

On the validity of minimin and minimax methods for Support Vector Regression with interval data

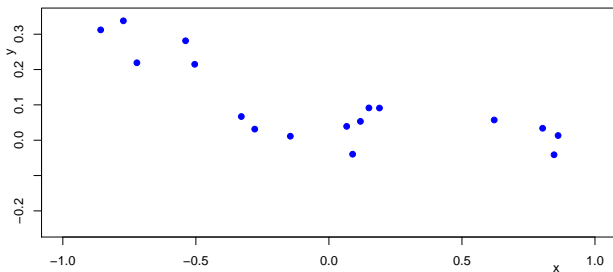
Andrea Wiencierz¹ Marco Cattaneo²

¹Department of Mathematics, University of York

²Department of Mathematics, University of Hull

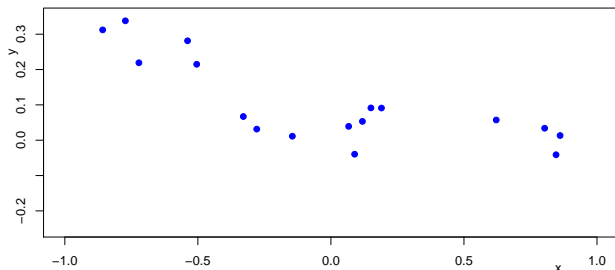
ISIPTA '15, Pescara, Italy
22 July 2015

support vector regression (SVR)



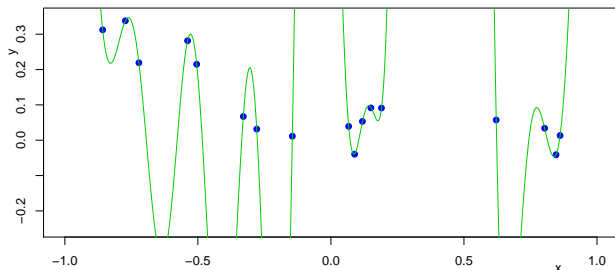
- ▶ precise data: $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^d \times \mathbb{R}$

support vector regression (SVR)



- ▶ **precise data:** $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{X} \times \mathcal{Y}^{\text{compact}} \subset \mathbb{R}^d \times \mathbb{R}$
- ▶ **reproducing kernel Hilbert space:** set \mathcal{F} of functions $f : \mathcal{X} \rightarrow \mathcal{Y}$, e.g., with Gaussian kernel, \mathcal{F} is dense in the space $\mathcal{C}(\mathcal{X})$ of continuous functions

support vector regression (SVR)

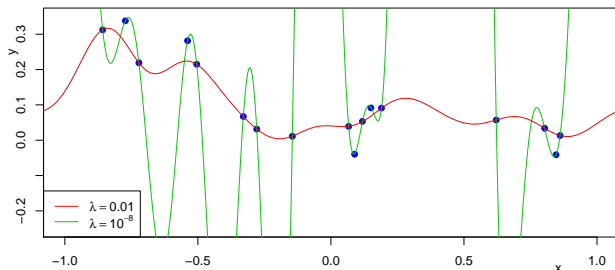


- ▶ precise data: $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{X} \times \mathcal{Y}^{\text{compact}} \subset \mathbb{R}^d \times \mathbb{R}$
- ▶ reproducing kernel Hilbert space: set \mathcal{F} of functions $f : \mathcal{X} \rightarrow \mathcal{Y}$, e.g., with Gaussian kernel, \mathcal{F} is dense in the space $\mathcal{C}(\mathcal{X})$ of continuous functions
- ▶ regression function: $f \in \mathcal{F}$ minimizing

$$\mathcal{E}(f) = \frac{1}{n} \sum_{i=1}^n \psi(|y_i - f(x_i)|),$$

where $\psi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is convex with $\psi(0) = 0$, e.g., linear loss $\psi : r \mapsto r$

support vector regression (SVR)

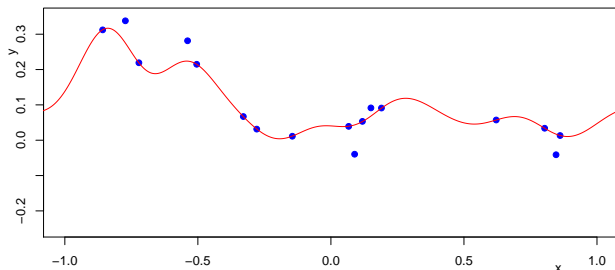


- ▶ **precise data:** $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{X} \times \mathcal{Y}^{\text{compact}} \subset \mathbb{R}^d \times \mathbb{R}$
- ▶ **reproducing kernel Hilbert space:** set \mathcal{F} of functions $f : \mathcal{X} \rightarrow \mathcal{Y}$, e.g., with Gaussian kernel, \mathcal{F} is dense in the space $\mathcal{C}(\mathcal{X})$ of continuous functions
- ▶ **regression function:** $f \in \mathcal{F}$ minimizing the regularized risk

$$\mathcal{E}(f) = \frac{1}{n} \sum_{i=1}^n \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2,$$

where $\psi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is convex with $\psi(0) = 0$, e.g., linear loss $\psi : r \mapsto r$, and $\lambda \in \mathbb{R}_{> 0}$

representer theorem (RT)

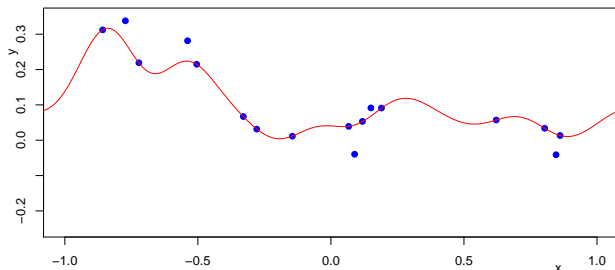


- ▶ the regression function minimizing $\mathcal{E}(f)$ exists, is unique, and has the form

$$f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j),$$

where $\alpha_j \in \mathbb{R}$, and κ is the kernel of \mathcal{F}

representer theorem (RT)



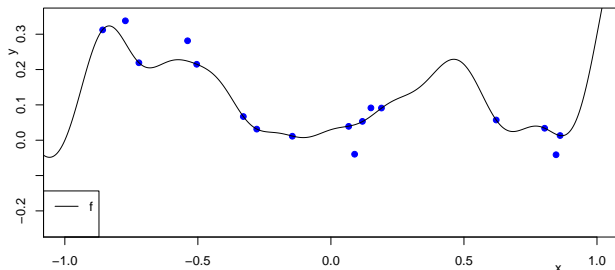
- ▶ the regression function minimizing $\mathcal{E}(f)$ exists, is unique, and has the form

$$f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j),$$

where $\alpha_j \in \mathbb{R}$, and κ is the kernel of \mathcal{F}

- ▶ the minimization of $\mathcal{E}(f)$ becomes a **convex optimization** task in n variables $\alpha_1, \dots, \alpha_n$: i.e., the RT makes the theoretical idea practically feasible

representer theorem (RT)



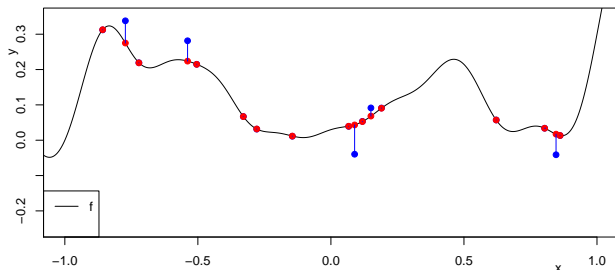
- ▶ the regression function minimizing $\mathcal{E}(f)$ exists, is unique, and has the form

$$f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j),$$

where $\alpha_j \in \mathbb{R}$, and κ is the kernel of \mathcal{F}

- ▶ the minimization of $\mathcal{E}(f)$ becomes a **convex optimization** task in n variables $\alpha_1, \dots, \alpha_n$: i.e., the RT makes the theoretical idea practically feasible
- ▶ core of the **proof**: the structure of \mathcal{F} implies that for each f , the orthogonal projection $f' = \sum_{j=1}^n \alpha'_j \kappa(\cdot, x_j)$ of f on the subspace spanned by the functions $\kappa(\cdot, x_j)$ satisfies $f'(x_i) = f(x_i)$, and therefore $\mathcal{E}(f') \leq \mathcal{E}(f)$

representer theorem (RT)



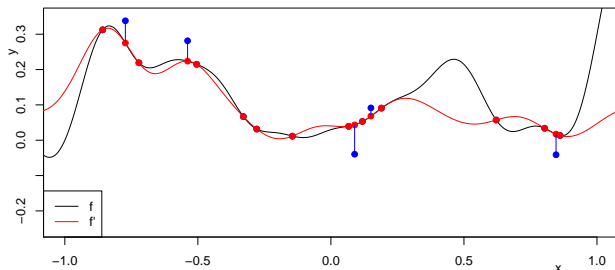
- ▶ the regression function minimizing $\mathcal{E}(f)$ exists, is unique, and has the form

$$f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j),$$

where $\alpha_j \in \mathbb{R}$, and κ is the kernel of \mathcal{F}

- ▶ the minimization of $\mathcal{E}(f)$ becomes a **convex optimization** task in n variables $\alpha_1, \dots, \alpha_n$: i.e., the RT makes the theoretical idea practically feasible
- ▶ core of the **proof**: the structure of \mathcal{F} implies that for each f , the orthogonal projection $f' = \sum_{j=1}^n \alpha'_j \kappa(\cdot, x_j)$ of f on the subspace spanned by the functions $\kappa(\cdot, x_j)$ satisfies $f'(x_i) = f(x_i)$, and therefore $\mathcal{E}(f') \leq \mathcal{E}(f)$

representer theorem (RT)



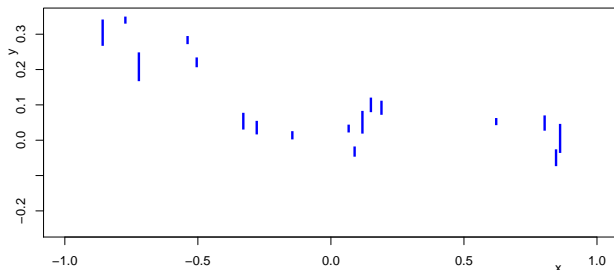
- ▶ the regression function minimizing $\mathcal{E}(f)$ exists, is unique, and has the form

$$f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j),$$

where $\alpha_j \in \mathbb{R}$, and κ is the kernel of \mathcal{F}

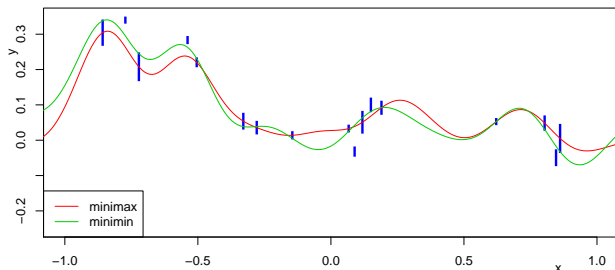
- ▶ the minimization of $\mathcal{E}(f)$ becomes a **convex optimization** task in n variables $\alpha_1, \dots, \alpha_n$: i.e., the RT makes the theoretical idea practically feasible
- ▶ core of the **proof**: the structure of \mathcal{F} implies that for each f , the orthogonal projection $f' = \sum_{j=1}^n \alpha'_j \kappa(\cdot, x_j)$ of f on the subspace spanned by the functions $\kappa(\cdot, x_j)$ satisfies $f'(x_i) = f(x_i)$, and therefore $\mathcal{E}(f') \leq \mathcal{E}(f)$

interval data



- ▶ instead of the values y_i , only the intervals $[\underline{y}_i, \bar{y}_i]$ are observed, with $y_i \in [\underline{y}_i, \bar{y}_i]$

interval data

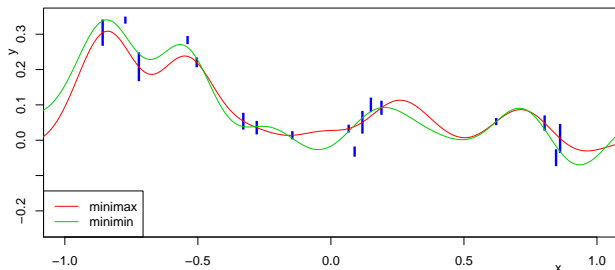


- ▶ instead of the values y_i , only the intervals $[\underline{y}_i, \bar{y}_i]$ are observed, with $y_i \in [\underline{y}_i, \bar{y}_i]$
- ▶ **minimin** and **minimax** SVR: $f \in \mathcal{F}$ minimizing respectively

$$\underline{\mathcal{E}}(f) = \frac{1}{n} \sum_{i=1}^n \min_{y_i \in [\underline{y}_i, \bar{y}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2$$

$$\bar{\mathcal{E}}(f) = \frac{1}{n} \sum_{i=1}^n \max_{y_i \in [\underline{y}_i, \bar{y}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2$$

interval data



- ▶ instead of the values y_i , only the intervals $[\underline{y}_i, \bar{y}_i]$ are observed, with $y_i \in [\underline{y}_i, \bar{y}_i]$
- ▶ **minimin and minimax SVR**: $f \in \mathcal{F}$ minimizing respectively

$$\underline{\mathcal{E}}(f) = \frac{1}{n} \sum_{i=1}^n \min_{y_i \in [\underline{y}_i, \bar{y}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2$$

$$\bar{\mathcal{E}}(f) = \frac{1}{n} \sum_{i=1}^n \max_{y_i \in [\underline{y}_i, \bar{y}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2$$

- ▶ **RT**: the regression functions minimizing $\underline{\mathcal{E}}(f)$ and $\bar{\mathcal{E}}(f)$ exist, are unique, and have the form $f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j)$, so that the minimizations of $\underline{\mathcal{E}}(f)$ and $\bar{\mathcal{E}}(f)$ become convex optimization tasks in n variables $\alpha_1, \dots, \alpha_n$

conclusion

- ▶ the main contribution of the paper is the generalization of the RT to the case with interval data $[\underline{y}_i, \bar{y}_i] \subset \mathbb{R}$

conclusion

- ▶ the main contribution of the paper is the generalization of the RT to the case with interval data $[y_i, \bar{y}_i] \subset \mathbb{R}$
- ▶ unfortunately, the RT cannot be directly generalized to the case with interval data $[x_i, \bar{x}_i] \subset \mathbb{R}^d$, in which

$$\begin{aligned}\underline{\mathcal{E}}(f) &= \frac{1}{n} \sum_{i=1}^n \min_{x_i \in [x_i, \bar{x}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2 \\ \bar{\mathcal{E}}(f) &= \frac{1}{n} \sum_{i=1}^n \max_{x_i \in [x_i, \bar{x}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2\end{aligned}$$

are minimized:

conclusion

- ▶ the main contribution of the paper is the generalization of the RT to the case with interval data $[\underline{y}_i, \bar{y}_i] \subset \mathbb{R}$
- ▶ unfortunately, the RT cannot be directly generalized to the case with interval data $[\underline{x}_i, \bar{x}_i] \subset \mathbb{R}^d$, in which

$$\underline{\mathcal{E}}(f) = \frac{1}{n} \sum_{i=1}^n \min_{x_i \in [\underline{x}_i, \bar{x}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2$$
$$\bar{\mathcal{E}}(f) = \frac{1}{n} \sum_{i=1}^n \max_{x_i \in [\underline{x}_i, \bar{x}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2$$

are minimized:

- ▶ a regression function minimizing $\underline{\mathcal{E}}(f)$ would have the form $f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j)$, where $\alpha_j \in \mathbb{R}$ and $x_j \in [\underline{x}_j, \bar{x}_j]$, but in general $\underline{\mathcal{E}}$ is **not** convex

conclusion

- ▶ the main contribution of the paper is the generalization of the RT to the case with interval data $[\underline{y}_i, \bar{y}_i] \subset \mathbb{R}$
- ▶ unfortunately, the RT cannot be directly generalized to the case with interval data $[\underline{x}_i, \bar{x}_i] \subset \mathbb{R}^d$, in which

$$\begin{aligned}\underline{\mathcal{E}}(f) &= \frac{1}{n} \sum_{i=1}^n \min_{x_i \in [\underline{x}_i, \bar{x}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2 \\ \bar{\mathcal{E}}(f) &= \frac{1}{n} \sum_{i=1}^n \max_{x_i \in [\underline{x}_i, \bar{x}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2\end{aligned}$$

are minimized:

- ▶ a regression function minimizing $\underline{\mathcal{E}}(f)$ would have the form $f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j)$, where $\alpha_j \in \mathbb{R}$ and $x_j \in [\underline{x}_j, \bar{x}_j]$, but in general $\underline{\mathcal{E}}$ is **not** convex
- ▶ by contrast, $\bar{\mathcal{E}}$ is convex, but a regression function minimizing $\bar{\mathcal{E}}(f)$ does **not** necessarily have the form $f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j)$, where $\alpha_j \in \mathbb{R}$ and $x_j \in [\underline{x}_j, \bar{x}_j]$

conclusion

- ▶ the main contribution of the paper is the generalization of the RT to the case with interval data $[\underline{y}_i, \bar{y}_i] \subset \mathbb{R}$
- ▶ unfortunately, the RT cannot be directly generalized to the case with interval data $[\underline{x}_i, \bar{x}_i] \subset \mathbb{R}^d$, in which

$$\begin{aligned}\underline{\mathcal{E}}(f) &= \frac{1}{n} \sum_{i=1}^n \min_{x_i \in [\underline{x}_i, \bar{x}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2 \\ \bar{\mathcal{E}}(f) &= \frac{1}{n} \sum_{i=1}^n \max_{x_i \in [\underline{x}_i, \bar{x}_i]} \psi(|y_i - f(x_i)|) + \lambda \|f\|_{\mathcal{F}}^2\end{aligned}$$

are minimized:

- ▶ a regression function minimizing $\underline{\mathcal{E}}(f)$ would have the form $f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j)$, where $\alpha_j \in \mathbb{R}$ and $x_j \in [\underline{x}_j, \bar{x}_j]$, but in general $\underline{\mathcal{E}}$ is **not** convex
- ▶ by contrast, $\bar{\mathcal{E}}$ is convex, but a regression function minimizing $\bar{\mathcal{E}}(f)$ does **not** necessarily have the form $f = \sum_{j=1}^n \alpha_j \kappa(\cdot, x_j)$, where $\alpha_j \in \mathbb{R}$ and $x_j \in [\underline{x}_j, \bar{x}_j]$
- ▶ the more general case with interval data $[\underline{x}_i, \bar{x}_i] \times [\underline{y}_i, \bar{y}_i] \subset \mathbb{R}^d \times \mathbb{R}$ also presents the above difficulties