Adaptive dimension reduction for regression

Quentin PARIS

ENS Cachan, antenne de Bretagne (Rennes)

August 27, 2013

Aussois

Regression estimation

Let (X, Y) be an $\mathbb{R}^p \times \mathbb{R}$-valued random variable. We consider the estimation of the regression function

$$r(x) := \mathbb{E}\left[Y|X = x\right],$$

based on a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of i.i.d. random variables with same distribution as (X, Y).

The performance of an estimate $\hat{r}$ is measured by

$$\mathbb{E}\|\hat{r} - r\|^2 := \mathbb{E}\left[(\hat{r}(X) - r(X))^2\right].$$

Curse of dimensionality

Let $\mathcal{F}$ be the class of all functions $f : \mathbb{R}^p \to \mathbb{R}$ satisfying

$$|f(x) - f(x')| \leq \|x - x'\|^\beta.$$

In a word

Suppose $r \in \mathcal{F}$. Then, classical estimates (kernel, nearest neighbors, least squares) satisfy

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(n^{-\frac{2\beta}{2\beta+p}}\right).$$

Minimax point of view

More precisely, we have

$$\inf_{\hat{r}} \sup_{r \in \mathcal{F}} \mathbb{E}\|\hat{r} - r\|^2 \asymp n^{-\frac{2\beta}{2\beta+p}}.$$

Curse of dimensionality

Let $\mathcal{F}$ be the class of all functions $f : \mathbb{R}^p \to \mathbb{R}$ satisfying

$$|f(x) - f(x')| \leq \|x - x'\|^{\beta}.$$

In a word

Suppose $r \in \mathcal{F}$. Then, classical estimates (kernel, nearest neighbors, least squares) satisfy

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(n^{-\frac{2\beta}{2\beta+p}}\right).$$

Minimax point of view

More precisely, we have

$$\inf_{\hat{r}} \sup_{r\in\mathcal{F}} \mathbb{E}\|\hat{r} - r\|^2 \asymp n^{-\frac{2\beta}{2\beta+p}}.$$

Curse of dimensionality

Let $\mathcal{F}$ be the class of all functions $f : \mathbb{R}^p \to \mathbb{R}$ satisfying

$$|f(x) - f(x')| \leq \|x - x'\|^{\beta}.$$

In a word

Suppose $r \in \mathcal{F}$. Then, classical estimates (kernel, nearest neighbors, least squares) satisfy

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(n^{-\frac{2\beta}{2\beta+p}}\right).$$

Minimax point of view

More precisely, we have

$$\inf_{\hat{r}} \sup_{r \in \mathcal{F}} \mathbb{E}\|\hat{r} - r\|^2 \asymp n^{-\frac{2\beta}{2\beta+p}}.$$

A natural approach for dimension reduction

Assume r satisfies structural assumptions in addition to classical regularity conditions.

Example 1

Projection pursuit (Friedman and Stuetzle, 1981) :

$$r(x) = \sum_{j=1}^{K} r_j(\alpha_j^T x), \quad \alpha_j \in \mathbb{R}^p,$$

where each $r_j : \mathbb{R} \to \mathbb{R}$ is assumed to be β-regular.

- If $K = p$ and $\alpha_j = j^{th}$-coordinate vector $\Rightarrow$ Additive regression model.
- If $K = 1 \Rightarrow$ Single index model.

Theorem (Györfi, Kohler, Krzyżak and Walk, 2002)

A least-squares procedure leads to an estimate $\hat{r}$ which satisfies

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(\left(\frac{\log n}{n}\right)^{\frac{2\beta}{2\beta+1}}\right).$$

Example 1

Projection pursuit (Friedman and Stuetzle, 1981) :

$$r(x) = \sum_{j=1}^{K} r_j(\alpha_j^T x), \quad \alpha_j \in \mathbb{R}^p,$$

where each $r_j : \mathbb{R} \to \mathbb{R}$ is assumed to be β-regular.

- If $K = p$ and $\alpha_j = j^{th}$-coordinate vector $\Rightarrow$ Additive regression model.
- If $K = 1 \Rightarrow$ Single index model.

Theorem (Györfi, Kohler, Krzyżak and Walk, 2002)

A least-squares procedure leads to an estimate $\hat{r}$ which satisfies

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(\left(\frac{\log n}{n}\right)^{\frac{2\beta}{2\beta+1}}\right).$$

Example 2

Sufficient dimension reduction (Härdle and Stoker, 1989 ; Li, 1991 ; Cook, 2007) :

$$r(x) = m(\Lambda x), \quad \Lambda \in \mathcal{M}_p(\mathbb{R}),$$

where m is assumed to be β-regular. In other words,

$$\mathbb{E}[Y|X] = \mathbb{E}[Y|\Lambda X] \quad \text{and} \quad \mathbb{E}[Y|\Lambda X = .] \text{ is } \beta\text{-regular.}$$

Theorem (Cadre and Dong, 2010)

Through the estimation of m and Λ, we can construct an estimate $\hat{r}$ which satisfies

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(n^{-\frac{2\beta}{2\beta + rank(\Lambda)}}\right).$$

Example 2

Sufficient dimension reduction (Härdle and Stoker, 1989 ; Li, 1991 ; Cook, 2007) :

$$r(x) = m(\Lambda x), \quad \Lambda \in \mathcal{M}_p(\mathbb{R}),$$

where m is assumed to be β-regular. In other words,

$$\mathbb{E}[Y|X] = \mathbb{E}[Y|\Lambda X] \quad \text{and} \quad \mathbb{E}[Y|\Lambda X = .] \text{ is } \beta\text{-regular.}$$

Theorem (Cadre and Dong, 2010)

Through the estimation of m and Λ, we can construct an estimate $\hat{r}$ which satisfies

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(n^{-\frac{2\beta}{2\beta + rank(\Lambda)}}\right).$$

Example 2

Sufficient dimension reduction (Härdle and Stoker, 1989 ; Li, 1991 ; Cook, 2007) :

$$r(x) = m(\Lambda x), \quad \Lambda \in \mathcal{M}_p(\mathbb{R}),$$

where m is assumed to be β-regular. In other words,

$$\mathbb{E}[Y|X] = \mathbb{E}[Y|\Lambda X] \quad \text{and} \quad \mathbb{E}[Y|\Lambda X = .] \text{ is } \beta\text{-regular.}$$

Theorem (Cadre and Dong, 2010)

Through the estimation of m and Λ, we can construct an estimate $\hat{r}$ which satisfies

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(n^{-\frac{2\beta}{2\beta + rank(\Lambda)}}\right).$$

A general model

We assume that $r \in \mathcal{F}$ where

$$\mathcal{F} := \Big\{ g \circ h : g \in \mathcal{G}, h \in \mathcal{H} \Big\}.$$

- $\mathcal{H}$: a (parametric) class of functions $h : \mathbb{R}^p \to \mathbb{R}^p$.
- $\mathcal{G}$: class of functions $g : \mathbb{R}^p \to \mathbb{R}$ such that

$$|g(u) - g(u')| \leq C \|u - u'\|^{\beta}.$$

In other words

We assume that there exists (at least) a function $h \in \mathcal{H}$ such that

$$\mathbb{E}[Y|h(X)] = \mathbb{E}[Y|X] \quad \text{and} \quad \mathbb{E}[Y|h(X) = .] \in \mathcal{G}.$$

A general model

We assume that $r \in \mathcal{F}$ where

$$\mathcal{F} := \Big\{ g \circ h : g \in \mathcal{G}, h \in \mathcal{H} \Big\}.$$

- $\mathcal{H}$: a (parametric) class of functions $h : \mathbb{R}^p \to \mathbb{R}^p$.
- $\mathcal{G}$: class of functions $g : \mathbb{R}^p \to \mathbb{R}$ such that

$$|g(u) - g(u')| \leq C\|u - u'\|^{\beta}.$$

In other words

We assume that there exists (at least) a function $h \in \mathcal{H}$ such that

$$\mathbb{E}[Y|h(X)] = \mathbb{E}[Y|X] \quad \text{and} \quad \mathbb{E}[Y|h(X) = .] \in \mathcal{G}.$$

A general model

We assume that $r \in \mathcal{F}$ where

$$\mathcal{F} := \left\{ g \circ h : g \in \mathcal{G}, h \in \mathcal{H} \right\}.$$

- $\mathcal{H}$: a (parametric) class of functions $h : \mathbb{R}^p \to \mathbb{R}^p$.
- $\mathcal{G}$: class of functions $g : \mathbb{R}^p \to \mathbb{R}$ such that

$$|g(u) - g(u')| \leq C\|u - u'\|^\beta.$$

In other words

We assume that there exists (at least) a function $h \in \mathcal{H}$ such that

$$\mathbb{E}[Y|h(X)] = \mathbb{E}[Y|X] \quad \text{and} \quad \mathbb{E}[Y|h(X) = .] \in \mathcal{G}.$$

A general model

We assume that $r \in \mathcal{F}$ where

$$\mathcal{F} := \Big\{ g \circ h : g \in \mathcal{G}, h \in \mathcal{H} \Big\}.$$

- $\mathcal{H}$: a (parametric) class of functions $h : \mathbb{R}^p \to \mathbb{R}^p$.
- $\mathcal{G}$: class of functions $g : \mathbb{R}^p \to \mathbb{R}$ such that

$$|g(u) - g(u')| \leq C \|u - u'\|^{\beta}.$$

In other words

We assume that there exists (at least) a function $h \in \mathcal{H}$ such that

$$\mathbb{E}[Y|h(X)] = \mathbb{E}[Y|X] \quad \text{and} \quad \mathbb{E}[Y|h(X) = .] \in \mathcal{G}.$$

The reduced dimension

For all $\ell = 1, \ldots, p$, let

$$\mathcal{H}_\ell := \Big\{ h \in \mathcal{H} : \dim \mathrm{Vect}(h(\mathbb{R}^p)) \leq \ell \Big\},$$

and

$$\mathcal{F}_\ell := \Big\{ g \circ h : g \in \mathcal{G}, h \in \mathcal{H}_\ell \Big\}.$$

Then, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_p = \mathcal{F}$.

Definition

The reduced dimension is defined by

$$d := \min \Big\{ \ell = 1, \ldots, p : r \in \mathcal{F}_\ell \Big\}.$$

The reduced dimension

For all $\ell = 1, \dots, p$, let

$$\mathcal{H}_\ell := \left\{ h \in \mathcal{H} : \dim \mathrm{Vect}(h(\mathbb{R}^p)) \leq \ell \right\},$$

and

$$\mathcal{F}_\ell := \left\{ g \circ h : g \in \mathcal{G}, h \in \mathcal{H}_\ell \right\}.$$

Then, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_p = \mathcal{F}$.

Definition

The reduced dimension is defined by

$$d := \min \left\{ \ell = 1, \dots, p : r \in \mathcal{F}_\ell \right\}.$$

The reduced dimension

For all $\ell = 1, \dots, p$, let

$$\mathcal{H}_\ell := \left\{ h \in \mathcal{H} : \dim \mathrm{Vect}(h(\mathbb{R}^p)) \leq \ell \right\},$$

and

$$\mathcal{F}_\ell := \left\{ g \circ h : g \in \mathcal{G}, h \in \mathcal{H}_\ell \right\}.$$

Then, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_p = \mathcal{F}$.

Definition

The reduced dimension is defined by

$$d := \min \left\{ \ell = 1, \dots, p : r \in \mathcal{F}_\ell \right\}.$$

The reduced dimension

For all $\ell = 1, \dots, p$, let

$$\mathcal{H}_\ell := \left\{ h \in \mathcal{H} : \dim \mathrm{Vect}(h(\mathbb{R}^p)) \leq \ell \right\},$$

and

$$\mathcal{F}_\ell := \left\{ g \circ h : g \in \mathcal{G}, h \in \mathcal{H}_\ell \right\}.$$

Then, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_p = \mathcal{F}$.

Definition

The reduced dimension is defined by

$$d := \min \left\{ \ell = 1, \dots, p : r \in \mathcal{F}_\ell \right\}.$$

Fundamental inequality

For all $\ell = 1, \ldots, p$, let

$$\hat{r}_\ell \in \arg\min_{f \in \mathcal{F}_\ell} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2 .$$

For any estimate $\hat{d}$ of the reduced dimension d, let

$$\hat{r} := \hat{r}_{\hat{d}} .$$

Proposition

If all functions in $\mathcal{F}$ are bounded by $L > 0$,

$$\mathbb{E}\|\hat{r} - r\|^2 \leq \mathbb{E}\|\hat{r}_d - r\|^2 + 4L^2 \mathbb{P}\left(\hat{d} \neq d\right) .$$

Fundamental inequality

For all $\ell = 1, \ldots, p$, let

$$\hat{r}_\ell \in \arg\min_{f \in \mathcal{F}_\ell} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - f(X_i)\right)^2 .$$

For any estimate $\hat{d}$ of the reduced dimension d, let

$$\hat{r} := \hat{r}_{\hat{d}}.$$

Proposition

If all functions in $\mathcal{F}$ are bounded by $L > 0$,

$$\mathbb{E}\|\hat{r} - r\|^2 \leq \mathbb{E}\|\hat{r}_d - r\|^2 + 4L^2 \mathbb{P}\left(\hat{d} \neq d\right).$$

Fundamental inequality

For all $\ell = 1, \ldots, p$, let

$$\hat{r}_\ell \in \arg\min_{f \in \mathcal{F}_\ell} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2 .$$

For any estimate $\hat{d}$ of the reduced dimension d, let

$$\hat{r} := \hat{r}_{\hat{d}}.$$

Proposition

If all functions in $\mathcal{F}$ are bounded by $L > 0$,

$$\mathbb{E}\|\hat{r} - r\|^2 \leq \mathbb{E}\|\hat{r}_d - r\|^2 + 4L^2 \mathbb{P}\left(\hat{d} \neq d\right).$$

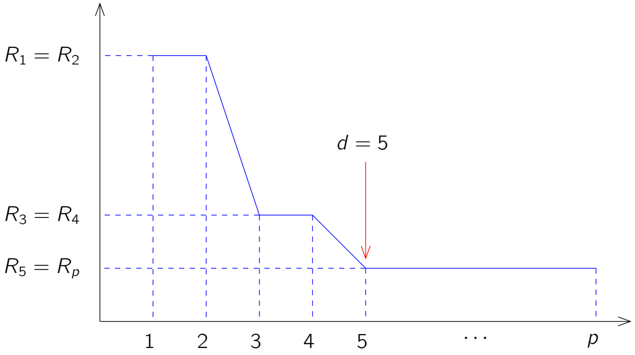
Estimation of the reduced dimension d (1/2)

First observation : For all $\ell = 1, \ldots, p$, let

$$R_\ell := \inf_{f \in \mathcal{F}_\ell} \mathbb{E}\left[Y - f(X)\right]^2 .$$

Then,

$$d = \min\left\{\ell = 1, \ldots, p : R_\ell = R_p\right\}.$$



Estimation of the reduced dimension d (2/2)

The estimate : For all $\ell = 1, \ldots, p$, let

$$\hat{R}_\ell := \inf_{f \in \mathcal{F}_\ell} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2 .$$

We define,

$$\hat{d} = \min \left\{ \ell = 1, \ldots, p : \hat{R}_\ell \leq \hat{R}_p + \frac{1}{n^\gamma} \right\}.$$

Theorem

Suppose that there exists $\tau > 0$ such that $\mathbb{E} e^{\tau |Y|} < +\infty$ and that $0 < \gamma < \frac{2\beta}{2\beta + p}$. Suppose, in addition, that there exist $C > 0$ and $0 < s \leq p/\beta$ such that

$$\sup_Q \ln N \left(\varepsilon, \mathcal{H}, \mathbb{L}^2(Q) \right) \leq C \varepsilon^{-s}.$$

Then, for all $\alpha > 0$,

$$n^\alpha \mathbb{P} \left(\hat{d} \neq d \right) \underset{n \to +\infty}{\longrightarrow} 0.$$

Study of the term $\mathbb{E}\|\hat{r}_d - r\|^2$

Recall that, for $\hat{r} := \hat{r}_{\hat{d}}$, we have $\mathbb{E}\|\hat{r} - r\|^2 \leq \mathbb{E}\|\hat{r}_d - r\|^2 + 4L^2\mathbb{P}(\hat{d} \neq d)$.

Proposition

Suppose that there exists $\tau > 0$ such that $\mathbb{E}e^{\tau|Y|} < +\infty$. Suppose, in addition, that there exist $C > 0$ and $0 < s \leq d/\beta$ such that

$$\sup_Q \ln N\left(\varepsilon, \mathcal{H}_d, \mathbb{L}^2(Q)\right) \leq C\varepsilon^{-s}.$$

Then, we have

$$\mathbb{E}\|\hat{r}_d - r\|^2 = O\left(\left(\frac{\log^3 n}{n}\right)^{\frac{2\beta}{2\beta+d}}\right).$$

Theorem

Under the assumptions of the two former results, we have

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(\left(\frac{\log^3 n}{n}\right)^{\frac{2\beta}{2\beta+d}}\right).$$

Optimality

Recall that $d = d(r) = \min\{\ell = 1, \ldots, p : r \in \mathcal{F}_\ell\}$.

Theorem (Upper bound)

The estimate $\hat{r} = \hat{r}_{\hat{d}}$ *satisfies*

$$\limsup_{n \to +\infty} \sup_{r \in \mathcal{F}} \left(\frac{n}{\log^3 n} \right)^{\frac{2\beta}{2\beta + d(r)}} \mathbb{E}\|\hat{r} - r\|^2 < +\infty.$$

Here, the supremum is taken over all distributions of (X, Y) *for which the two conditions* (i) $\mathbb{E}e^{\tau|Y|} < +\infty$ *and* (ii) $r \in \mathcal{F}$ *are satisfied.*

Theorem (Lower bound)

We have

$$\liminf_{n \to +\infty} n^{\frac{2\beta}{2\beta + d}} \inf_{\hat{r}} \sup_{r \in \mathcal{F}_d} \mathbb{E}\|\hat{r} - r\|^2 > 0.$$

Thank you for your attention