A Coverage Theory for Least Squares

Algo Carè
CWI, Amsterdam

with S. Garatti\textsuperscript{a} and M.C. Campi\textsuperscript{b}

\textsuperscript{a}Politecnico di Milano \textsuperscript{b}University of Brescia
Least Squares

\((A_i, b_i), \ i = 1, \ldots, N\) \quad \text{i.i.d.}
Example (I)

$$(A_i, b_i), \ i = 1, \ldots, N \quad \text{i.i.d.}$$
Example (I)

$$(A_i, b_i), \ i = 1, \ldots, N$$  \hspace{1cm} \text{i.i.d.}$$
Example (I)

\[(A_i, b_i), \; i = 1, \ldots, N\quad \text{i.i.d.}\]

\[A_i x \approx b_i\]
Example (I)

\[(A_i, b_i), \ i = 1, \ldots, N\]  
i.i.d.

\[(A_i x - b_i)^2\]
Example (I)

\((A_i, b_i), \ i = 1, \ldots, N\) i.i.d.

\[ \sum_{i=1}^{N} (A_i x - b_i)^2 \]
Example (I)

\[(A_i, b_i), \ i = 1, \ldots, N\] i.i.d.

\[
\min_{x \in \mathbb{R}^d} \sum_{i=1}^{N} (A_i x - b_i)^2
\]
Least Squares

$$(A_i, b_i), \ i = 1, \ldots, N \quad \text{i.i.d.}$$

$m \times d$
Least Squares

\[(A_i, b_i), \ i = 1, \ldots, N\] i.i.d.
Least Squares

\[(A_i, b_i), \ i = 1, \ldots, N\] i.i.d.

\[m \times d \quad d \times 1\]

\[A_i x \approx b_i\]
Least Squares

\[(A_i, b_i), \ i = 1, \ldots, N \text{ i.i.d.}\]

\[A_i x - b_i\]
Least Squares

\[(A_i, b_i), \ i = 1, \ldots, N \text{ \ i.i.d.}\]

\[\|A_i x - b_i\|^2\]
Least Squares

\[(A_i, b_i), \ i = 1, \ldots, N\] 

i.i.d.

\[m \times d \quad d \times 1\]

\[\sum_{i=1}^{N} \| A_i x - b_i \|^2\]
Least Squares

\[(A_i, b_i), \; i = 1, \ldots, N \quad \text{i.i.d.}\]

\[m \times d \quad d \times 1\]

\[
\min_{x \in \mathbb{R}^d} \sum_{i=1}^{N} \|A_i x - b_i\|^2
\]
Least Squares

\((A_i, b_i), \ i = 1, \ldots, N\) \ i.i.d.

\[ \text{min}_{x \in \mathbb{R}^d} \sum_{i=1}^{N} (x - c_i)^T K_i (x - c_i) + h_i \]

\[ K_i = A_i^T A_i \quad c_i = A_i^\dagger b_i \quad h_i = \|A_i c_i - b_i\|^2 \]
Example (II)

\[(A_i, b_i), \ i = 1, \ldots, N \quad \text{i.i.d.}\]
Example (II)

\[(A_i, b_i), \ i = 1, \ldots, N \quad \text{i.i.d.}\]

\[\begin{align*}
A_i &= I \\
\mathbf{x} &= b_i = p_i \\
\sum_{i=1}^{N} \| x - p_i \|^2
\end{align*}\]

\[x^*_{LS} \quad \text{Centre of mass}\]
Facility Location

locate a facility to serve a population
Facility Location

locate a facility to serve a population

$$\sum_{i=1}^{N} \left\| x - p_i \right\|^2$$
Facility Location

locate a facility to serve a population

\[ \sum_{i=1}^{N} \| x - p_i \|^2 \]
Facility Location

locate a facility to serve a population

$\alpha$ weight
Facility Location

locate a facility to serve a population

$$\alpha \quad \text{weight}$$

$$\sum_{i=1}^{N} \alpha_i \| x - p_i \|^2$$
A Coverage Theory

\[ \| A_i x - b_i \|^2 \]

\[ x^* \]

\[ x_{LS} \]
A Coverage Theory

\[ \frac{1}{N} \sum_{i=1}^{N} \| A_i x_{LS}^* - b_i \|^2 \]

\[ \| A_i x - b_i \|^2 \]
A Coverage Theory

\[ \mathbb{E}[\|Ax^*_{LS} - b\|^2] \]

\[ \frac{1}{N} \sum_{i=1}^{N} \|A_i x^*_{LS} - b_i\|^2 \]

\[ \|A_ix - b_i\|^2 \]
A Coverage Theory

\[ \mathbb{E}[\|Ax_{LS}^* - b\|^2] \]

\[ \frac{1}{N} \sum_{i=1}^{N} \|A_i x_{LS}^* - b_i\|^2 \]

\[ \|A_i x - b_i\|^2 \]

\[ x_{LS}^* \]

\[ x_{LS}^* \]
A Coverage Theory

\[ \mathbb{E}[\|A x^*_L S - b\|^2] \]

\[ \frac{1}{N} \sum_{i=1}^{N} \|A_i x^*_L S - b_i\|^2 \]

\[ \|A_{N+1} x - b_{N+1}\|^2 \]
A Coverage Theory

\[ \| A_{N+1} x^*_\text{LS} - b_{N+1} \|^2 > c \]

\[ \mathbb{E}[\| A x^*_\text{LS} - b \|^2] \]

\[ \frac{1}{N} \sum_{i=1}^{N} \| A_i x^*_\text{LS} - b_i \|^2 \]

\[ \| A_{N+1} x - b_{N+1} \|^2 \]
A Coverage Theory

\[ \| A_{N+1} x_{LS}^* - b_{N+1} \|^2 \leq c \]
A Coverage Theory

\[ \| A_{N+1} x^*_{LS} - b_{N+1} \|^2 \leq c \]

with probability at least \( p^{0\%} \)
Facility Location

locate a facility to serve a population
Facility Location

locate a facility to serve a population
Facility Location

locate a facility to serve a population

\[ \alpha(p) \text{ weight} \]
Facility Location

locate a facility to serve a population

$$\sum_{i=1}^{N} \alpha_i \| x - p_i \|^2$$
Facility Location

locate a facility to serve a population

\[ \sum_{i=1}^{N} \alpha_i \| x - p_i \|^2 \]
Facility Location

locate a facility to serve a population

\[ \sum_{i=1}^{N} \alpha_i \| x - p_i \|^2 \]

\[ \text{cost} = \alpha_{N+1} \| p_{N+1} - x_{LS}^* \|^2 \]
\[ \text{Prob}\{\text{cost } \leq c_i\} = \frac{i}{N + 1} = \frac{17}{20} \]
\[ \text{Prob}\{\text{cost} \leq c_i\} = \frac{i}{N + 1} = \frac{17}{20} \] FALSE!!
$$\|A_i x - b_i\|^2 = (x - v_i)^T K_i (x - v_i) + h_i$$
\[ \|A_i x - b_i\|^2 = (x - v_i)^T K_i (x - v_i) + h_i \]

\[ (x - v_i)^T \bar{K}_i (x - v_i) + h_i \]

\[ \bar{K}_i := K_i + 6K_i \left( \sum_{\ell=1}^{N} K_{\ell} \right)^{-1} K_i \]
Theorem

Relation

$$\text{Prob}\{\text{cost} \leq \bar{c}_i\} \geq \frac{i}{N + 1}, \quad i = 1, \ldots, N$$

holds true for any probability distribution over \((A, b)\).
Facility Location
Facility Location

\[ \text{Prob}\{\text{cost} \leq \bar{c}_i\} \geq \frac{i}{N + 1}, \quad i = 1, \ldots, N \]
Facility Location

\[ \text{Prob}\{\text{cost} \leq \bar{c}_i\} \geq \frac{i}{N + 1}, \quad i = 1, \ldots, N \]
A notable case:

\[ \alpha(p_i) = 1 \quad i = 1, \ldots, N \]

\[ \bar{C}_i = \frac{N + 5}{N - 1} C_i \quad (N \geq 8) \]

\[ \text{Prob}\{\text{cost} \leq \bar{c}_i\} \geq \frac{i}{N + 1}, \quad i = 1, \ldots, N \]
margins: \textit{small enough to be practically useful}

margins \to 0 \text{ as } N \to \infty
Another notable case:
Another notable case:

$$\bar{C}_i = C_i \quad (N \geq 8)$$

$$\text{Prob}\{\text{cost} \leq \bar{c}_i\} \geq \frac{i}{N + 1}, \quad i = 1, \ldots, N$$
Another notable case:

\[ \bar{C}_i = C_i \quad (N \geq 8) \]

\[ \text{Prob}\{\text{cost} \leq \bar{c}_i\} \geq \frac{i}{N + 1}, \quad i = 1, \ldots, N \]

Distribution-free \quad \Longrightarrow \quad \text{conservative}
Another notable case:

\[ \overline{C}_i = C_i \quad (N \geq 8) \]

\[ \text{Prob}\{\text{cost } \leq \overline{c}_i\} \geq \frac{i}{N + 1}, \quad i = 1, \ldots, N \]

Distribution-free \hspace{1cm} \rightarrow \hspace{1cm} \text{conservative}

data-dependent margins allow us to “squeeze data dry”
Idea behind the technical result

\((A_1, b_1), \ldots, (A_N, b_N), (A_{N+1}, b_{N+1})\)
Idea behind the technical result

\((A_1, b_1), \ldots, (A_N, b_N)(A_{N+1}, b_{N+1})\)
Idea behind the technical result

\[(A_1, b_1), \ldots, (A_N, b_N)(A_{N+1}, b_{N+1})\]

\[
\begin{align*}
\mathbf{x}_{LS}^* & \quad c_i
\end{align*}
\]
Idea behind the technical result

\[(A_1, b_1), \ldots, (A_N, b_N)(A_{N+1}, b_{N+1})\]

\[x^*_\text{LS} \quad c_i\]

\[\|A_{N+1}x^*_\text{LS} - b_{N+1}\|^2 \leq c_i \quad ?\]
Idea behind the technical result

\((A_1, b_1), \ldots, (A_N, b_N)(A_{N+1}, b_{N+1})\)

\[ x^*_L S \quad c_i \]

\[ \| A_{N+1} x^*_L S - b_{N+1} \|^2 \leq c_i \quad ? \]
Idea behind the technical result

\((A_1, b_1), \ldots, (A_N, b_N), (A_{N+1}, b_{N+1})\)

\[
\|A_N x^*_\text{LS} - b_N \|^2 \leq c_i
\]

\[
\square \quad \ldots \quad \square \quad \square \quad \square \quad \square \quad \checkmark
\]
Idea behind the technical result

\[(A_1, b_1), \ldots, (A_N, b_N), (A_{N+1}, b_{N+1})\]

\[x^*_\text{LS}, \quad c_i\]

\[\|A_N x^*_\text{LS} - b_N\|^2 \leq c_i \quad ?\]
Idea behind the technical result

\[(A_1, b_1), \ldots, (A_N, b_N), (A_{N+1}, b_{N+1})\]

\[x^*_\text{LS} \quad c_i\]

\[\|A_1 x^*\text{LS} - b_1\|^2 \leq c_i\]

\[\times \quad \ldots \quad \times \checkmark \times \times \times \times \checkmark\]
Idea behind the technical result

\[(A_1, b_1), \ldots, (A_N, b_N), (A_{N+1}, b_{N+1})\]

\[\begin{align*}
& x_{LS}^* \quad c_i \\
& \|A_1 x_{LS}^* - b_1\|^2 \leq c_i
\end{align*}\]

\[
\begin{array}{cccccc|c}
\times & \ldots & \times & \checkmark & \times & \times & \checkmark \\
\hline
\geq & \frac{i}{N + 1}
\end{array}
\]
Idea behind the technical result

\[(A_1, b_1), \ldots, (A_N, b_N), (A_{N+1}, b_{N+1})\]

\[x_{LS}^* \leq c_i\]

\[
\|A_1 x_{LS}^* - b_1\|^2 \leq c_i
\]

\[\begin{array}{cccccc}
\times & \ldots & \times & \times & \times & \times
\end{array}
\]

\[\begin{array}{cccccc}
\times & \times & \times & \times & \times & \times
\end{array}\]

\[\geq \frac{i}{N + 1}\]
Idea behind the technical result

\[(A_1, b_1), \ldots, (A_N, b_N)(A_{N+1}, b_{N+1})\]

\[x^*_{LS} \Rightarrow \bar{c}_i\]

\[\|A_{N+1}x^*_{LS} - b_{N+1}\|^2 \leq \bar{c}_i\]

...
Idea behind the technical result

The bound is naturally satisfied in min-max convex problems where only a few observations (support scenarios) determine \( x^*, C_i \).

Here, instead, we have a problem where all the observations affect \( x_{LS}^* \). 

→ a measure has been designed and used to rank the observations based on their influence on \( x_{LS}^* \) and \( C_i \) (related to the concept of “conformity score” in Machine Learning by V. Vovk et al.)
A. Carè, S. Garatti, and M.C. Campi,
“A Coverage Theory for Least Squares,”

See also:

https://seriesblog.net

Thank you!

www.algocare.com  algocare@gmail.com