ICML | 2022

Thirty-ninth International Conference on Machine Learning

The Teaching Dimension of Regularized Kernel Learners

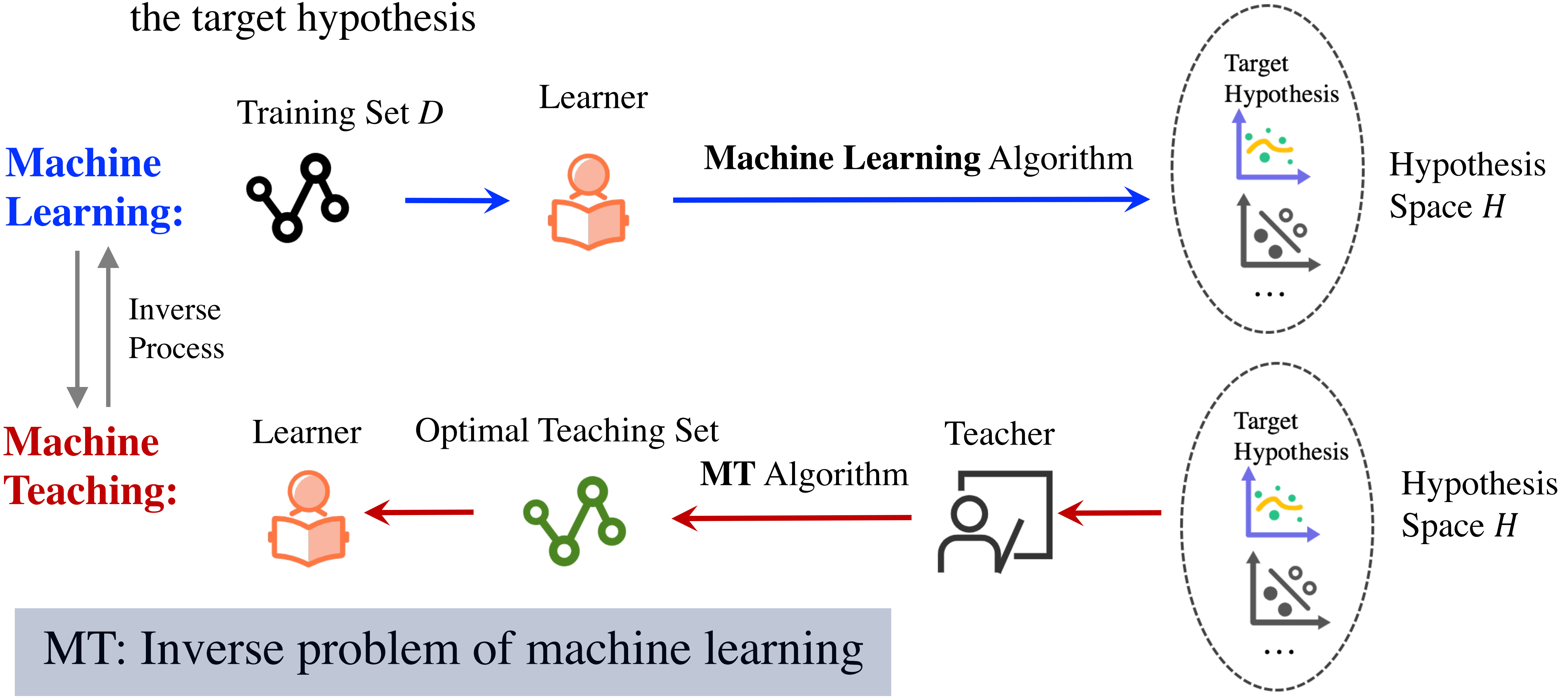
Hong Qian, Xu-Hui Liu, Chen-Xi Su, Aimin Zhou, Yang Yu

East China Normal University, Shanghai, China

Nanjing University, Nanjing, China

What is Machine Teaching (MT)?

MT: Design an **optimal** teaching set to steer a learner (student) towards the target hypothesis



MT: Inverse problem of machine learning

Why Machine Teaching (MT)?

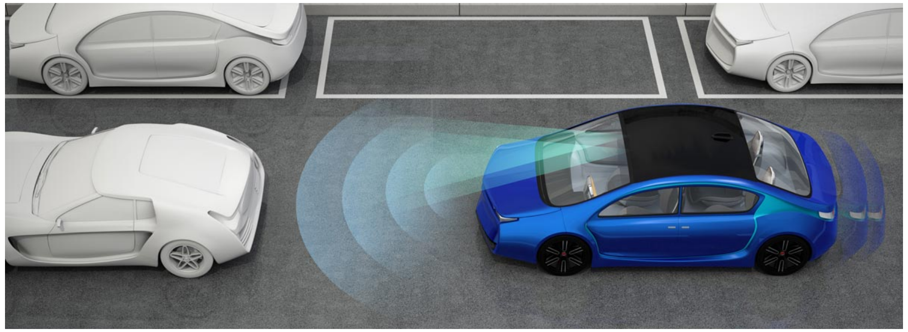
Sometimes, a teacher knows the target hypothesis, but she cannot telepathize it into the learner's mind directly

Example A: Teaching students to categorize flowers

Rosa Chinensis

Rose

Example B: Autonomous driving in reinforcement learning (RL)



Many applications

- **Education** [Patil et al., 2014]
- **RL** [Kamalaruban et al., 2019]
- **Trustworthy AI** [Zhang et al., 2018]
- **Cognitive Psychology** [Shafto et al., 2014]

Convey the target hypothesis via the designed optimal teaching set

Motivation: High Teaching Dimension in MT

Teaching dimension (TD): Measure the teaching complexity

The minimum number of teaching examples required to teach the target hypothesis to a learner

For teaching the empirical risk minimization (ERM) learners

- Liu et al. analyze linear learners

- Kumar et al. generalize them to non-linear learners by introducing kernels

  Suffer from **high TD**

  Only consider polynomial and Gaussian kernels for non-linear cases

Our goal: to reduce TD, to analyze any type of kernels

[Liu et al. The Teaching Dimension of Linear Learners. ICML 2016, JMLR 2016.]
[Kumar et al. The Teaching Dimension of Kernel Perceptron. AISTATS 2021.]

Method

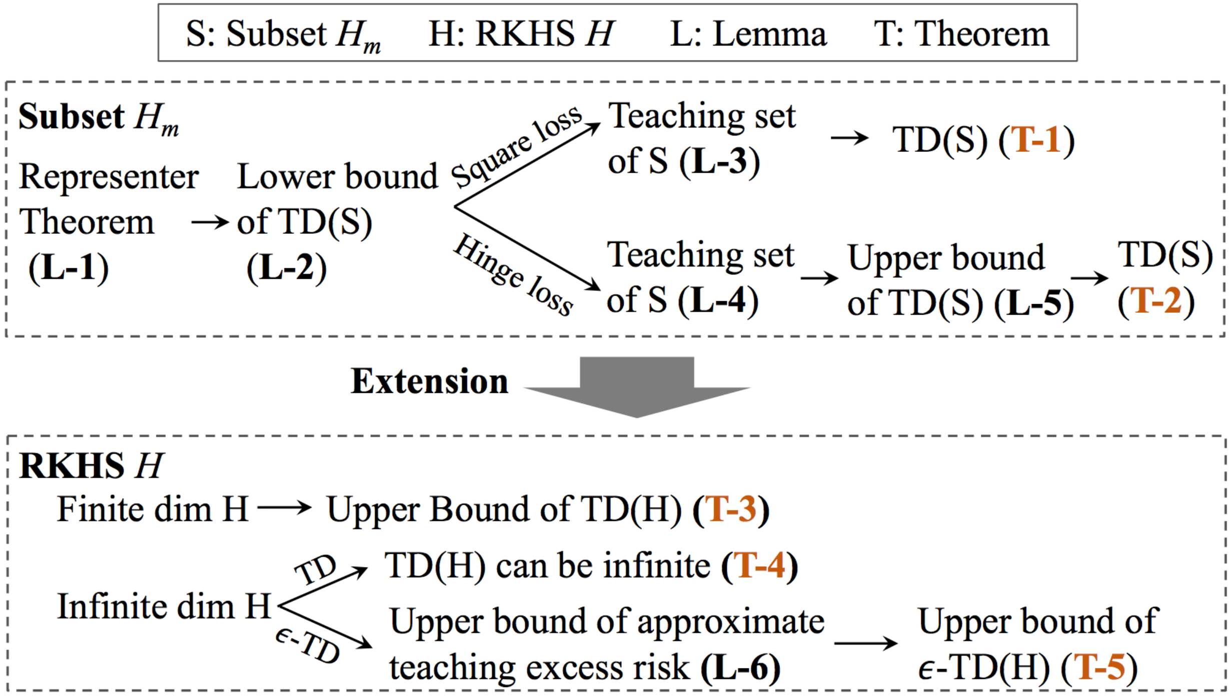
Inspired by machine learning, adding regularization to reduce the teaching complexity

Regularized ERM kernel learners:

$$\mathcal{A}(D) = \arg\min_{\boldsymbol{\theta} \in H} \sum_{i=1}^{n} \ell(\langle \boldsymbol{\theta}, k(\mathbf{x}_i, \cdot) \rangle, y_i) + \Omega(||\boldsymbol{\theta}||_H^2)$$

We propose a unified theoretical framework STARKE for teaching the regularized ERM kernel learners

- TD could be (significantly) reduced
- Can analyze any type of kernels



The STARKE framework

Results

Theoretical result

Kernel Type (TD Type)	*With* Regularization (This Paper)	*Without* Regularization (Kumar et al., 2021)
Linear (TD)	1	$\Theta(d)$
Polynomial (TD)	$\mathcal{G}^*(d,p)$	$\mathrm{TD} \geq \binom{d+p-1}{p}$
Gaussian (TD)	∞	∞
Gaussian (ϵ-TD)	$O(1/\epsilon^2)$	$d^{O(\log^2(1/\epsilon))}$

ϵ-TD tolerates ϵ excess error to handle the infinite TD scenarios

d: dimension of input space

p: degree of polynomial kernel

[Kumar et al. The Teaching Dimension of Kernel Perceptron. AISTATS 2021.]

Empirical result

The difference between ϵ-TD without regularization and ϵ-TD with regularization

∘: unregularized learner cannot reach such a ratio within 60 samples

×: both regularized and unregularized learners cannot reach such a ratio within 60 samples

Dataset	$\bar{\Lambda} = 100\%$	80%	60%	40%	20%	0%
Sin	∘	∘	∘	∘	∘	∘
MR	4	5	4	9	8	×
MPG	23	25	27	29	∘	×
Eunite	∘	∘	∘	∘	∘	×
Circle	∘	∘	∘	∘	∘	∘
Moon	25	26	∘	∘	∘	∘
Adult	∘	∘	∘	×	×	×
Sonar	∘	∘	∘	∘	×	×

Excess risk ratio: Normalized excess risk

- Sin, Eunite, Circle, Moon, Sonar: hard for teaching without regularization
- MR and MPG: easy (MR is easier)
- Adult: hard for both

Summary and Future Work

Take-home message

Regularization is able to reduce the teaching complexity (TD) in machine teaching

STARKE can analyze the regularized ERM learners with any type of kernels

Future work: bridging between VC dimension and TD in non-linear cases

Machine Learning ⟷ (relationship) ⟷ **Machine Teaching**

VC dimension ⟷ Teaching dimension

Machine Learning	Machine Teaching
VC dimension	Teaching dimension
• Measure the complexity of learning problems	• Measure the complexity of teaching problems

Thank You !





Hong QIAN
@ECNU
Equal contribution

Xu-Hui LIU
@NJU
Equal contribution

Chen-Xi SU
@ECNU

Aimin ZHOU
@ECNU

Yang YU
@NJU

Contact: hqian@cs.ecnu.edu.cn