Meta-Learning Constitutive Laws-Method Development #1

March 24, 2022

Ozgur Taylan Turan

Meta-Learning with Kernel Models

- Comments on the writing!
- My ideas going forward!

Comments on writing?

Going Froward

- Latent Variable Models (Maybe a bit later?)
- Meta-learning with KernelRidge?

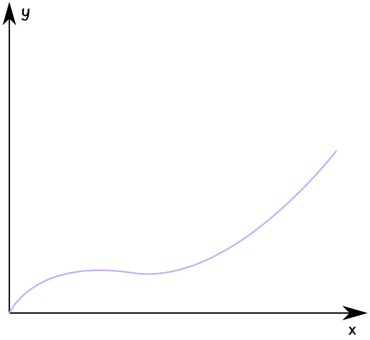
Meta-learning with Kernel Ridge

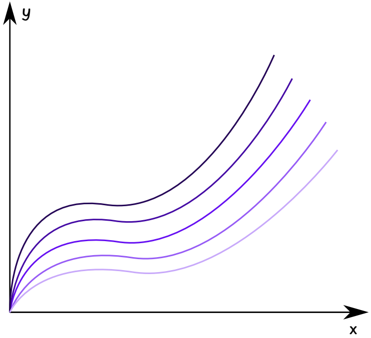
- We are after $f \in \mathcal{F}$, where $\mathcal{F}$ is a class of functions
- Non-parametric Representer Theorem states that f that minimizes regularized risk functional $\mathcal{L} + g(||f||)$ (note, $||\cdot||$ is in RKHS) is given by $f(\cdot) = \sum_{i=1}^{N} \alpha_i k(\cdot, \mathbf{x}_i)$, where N is the number of training samples $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$

Meta-learning with Kernel Ridge

- We are after $f \in \mathcal{F}$, where $\mathcal{F}$ is a class of functions
- Non-parametric Representer Theorem states that f that minimizes regularized risk functional $\mathcal{L} + \Omega(||f||)$ (note, $||\cdot||$ is in RKHS) is given by $f(\cdot) = \sum_{i=1}^{N} \alpha_i k(\cdot, \mathbf{x}_i)$, where N is the number of training samples $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$

What I was trying to do?





2-Options[1]

Prior Knowledge by Parametric Expansions

Theorem 4.3 (Semiparametric Representer Theorem) *Suppose that in addition to the assumptions of the previous theorem we are given a set of M real-valued functions* $\{\psi_p\}_{p=1}^M : \mathcal{X} \to \mathbb{R}$, *with the property that the* $m \times M$ *matrix* $(\psi_p(x_i))_{ip}$ *has rank M. Then any* $\tilde{f} := f + h$, *with* $f \in \mathcal{H}$ *and* $h \in \operatorname{span}\{\psi_p\}$, *minimizing the regularized risk*

$$c\left((x_1, y_1, \tilde{f}(x_1)), \ldots, (x_m, y_m, \tilde{f}(x_m))\right) + \Omega\left(\|f\|_{\mathcal{H}}\right) \tag{4.10}$$

admits a representation of the form

$$\tilde{f}(x) = \sum_{i=1}^{m} \alpha_i k(x_i, x) + \sum_{p=1}^{M} \beta_p \psi_p(x), \tag{4.11}$$

with $\beta_p \in \mathbb{R}$ *for all* $p \in [M]$.

We will discuss applications of the semiparametric extension in Section 4.8.

Bias

Remark 4.4 (Biased Regularization) *Another extension of the representer theorems can be obtained by including a term* $-\langle f_0, f\rangle$ *in (4.4) or (4.10), where* $f_0 \in \mathcal{H}$. *In this case, if a solution to the minimization problem exists, it admits an expansion which differs from those described above in that it additionally contains a multiple of* f_0. *To see this, decompose* $f_\perp(\cdot)$ *used in the proof of Theorem 4.2 into a part orthogonal to* f_0 *and the remainder.*

[1] B. Schölkopf (2002). “Learning with kernels”. In: *Proceedings of 2002 International Conference on Machine Learning and Cybernetics* 1. DOI: 10.7551/mitpress/4175.001.0001

2-Options[2]

- Biased regularization framework similar to [1] in a nonlinear setting. (They just mention that it is doable and do not show explicitely.)
- Semi-parametric definition is also interesting.

[1] G. Denevi, C. Ciliberto, D. Stamos, and M. Pontil (2018). "Incremental learning-to-learn with statistical guarantees". In: *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018* 1, pp. 457–466. arXiv: 1803.08089

[2] B. Schölkopf (2002). "Learning with kernels". In: *Proceedings of 2002 International Conference on Machine Learning and Cybernetics* 1. DOI: 10.7551/mitpress/4175.001.0001