Minimum Volume Conformal Sets for Multivariate Regression

Check the paper :

Sacha Braun[1] Liviu Aolaritei[2] Michael I. Jordan[1,2] Francis Bach[1]

[1] Inria, ENS, PSL University [2] UC Berkeley

Abstract

Goal: Multivariate prediction sets with target coverage that adapt to the output geometry in multivariate regression.

Innovation: Novel volume-minimizing loss and data-driven nonconformity scores, adaptive to both covariates and residuals.

Minimum volume covering set

Initial problem (MVCS)

- Given : Set of n points $(y_i)_{1 \le i \le n}$ in $\mathbb{R}^k$.
- A set : $\mathbb{B}(p, M, \mu) := \{y \in \mathbb{R}^k \mid \|M(y-\mu)\|_p \le 1\}$.
- Goal : Find the smallest set that contains $n - r + 1$ of them.
- Problem :

$$\begin{aligned} \min \quad & \mathrm{Vol}(\mathbb{B}(p, M, \mu)) \\ \text{s.t.} \quad & M \succcurlyeq 0, \mu \in \mathbb{R}^k, p > 0, \\ & \mathrm{Card}\left\{ i \in [n] \mid \|M(y_i - \mu)\|_p \le 1 \right\} \ge n - r + 1. \end{aligned}$$

Combinatorial problem NP-hard

Reformulation (exact)

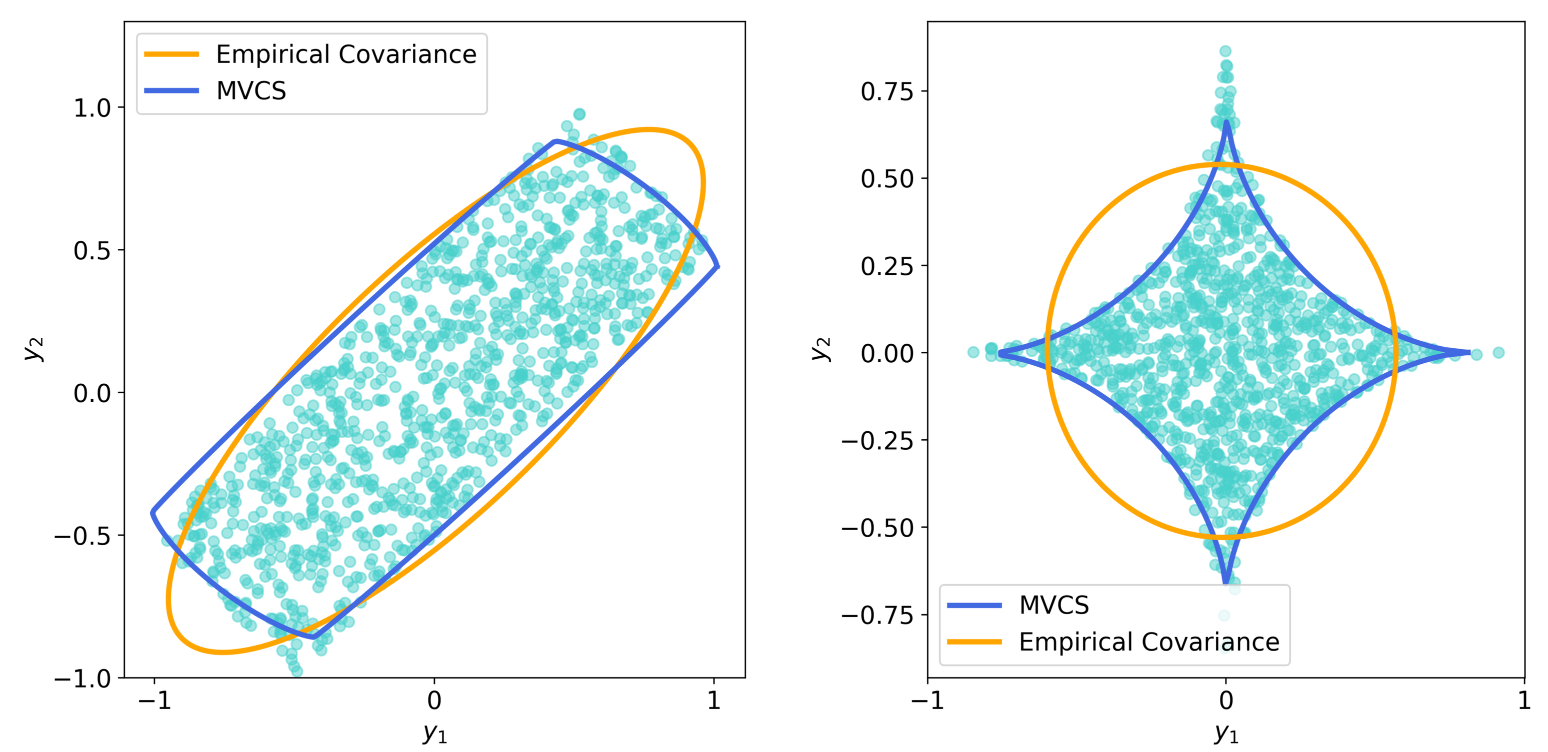
$$\begin{aligned} \min \quad & -\log\det(\Lambda) + \sigma_r\left\{ \|\Lambda y_i + \eta\|_p \right\} + \log \lambda(B_p(1)) \\ \text{s.t.} \quad & \Lambda \succcurlyeq 0, \eta \in \mathbb{R}^k, p > 0, \end{aligned}$$

Where $\sigma_r\{a_i\}$ is the r-th largest element of a set $\{a_i\}_{i=1}^n$ with $a_i \in \mathbb{R}$.

We can : Use first-order optimization, Write in a difference of convex, Derive a convex relaxation.

(But still NP-hard so no convergence guarantees).

Results



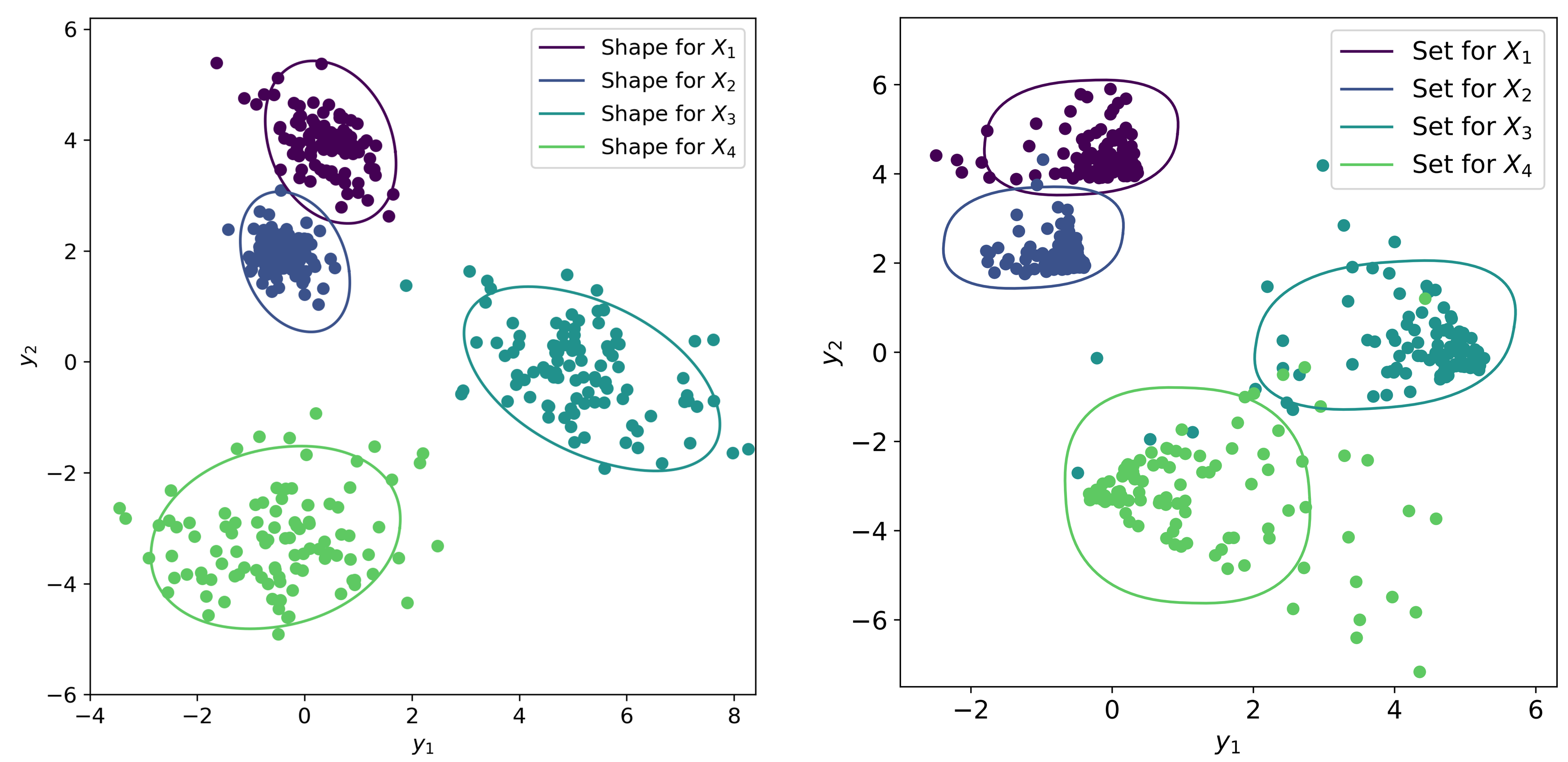
Get covariate-dependent sets

Probabilistic problem

$$\begin{aligned} \min \quad & \mathbb{E}\left[\mathrm{Vol}(\mathbb{B}(p, M_\phi(x), f_\theta(x)))\right] \\ \text{s.t.} \quad & \mathrm{Prob}\left\{ Y \in \mathbb{B}(p, M_\phi(x), f_\theta(x)) \right\} \ge 1 - \alpha\,. \end{aligned}$$

Approximation with the training data

$$\begin{aligned} \min \quad & \log\left(\sum_{i=1}^n \frac{1}{\det(\Lambda_\phi(x_i))}\right) + k \log \sigma_r\left\{ \|\Lambda_\phi(x_i)(y_i - f_\theta(x_i))\|_p \right\} + \log \lambda(B_p(1)) \\ \text{s.t.} \quad & \Lambda_\phi(\cdot) \succcurlyeq 0, p > 0, \theta \in \Theta\,. \end{aligned}$$



Conformalize the sets

Adaptive score function

- Score function : $S(X, Y) = \|M_\phi(X)(Y - f_\theta(X))\|_p$.
- Given : n samples i.i.d $(X_i, Y_i) \sim \mathbb{P}$ $\rightarrow$ split in two
  - $\mathcal{D}_1$ training set with $\mathrm{Card}(\mathcal{D}_1) = n_1$,
  - $\mathcal{D}_2$ calibration set with $\mathrm{Card}(\mathcal{D}_2) = n_2$.
- $\hat{q}_\alpha = \lceil (1-\alpha)(n_2+1) \rceil$-smallest value of $S(X_i, Y_i)$, for $i \in [n_2]$.

(Proposition) Let (X_{n+1}, Y_{n+1}) be a test sample from $\mathbb{P}$, independent of the calibration samples :

$$\mathrm{Prob}\left\{ Y_{n+1} \in \mathbb{B}\left(p, \frac{M_\phi(X_{n+1})}{\hat{q}_\alpha}, f_\theta(X_{n+1})\right) \middle| \{(X_i, Y_i)\}_{i \in \mathcal{D}_2} \right\} \in \left[1-\alpha, 1-\alpha+\frac{1}{n_2+1}\right).$$

Results - (Volume$^{1/d}$)

Dataset	Naïve QR	Emp. Cov.	Loc. Emp. Cov.	MVCS
Bias correction	1.29 ± 0.02	$\mathbf{1.26} \pm 0.03$	1.45 ± 0.10	1.33 ± 0.24
CASP	1.40 ± 0.01	1.52 ± 0.02	1.44 ± 0.02	$\mathbf{1.32} \pm 0.02$
Energy	1.28 ± 0.11	1.10 ± 0.16	1.10 ± 0.16	$\mathbf{0.97} \pm 0.13$
House	1.37 ± 0.02	1.39 ± 0.02	1.38 ± 0.02	$\mathbf{1.33} \pm 0.02$
rf1	0.43 ± 0.02	0.44 ± 0.02	0.64 ± 0.03	$\mathbf{0.39} \pm 0.05$
rf2	0.61 ± 0.01	0.42 ± 0.02	0.44 ± 0.02	$\mathbf{0.35} \pm 0.01$
scm1d	2.71 ± 0.09	1.74 ± 0.06	1.74 ± 0.06	$\mathbf{1.47} \pm 0.08$
scm20d	3.45 ± 0.47	2.64 ± 0.49	2.64 ± 0.49	$\mathbf{1.51} \pm 0.03$
Taxi	3.48 ± 0.02	3.42 ± 0.04	3.35 ± 0.03	$\mathbf{3.18} \pm 0.02$