scch {
software
competence
center
hagenberg
}

scch.at

Operator-Theoretic and Complexity-Based Synthesis of a Gradient-Free Federated Kernel Learner

SCML-2026, Hagenberg, Austria

July 6-8, 2026

Mohit Kumar, Bernhard Moser, Manuela Geiß

scch {
software
competence
center
hagenberg
}

CONTENTS

scch {
software
competence
center
hagenberg
}

WHY A NEW LEARNER?

Statistically heterogeneous data distributed across clients

- Data across clients are often non-IID, which can degrade the performance of a one-size-fits-all model.

Computation and communication budget

- Devices may have limited computation and communication budgets, making frequent or large exchanges infeasible.

Privacy and security

- Strict privacy and security guarantees (e.g., via differential privacy or encryption) must be enforced without significantly degrading the model's learning performance and computation-efficiency.

R1 (Hypothesis Space for Learning from Heterogeneous Distributed Data)

Provide a mathematical framework that, without imposing parametric form or homogeneity assumptions on the client data distributions, determines a suitable hypothesis space for task learning in a distributed setting setting.

R2 (Theoretical Guarantees)

Calculate theoretically the error bounds and evaluate the task learning solution in terms of

- robustness of the prediction error against the disturbances arising from uncertainties and data noise, and
- accuracy of the solution and asymptotic upper bound on the approximation error.

R3 (Communication Efficiency)

Ensure that the analytically derived task learning solution can be implemented in a federated setting with communication and computational efficiency. Specifically, the optimization of the global model should not require multiple rounds of communication between the server and clients.

R4 (Efficient Differentially Private Knowledge Transfer Across Global and Local Models)

Define a novel differentially private metric, that allows for knowledge transfer from clients to server by solving the global model optimization problem without requiring exchange of gradients or model parameters (that may be computationally challenging and not optimal for privacy preservation), while optimizing the utility-privacy tradeoff.

R5 (Computationally Efficient FHE-Based Secure Federated Learning)

Rather than transmitting high-dimensional gradient or parameter update vectors, which entail substantial computational overhead when transmitted and operated on within encrypted domains, define novel low-dimensional attributes. These attributes must enable the inference of the global model with low communication overhead and latency, thereby supporting a computationally efficient realization of fully homomorphically encrypted inference of the global model.

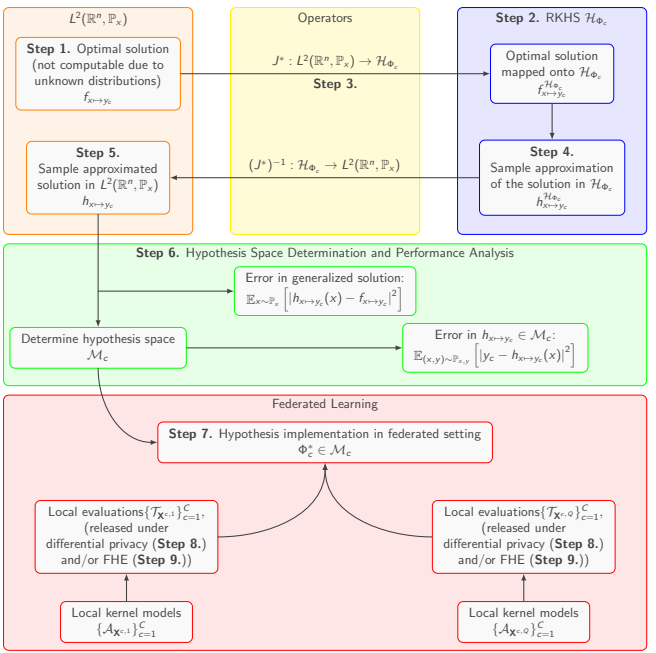
◎ Goal

Operator-Theoretic and Complexity-Based Federated Kernel Learning Framework that addresses all five requirements (R1-R5) within one mathematical pipeline.

scch {
software
competence
center
hagenberg
}

LEARNING FRAMEWORK

Learning Framework ▶ Overall Picture



Operator-Theoretic and Complexity-Based Federated Kernel Learning Framework

1. consider the optimal learning solution in $L^2(\mathbb{R}^n, \mathbb{P}_x)$,
2. map the optimal solution onto a RKHS using an operator,
3. approximate the optimal solution using available data samples in RKHS,
4. map the sample approximated solution onto $L^2(\mathbb{R}^n, \mathbb{P}_x)$ using the inverse-operator,
5. analyze the sample approximated solution in $L^2(\mathbb{R}^n, \mathbb{P}_x)$ and identifying conditions on kernel choice to define hypothesis space,
6. implement a suitable hypothesis with the minimum computational and communication cost in the federated setting using kernel models.

- Our goal is to learn a function $f_{x \mapsto y_c} : \mathbb{R}^n \to \mathbb{R}$ that minimizes the mean squared error:

$$f_{x \mapsto y_c} := \operatorname*{argmin}_{g \in L^2(\mathbb{R}^n, \mathbb{P}_x)} \mathop{\mathbb{E}}_{(x,y) \sim \mathbb{P}_{x,y}} \left[|y_c - g(x)|^2 \right] \tag{1}$$

$$= \operatorname*{argmin}_{g \in L^2(\mathbb{R}^n, \mathbb{P}_x)} \left(\int_{\mathbb{R}^n \times \{0,1\}^C} |y_c - g(x)|^2 \, \mathrm{d}\mathbb{P}_{x,y}(x,y) \right). \tag{2}$$

- It is well-known that the conditional expectation, also known as regression function, minimizes the mean squared error. That is,

$$f_{x \mapsto y_c}(x) = \mathop{\mathbb{E}}_{y \sim \mathbb{P}_{y|x}} [y_c | x] \,, \tag{3}$$

where we have made the following regularity assumption:

$$\mathop{\mathbb{E}}_{y \sim \mathbb{P}_{y|x}} [y_c | x] \in L^2(\mathbb{R}^n, \mathbb{P}_x). \tag{4}$$

Learning Framework ▶ A Generalized Kernel Function

- We consider a generalized kernel function such that for each class $c \in \{1, 2, \cdots, C\}$,

$$\mathcal{K}_{\Phi_c}(x, x') := \Phi_c(x)\Phi_c(x'), \tag{5}$$

  where $\Phi_c : \mathbb{R}^n \to [0, 1]$ is the feature-map (which will be determined based on theoretical analysis to estimate class posterior probability).

- Now the RKHS associated to $\mathcal{K}_{\Phi_c}$ is given as

$$\mathcal{H}_{\Phi_c} := \left\{ f = \sum_{i=1}^{\infty} \alpha_i \mathcal{K}_{\Phi_c}(\cdot, x^i) \;\mid\; \alpha_i \in \mathbb{R},\; x^i \in \mathbb{R}^n,\; \|f\|^2_{\mathcal{H}_{\Phi_c}} := \sum_{i,j=1}^{\infty} \alpha_i \alpha_j \mathcal{K}_{\Phi_c}(x^i, x^j) < \infty \right\} \tag{6}$$

  with inner product for any $f = \sum_{i=1}^{N} a_i \mathcal{K}_{\Phi_c}(\cdot, s^i) \in \mathcal{H}_{\Phi_c}$ and $g = \sum_{j=1}^{M} b_j \mathcal{K}_{\Phi_c}(\cdot, t^j) \in \mathcal{H}_{\Phi_c}$ defined as

$$\langle f, g \rangle_{\mathcal{H}_{\Phi_c}} := \sum_{i=1}^{N} \sum_{j=1}^{M} a_i b_j \mathcal{K}_{\Phi_c}(s^i, t^j). \tag{7}$$

Learning Framework ▶ Operators between L^2 and RKHS

- We define an operator, $J^* : L^2(\mathbb{R}^n, \mathbb{P}_x) \to \mathcal{H}_{\Phi_c}$, as

$$(J^* g)(x) := \mathop{\mathbb{E}}_{x' \sim \mathbb{P}_x} [\mathcal{K}_{\Phi_c}(x', x) g(x')] \tag{8}$$

$$= \int_{\mathbb{R}^n} \mathcal{K}_{\Phi_c}(x', x) g(x') \, \mathrm{d}\mathbb{P}_x(x'). \tag{9}$$

- The inverse of J^*, $(J^*)^{-1} : \mathcal{H}_{\Phi_c} \to L^2(\mathbb{R}^n, \mathbb{P}_x)$, is given as

$$\left((J^*)^{-1} f\right)(x) := \frac{f(x)}{\|\Phi_c\|^2_{L^2(\mathbb{R}^n, \mathbb{P}_x)}}, \tag{10}$$

where $\Phi_c : \mathbb{R}^n \to [0, 1]$ characterizes the kernel $\mathcal{K}_{\Phi_c}$.

Learning Framework ▶ RKHS Learning Solution

- The optimal solution (3), $f_{x\mapsto y_c} \in L^2(\mathbb{R}^n, \mathbb{P}_x)$, is mapped onto $\mathcal{H}_{\Phi_c}$ through the operator J^*, i.e.,

$$f_{x\mapsto y_c}^{\mathcal{H}_{\Phi_c}} := J^* f_{x\mapsto y_c}. \tag{11}$$

- To approximate $f_{x\mapsto y_c}^{\mathcal{H}_{\Phi_c}}$ in RKHS, a natural approach is of approximating J^* in using available data samples. This leads to

$$h_{x\mapsto y_c}^{\mathcal{H}_{\Phi_c}} = (\widehat{S}^*_{(x^i)_{i=1}^N} \mathrm{Ev}_{(x^i)_{i=1}^N}) f_{x\mapsto y_c} + \widehat{S}^*_{(x^i)_{i=1}^N}(\xi_c(x^1, y^1), \cdots, \xi_c(x^N, y^N)), \tag{12}$$

where $\widehat{S}^*_{(x^i)_{i=1}^N} : (\mathbb{R}^N, \langle\cdot,\cdot\rangle_{\mathbb{R}^N}) \to \mathcal{H}_{\Phi_c}$ and $\mathrm{Ev}_{(x^i)_{i=1}^N} : L^2(\mathbb{R}^n, \mathbb{P}_x) \to (\mathbb{R}^N, \langle\cdot,\cdot\rangle_{\mathbb{R}^N})$ are given as

$$\widehat{S}^*_{(x^i)_{i=1}^N}(u_1, \cdots, u_N) := \frac{1}{N}\sum_{i=1}^N u_i \mathcal{K}_{\Phi_c}(x^i, \cdot) \tag{13}$$

$$\mathrm{Ev}_{(x^i)_{i=1}^N} g := (g(x^1), \cdots, g(x^N)). \tag{14}$$

$\widehat{S}^*_{(x^i)_{i=1}^N}$ is viewed as the sample approximation of J^*.

Theorem (Risk for Sample Approximation of the Optimal Solution in RKHS)

The following holds with probability at least $1-\delta$ *for any* $\delta \in (0,1)$*:*

$$\mathop{\mathbb{E}}_{x\sim\mathbb{P}_x}\left[\left|h^{\mathcal{H}_{\Phi_c}}_{x\mapsto y_c}(x) - f^{\mathcal{H}_{\Phi_c}}_{x\mapsto y_c}(x)\right|^2\right] \leq \frac{3}{\sqrt{N}} + \sqrt{\frac{8\log(1/\delta)}{N}}. \qquad (15)$$

Learning Framework ▶ Learning Solution in L^2

- The learning solution $h^{\mathcal{H}_{\Phi_c}}_{x\mapsto y_c} \in \mathcal{H}_{\Phi_c}$ is mapped onto $L^2(\mathbb{R}^n, \mathbb{P}_x)$ using inverse-operator $(J^*)^{-1}$, i.e.,

$$h_{x\mapsto y_c} := (J^*)^{-1} h^{\mathcal{H}_{\Phi_c}}_{x\mapsto y_c}. \tag{16}$$

Theorem (Risk for Generalized Learning Solution)

The following holds with probability at least $1-\delta$ *for any* $\delta \in (0,1)$*:*

$$\mathop{\mathbb{E}}_{x\sim\mathbb{P}_x}\left[\left|h_{x\mapsto y_c}(x) - f_{x\mapsto y_c}(x)\right|^2\right] \leq \frac{1}{\left(\|\Phi_c\|^2_{L^2(\mathbb{R}^n,\mathbb{P}_x)}\right)^2}\left(\frac{3}{\sqrt{N}} + \sqrt{\frac{8\log(1/\delta)}{N}}\right). \tag{17}$$

Theorem (Prediction Error Bound for Generalized Learning Solution)

The following holds with probability at least $1-\delta$ *for any* $\delta \in (0,1)$*:*

$$\underbrace{\mathop{\mathbb{E}}_{(x,y)\sim\mathbb{P}_{x,y}}\left[\left|y_c - h_{x\mapsto y_c}(x)\right|^2\right]}_{\text{mean-squared prediction error}} \leq \underbrace{\mathop{\mathbb{E}}_{(x,y)\sim\mathbb{P}_{x,y}}\left[\left|y_c - \mathop{\mathbb{E}}_{y\sim\mathbb{P}_{y|x}}[y_c|x]\right|^2\right]}_{\text{mean-squared disturbance magnitude}} + \frac{1}{\left(\|\Phi_c\|^2_{L^2(\mathbb{R}^n,\mathbb{P}_x)}\right)^2}\left(\frac{3}{\sqrt{N}} + \sqrt{\frac{8\log(1/\delta)}{N}}\right). \quad (18)$$

Theorem (Approximation Error Bound for Generalized Learning Solution)

The following holds with probability at least $1-\delta$ *for any* $\delta \in (0,1)$*:*

$$\underbrace{\mathop{\mathbb{E}}_{x\sim\mathbb{P}_x}\left[\left|h_{x\mapsto y_c}(x)-\mathbb{P}_{y|x}(y_c=1|x)\right|^2\right]}_{\text{mean-squared approximation error}} \leq \frac{1}{\left(\|\Phi_c\|^2_{L^2(\mathbb{R}^n,\mathbb{P}_x)}\right)^2}\left(\frac{3}{\sqrt{N}}+\sqrt{\frac{8\log(1/\delta)}{N}}\right). \quad (19)$$

- Since $h_{x \mapsto y_c}$ is viewed as the predictor of c^{th} class label $y_c \in \{0, 1\}$, we ensure that $h_{x \mapsto y_c}(x) \in [0, 1]$ by constraining the feature-map Φ_c as

$$\|\Phi_c\|^2_{L^2(\mathbb{R}^n, \mathbb{P}_x)} = \frac{N_c}{N}. \tag{20}$$

- With the kernel feature-map normalization condition (20), we define our hypothesis space as

$$\mathcal{M}_c := \left\{ h_{x \mapsto y_c} = \frac{\Phi_c(\cdot)}{N\|\Phi_c\|^2_{L^2(\mathbb{R}^n, \mathbb{P}_x)}} \sum_{i=1}^{N_c} \Phi_c(x^{\mathrm{I}_i^c}) \;\mid\; \Phi_c : \mathbb{R}^n \to [0, 1],\ \|\Phi_c\|^2_{L^2(\mathbb{R}^n, \mathbb{P}_x)} = \frac{N_c}{N} \right\}. \tag{21}$$

- For a given data set $\mathcal{D}$, the empirical Rademacher complexity of the hypothesis space $\mathcal{M}_c$ is given as

$$\widehat{\mathcal{R}}_{\mathcal{D}}(\mathcal{M}_c) = \frac{1}{N}\,\mathbb{E}_{\sigma}\left[\sup_{h_{x\mapsto y_c}\in\mathcal{M}_c}\sum_{i=1}^{N}\sigma_i\, h_{x\mapsto y_c}(x^i)\right], \tag{22}$$

  where $\sigma = (\sigma_1, \cdots, \sigma_N)$ with $\sigma_1, \cdots, \sigma_N$ as the independent random variables drawn from the Rademacher distribution.

Theorem (Bound on Rademacher Complexity of the Hypothesis Space)

Given a dataset $\mathcal{D}$, we have

$$\mathbb{E}_{\mathcal{D}\sim(\mathbb{P}_{x,y})^N}\left[\widehat{\mathcal{R}}_{\mathcal{D}}(\mathcal{M}_c)\right] \leq \frac{1}{\sqrt{N}}. \tag{23}$$

Theorem (Prediction Error Bound for Hypothesis Space)

Given a data set $\mathcal{D} = \{(x^i, y^i) \mid i \in \{1, 2, \cdots, N\}\} \sim (\mathbb{P}_{x,y})^N$, *for any* $h_{x \mapsto y_c} \in \mathcal{M}_c$, *we have with probability at least* $1 - \delta$ *for any* $\delta \in (0, 1)$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(x,y)\sim\mathbb{P}_{x,y}} \left[|y_c - h_{x\mapsto y_c}(x)|^2 \right] \\
&\leq \min \left(\frac{1}{N} \sum_{i=1}^{N} |y_c^i - h_{x\mapsto y_c}(x^i)|^2 + \frac{4}{\sqrt{N}} + \sqrt{\frac{\log(1/\delta)}{2N}}, \right. \\
&\qquad \left. \mathop{\mathbb{E}}_{(x,y)\sim\mathbb{P}_{x,y}} \left[\left| y_c - \mathop{\mathbb{E}}_{y\sim\mathbb{P}_{y|x}} [y_c|x] \right|^2 \right] + \frac{1}{(N_c/N)^2} \left(\frac{3}{\sqrt{N}} + \sqrt{\frac{8\log(1/\delta)}{N}} \right) \right). \qquad (24)
\end{aligned}$$

Theorem (Approximation Error Bound for Hypothesis Space)

Given a data set $\mathcal{D} = \{(x^i, y^i) \mid i \in \{1, 2, \cdots, N\}\} \sim (\mathbb{P}_{x,y})^N$, *for any* $h_{x \mapsto y_c} \in \mathcal{M}_c$, *we have with probability at least* $1 - \delta$ *for any* $\delta \in (0, 1)$,

$$\mathop{\mathbb{E}}_{x \sim \mathbb{P}_x} \left[\left| h_{x \mapsto y_c}(x) - \mathbb{P}_{y|x}(y_c = 1|x) \right|^2 \right] \leq \min \left(\frac{1}{N} \sum_{i=1}^{N} |y_c^i - h_{x \mapsto y_c}(x^i)|^2 + \frac{4}{\sqrt{N}} + \sqrt{\frac{\log(1/\delta)}{2N}}, \right.$$
$$\left. \frac{1}{(N_c/N)^2} \left(\frac{3}{\sqrt{N}} + \sqrt{\frac{8 \log(1/\delta)}{N}} \right) \right). \qquad (25)$$

scch {
software
competence
center
hagenberg
}

FEDERATED ML

Federated ML ▶ Geometric Kernel Machines

- We have recently introduced geometrically inspired kernel machines, KAHM (Kernel Affine Hull Machines) [1], for efficient data representations learning beyond gradient-descent.

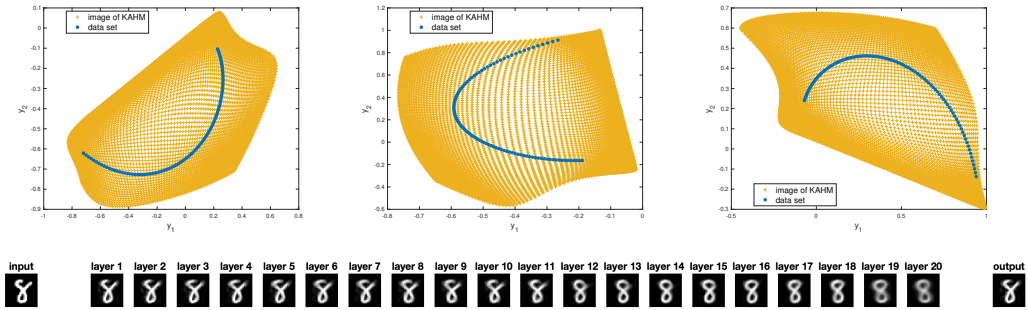
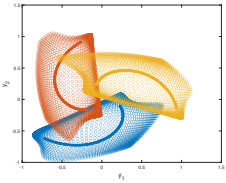
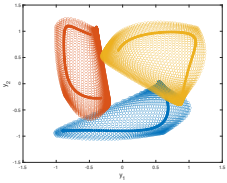


Figure: The deep KAHM discovers layers of increasingly abstract data representation with lowest-level data features being modeled by first layer and the highest-level by end layer.

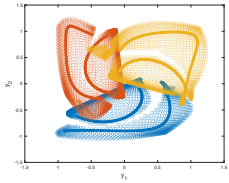
Learning in Federated Setting [2]



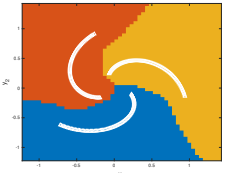
(a) Local dataset 1

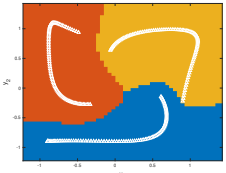


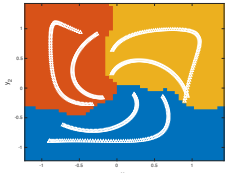
(b) Local dataset 2



(c) Global KAHM







Definition (Space Folding Measure)

To evaluate the amount of folding required for an arbitrary point $x \in \mathbb{R}^n$ to map it (by the KAHM $\mathcal{A}_{\mathbf{X}}$) to a point closer to data samples $\mathbf{X}$, we define a space folding measure, $\mathcal{T}_{\mathbf{X}} : \mathbb{R}^n \to [0,1]$, associated to data samples $\mathbf{X}$, as

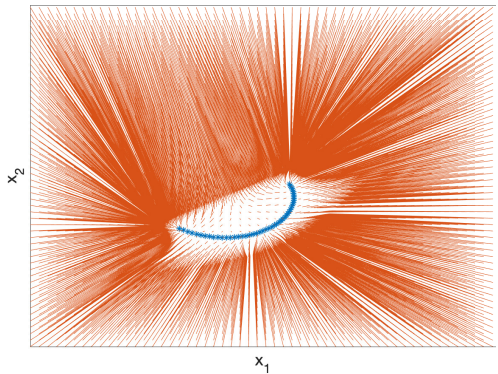
$$\mathcal{T}_{\mathbf{X}}(x) := \begin{cases} \sqrt{\frac{1}{2}\left(\left|\mathcal{T}_{\mathbf{X}}^{\text{Euc}}(x)\right|^2 + \left|\mathcal{T}_{\mathbf{X}}^{\text{Cos}}(x)\right|^2\right)} & \text{option 1} \\ \mathcal{T}_{\mathbf{X}}^{\text{Euc}}(x)\mathcal{T}_{\mathbf{X}}^{\text{Cos}}(x) & \text{option 2} \\ \min\left(\mathcal{T}_{\mathbf{X}}^{\text{Euc}}(x), \mathcal{T}_{\mathbf{X}}^{\text{Cos}}(x)\right) & \text{option 3} \\ \max\left(\mathcal{T}_{\mathbf{X}}^{\text{Euc}}(x), \mathcal{T}_{\mathbf{X}}^{\text{Cos}}(x)\right) & \text{option 4,} \end{cases} \tag{26}$$

where

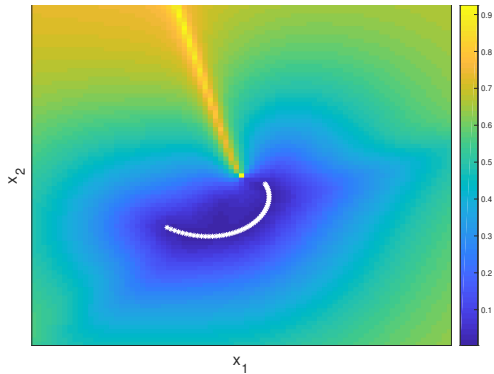
$$\mathcal{T}_{\mathbf{X}}^{\text{Euc}}(x) := 1 - \exp\left(-\|x - \mathcal{A}_{\mathbf{X}}(x)\|\right) \tag{27}$$

$$\mathcal{T}_{\mathbf{X}}^{\text{Cos}}(x) := \frac{1}{\pi}\arccos\left(\frac{\left(\mathcal{A}_{\mathbf{X}}(x)\right)^T x}{\|\mathcal{A}_{\mathbf{X}}(x)\|\|x\|}\right). \tag{28}$$

Federated ML ▶ Geometric Kernel Machines (Cont.)

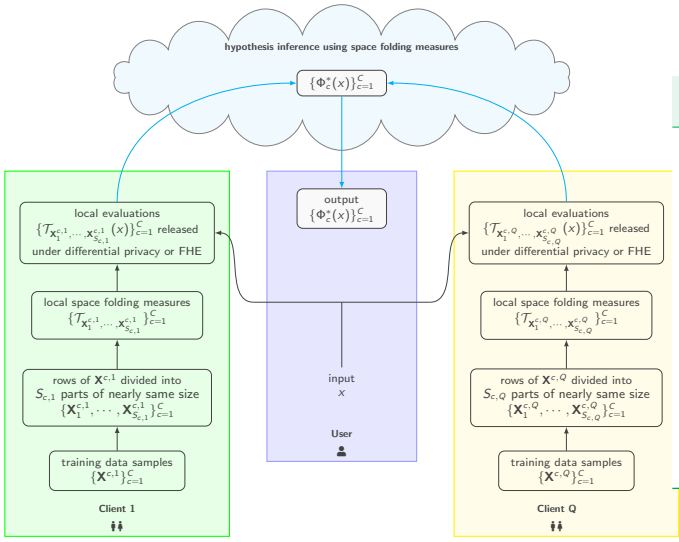


(g) The given 2-dimensional data samples $\mathbf{X}$ have been marked in blue color as "*", and a red line connects a point to its image under KAHM (i.e. x is connected to $\mathcal{A}_{\mathbf{X}}(x)$ through a line).



(h) The color plot of the space folding measure function $\mathcal{T}_{\mathbf{X}}$ (option 1). The given 2-dimensional data samples $\mathbf{X}$ have been marked in white color as "*".

Federated ML ▶ Protocol



Beyond Gradient-Descent

KAHMs allows for federated learning with

- no multiple epochs of local optimisation using stochastic gradient descent, and
- no rounds of communication between client/server for optimising the global model,

while addressing all of target requirements R1-R5.

scch {
software
competence
center
hagenberg
}

CONCLUDING REMARKS

Concluding Remarks ▶

1. **Operator theory gives the learning problem a principled route:** L^2 optimum $\rightarrow$ RKHS approximation $\rightarrow$ inverse-mapped generalized solution.
2. **The hypothesis space is not arbitrary:** kernel normalization and Rademacher analysis yield finite-sample prediction and approximation bounds of order $O(N^{-1/2})$.
3. **KAHM space folding makes the theory deployable:** scalar summaries enable gradient-free FL, compatible with differential privacy and FHE inference.

Based on: Kumar et al., *Operator-Theoretic Framework for Gradient-Free Federated Learning*, arXiv:2512.01025v1.

scch {
software
competence
center
hagenberg
}

Contact

Mohit Kumar, Bernhard Moser, Manuela Geiß

- Looking forward to your questions...

Email: mohit.kumar@scch.at

scch {
software
competence
center
hagenberg
}

BIBLIOGRAPHY

References I

Mohit Kumar, Bernhard A. Moser, and Lukas Fischer.

On mitigating the utility-loss in differentially private learning: A new perspective by a geometrically inspired kernel approach.

Journal of Artificial Intelligence Research, 79:515–567, 2024.

Mohit Kumar, Alexander Valentinitsch, Magdalena Fuchs, Mathias Brucker, Juliana Bowles, Adnan Husakovic, Ali Abbas, and Bernhard A. Moser.

Geometrically inspired kernel machines for collaborative learning beyond gradient descent.

Journal of Artificial Intelligence Research, 83, July 2025.