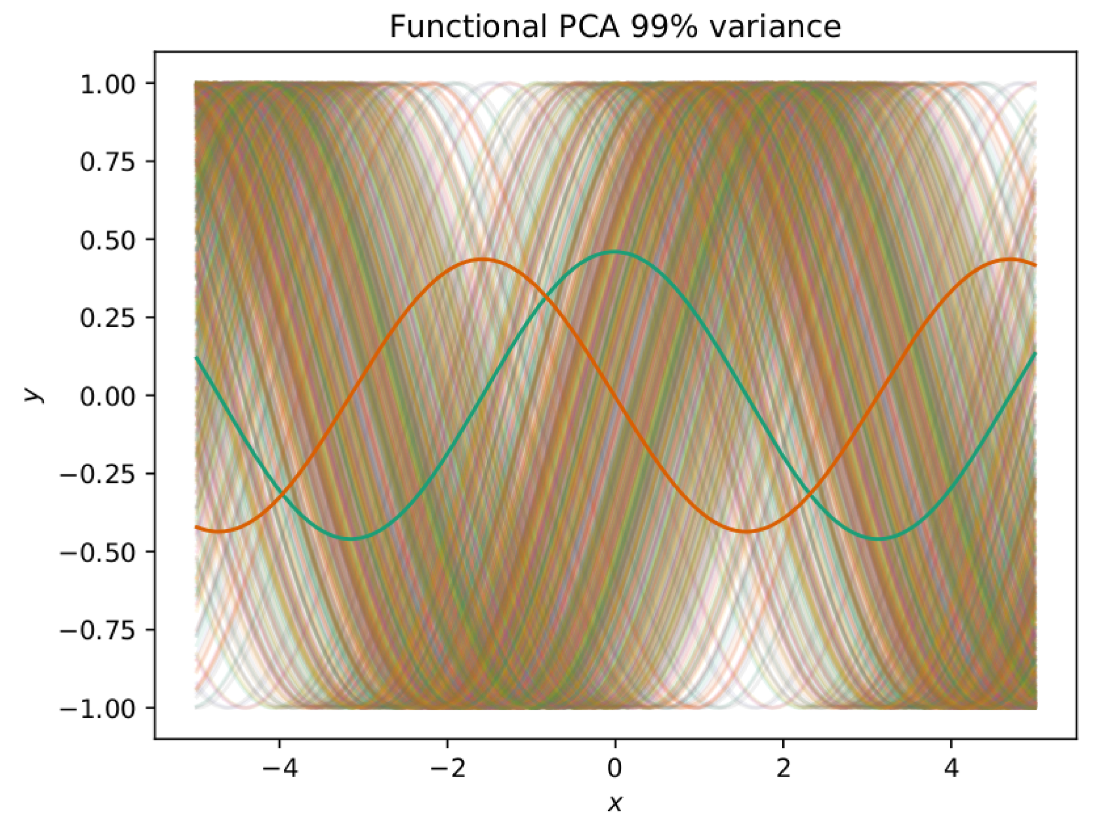
has the form $\tilde{f}(\cdot) = \sum_i^N \alpha_i k(\cdot, \mathbf{x}_i) + \sum_j^M \beta_j \psi_j(\cdot)$. (Semi-Parametric Representer Theorem!)

3. Setup

$$y = f_a(\mathbf{x}) + \varepsilon, \tag{1}$$

where $y \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^d$, and ε is taken to be standard normal.

Meta-learning tasks are determined by p_f!

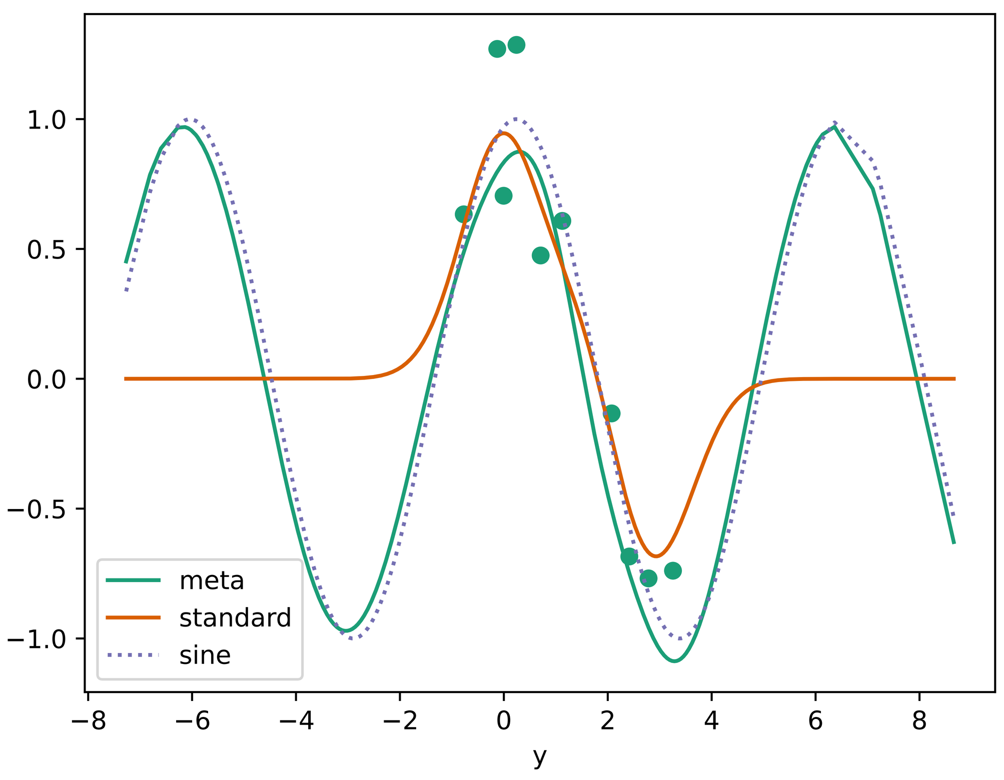


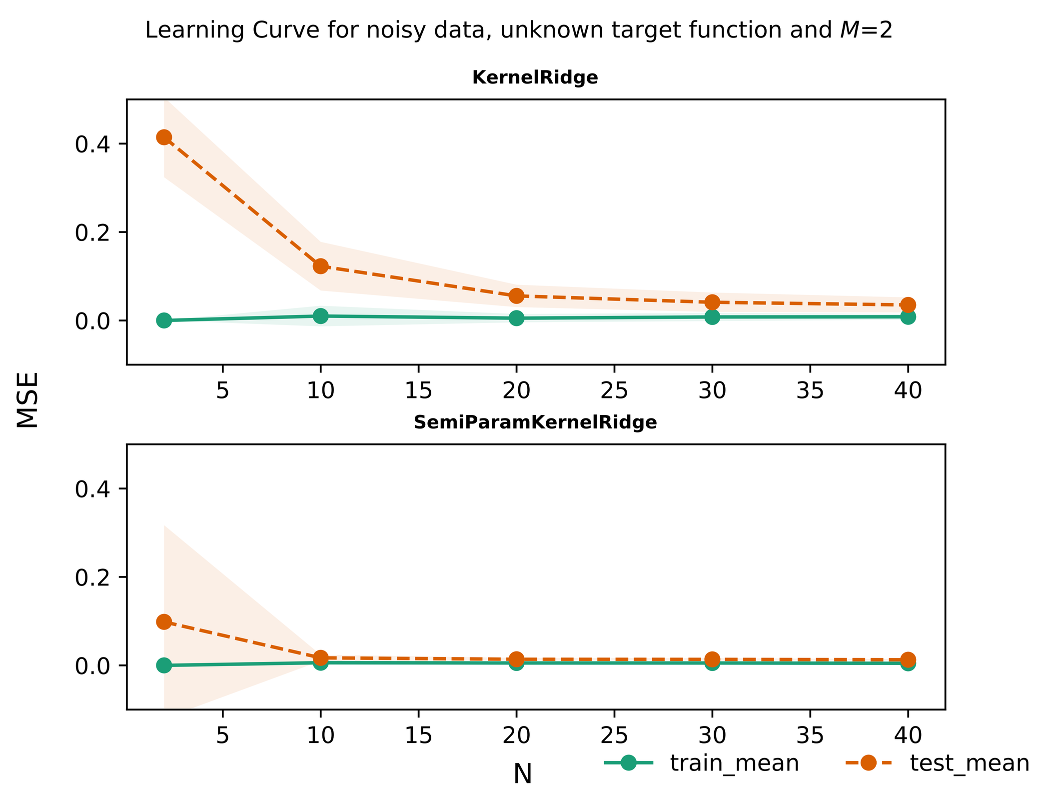
4. Synthetic Problem

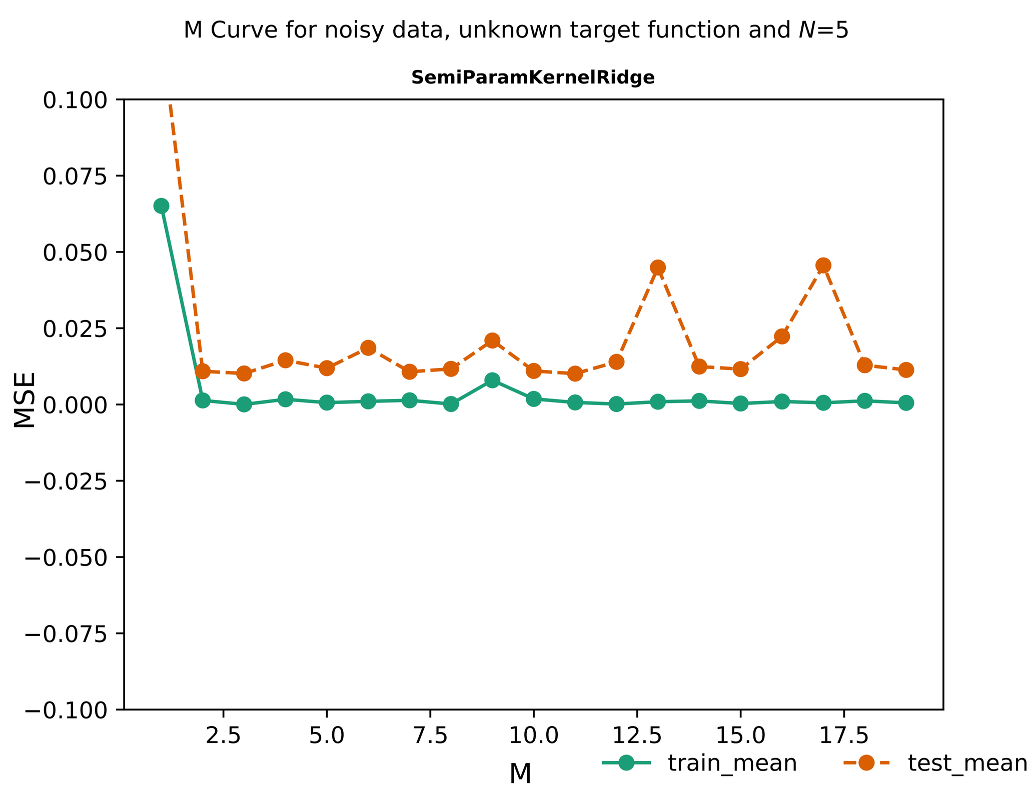
Let us assume we are trying to solve a problem for a 1-dimensional regression problem (D=1),

$$y = \sin(\mathbf{x} + \phi_a) + \varepsilon, \tag{4}$$

where $y \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^D$, $\phi_a \sim \mathcal{N}(\mathbf{0}, c\mathbf{I})$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.



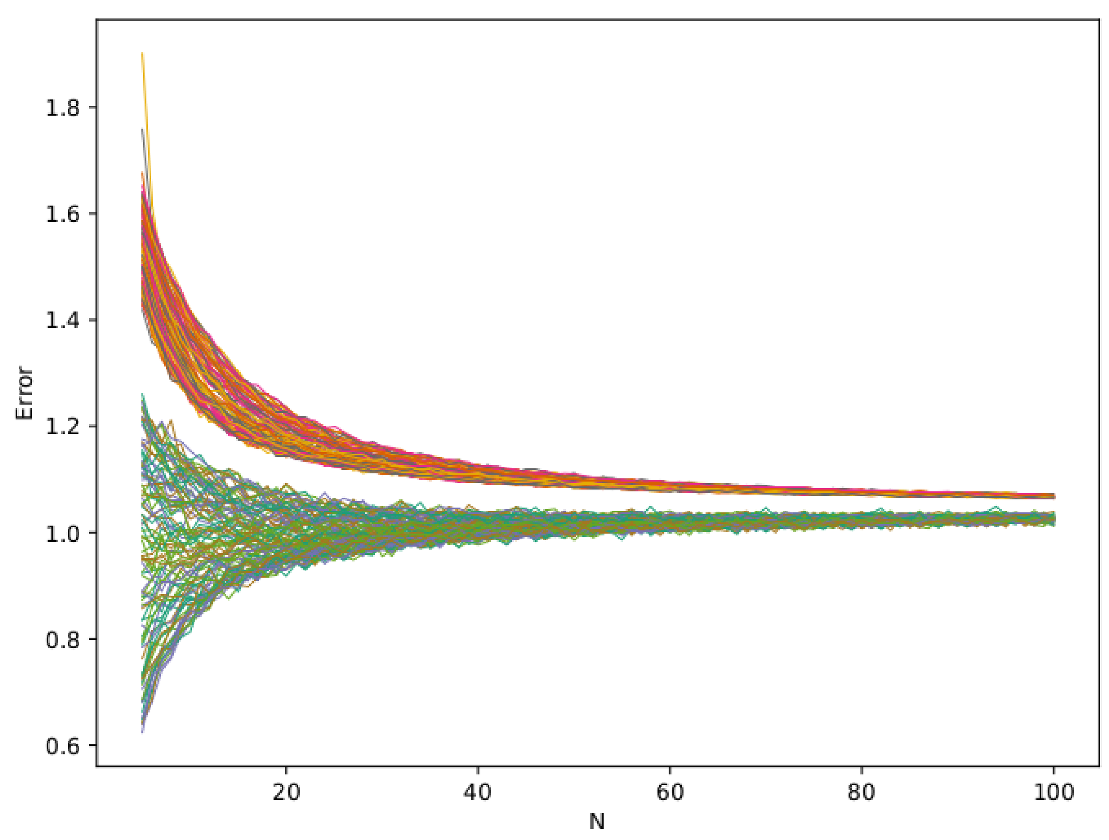




3. Open Questions

- Regularization of β
- Can we just improve our kernels with $\{\psi_i\}_{i=1}^{N}$
- Functional PCA utilization!

5.Learning Curve Application



- Bunch of learning curves for varying regularization λ
- Work in progress!!!