On kernel-based statistical learning in the mean field limit

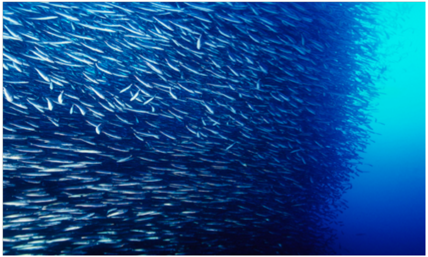
NeurIPS 2023

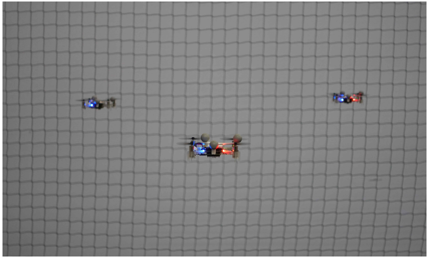
Christian Fiedler[1], Michael Herty[2], Sebastian Trimpe[3]
[1,3]Institute for Data Science in Mechanical Engineering (DSME)
fiedler@dsme.rwth-aachen.de, trimpe@dsme.rwth-aachen.de
[2]Institute for Geometry and Practical Mathematics (IGPM)
herty@igpm.rwth-aachen.de

RWTH Aachen University

Version/Date: November 14, 2023

Motivation: Multiagent Systems (MAS)





Sources: Top left Wiki Commons, top right Wiki Commons, lower left U.S. Fish and Wildlife Service, lower right DSME RWTH Aachen

Motivation: Learning state-dependent features of MAS

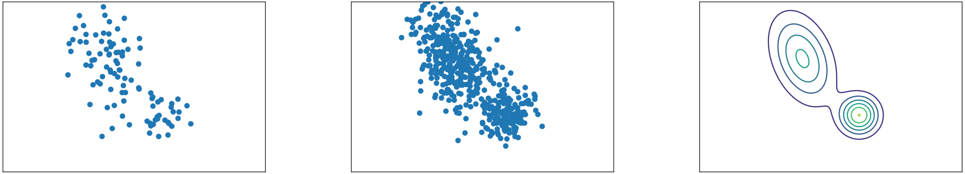
Consider a system of $M \in \mathbb{N}_+$ agents or particles

- $\vec{x}^{[M]} \in X^M$ pointwise-in-time state of system, $\left(\vec{x}^{[M]}\right)_m$ state of agent m
- State-dependent feature of system with state-to-feature mapping $f_M : X^M \to \mathbb{R}$
- Goal: Learn f_M from data $(\vec{x}_1^{[M]}, y_1^{[M]}), \ldots, (\vec{x}_N^{[M]}, y_N^{[M]})$, assuming $y_n^{[M]} = f_M(\vec{x}_n^{[M]}) + \eta_n$
  $\rightsquigarrow$ Standard supervised learning problem (regression)
- Most kernel methods lead to estimate of the form

$$\hat{f}_M = \sum_{n=1}^{N} \alpha_n^{[M]} k_M(\cdot, \vec{x}_n^{[M]}),$$

where $k_M : X^M \times X^M \to \mathbb{R}$ is the kernel

Microscopic to mesoscopic



Mean field limit

$$X^M \ni \vec{x} \equiv \frac{1}{M}\sum_{m=1}^{M} \delta_{x_m} \in \mathcal{P}(X) \quad \xrightarrow{M\to\infty} \quad \mu \in \mathcal{P}(X)$$

Mean field limit of functions

Definition

Sequence $g_M : X^M \to \mathbb{R}$, $M \in \mathbb{N}_+$, has *mean field limit (MFL)* $g : \mathcal{P}(X) \to \mathbb{R}$, denoted by $g_M \xrightarrow{\mathcal{P}_1} g$, if

$$\lim_{M\to\infty} \sup_{\vec{x}\in X^M} |g_M(\vec{x}) - g(\hat{\mu}[\vec{x}])| = 0,$$

where $\hat{\mu}[\vec{x}] = \frac{1}{M}\sum_{m=1}^M \delta_{x_m}$ is the empiricial measure with atoms $x_1, \ldots, x_M$.

- Reasonable assumption: $f_M \xrightarrow{\mathcal{P}_1} f$, where MFL $f : \mathcal{P}(X) \to \mathbb{R}$ is state-to-feature map on mesoscopic level
- Learning on mesoscopic level: Data now $(\mu_1, y_1), \ldots, (\mu_N, y_N)$, where $y_n = f(\mu_n) + \eta_n$
- In this situation, kernel methods need kernel $k : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R}$

Questions

Central question How are the learning problems on the microscopic and mesoscopic level related?

1. Mean field limit of kernels $k_M : X^M \times X^M \to \mathbb{R}$? Mean field limit RKHS?
2. Mean field limit in representer theorem?
3. Mean field limit of statistical learning problems? Convergence of risks and (regularized) Empirical Risk Minimizers?

Institute for Data Science in Mechanical Engineering
RWTH AACHEN UNIVERSITY

Previous work: Mean field limit of kernels

Definition

$k_M : X^M \times X^M \to \mathbb{R}$, $M \in \mathbb{N}_+$, kernels on X^M have *mean field limit* $k : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R}$, denoted by $k_M \xrightarrow{\mathcal{P}_1} k$, if

$$\lim_{M\to\infty} \sup_{\vec{x},\vec{x}' \in X^M} |k_M(\vec{x}, \vec{x}') - k(\hat{\mu}[\vec{x}], \hat{\mu}[\vec{x}'])| = 0,$$

Theorem (informal, cf. F., Herty, Rom, Segala, Trimpe '23)

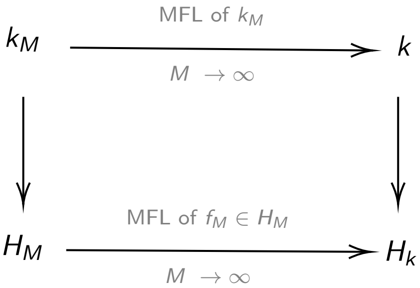
If $(k_M)_M$ is a sequence of permutation-invariant, uniformly bounded, and uniformly Lipschitz-continuous (w.r.t. the Monge-Kantorowich metric) kernels on X^M, where X is a compact metric space, then there exists a subsequence that has a mean field limit k, which is again a kernel.

Mean field limit of RKHSs

Theorem (informal, Thm 2.3 in the paper)

k mean field limit kernel of $(k_M)_M$, H_M and H_k the associated RKHSs.

- Every RKHS function $f \in H_k$ arises as a mean field limit of functions $f_M \in H_M$.
- Every uniformly norm-bounded sequence $f_M \in H_M$ has a mean field limit f that is in H_k and shares the same norm bound.



RKHS H_k is the mean field limit of the RKHSs H_M

Representer theorem in the mean field limit

Theorem (informal, Thm 3.3 in the paper)

Assume $\hat{\mu}[\vec{x}_n^{[M]}] \to \mu_n$ for $M \to \infty$, $n = 1, \ldots, N$, let $L : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ be continuous and strictly convex, and $\lambda > 0$. For each $M \in \mathbb{N}_+$, the problem

$$\min_{f \in H_M} L(f(\vec{x}_1^{[M]}), \ldots, f(\vec{x}_N^{[M]})) + \lambda \|f\|_M, \tag{1}$$

has a unique solution $f_M^* = \sum_{n=1}^N \alpha_n^{[M]} k_M(\cdot, \vec{x}_n^{[M]}) \in H_M$,

$$\min_{f \in H_k} L(f(\mu_1), \ldots, f(\mu_N)) + \lambda \|f\|_k. \tag{2}$$

has a unique solution $f^* = \sum_{n=1}^N \alpha_n k(\cdot, \mu_n) \in H_k$, and $f_M^* \xrightarrow{\mathcal{P}_1} f^*$ for $M \to \infty$.

Regularized empirical risk minimization in the mean field limit

Proposition (informal)

Loss functions $\ell_M : X^M \times Y \times \mathbb{R} \to \mathbb{R}_{\geq 0}$, $M \in \mathbb{N}_+$, have a mean field limit $\ell : \mathcal{P}(X) \times Y \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ under reasonable assumptions.

Definition

Data sets $\mathcal{D}_N^{[M]} = ((\vec{x}_1^{[M]}, y_1^{[M]}), \ldots, (\vec{x}_1^{[M]}, y_1^{[M]}))$ have mean field limit $\mathcal{D}_N = ((\mu_1, y_1), \ldots, (\mu_N, y_N))$ if $\hat{\mu}[\vec{x}_n^{[M]}] \to \mu_n$ and $y_n^{[M]} \to y_n$ for $M \to \infty$, for all $n = 1, \ldots, N$.

Proposition (informal, Prop. 4.3 in the paper)

Regularized empirical risk minimizer for data $\mathcal{D}_N$ is mean field limit of the regularized empirical risk minimizers for data $\mathcal{D}_N^{[M]}$, and the (empirical) risks also converge.

Regularized risk minimization in the mean field limit

Definition

Probability distributions P_M on $X^M \times Y$, $M \in \mathbb{N}_+$, converge in mean field to probability distribution P on $\mathcal{P}(X) \times Y$ if

$$\int_{X^M \times Y} f(\hat{\mu}[\vec{x}], y) \mathrm{d}P^{[M]}(\vec{x}, y) \to \int_{\mathcal{P}(X) \times Y} f(\mu, y) \mathrm{d}P(\mu, y).$$

for all continuous and bounded $f : \mathcal{P}(X) \to \mathbb{R}$.

Proposition (informal, Prop. 4.5 in the paper)

If P is the mean field limit of P_M, and ℓ the mean field limit of ℓ_M, then the regularized risk minimizer w.r.t. P and ℓ is the mean field limit of the regularized risk minimizers w.r.t. P_M and ℓ_M.

Institute for Data Science in Mechanical Engineering
RWTH AACHEN UNIVERSITY

Summary

Results

- Mean field limit of RKHSs (increasing number of inputs of kernels)
- Representer theorem in mean field limit
- Mean field limit for statistical learning theory setup
- Convergence of regularized (empirical) risks and mean field convergence of minimizers

Relevance

- New large-scale limit in theory of machine learning
- Theoretical foundation for new learning tasks on multiagent systems

Institute for Data Science in Mechanical Engineering
RWTH AACHEN UNIVERSITY