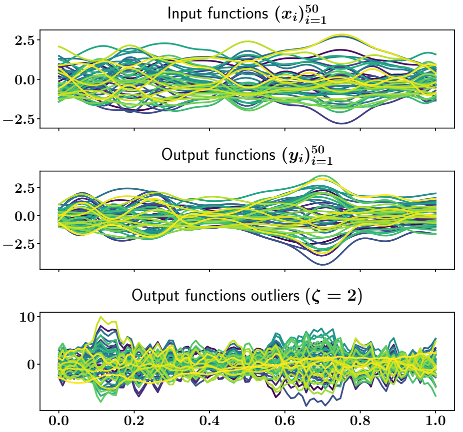
Functional Output Regression with Infimal Convolution: Exploring the Huber and ϵ-insensitive Losses

Alex Lambert, Dimitri Bouche, Zoltán Szabó, Florence d'Alché-Buc

ICML 2022

Challenges in functional output regression

- Regression when the target variable is a function [Kad+16], in presence of outliers



Challenges in functional output regression

Square loss fails to handle outliers.

Goal

Go beyond the square loss in functional output regression

Challenges in functional output regression

Square loss fails to handle outliers.

Goal

Go beyond the square loss in functional output regression

Key idea: use convoluted losses [Laf+20] of the form

$$L = \frac{1}{2}\|\cdot\|^2 \,\square\, g$$

where g is to be designed to capture outliers or impose sparsity

Regularized empirical risk minimization

Regularized empirical risk minimization in vv-RKHSs:

$$\inf_{h \in \mathcal{H}_K} \frac{1}{n} \sum_{i=1}^{n} L(y_i - h(x_i)) + \frac{\lambda}{2} \|h\|^2_{\mathcal{H}_K}$$

Regularized empirical risk minimization in vv-RKHSs:

$$\inf_{h \in \mathcal{H}_K} \frac{1}{n} \sum_{i=1}^{n} L(y_i - h(x_i)) + \frac{\lambda}{2} \|h\|_{\mathcal{H}_K}^2$$

- Modelization choice $(y_i)_{i=1}^n \in \mathcal{Y} := L^2[\Theta, \mu]$

Regularized empirical risk minimization

Regularized empirical risk minimization in vv-RKHSs:

$$\inf_{h \in \mathcal{H}_K} \frac{1}{n} \sum_{i=1}^{n} L(y_i - h(x_i)) + \frac{\lambda}{2} \|h\|^2_{\mathcal{H}_K}$$

- Modelization choice $(y_i)_{i=1}^n \in \mathcal{Y} := L^2[\Theta, \mu]$
- Suitable kernel: $K = k_{\mathcal{X}} \cdot T_{k_\Theta}$, where
  - $k_{\mathcal{X}} \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ scalar-valued kernel on input data
  - $k_\Theta \colon \Theta \times \Theta \to \mathbb{R}$ scalar-valued kernel
  - $T_{k_\Theta} \in \mathcal{L}(\mathcal{Y})$ integral operator associated to k_Θ

Exploiting duality with convoluted losses

Why convoluted losses: easy Fenchel-Legendre conjugate

$$\left(\frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2 \mathbin{\square} g\right)^\star = \frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2 + g^\star$$

Exploiting duality with convoluted losses

Why convoluted losses: easy Fenchel-Legendre conjugate

$$\left(\frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2 \,\square\, g\right)^\star = \frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2 + g^\star$$

Dual problem:

$$\inf_{(\alpha_i)_{i=1}^n \in \mathcal{Y}^n} \sum_{i=1}^n \left[\frac{1}{2}\|\alpha_i\|_{\mathcal{Y}}^2 - \langle \alpha_i, y_i\rangle_{\mathcal{Y}} + g^\star(\alpha_i)\right] + \frac{1}{2\lambda n}\sum_{i,j=1}^n k_{\mathcal{X}}(x_i, x_j)\left\langle \alpha_i, T_{k_\Theta}\alpha_j\right\rangle_{\mathcal{Y}}$$

with estimator $h = \frac{1}{\lambda n}\sum_{i=1}^n k_{\mathcal{X}}(\cdot, x_i) T_{k_\Theta}\alpha_i$.

Exploiting duality with convoluted losses

Why convoluted losses: easy Fenchel-Legendre conjugate

$$\left(\frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2 \,\square\, g\right)^\star = \frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2 + g^\star$$

Dual problem:

$$\inf_{(\alpha_i)_{i=1}^n \in \mathcal{Y}^n} \sum_{i=1}^n \left[\frac{1}{2}\|\alpha_i\|_{\mathcal{Y}}^2 - \langle \alpha_i, y_i\rangle_{\mathcal{Y}} + g^\star(\alpha_i)\right] + \frac{1}{2\lambda n}\sum_{i,j=1}^n k_{\mathcal{X}}(x_i, x_j)\left\langle \alpha_i, T_{k_\Theta}\alpha_j\right\rangle_{\mathcal{Y}}$$

with estimator $h = \frac{1}{\lambda n}\sum_{i=1}^n k_{\mathcal{X}}(\cdot, x_i) T_{k_\Theta}\alpha_i$.

Challenges: representing α_i, computability

The extended Huber loss

Huber loss with parameters ($\kappa \geqslant 0, p \in [1, +\infty]$):

$$H_\kappa^p := \frac{1}{2} \|\cdot\|_y^2 \,\square\, \kappa \|\cdot\|_p$$

The extended Huber loss

Huber loss with parameters ($\kappa \geqslant 0, p \in [1, +\infty]$):

$$H_\kappa^p := \frac{1}{2} \|\cdot\|_y^2 \,\square\, \kappa \|\cdot\|_p$$

Fenchel conjugate of p-norm is indicator of q-ball:

$$\|\cdot\|_p^\star = \iota_{\mathcal{B}_1^q}(\cdot) \quad \text{where} \quad \frac{1}{p} + \frac{1}{q} = 1$$

The extended Huber loss

Huber loss with parameters ($\kappa \geqslant 0, p \in [1, +\infty]$):

$$H_\kappa^p := \frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2 \,\square\, \kappa\|\cdot\|_p$$

Fenchel conjugate of p-norm is indicator of q-ball:

$$\|\cdot\|_p^\star = \iota_{\mathcal{B}_1^q}(\cdot) \quad \text{where} \quad \frac{1}{p} + \frac{1}{q} = 1$$

Dual problem becomes

$$\inf_{(\alpha_i)_{i=1}^n \in \mathcal{Y}^n} \sum_{i=1}^n \left[\frac{1}{2}\|\alpha_i\|_{\mathcal{Y}}^2 - \langle \alpha_i, y_i \rangle_{\mathcal{Y}}\right] + \frac{1}{2\lambda n} \sum_{i,j=1}^n k_{\mathcal{X}}(x_i, x_j) \langle \alpha_i, T_{k_\Theta} \alpha_j \rangle_{\mathcal{Y}}$$

$$\text{s.t } \|\alpha_i\|_q \leqslant \kappa, \quad 1 \leqslant i \leqslant n$$

Approximate problem

Representation choice for α_i: linear splines with anchors $(\theta_{ij})_{j=1}^{m}$ i.i.d. as μ

- Easy parameterization: vector in $\mathbb{R}^m$
- Suitable with Monte-Carlo approximation

Approximate problem

Representation choice for α_i: linear splines with anchors $(\theta_{ij})_{j=1}^{m}$ i.i.d. as μ

- Easy parameterization: vector in $\mathbb{R}^m$
- Suitable with Monte-Carlo approximation

Approximate problem

$$\inf_{\mathsf{A}\in\mathbb{R}^{n\times m}} \mathrm{Tr}\left(\frac{1}{2}\mathsf{A}\mathsf{A}^\top - \mathsf{A}\mathsf{Y}^\top + \frac{1}{2\lambda nm}\mathsf{K}_{\mathcal{X}}\mathsf{A}\mathsf{K}_{\Theta}\mathsf{A}^\top\right)$$
$$s.t.\ \ \|\mathsf{A}\|_{q,\infty} \leqslant m^{\frac{1}{q}}\kappa$$

with estimator

$$h = \frac{1}{\lambda nm}\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} k_{\mathcal{X}}(\cdot, x_i) k_{\Theta}(\cdot, \theta_j)$$

Optimization algorithm

”Smooth + nonsmooth” optimization problem solvable with proximal gradient descent

Optimization algorithm

"Smooth + nonsmooth" optimization problem solvable with proximal gradient descent

Amounts to knowing how to project on q-balls

$$\mathrm{prox}_{\iota_{\mathcal{B}_\kappa^q}} = \mathrm{Proj}_{\mathcal{B}_\kappa^q}$$

Closed-form available when $q \in \{2, +\infty\}$, corresponding to initial choices $p \in \{1, 2\}$

Contamination scenario

Diversity of outliers in the functional setting

- Local: only a few measurements are compromised
- Global: the whole function is corrupted

Contamination scenario

Diversity of outliers in the functional setting

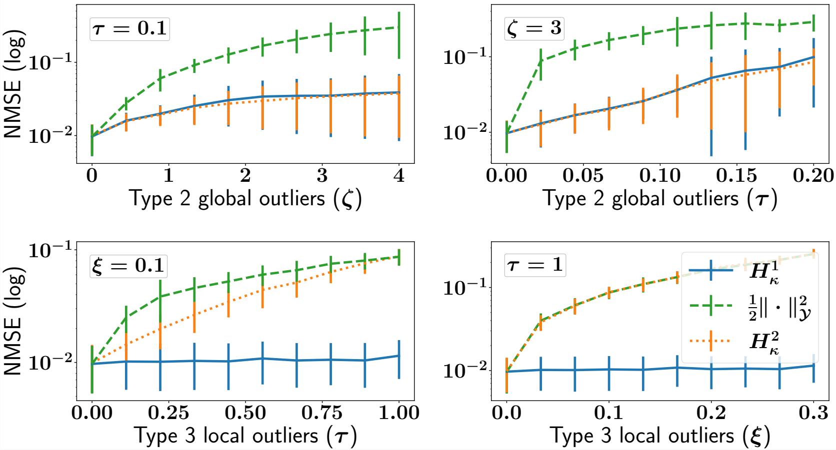
- **Local**: only a few measurements are compromised
- **Global**: the whole function is corrupted

Experimental setup:

- Contaminate a synthetic dataset
- Learn with losses $\frac{1}{2}\|\cdot\|_{\mathcal{Y}}^2$, H_κ^1, H_κ^2
- Compare with NMSE metric

$$\mathrm{NMSE} := \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}[y_i(\theta_{ij}) - \widehat{y}_i(\theta_{ij})]^2$$

Experimental results



Take home message

- In high dimension, extending the Huber loss with p-norms allows to be robust to a larger class of outliers

References

Hachem Kadri et al. “Operator-valued Kernels for Learning from Functional Response Data”. In: *Journal of Machine Learning Research* 17.20 (2016), pp. 1–54 (cit. on p. 2).

Pierre Laforgue et al. “Duality in RKHSs with Infinite Dimensional Outputs: Application to Robust Losses”. In: *International Conference on Machine Learning (ICML)*. 2020, pp. 5598–5607 (cit. on pp. 3, 4).