A Kernel-based Consensual Aggregation for Regression

Sothea HAS

LPSM, Sorbonne Université (Paris 6)
sothea.has@lpsm.paris

23 février 2022

Overview

Some studies

- [Mojirsheibani, 1999] : binary classification.
  Example :
  - $\mathbf{C} = (C_1, C_2, C_3, C_4)$: 4 classifiers.
  - A new point x with predictions $\mathbf{C}(x) = (1, 1, 0, 1)$.

ID	C_1	C_2	C_3	C_4	y
1	1	1	0	1	1
2	0	0	0	1	0
3	1	1	0	1	0
4	1	1	0	0	1
5	1	1	0	1	1
6	0	1	1	0	1
x	1	1	0	1	$\hat{y}$

Table – Table of predictions.

Some studies

- [Mojirsheibani, 1999] : binary classification.
  Example :
  - $\mathbf{C} = (C_1, C_2, C_3, C_4)$: 4 classifiers.
  - A new point x with predictions $\mathbf{C}(x) = (1, 1, 0, 1)$.

ID	C_1	C_2	C_3	C_4	y
1	1	1	0	1	1
2	0	0	0	1	0
3	1	1	0	1	0
4	1	1	0	0	1
5	1	1	0	1	1
6	0	1	1	0	1
x	1	1	0	1	$\hat{y}$

Table – Table of predictions.

- Thus $\hat{y} = 1$.

Some studies

- [Mojirsheibani, 2000] : exponential kernel-based version. Example :
  - $\mathbf{C} = (C_1, C_2, C_3, C_4)$: 4 classifiers.
  - A new point x with predictions $\mathbf{C}(x) = (1, 1, 0, 1)$.

ID	C_1	C_2	C_3	C_4	y
1	1	1	0	1	1
2	0	0	**0**	**1**	0
3	1	1	0	1	0
4	**1**	**1**	**0**	0	1
5	1	1	0	1	1
6	0	**1**	1	0	0
x	1	1	0	1	$\hat{y}$

Table – Table of predictions.

Some studies

- [Mojirsheibani, 2000] : exponential kernel-based version. Example :
  - $\mathbf{C} = (C_1, C_2, C_3, C_4)$: 4 classifiers.
  - A new point x with predictions $\mathbf{C}(x) = (1, 1, 0, 1)$.

ID	C_1	C_2	C_3	C_4	y
1	1	1	0	1	1
2	0	0	**0**	**1**	0
3	1	1	0	1	0
4	**1**	**1**	**0**	0	1
5	1	1	0	1	1
6	0	**1**	1	0	0
x	1	1	0	1	$\hat{y}$

Table – Table of predictions.

- Strongly agree, larger exponential kernel-based weight.

Regression configuration

Setting :

- $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$: input-out data.
- $\mathcal{D}_n = \{(X_i, Y_i)_{i=1}^n\}$: training data of *iid* copies of (X, Y).
- $\mathcal{D}_k = \{(X_i^{(k)}, Y_i^{(k)})_{i=1}^k\}, \mathcal{D}_\ell = \{(X_i^{(\ell)}, Y_i^{(\ell)})_{i=1}^\ell\} \subset \mathcal{D}_n$ such that $\mathcal{D}_k \cup \mathcal{D}_\ell = \mathcal{D}_n$ and $\mathcal{D}_k \cap \mathcal{D}_\ell = \varnothing$.
- $\mathbf{r}_k = (r_{k,1}, ..., r_{k,M})$: M regression estimators constructed using $\mathcal{D}_k$.
- $g^*(X) = \mathbb{E}[Y|X]$: the regression function over X.
- $g^*(\mathbf{r}_k(X)) = \mathbb{E}[Y|\mathbf{r}_k(X)]$: the regression function over $\mathbf{r}_k(X)$.

Quadratic Risk :

$$\mathcal{R}_X(f) = \mathbb{E}[(f(X) - g^*(X))^2].$$

COBRA method

- [Biau et al., 2016] : regression configuration of [Mojirsheibani, 1999].
- The combination :

$$g_n(\mathbf{r}_k(x)) = \sum_{i=1}^{\ell} W_{n,i}(x) Y_i^{(\ell)}$$

where the weight is defined by,

$$W_{n,i}(x) = \frac{\prod_{m=1}^{M} \mathbb{1}_{\{|r_{k,m}(X_i^{(\ell)}) - r_{k,m}(x)| < \varepsilon\}}}{\sum_{j=1}^{\ell} \prod_{m=1}^{M} \mathbb{1}_{\{|r_{k,m}(X_j^{(\ell)}) - r_{k,m}(x)| < \varepsilon\}}}, i = 1, 2, ..., n.$$

for some smoothing parameter $\varepsilon > 0$.

Motivation of the present method

- [Has et al., 2021] : "KFC : A clusterwise supervised learning procedure based one aggregation of distances".

Motivation of the present method

- [Has et al., 2021] : "KFC : A clusterwise supervised learning procedure based one aggregation of distances".
- K/F/C = K-means/Fitting/Consensual Aggregation.

Motivation of the present method

- [Has et al., 2021] : “KFC : A clusterwise supervised learning procedure based one aggregation of distances”.
- K/F/C = K-means/Fitting/Consensual Aggregation.
- Preprint : https://hal.archives-ouvertes.fr/hal-02884333v5.

The present method

- $K : \mathbb{R}^M \to \mathbb{R}$, a regular kernel satisfying :

$$\exists b, \kappa_0, \rho > 0 \text{ s.t} \begin{cases} \forall x \in \mathbb{R}^M : b\mathbb{1}_{B_M(0,\rho)}(x) \leq K(x) \leq 1 \\ \int_{\mathbb{R}^M} \sup_{u \in B_M(x,\rho)} K(u)dx = \kappa_0 < +\infty \end{cases}$$

where $B_M(x, r) = \{z \in \mathbb{R}^M : \|x - z\|_2 < r\}$, open ball of $\mathbb{R}^M$.

The present method

- $K : \mathbb{R}^M \to \mathbb{R}$, a regular kernel satisfying :

$$\exists b, \kappa_0, \rho > 0 \text{ s.t} \begin{cases} \forall x \in \mathbb{R}^M : b\mathbb{1}_{B_M(0,\rho)}(x) \leq K(x) \leq 1 \\ \int_{\mathbb{R}^M} \sup_{u \in B_M(x,\rho)} K(u)dx = \kappa_0 < +\infty \end{cases}$$

where $B_M(x, r) = \{z \in \mathbb{R}^M : \|x - z\|_2 < r\}$, open ball of $\mathbb{R}^M$.

- We propose the following weight :

$$W_{n,i}(x) = \frac{K_h(\mathbf{r}_k(X_i^{(\ell)}) - \mathbf{r}_k(x))}{\sum_{j=1}^{\ell} K_h(\mathbf{r}_k(X_j^{(\ell)}) - \mathbf{r}_k(x))}, i = 1, 2, ..., \ell,$$

where $K_h(x) = K(x/h)$ for some $h > 0$ with $0/0 = 0$.

The present method

- $K : \mathbb{R}^M \to \mathbb{R}$, a regular kernel satisfying :

$$\exists b, \kappa_0, \rho > 0 \text{ s.t} \begin{cases} \forall x \in \mathbb{R}^M : b\mathbb{1}_{B_M(0,\rho)}(x) \leq K(x) \leq 1 \\ \int_{\mathbb{R}^M} \sup_{u \in B_M(x,\rho)} K(u)dx = \kappa_0 < +\infty \end{cases}$$

where $B_M(x, r) = \{z \in \mathbb{R}^M : \|x - z\|_2 < r\}$, open ball of $\mathbb{R}^M$.

- We propose the following weight :

$$W_{n,i}(x) = \frac{K_h(\mathbf{r}_k(X_i^{(\ell)}) - \mathbf{r}_k(x))}{\sum_{j=1}^{\ell} K_h(\mathbf{r}_k(X_j^{(\ell)}) - \mathbf{r}_k(x))}, i = 1, 2, ..., \ell,$$

where $K_h(x) = K(x/h)$ for some $h > 0$ with $0/0 = 0$.

- Again, the combination is :

$$g_n(\mathbf{r}_k(x)) = \sum_{i=1}^{\ell} W_{n,i}(x) Y_i^{(\ell)}$$

Theoretical performance

Proposition.1

Let $\mathbf{r}_k = (r_{k,1}, r_{k,2}, ..., r_{k,M})$ be the collection of all basic estimators and $g_n(\mathbf{r}_k(x))$ be the combined estimator computed at point $x \in \mathbb{R}^d$. Then, for all distributions of (X, Y) with $\mathbb{E}[|Y|^2] < +\infty$,

$$\mathbb{E}\Big[|g_n(\mathbf{r}_k(X)) - g^*(X)|^2\Big] \leq \inf_{f \in \mathcal{G}} \mathbb{E}\Big[|f(\mathbf{r}_k(X)) - g^*(X)|^2\Big] + \mathbb{E}\Big[|g_n(\mathbf{r}_k(X)) - g^*(\mathbf{r}_k(X))|^2\Big],$$

where $\mathcal{G} = \{f : \mathbb{R}^M \to \mathbb{R}, \text{s.t } \mathbb{E}[|f(\mathbf{r}_k(X))|^2] < +\infty\}$. In particular,

$$\mathbb{E}\Big[|g_n(\mathbf{r}_k(X)) - g^*(X)|^2\Big] \leq \min_{1 \leq m \leq M} \mathbb{E}\Big[|r_{k,m}(X) - g^*(X)|^2\Big] + \mathbb{E}\Big[|g_n(\mathbf{r}_k(X)) - g^*(\mathbf{r}_k(X))|^2\Big].$$

Theoretical performance

Proposition.2

Assume that $r_{k,m}$ is bounded for all $m = 1, 2, .., M$. Let $h \to 0$ and $\ell \to +\infty$ such that $h^M \ell \to +\infty$. Then

$$\mathbb{E}\left[|g_n(\mathbf{r}_k(X)) - g^*(\mathbf{r}_k(X))|^2\right] \to 0 \text{ as } \ell \to +\infty$$

for all distribution of (X, Y) s.t $\mathbb{E}[|Y|^2] < +\infty$. Thus,

$$\limsup_{\ell \to +\infty} \mathbb{E}\left[|g_n(\mathbf{r}_k(X)) - g^*(X)|^2\right] \leq \inf_{f \in \mathcal{G}} \mathbb{E}\left[|f(\mathbf{r}_k(X)) - g^*(X)|^2\right].$$

And in particular,

$$\limsup_{\ell \to +\infty} \mathbb{E}\left[|g_n(\mathbf{r}_k(X)) - g^*(X)|^2\right] \leq \min_{1 \leq m \leq M} \mathbb{E}\left[|r_{k,m}(X) - g^*(X)|^2\right].$$

Theoretical performance

Theorem

Assume that

- Y and all the basic machines $r_{k,m}, m = 1, 2, ..., M$, are bounded by R.
- $\exists L > 0, \forall k \geq 1$:

$$|g^*(\mathbf{r}_k(x)) - g^*(\mathbf{r}_k(y))| \leq L\|\mathbf{r}_k(x) - \mathbf{r}_k(y)\|, \forall x, y \in \mathbb{R}^d.$$

- $\exists R_K, C_k > 0 : K(z)\|z\|^2 \leq \frac{C_K}{1+\|z\|^M}, \forall z \in \mathbb{R}^M$ such that $\|z\| \geq R_K$.

Then, with the choice of $h \propto \ell^{-\frac{M+2}{M^2+2M+4}}$, there exists $C > 0$ such that

$$\mathbb{E}[|g_n(\mathbf{r}_k(X)) - g^*(X)|^2] \leq \min_{1\leq m\leq M} \mathbb{E}[|r_{k,m}(X) - g^*(X)|^2] + C\ell^{-\frac{4}{M^2+2M+4}}.$$

Theoretical performance

Theorem

Assume that

- Y and all the basic machines $r_{k,m}, m = 1, 2, ..., M$, are bounded by R.
- $\exists L > 0, \forall k \geq 1$:

$$|g^*(\mathbf{r}_k(x)) - g^*(\mathbf{r}_k(y))| \leq L\|\mathbf{r}_k(x) - \mathbf{r}_k(y)\|, \forall x, y \in \mathbb{R}^d.$$

- $\exists R_K, C_k > 0 : K(z)\|z\|^2 \leq \frac{C_K}{1+\|z\|^M}, \forall z \in \mathbb{R}^M$ such that $\|z\| \geq R_K$.

Then, with the choice of $h \propto \ell^{-\frac{M+2}{M^2+2M+4}}$, there exists $C > 0$ such that

$$\mathbb{E}[|g_n(\mathbf{r}_k(X)) - g^*(X)|^2] \leq \min_{1\leq m\leq M} \mathbb{E}[|r_{k,m}(X) - g^*(X)|^2] + C\ell^{-\frac{4}{M^2+2M+4}}.$$

$*$ Remark : rate in [Biau et al., 2016] is of order $O(\ell^{-2/(M+2)})$. We can get as close as we want to this rate with exponential bound on the kernels.

Optimization : Gradient descent

- Motivation : convex-like curve of risk.

Optimization : Gradient descent

- Motivation : convex-like curve of risk.
- Objective function : κ-fold cross validation error,

$$\varphi^{\kappa}(h) = \frac{1}{\kappa}\sum_{p=1}^{\kappa}\sum_{(X_j,Y_j)\in F_p}[g_n(\mathbf{r}_k(X_j)) - Y_j]^2,$$

where $g_n(\mathbf{r}_k(X_j)) = \sum_{(X_i,Y_i)\in\mathcal{D}_\ell\setminus F_p} W_{n,i}(X_j)Y_i$.

Optimization : Gradient descent

- Motivation : convex-like curve of risk.
- Objective function : κ-fold cross validation error,

$$\varphi^{\kappa}(h) = \frac{1}{\kappa}\sum_{p=1}^{\kappa}\sum_{(X_j,Y_j)\in F_p}[g_n(\mathbf{r}_k(X_j)) - Y_j]^2,$$

where $g_n(\mathbf{r}_k(X_j)) = \sum_{(X_i,Y_i)\in\mathcal{D}_\ell\setminus F_p} W_{n,i}(X_j)Y_i$.

- Algorithm :

Gradient descent for estimating h^*

1. Initialization : h_0, a learning rate $\lambda > 0$, threshold $\delta > 0$ and the maximum number of iteration N.
2. For $k = 1, 2, ..., N$, **while** $\left|\frac{d}{dh}\varphi^{\kappa}(h_{k-1})\right| > \delta$ do :

$$h_k \leftarrow h_{k-1} - \lambda\frac{d}{dh}\varphi^{\kappa}(h_{k-1})$$

3. return h_k violating the **while** condition or h_N to be the estimation of h^*.

Kernels and basic estimators

- Kernels :

Kernel	Formula
Naive [1]	$K(x) = \prod_{i=1}^{d} \mathbb{1}_{\{\lvert x_i \rvert \leq 1\}}$
Epanechnikov	$K(x) = (1 - \lVert x \rVert^2) \mathbb{1}_{\{\lVert x \rVert \leq 1\}}$
Bi-weight	$K(x) = (1 - \lVert x \rVert^2)^2 \mathbb{1}_{\{\lVert x \rVert \leq 1\}}$
Tri-weight	$K(x) = (1 - \lVert x \rVert^2)^3 \mathbb{1}_{\{\lVert x \rVert \leq 1\}}$
Compact-support Gaussian	$K(x) = \exp\{-\lVert x \rVert^2/(2\sigma^2)\} \mathbb{1}_{\{\lVert x \rVert \leq \rho_1\}}, \sigma, \rho_1 > 0$
Gaussian	$K(x) = \exp\{-\lVert x \rVert^2/(2\sigma^2)\}, \sigma > 0$
4-exponential	$K(x) = \exp\{-\lVert x \rVert^4/(2\sigma^4)\}, \sigma > 0$

Table – Kernel functions used.

- Basic estimators : Ridge, Lasso, kNN, Pruned Tree and RF.

1. The naive kernel corresponds to the method by [Biau et al., 2016].

Uncorrelated case

- $X \sim \mathcal{U}[-1,1]^d$ iid, $d \in \{30, 50, 100, 300\}$ and $n \in \{500, 600, 700, 800\}$.
- Model 9 and 10 are high-dimensional cases where $d = 1000$ and $d = 1500$.
- Average MSEs and SDs over 100 runs are reported.

Mod	Las	Rid	kNN	Tr	RF	COBRA	Epan	Bi-wgt	Tri-wgt	C-Gaus	Gauss	Exp4
1.	0.156 (0.016)	0.134 (0.013)	0.144 (0.014)	**0.027** (0.004)	0.033 (0.004)	0.022 (0.004)	0.020 (0.003)	0.019 (0.003)	0.019 (0.003)	0.019 (0.003)	**0.018** (0.002)	0.019 (0.003)
2.	1.301 (0.216)	0.784 (0.110)	0.873 (0.123)	1.124 (0.165)	**0.707** (0.097)	0.722 (0.065)	0.718 (0.079)	0.712 (0.080)	0.715 (0.079)	0.712 (0.079)	**0.709** (0.078)	0.710 (0.079)
3.	0.664 (0.107)	0.669 (0.255)	1.477 (0.192)	0.797 (0.135)	**0.629** (0.091)	0.554 (0.069)	0.482 (0.062)	0.478 (0.060)	0.476 (0.060)	0.479 (0.063)	**0.475** (0.060)	0.483 (0.060)
4.	7.783 (1.121)	6.550 (1.115)	10.238 (1.398)	3.796 (0.840)	**3.774** (0.523)	3.608 (0.526)	3.231 (0.383)	3.185 (0.382)	3.153 (0.384)	3.189 (0.371)	**2.996** (0.384)	3.186 (0.464)
5.	0.508 (0.051)	0.518 (0.073)	0.699 (0.084)	0.575 (0.081)	**0.436** (0.051)	0.429 (0.035)	0.389 (0.031)	0.387 (0.030)	0.386 (0.030)	0.387 (0.030)	**0.383** (0.030)	0.387 (0.028)
6.	2.693 (0.537)	1.958 (0.292)	2.675 (0.349)	3.065 (0.475)	**1.826** (0.262)	1.574 (0.270)	1.274 (0.129)	1.259 (0.130)	**1.254** (0.130)	1.270 (0.125)	1.273 (0.130)	1.286 (0.130)
7.	1.971 (0.410)	0.796 (0.132)	1.074 (0.152)	0.737 (0.109)	**0.515** (0.073)	0.506 (0.063)	0.472 (0.049)	0.468 (0.048)	0.467 (0.049)	0.469 (0.049)	**0.451** (0.049)	0.477 (0.067)
8.	0.134 (0.016)	0.131 (0.020)	0.200 (0.020)	0.174 (0.034)	**0.127** (0.013)	0.104 (0.013)	0.092 (0.013)	**0.091** (0.013)	**0.091** (0.013)	**0.091** (0.013)	**0.091** (0.011)	0.094 (0.016)
9.	1.592 (0.219)	2.948 (0.436)	3.489 (0.516)	1.830 (0.373)	**1.488** (0.267)	1.130 (0.151)	0.929 (0.128)	0.918 (0.127)	0.914 (0.130)	0.918 (0.124)	**0.895** (0.126)	0.993 (0.186)
10.	2012.660 (284.391)	**1485.065** (210.816)	1778.955 (261.396)	3058.381 (486.504)	1618.977 (231.555)	1511.283 (129.796)	1462.509 (143.976)	1458.306 (142.988)	1459.558 (142.602)	1452.523 (141.168)	**1400.365** (143.330)	1414.316 (144.929)

Correlated case

- $X \sim \mathcal{N}(0, \Sigma)$ with $\Sigma_{ij} = 2^{-|i-j|}$ for $1 \leq i, j \leq d$.
- Model 10 : 1 unit $= 10^8$.

Mod	Las	Rid	kNN	Tr	RF	COBRA	Epan	Bi-wgt	Tri-wgt	C-Gaus	Gauss	Exp4
1.	2.294 (0.544	1.947 (0.507)	1.941 (0.487)	**0.320** (0.145)	0.542 (0.231)	0.307 (0.129)	0.304 (0.105)	0.301 (0.111)	0.288 (0.103)	0.297 (0.104)	**0.269** (0.092)	0.291 (0.098)
2.	14.273 (2.593)	8.442 (1.912)	8.572 (1.751)	6.796 (1.548)	**5.135** (1.372)	5.345 (1.194)	4.582 (0.941)	4.529 (0.934)	4.491 (0.922)	4.541 (0.896)	**4.377** (0.905)	4.910 (1.181)
3.	7.996 (3.393)	6.266 (3.296)	8.704 (3.523)	4.110 (2.894)	**3.722** (2.956)	3.327 (1.006)	2.598 (0.912)	2.536 (0.944)	2.444 (0.840)	2.554 (0.907)	**2.168** (0.680)	2.357 (0.756)
4.	61.474 (13.986)	42.351 (11.622)	46.934 (12.543)	**8.855** (3.480)	13.381 (5.549)	9.599 (4.125)	10.511 (2.961)	9.963 (3.101)	9.682 (2.860)	10.085 (2.904)	**9.056** (2.407)	9.713 (2.695)
4.	6.805 (3.685)	7.479 (5.336)	10.342 (5.425)	**4.000** (3.144)	4.880 (3.787)	3.225 (2.088)	2.640 (1.455)	2.401 (1.387)	2.235 (1.250)	2.412 (1.355)	**1.792** (0.913)	2.194 (1.242)
6.	4.221 (0.848)	2.087 (0.485)	4.461 (0.599)	3.408 (0.636)	**1.701** (0.288)	1.493 (0.326)	1.271 (0.149)	1.238 (0.146)	1.217 (0.143)	1.248 (0.148)	**1.097** (0.145)	1.270 (0.386)
7.	17.875 (5.632)	4.695 (1.318)	5.591 (1.418)	4.132 (1.360)	**3.081** (1.091)	3.304 (0.799	2.819 (0.636)	2.779 (0.614)	2.736 (0.605)	2.788 (0.623)	**2.640** (0.590)	2.979 (0.764)
8.	0.139 (0.016)	0.133 (0.020)	0.201 (0.019)	0.159 (0.035)	**0.121** (0.013)	0.102 (0.021)	0.100 (0.020)	0.100 (0.021)	0.100 (0.020)	0.100 (0.020)	**0.092** (0.021)	**0.092** (0.018)
9.	43.445 (12.210)	37.827 (12.201)	43.991 (12.920)	**15.258** (8.119)	16.957 (8.774)	13.505 (4.822)	11.303 (3.891)	11.007 (3.815)	11.067 (3.949)	11.206 (3.960)	**10.303** (3.634)	12.346 (5.014)
10.	7235.062 (1100.579)	**5244.843** (996.181)	7636.811 (1159.445)	13014.596 (2020.133)	7092.741 (1030.249)	5147.950 (835.384)	4717.225 (703.049)	4669.516 (696.027)	4663.430 (687.474)	4697.019 (681.370)	**4660.043** (764.363)	5073.591 (1022.894)

Real datasets

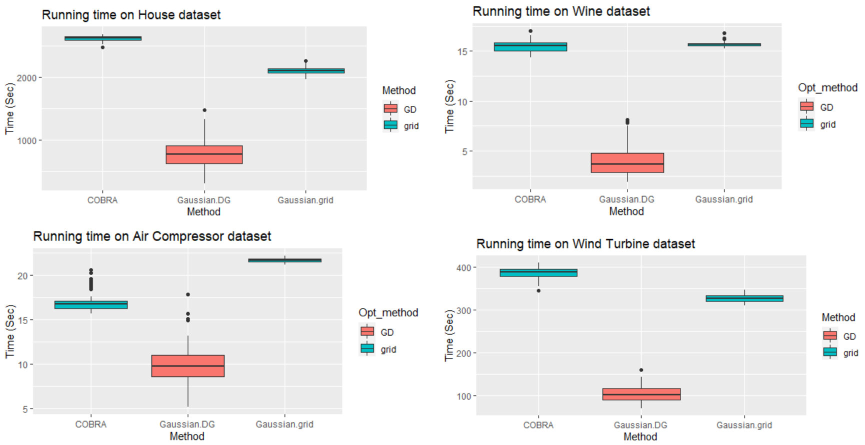
Average RMSEs and SDs over 100 runs are reported.

- **House** (\$$10^4$) : [Kaggle, 2016].
- **Wine** : [Dua and Graff, 2017b, Cortez et al., 2009].
- **Abalone** : [Dua and Graff, 2017a].
- **Air compressor** : [Cadet et al., 2005].
- **Wind turbine** : [Fischer et al., 2017].

Model	Las	Rid	*k*NN	Tr	RF	COBRA	Gauss
House	241083.959 (8883.107)	241072.974 (8906.332)	245153.608 (23548.367)	254099.652 (9350.885)	**205943.768** (7496.766)	223596.317 (13299.934)	**209955.276** (7815.623)
Wine	0.0.660 (0.029)	0.685 (0.053)	0.767 (0.031)	0.711 (0.030)	**0.623** (0.028)	0.650 (0.026)	**0.617** (0.020)
Abalone	2.204 (0.071)	2.215 (0.075)	2.175 (0.062)	2.397 (0.072)	**2.153** (0.060)	2.171 (0.081)	**2.128** (0.057)
Air	**163.099** (3.694)	164.230 (3.746)	241.657 (5.867)	351.317 (31.876)	174.836 (6.554)	172.858 (7.644)	**163.253** (3.333)
Turbine	70.051 (4.986)	68.987 (3.413)	44.516 (1.671)	81.714 (4.976)	**38.894** (1.506)	38.927 (1.561)	**37.135** (1.555)

Running times on some datasets

Running times over 100 runs are reported.



Thank you

Question ?

Conclusion

- We extend the theoretical result of [Biau et al., 2016] to a more general kernel-based framework.

Conclusion

- We extend the theoretical result of [Biau et al., 2016] to a more general kernel-based framework.
- In practical point of view :

Conclusion

- We extend the theoretical result of [Biau et al., 2016] to a more general kernel-based framework.
- In practical point of view :
    - The performance of the method is improved with the introduction of more smooth kernel functions.

Conclusion

- We extend the theoretical result of [Biau et al., 2016] to a more general kernel-based framework.
- In practical point of view :
  - The performance of the method is improved with the introduction of more smooth kernel functions.
  - The computational time is improved with gradient descent algorithm.

Thank you

Question ?

References I

Biau, G., Fischer, A., Guedj, B., and Malley, J. D. (2016).
COBRA : a combined regression strategy.
Journal of Multivariate Analysis, 146 :18–28.

Cadet, O., Harper, C., and Mougeot, M. (2005).
Monitoring energy performance of compressors with an innovative auto-adaptive approach.
In *Instrumentation System and Automation -ISA- Chicago*.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis., J. (2009).
Modeling wine preferences by data mining from physicochemical properties.
Decision Support Systems, Elsevier, 47 :547–553.

Devroye, L., Györfi, L., and Lugosi, G. (1997).
A Probabilistic Theory of Pattern Recognition.
Springer.

Dua, D. and Graff, C. (2017a).
UCI machine learning repository : Abalone data set.

Dua, D. and Graff, C. (2017b).
UCI machine learning repository : Wine quality data set.

Fischer, A., Montuelle, L., Mougeot, M., and Picard, D. (2017).
Statistical learning for wind power : A modeling and stability study towards forecasting.
Wiley Online Library, 20(12) :2037–2047.

References II

Fischer, A. and Mougeot, M. (2019).
Aggregation using input-output trade-off.
Journal of Statistical Planning and Inference, 200 :1–19.

Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002).
A Distribution-Free Theory of Nonparametric Regression.
Springer.

Has, S., Fischer, A., and Mougeot, M. (2021).
Kfc : A clusterwise supervised learning procedure based on aggregation of distances.
Journal of Statistical Computation and Simulation, 0(0) :1–21.

Kaggle (2016).
House sales in king county, usa.

Mojirsheibani, M. (1999).
Combined classifiers via disretization.
Journal of the American Statistical Association, 94(446) :600–609.

Mojirsheibani, M. (2000).
A kernel-based combined classification rule.
Journal of Statistics and Probability Letters, 48(4) :411–419.

Mojirsheibani, M. and Kong, J. (2016).
An asymptotically optimal kernel combined classifier.
Journal of Statistics and Probability Letters, 119 :91–100.