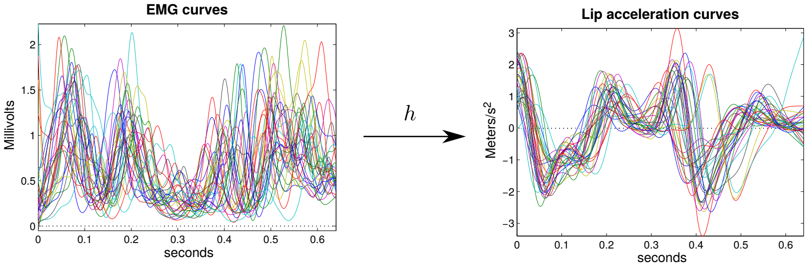
On the Dualization of Operator-Valued Kernel Machines

Pierre Laforgue*, Alex Lambert*, Luc Brogat-Motte, Florence d'Alché-Buc

* Equal contribution
LTCI, Télécom Paris, Institut Polytechnique de Paris, France

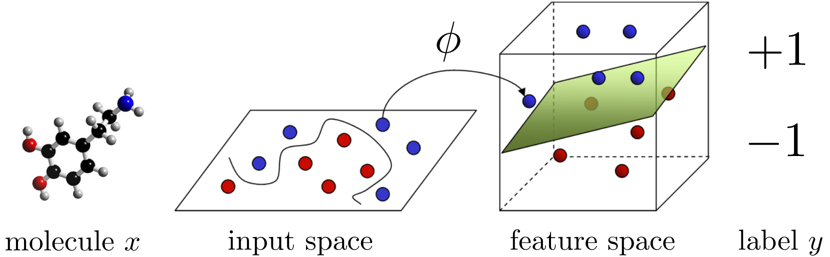
Outline

Function to Function Regression [Kadri et al., 2016]



Input functions $\{x_i\}_{i\leq n}$ Output functions $\{y_i\}_{i\leq n}$

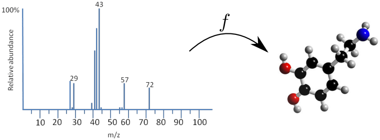
$$\underset{h\in\mathcal{H}_\mathcal{K}}{\operatorname{argmin}} \quad \frac{1}{2n}\sum_{i=1}^{n}\|y_i - h(x_i)\|_{L^2}^2 + \frac{\Lambda}{2}\|h\|^2$$



$$\min_{h \in \mathrm{RKHS}} \quad \frac{1}{n}\sum_{i=1}^{n} \ell_{\mathrm{hinge}}\left(h(x_i), y_i\right) + \frac{\Lambda}{2}\|h\|^2,$$

$$\min_{h \in \mathrm{RKHS}} \quad \frac{1}{n}\sum_{i=1}^{n} \ell_{\mathrm{hinge}}\left(\langle h, \phi(x_i)\rangle, y_i\right) + \frac{\Lambda}{2}\|h\|^2.$$

Metabolite Identification [Brouard et al., 2016]



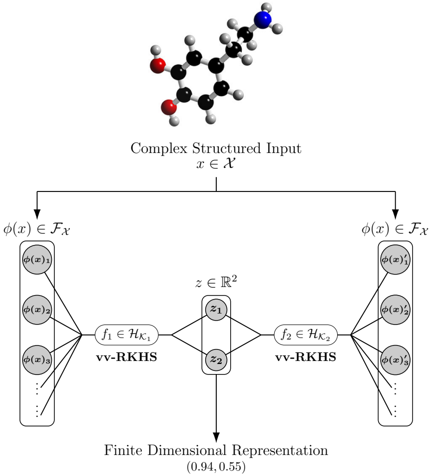
Mass spectrum x Molecule y

$$\begin{tikzcd}
 & \phi(y) \in \mathcal{F}_{\mathcal{Y}} \arrow[d, dashed, bend left, "\text{argmin}"] \\
x \in \mathcal{X} \arrow[ur, "\hat{h} \in \mathcal{H}_{\mathcal{K}}"] \arrow[r, "g"'] & y \in \mathcal{Y} \arrow[u, "\phi"]
\end{tikzcd}$$

$$\hat{h} = \underset{h \in \mathcal{H}_{\mathcal{K}}}{\text{argmin}} \ \frac{1}{2n} \sum_{i=1}^{n} \|\phi(y_i) - h(x_i)\|_{\mathcal{F}_{\mathcal{Y}}}^2 + \frac{\Lambda}{2} \|h\|_{\mathcal{H}_{\mathcal{K}}}^2,$$

$$g(x) = \underset{y \in \mathcal{Y}}{\text{argmin}} \ \left\|\phi(y) - \hat{h}(x)\right\|_{\mathcal{F}_{\mathcal{Y}}}.$$

Molecule Autoencoding [Laforgue et al., 2019a]



$$\frac{1}{2n}\sum_{i=1}^{n} \|\phi(x_i) - f_2 \circ f_1(\phi(x_i))\|^2_{\mathcal{F}_{\mathcal{X}}} + \Lambda \operatorname{Reg}(f_1, f_2).$$

Outline

Operator-Valued Kernels and vv-RKHSs

Scalar Kernels

- $k \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- $k(x, x') = k(x', x)$
- $\sum_{i,j} \alpha_i k(x_i, x_j) \alpha_j \geq 0$

- RKHS $\mathcal{H}_k \subset \mathcal{F}(\mathcal{X}, \mathbb{R})$

  $\mathcal{H}_k = \overline{\mathrm{Span}\left\{k(\cdot, x) : x \in \mathcal{X}\right\}}$
- $\forall h \in \mathcal{H}_k,\ h(x) = \langle h, k(\cdot, x) \rangle$

Operator-Valued Kernels and vv-RKHSs

Scalar Kernels

- $k \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- $k(x, x') = k(x', x)$
- $\sum_{i,j} \alpha_i k(x_i, x_j) \alpha_j \geq 0$

- RKHS $\mathcal{H}_k \subset \mathcal{F}(\mathcal{X}, \mathbb{R})$

  $\mathcal{H}_k = \overline{\text{Span}\left\{k(\cdot, x) : x \in \mathcal{X}\right\}}$
- $\forall h \in \mathcal{H}_k,\ h(x) = \langle h, k(\cdot, x) \rangle$

Operator-Valued Kernels

- $\mathcal{K} \colon \mathcal{X} \times \mathcal{X} \to \mathcal{L}(\mathcal{Y})$
- $\mathcal{K}(x, x') = \mathcal{K}(x', x)^*$
- $\sum_{i,j} \langle \alpha_i, \mathcal{K}(x_i, x_j) \alpha_j \rangle_{\mathcal{Y}} \geq 0$

- vv-RKHS $\mathcal{H}_{\mathcal{K}} \subset \mathcal{F}(\mathcal{X}, \mathcal{Y})$

  $\mathcal{H}_{\mathcal{K}} = \overline{\text{Span}\left\{\mathcal{K}(\cdot, x) y : x, y \in \mathcal{X} \times \mathcal{Y}\right\}}$
- $\forall h \in \mathcal{H}_{\mathcal{K}},\ \langle h(x), y \rangle = \langle h, \mathcal{K}(\cdot, x) y \rangle$

- Decomposable: $\mathcal{K}(x, x') = k(x, x') A$

The Representer Theorem

$$\hat{h} \in \underset{h \in \mathcal{H}_k}{\operatorname{argmin}} \; V\Big(\{h(x_i)\}_{i \le n}, \|h\|\Big),$$

with V increasing w.r.t. its last argument, then $\exists \hat{\boldsymbol{\alpha}} \in \mathbb{R}^n$ s.t.

$$\hat{h} = \sum_{i=1}^{n} k(\cdot, x_i)\hat{\alpha}_i$$

The Representer Theorem

$$\hat{h} \in \underset{h \in \mathcal{H}_k}{\operatorname{argmin}} \; V\Big(\{h(x_i)\}_{i \leq n}, \|h\|\Big),$$

with V increasing w.r.t. its last argument, then $\exists \hat{\boldsymbol{\alpha}} \in \mathbb{R}^n$ s.t.

$$\hat{h} = \sum_{i=1}^{n} k(\cdot, x_i)\hat{\alpha}_i$$

Proof. $E = \mathrm{Span}(k(\cdot, x_i), i \leq n)$, $\hat{h} = \bar{h} + h^{\perp}$ with $\bar{h}, h^{\perp} \in E \times E^{\perp}$.

$\hat{h}(x_i) = \langle \bar{h} + h^{\perp}, k(\cdot, x_i) \rangle = \bar{h}(x_i)$ and $\|\hat{h}\| \geq \|\bar{h}\|$, so $h^{\perp} = 0$.

$$\hat{h} \in \underset{h \in \mathcal{H}_k}{\operatorname{argmin}} \; V\Big(\{h(x_i)\}_{i \leq n}, \|h\|\Big),$$

with V increasing w.r.t. its last argument, then $\exists \hat{\boldsymbol{\alpha}} \in \mathbb{R}^n$ s.t.

$$\hat{h} = \sum_{i=1}^{n} k(\cdot, x_i)\hat{\alpha}_i$$

Proof. $E = \mathrm{Span}(k(\cdot, x_i), i \leq n)$, $\hat{h} = \bar{h} + h^{\perp}$ with $\bar{h}, h^{\perp} \in E \times E^{\perp}$.

$\hat{h}(x_i) = \langle \bar{h} + h^{\perp}, k(\cdot, x_i)\rangle = \bar{h}(x_i)$ and $\|\hat{h}\| \geq \|\bar{h}\|$, so $h^{\perp} = 0$.

OVK version: $\exists (\hat{\alpha}_i)_{i \leq n} \in \mathcal{Y}^n, \qquad \hat{h} = \sum_{i=1}^{n} \mathcal{K}(\cdot, x_i)\hat{\alpha}_i$

OVK Ridge Regression [Micchelli and Pontil, 2005]

$$\min_{f \in \mathcal{H}_{\mathcal{K}}} \quad \frac{1}{2n} \sum_{i=1}^{n} \|y_i - h(x_i)\|_{\mathcal{Y}}^2 + \frac{\Lambda}{2} \|h\|_{\mathcal{H}_{\mathcal{K}}}^2$$

OVK Ridge Regression [Micchelli and Pontil, 2005]

$$\min_{f \in \mathcal{H}_\mathcal{K}} \quad \frac{1}{2n}\sum_{i=1}^{n} \|y_i - h(x_i)\|_\mathcal{Y}^2 + \frac{\Lambda}{2}\|h\|_{\mathcal{H}_\mathcal{K}}^2$$

The Representer Theorem applies: $h = \sum_i \mathcal{K}(\cdot, x_i)\alpha_i$.

Plugging and differentiating w.r.t. the $(\alpha_i)_{i \leq n}$

$$\begin{cases} \sum_{i=1}^{n} \left(\mathcal{K}(x_1, x_i) + \Lambda n \delta_{1i} \mathbf{I}_\mathcal{Y}\right) \hat{\alpha}_i = y_1, \\ \qquad \qquad \dots \\ \sum_{i=1}^{n} \left(\mathcal{K}(x_n, x_i) + \Lambda n \delta_{ni} \mathbf{I}_\mathcal{Y}\right) \hat{\alpha}_i = y_n. \end{cases}$$

OVK Ridge Regression [Micchelli and Pontil, 2005]

$$\min_{f \in \mathcal{H}_{\mathcal{K}}} \quad \frac{1}{2n} \sum_{i=1}^{n} \|y_i - h(x_i)\|_{\mathcal{Y}}^2 + \frac{\Lambda}{2} \|h\|_{\mathcal{H}_{\mathcal{K}}}^2$$

The Representer Theorem applies: $h = \sum_i \mathcal{K}(\cdot, x_i)\alpha_i$.

Plugging and differentiating w.r.t. the $(\alpha_i)_{i \leq n}$

$$\begin{cases} \sum_{i=1}^{n} \left(\mathcal{K}(x_1, x_i) + \Lambda n \delta_{1i} \mathbf{I}_{\mathcal{Y}}\right) \hat{\alpha}_i = y_1, \\ \qquad \qquad \dots \\ \sum_{i=1}^{n} \left(\mathcal{K}(x_n, x_i) + \Lambda n \delta_{ni} \mathbf{I}_{\mathcal{Y}}\right) \hat{\alpha}_i = y_n. \end{cases}$$

$\mathcal{K}(x, x') = k(x, x')\mathbf{I}_{\mathcal{Y}}$, then $\hat{\alpha}_i = \sum_j M_{ij} y_j$ with $M = (K + \Lambda n \mathbf{I}_n)^{-1}$

Outline

The OVK Dual Problem [Sangnier et al., 2017]

The solution to

$$\min_{h \in \mathcal{H}_\mathcal{K}} \quad \frac{1}{n}\sum_{i=1}^{n} \ell_i(h(x_i)) + \frac{\Lambda}{2}\|h\|^2_{\mathcal{H}_\mathcal{K}}$$

is given by

$$\hat{h} = \frac{1}{\Lambda n}\sum_{i=1}^{n} \mathcal{K}(\cdot, x_i)\hat{\alpha}_i,$$

with $(\hat{\alpha}_i)_{i=1}^n \in \mathcal{Y}^n$ the solutions to the **dual problem**

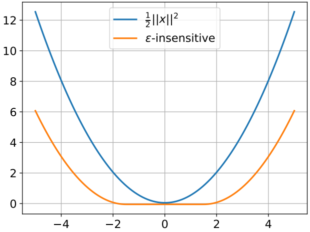
$$\min_{(\alpha_i)_{i=1}^n \in \mathcal{Y}^n} \quad \sum_{i=1}^{n} \ell_i^\star(-\alpha_i) + \frac{1}{2\Lambda n}\sum_{i,j=1}^{n} \langle \alpha_i, \mathcal{K}(x_i, x_j)\alpha_j \rangle_\mathcal{Y},$$

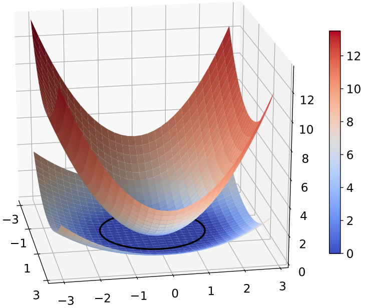
where $f^\star : \alpha \in \mathcal{Y} \mapsto \sup_{y \in \mathcal{Y}} \langle \alpha, y \rangle_\mathcal{Y} - f(y)$ denotes the Fenchel-Legendre transform of a function $f : \mathcal{Y} \to \mathbb{R}$.

The ϵ-insensitive Losses

Let $\ell : \mathcal{Y} \to \mathbb{R}$ be a convex loss with unique minimum at 0, and $\epsilon > 0$. The ϵ-insensitive version of ℓ, denoted ℓ_ϵ, is defined by:

$$\forall y \in \mathcal{Y}, \quad \ell_\epsilon(y) = \begin{cases} \ell(0) & \text{if } \|y\|_{\mathcal{Y}} \leq \epsilon \\ \inf\limits_{\|d\|_{\mathcal{Y}} \leq 1} \ell(y - \epsilon d) & \text{otherwise} \end{cases}$$





Infimal Convolution and Fenchel-Legendre Transforms

$$\left(\frac{1}{2}\|\cdot\|^2\right)_\epsilon = \frac{1}{2}\|\cdot\|^2 \,\square\, \chi_{\mathcal{B}_\epsilon},$$

with $\square$ the infimal-convolution operator [Bauschke et al., 2011] such that $(f \,\square\, g)(x) = \inf_y f(y) + g(x-y)$,

and χ_S the indicator function of S: $\chi_S(x) = \begin{cases} 0 & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases}$.

$$\left(\frac{1}{2}\|\cdot\|^2\right)^\star_\epsilon = \left(\frac{1}{2}\|\cdot\|^2 \,\square\, \chi_{\mathcal{B}_\epsilon}\right)^\star = \left(\frac{1}{2}\|\cdot\|^2\right)^\star + \chi^\star_{\mathcal{B}_\epsilon} = \frac{1}{2}\|\cdot\|^2 + \epsilon\|\cdot\|.$$

ϵ-Ridge Dual Problem (1/2)

The solution to

$$\min_{h \in \mathcal{H}_{\mathcal{K}}} \quad \frac{1}{n} \sum_{i=1}^{n} \|y_i - h(x_i)\|_{\epsilon}^2 + \frac{\Lambda}{2} \|h\|_{\mathcal{H}_{\mathcal{K}}}^2$$

is given by

$$\hat{h} = \frac{1}{\Lambda n} \sum_{i=1}^{n} \mathcal{K}(\cdot, x_i) \hat{\alpha}_i,$$

with $(\hat{\alpha}_i)_{i=1}^n \in \mathcal{Y}^n$ minimizing

$$\frac{1}{2\Lambda n} \sum_{i,j=1}^{n} \left\langle \alpha_i, \left(\mathcal{K}(x_i, x_j) + \Lambda n \mathbf{I}_{\mathcal{Y}}\right) \alpha_j \right\rangle_{\mathcal{Y}} + \epsilon \sum_{i=1}^{n} \|\alpha_i\|_{\mathcal{Y}} - \sum_{i=1}^{n} \langle \alpha_i, y_i \rangle_{\mathcal{Y}}$$

ϵ-Ridge Dual Problem (2/2)

$$\frac{1}{2\Lambda n}\sum_{i,j=1}^{n}\left\langle \alpha_i, \left(\mathcal{K}(x_i, x_j) + \Lambda n \mathbf{I}_{\mathcal{Y}}\right)\alpha_j \right\rangle_{\mathcal{Y}} + \epsilon \sum_{i=1}^{n} \|\alpha_i\|_{\mathcal{Y}} - \sum_{i=1}^{n} \langle \alpha_i, y_i \rangle_{\mathcal{Y}}$$

ϵ-Ridge Dual Problem (2/2)

$$\frac{1}{2\Lambda n}\sum_{i,j=1}^{n}\left\langle \alpha_i, \left(\mathcal{K}(x_i,x_j)+\Lambda n \mathbf{I}_{\mathcal{Y}}\right)\alpha_j \right\rangle_{\mathcal{Y}} + \epsilon\sum_{i=1}^{n}\|\alpha_i\|_{\mathcal{Y}} - \sum_{i=1}^{n}\langle \alpha_i, y_i\rangle_{\mathcal{Y}}$$

$$Y = \mathrm{Span}(y_i, i \leq n), \qquad \forall i \quad \alpha_i = \alpha_i^{Y} + \alpha_i^{\perp} \in Y \times Y^{\perp}$$

ϵ-Ridge Dual Problem (2/2)

$$\frac{1}{2\Lambda n}\sum_{i,j=1}^{n}\left\langle \alpha_i, \left(\mathcal{K}(x_i,x_j)+\Lambda n \mathbf{I}_{\mathcal{Y}}\right)\alpha_j\right\rangle_{\mathcal{Y}} + \epsilon\sum_{i=1}^{n}\|\alpha_i\|_{\mathcal{Y}} - \sum_{i=1}^{n}\langle \alpha_i, y_i\rangle_{\mathcal{Y}}$$

$$Y = \mathrm{Span}(y_i, i \leq n), \qquad \forall i \quad \alpha_i = \alpha_i^Y + \alpha_i^\perp \in Y \times Y^\perp$$

If for all $i, j \leq n$, Y is an invariant subspace of $\mathcal{K}(x_i, x_j)$, then $\hat{\alpha}_i \in Y$ for all $i \leq n$.

Reparametrization $\hat{\alpha}_i = \sum_j \omega_{ij} y_j \rightarrow$ dual problem in Ω computable.

The Double Representer Theorem [Laforgue et al., 2019b]

- $\forall i \leq n,\ \forall (\alpha^Y, \alpha^\perp) \in Y \times Y^\perp, \quad \ell_i^\star(\alpha^Y) \leq \ell_i^\star(\alpha^Y + \alpha^\perp)$.
- $\forall i \leq n, \exists L_i : \mathbb{R}^{n+n^2} \to \mathbb{R}, \forall\ \boldsymbol{\omega} \in \mathbb{R}^n, \quad \ell_i^\star(\sum_j \omega_j\ y_j) = L_i(\boldsymbol{\omega}, K^Y)$.
- $\forall i, j \leq n, \quad Y$ is an invariant subspace of $\mathcal{K}(x_i, x_j)$.

Then, $\hat{h} = \underset{\mathcal{H}_\mathcal{K}}{\operatorname{argmin}}\ \frac{1}{n}\sum_i \ell(h(x_i), y_i) + \frac{\Lambda}{2}\|h\|^2_{\mathcal{H}_\mathcal{K}}$ is given by

$$\hat{h} = \frac{1}{\Lambda n}\sum_{i,j=1}^{n} \mathcal{K}(\cdot, x_i)\ \hat{\omega}_{ij}\ y_j,$$

with $\hat{\Omega} = [\hat{\omega}_{ij}] \in \mathbb{R}^{n \times n}$ the solution to

$$\min_{\Omega \in \mathbb{R}^{n \times n}} \ \sum_{i=1}^{n} L_i\left(\Omega_{i:}, K^Y\right) + \frac{1}{2\Lambda n}\mathbf{Tr}\left(\tilde{M}^\top(\Omega \otimes \Omega)\right),$$

with $\tilde{M}$ the $n^2 \times n^2$ matrix writing of M s.t. $M_{ijkl} = \langle y_k, \mathcal{K}(x_i, x_j) y_l\rangle_\mathcal{Y}$.

The Double Representer Theorem (cont.)

If $\mathcal{K}$ further satisfies $\mathcal{K}(x, x') = \sum_t k_t(x, x') A_t$, then tensor M simplifies to $M_{ijkl} = \sum_t [K_t^X]_{ij} [K_t^Y]_{kl}$ and the problem rewrites

$$\min_{\Omega \in \mathbb{R}^{n \times n}} \sum_{i=1}^{n} L_i \left(\Omega_{i:}, K^Y \right) + \frac{1}{2\Lambda n} \sum_{t=1}^{T} \mathbf{Tr} \left(K_t^X \Omega K_t^Y \Omega^\top \right).$$

Rmk. Only need the n^4 tensor $\langle y_k, \mathcal{K}(x_i, x_j) y_l \rangle_{\mathcal{Y}}$ to learn OVKMs.

Simplifies to 2 n^2 matrices $K_{ij}^X K_{kl}^Y$ if $\mathcal{K}$ is decomposable.

Admissible Losses

- $\ell_i(y) = f(\langle y, z_i \rangle)$, $z_i \in Y$ and $f : \mathbb{R} \to \mathbb{R}$ convex. (MMR)
- $\ell(y) = f(\|y\|)$, $f : \mathbb{R}_+ \to \mathbb{R}$ cvx inc. s.t. $t \mapsto \frac{f'(t)}{t}$ is cont.
- $\forall \lambda > 0$, with $\mathcal{B}_\lambda$ the centered ball of radius λ,
  - $\ell(y) = \lambda\|y\|$,
  - $\ell(y) = \lambda\|y\| \log(\|y\|)$,
  - $\ell(y) = \chi_{\mathcal{B}_\lambda}(y)$,
  - $\ell(y) = \lambda(\exp(\|y\|) - 1)$.
- $\ell_i(y) = f(y - y_i)$ for f above.
- Any infimal convolution of previous functions

Back to ϵ-Ridge Regression

If $\mathcal{K} = k\mathbf{I}_{\mathcal{Y}}$, the solution to the ϵ-Ridge regression problem

$$\min_{h \in \mathcal{H}_{\mathcal{K}}} \quad \frac{1}{2n} \sum_{i=1}^{n} \max\Big(\|y_i - h(x_i)\|_{\mathcal{Y}} - \epsilon, 0 \Big)^2 + \frac{\Lambda}{2} \|h\|^2_{\mathcal{H}_{\mathcal{K}}},$$

is given by the *Double Representer Theorem*, with $\hat{\Omega} = \hat{W} V^{-1}$, and $\hat{W}$ the solution to the **Multi-Task Lasso problem**

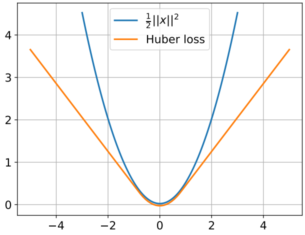
$$\boxed{\min_{W \in \mathbb{R}^{n \times n}} \quad \frac{1}{2} \|AW - B\|^2_{\mathrm{Fro}} + \epsilon \, \|W\|_{2,1},}$$

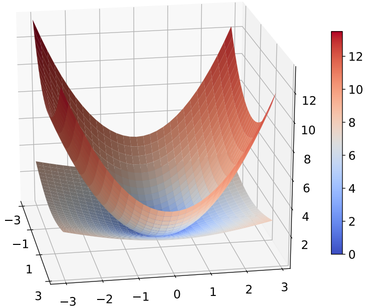
with V, A, B s.t. $K^Y = VV^\top$, $\frac{K^X}{\Lambda n} + \mathbf{I}_n = A^\top A$, and $V = A^\top B$.

The Huber Loss

For $\kappa > 0$, the Huber loss is $\ell_{H,\kappa}(y) = \left(\kappa\|\cdot\|_{\mathcal{Y}} \,\square\, \frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2\right)(y)$, or:

$$\forall y \in \mathcal{Y}, \quad \ell_{H,\kappa}(y) = \begin{cases} \frac{1}{2}\|y\|_{\mathcal{Y}}^2 & \text{if } \|y\|_{\mathcal{Y}} \leq \kappa \\ \kappa\left(\|y\|_{\mathcal{Y}} - \frac{\kappa}{2}\right) & \text{otherwise} \end{cases}$$





The Huber Dual

If $\mathcal{K} = k\mathbf{I}_{\mathcal{Y}}$, the solution to the Huber regression problem

$$\min_{h \in \mathcal{H}_{\mathcal{K}}} \frac{1}{n} \sum_{i=1}^{n} \ell_{H,\kappa}(y_i - h(x_i)) + \frac{\Lambda}{2} \|h\|_{\mathcal{H}_{\mathcal{K}}}^2,$$

is given by the *Double Representer Theorem*, with $\hat{\Omega} = \hat{W} V^{-1}$, and $\hat{W}$ the solution to the **constrained least squares problem**

$$\boxed{\begin{aligned} \min_{W \in \mathbb{R}^{n \times n}} \quad & \frac{1}{2} \|AW - B\|_{\text{Fro}}^2, \\ \text{s.t.} \quad & \|W\|_{2,\infty} \leq \kappa, \end{aligned}}$$

with V, A, and B as in the ϵ-insensitive Regression.

Outline

Surrogate Approaches

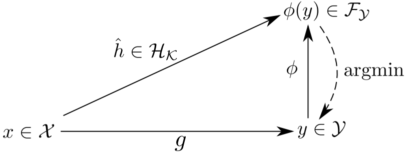


Figure 1: Structured Prediction (Output Kernel Regression)

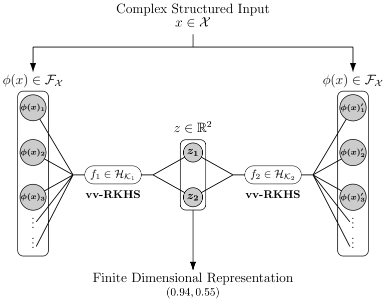


Figure 2: Representation Learning (2-Layer Kernel Autoencoder)

Results

Figure 3: MSEs w.r.t. ϵ.

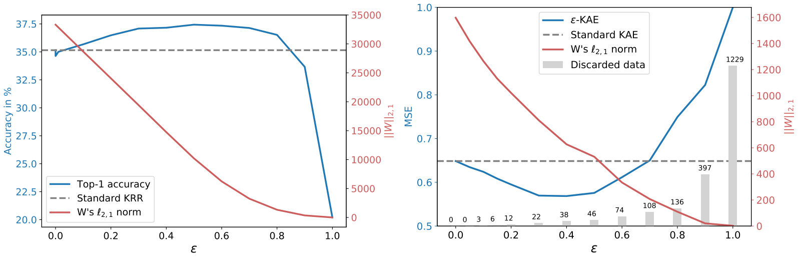


Table 1: Huber test accuracies (%) with respect to κ

κ	Top 1	Top 10	Top 20	$\|W\|_{2,1}$
0.5	38.0	83.5	89.6	2789.6
1.0	**38.9**	**83.8**	**89.9**	5572.4
1.5	38.6	83.7	89.8	8231.9

Outline

Conclusion

- OVK and vv-RKHSs tailored to infinite dimensional outputs y
- RT: expansion with no information on the coefficients
- Duality: coefficients solutions to the (infinite) dual problem

Conclusion

- OVK and vv-RKHSs tailored to infinite dimensional outputs y
- RT: expansion with no information on the coefficients
- Duality: coefficients solutions to the (infinite) dual problem
- Double RT: coefficients linear combinations of the outputs
- Allows to cope with many losses (ϵ, Huber) and kernels
- Empirical improvements on surrogate approaches

Conclusion

- OVK and vv-RKHSs tailored to infinite dimensional outputs y
- RT: expansion with no information on the coefficients
- Duality: coefficients solutions to the (infinite) dual problem
- Double RT: coefficients linear combinations of the outputs
- Allows to cope with many losses (ϵ, Huber) and kernels
- Empirical improvements on surrogate approaches
- Integral Losses: $\ell(y, f(x)) = \int_\Theta \ell_\theta\Big(y(\theta), f(x)(\theta)\Big) d\mu(\theta)$
- Preprint available at: arxiv.org/1910.04621

References I

Bauschke, H. H., Combettes, P. L., et al. (2011).
***Convex analysis and monotone operator theory in Hilbert spaces*, volume 408.**
Springer.

Brouard, C., Szafranski, M., and d'Alché-Buc, F. (2016).
Input output kernel regression: Supervised and semi-supervised structured output prediction with operator-valued kernel.
Journal of Machine Learning Research, 17:176:1–176:48.

References II

Kadri, H., Duflos, E., Preux, P., Canu, S., Rakotomamonjy, A., and Audiffren, J. (2016).
Operator-valued kernels for learning from functional response data.
Journal of Machine Learning Research, 17:20:1–20:54.

Laforgue, P., Clémençon, S., and d'Alché-Buc, F. (2019a).
Autoencoding any data through kernel autoencoders.
In *Artificial Intelligence and Statistics*, pages 1061–1069.

Laforgue, P., Lambert, A., Motte, L., and d'Alché Buc, F. (2019b).
On the dualization of operator-valued kernel machines.
arXiv preprint arXiv:1910.04621.

Micchelli, C. A. and Pontil, M. (2005).
On learning vector-valued functions.
Neural computation, 17(1):177–204.

Sangnier, M., Fercoq, O., and d'Alché-Buc, F. (2017).
Data sparse nonparametric regression with ϵ-insensitive losses.
In *Asian Conference on Machine Learning*, pages 192–207.