Scenario Optimization:

Data-Driven Goal-Oriented Designs with Certified Reliability

Marco C. Campi

University of Brescia, Italy

(*joint work with Simone Garatti and Algo Carè*)

SPS Seminar – May 2024

1. a quandary: indirect or direct?

→ stochastic programming deals with optimization problems wherein uncertainty is described by means of probability distributions

- → stochastic programming deals with optimization problems wherein uncertainty is described by means of probability distributions

- → but, should we use the data to obtain such a "description"? or, rather, should we aim at directly optimizing?

⟶ stochastic programming deals with optimization problems wherein uncertainty is described by means of probability distributions

⟶ but, should we use the data to obtain such a “description”? or, rather, should we aim at directly optimizing?

⟶ especially important in today’s world as we confront increasingly complex problems

example: portfolio optimization

example: portfolio optimization

Black and Scholes → log-normal distribution

example: portfolio optimization

Black and Scholes → log-normal distribution

d assets

p_k = percentage of capital invested on asset k

$\theta = [p_1 \cdots p_d]^T$

$$J(\theta, \delta) = \sum_{k=1}^{d} p_k R_k \qquad \delta = [R_1 \cdots R_d]^T$$

example: portfolio optimization

Black and Scholes → log-normal distribution

d assets

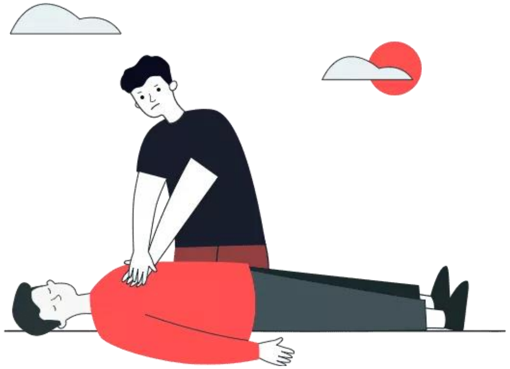
p_k = percentage of capital invested on asset k

$\theta = [p_1 \cdots p_d]^T$

$$J(\theta, \delta) = \sum_{k=1}^{d} p_k R_k \qquad \delta = [R_1 \cdots R_d]^T$$

d = 100 results in 5150 parameters!

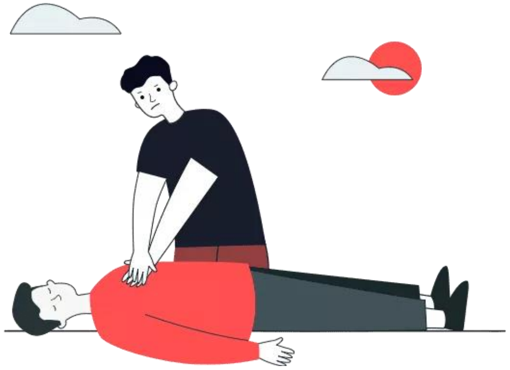
example: defibrillation



goal: predict whether the defibrillator shock will be effective

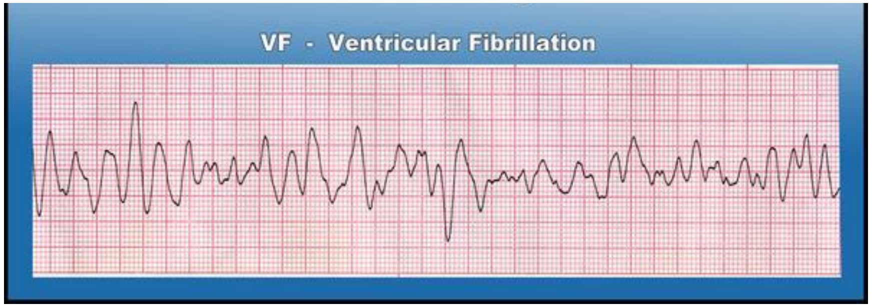
[with A. Caré]

example: defibrillation



goal: predict whether the defibrillator shock will be effective

$$\delta_i = (u_i, y_i)$$

$u_i =$ 

$$y_i = \begin{cases} 0, & \text{uneffective shock} \\ 1, & \text{effective shock} \end{cases}$$

[with A. Caré]

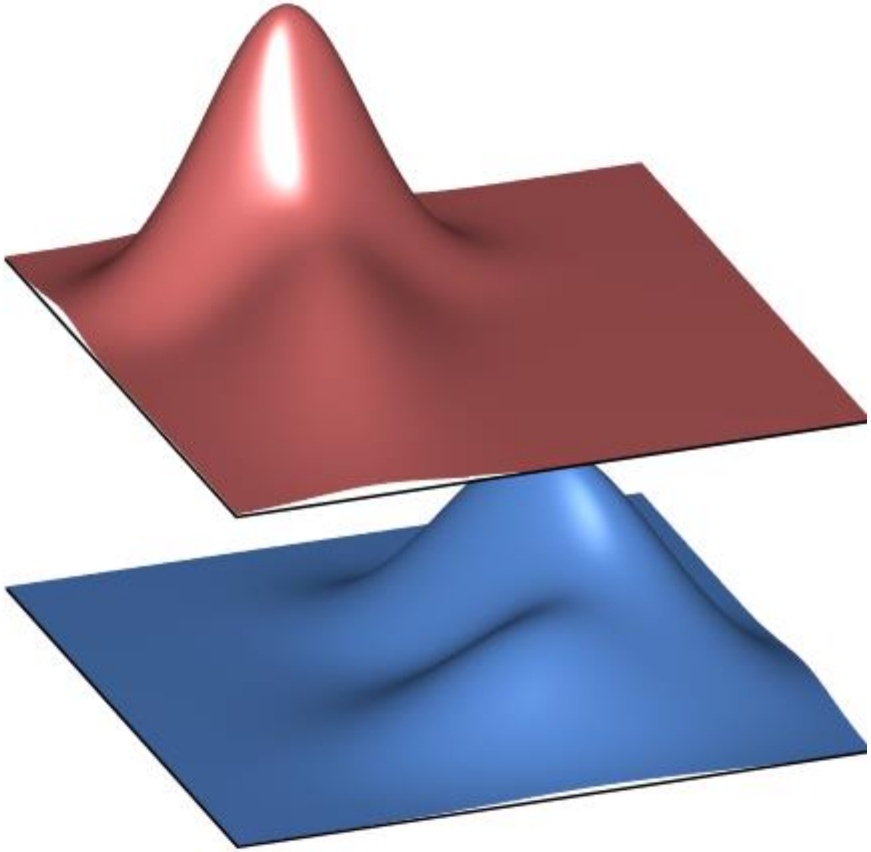
example: defibrillation

ECG = infinite dimentional object → extract finitely many features

- *peak-to-peak (PTP)*
- *maximum amplitude (Amax)*
- *minimum amplitude (Amin)*
- *wave amplitude (WA)*
- *root mean square (RMS)*
- *dominant frequency (DF)*
- *centroid frequency (CF)*
- *edge frequency (EF)*
- *amplitude spectral area (AMSA)*

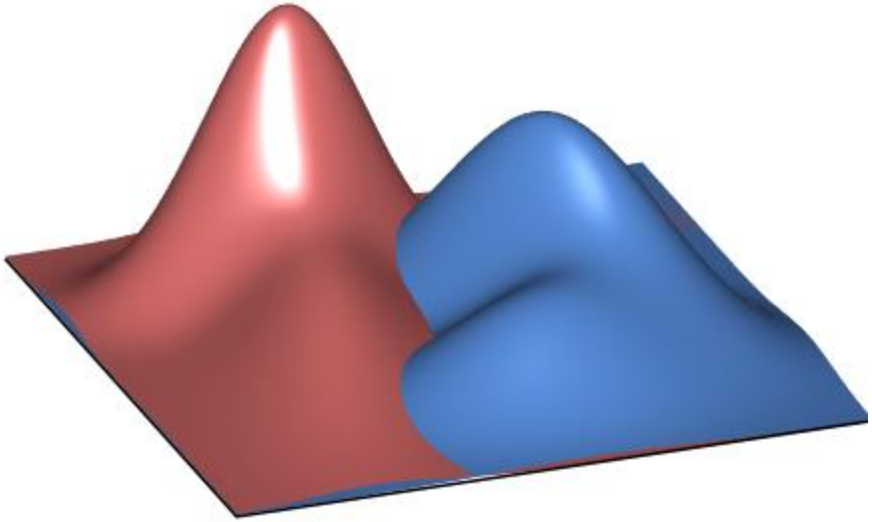
[with A. Caré]

example: defibrillation



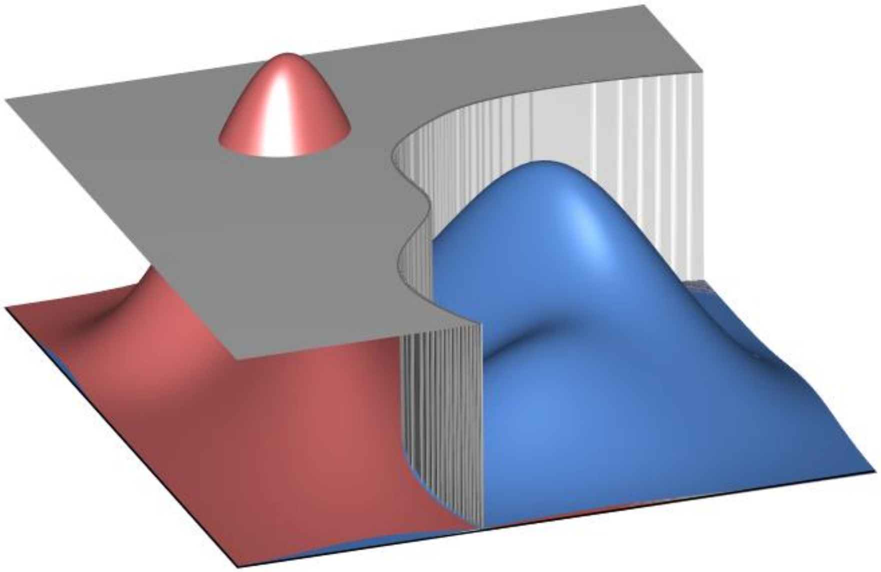
[with A. Caré]

example: defibrillation



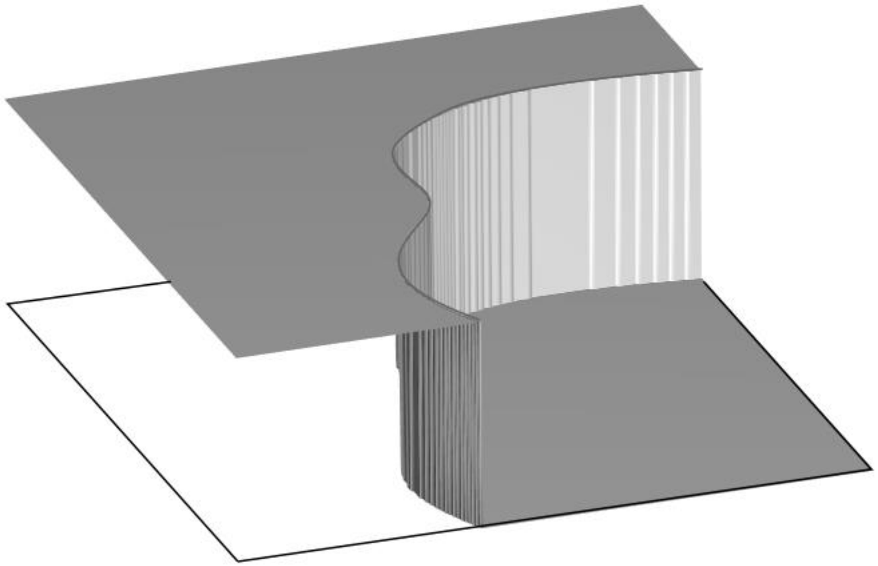
[with A. Caré]

example: defibrillation



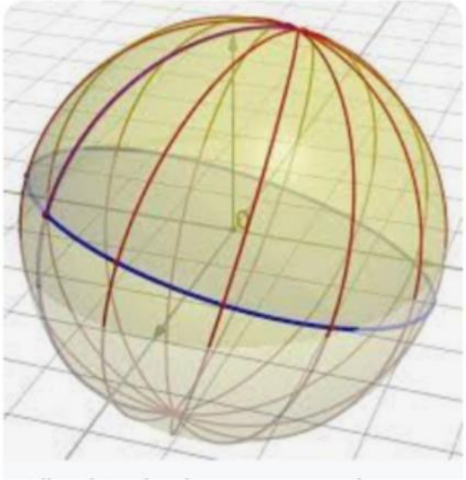
[with A. Caré]

example: defibrillation



[with A. Caré]

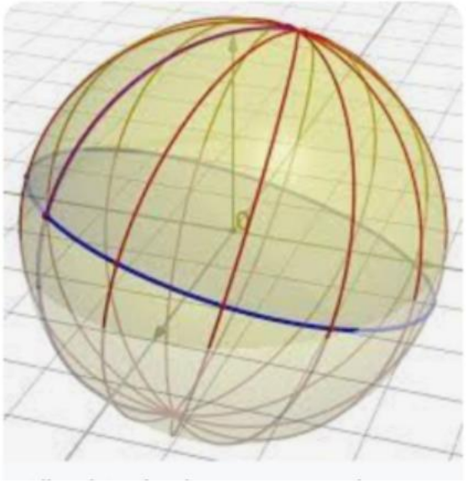
The Distributionally-Robust approach



space of distributions

Wasserstein ball

The Distributionally-Robust approach



space of distributions

Wasserstein ball

⟶ can be conservative

So:

→ for simple problems, the indirect path works well

→ however, as the complexity of the problem increases, the indirect path falls short

→ therefore, in real problems it may be convenient to aim for a direct path: from data to decisions

2. certified data-driven optimization

⟶ we should aim to develop a science for data-driven optimization in the presence of uncertainty

⟶ we should aim to develop a science for data-driven optimization in the presence of uncertainty

→ *provide guarantees (certificates of quality)*

⟶ we should aim to develop a science for data-driven optimization in the presence of uncertainty

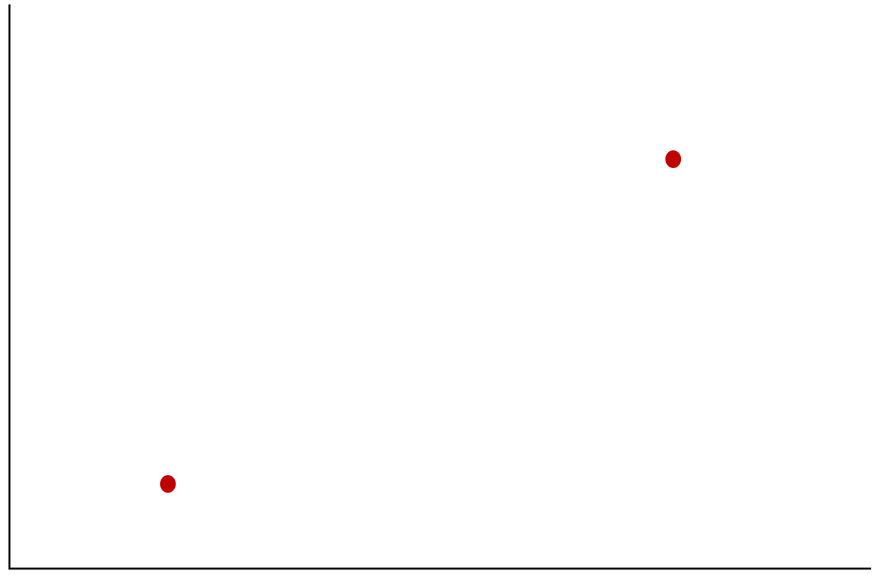
→ *provide guarantees (certificates of quality)*

→ *use priors*

⟶ we should aim to develop a science for data-driven optimization in the presence of uncertainty

→ *provide guarantees (certificates of quality)*

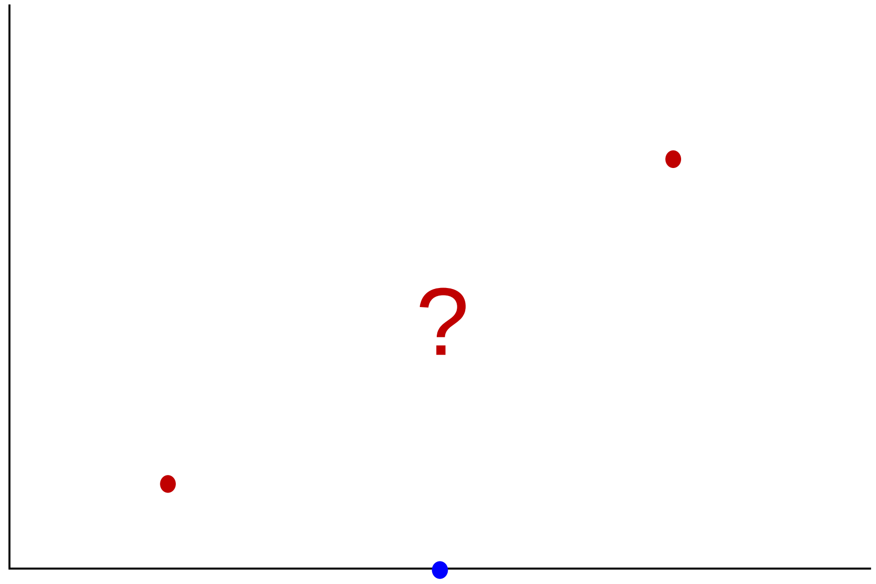
→ *use priors*



⟶ we should aim to develop a science for data-driven optimization in the presence of uncertainty

→ *provide guarantees (certificates of quality)*

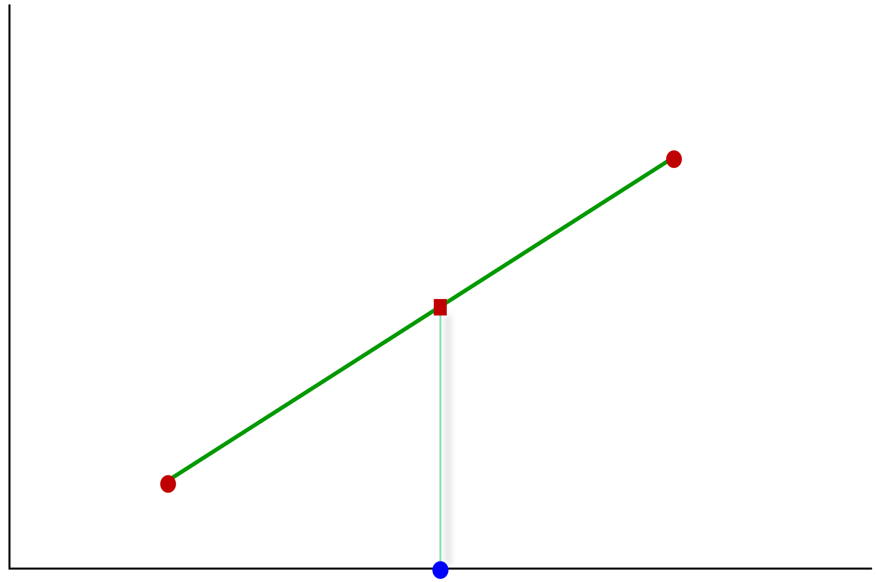
→ *use priors*



⟶ we should aim to develop a science for data-driven optimization in the presence of uncertainty

→ *provide guarantees (certificates of quality)*

→ *use priors*



⟶ we should aim to develop a science for data-driven optimization in the presence of uncertainty

→ *provide guarantees (certificates of quality)*

→ *use priors*

should we aim higher?

→ we should aim to develop a science for data-driven optimization in the presence of uncertainty

→ *provide guarantees (certificates of quality)*

→ *use priors*

should we aim higher?

prior + data → design

~~prior~~ + data → guarantees

⟶ we should aim to develop a science for data-driven optimization in the presence of uncertainty

→ *provide guarantees (certificates of quality)*

→ *use priors*

<u>should we aim higher?</u>

prior + data → design

~~prior~~ + data → guarantees

is this at all possible?

the Scenario Approach: an eminently direct approach

robust optimization

$$\begin{array}{rl} \min\limits_{x} & c(x) \\ \text{subject to:} & x \in \mathcal{X}_{\delta_i}, \quad i = 1, \ldots, N \end{array}$$

robust optimization

$$\min_{x} \quad c(x)$$

$$\text{subject to:} \quad x \in \mathcal{X}_{\delta_i}, \quad i = 1, \ldots, N$$

$$\longrightarrow \quad \text{Risk} = \mathbb{P}\{\delta : x^* \notin \mathcal{X}_\delta\}$$

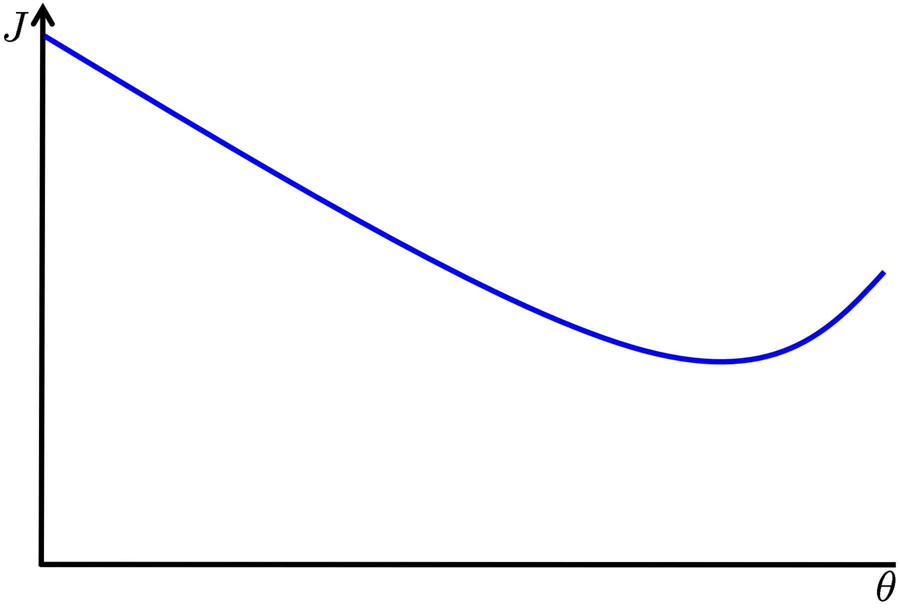
example: minimize $J(\theta, \delta)$

example: minimize $J(\theta, \delta)$

$$\delta_1, \delta_2, \ldots, \delta_N \quad \longrightarrow \quad J(\theta, \delta_1), J(\theta, \delta_2), \ldots, J(\theta, \delta_N)$$

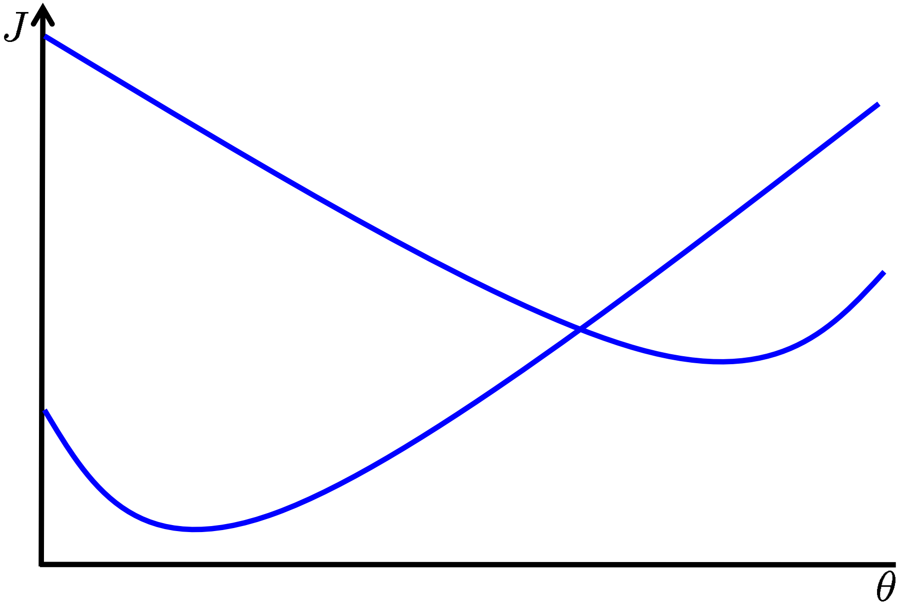
example: minimize $J(\theta, \delta)$

$$\delta_1, \delta_2, \ldots, \delta_N \quad \longrightarrow \quad J(\theta, \delta_1), J(\theta, \delta_2), \ldots, J(\theta, \delta_N)$$



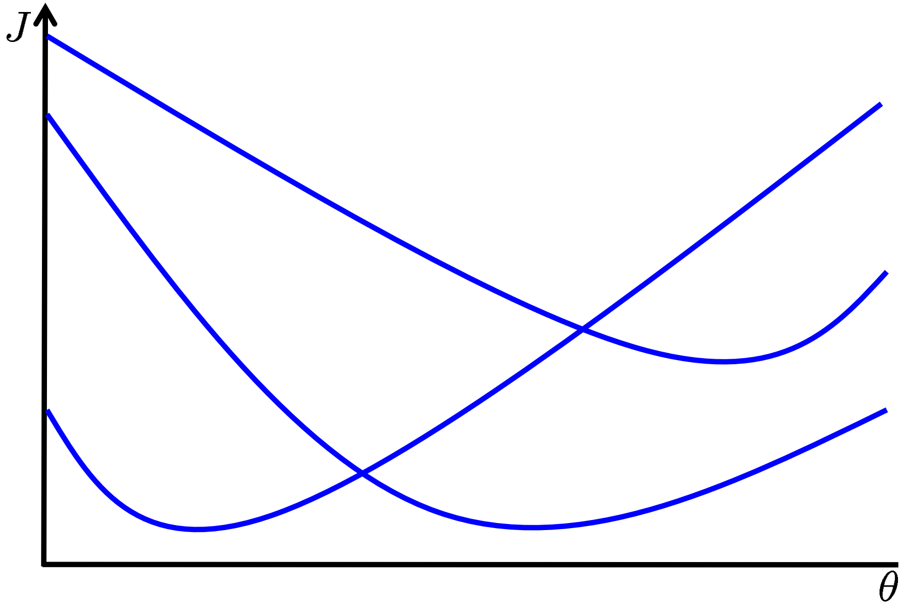
example: minimize $J(\theta, \delta)$

$$\delta_1, \delta_2, \ldots, \delta_N \quad \longrightarrow \quad J(\theta, \delta_1), J(\theta, \delta_2), \ldots, J(\theta, \delta_N)$$



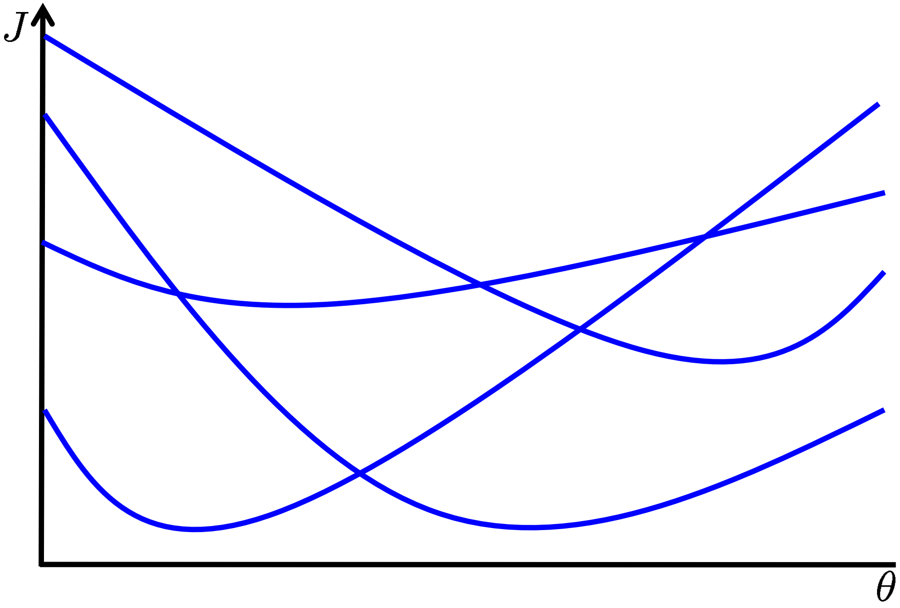
example: minimize $J(\theta, \delta)$

$$\delta_1, \delta_2, \ldots, \delta_N \longrightarrow J(\theta, \delta_1), J(\theta, \delta_2), \ldots, J(\theta, \delta_N)$$



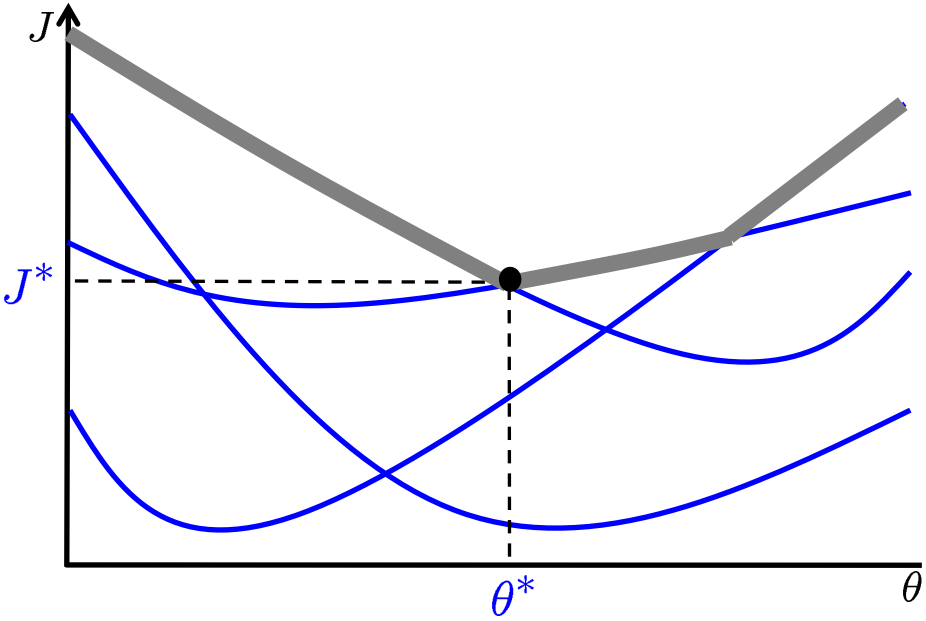
example: minimize $J(\theta, \delta)$

$$\delta_1, \delta_2, \ldots, \delta_N \longrightarrow J(\theta, \delta_1), J(\theta, \delta_2), \ldots, J(\theta, \delta_N)$$



example: minimize $J(\theta,\delta)$

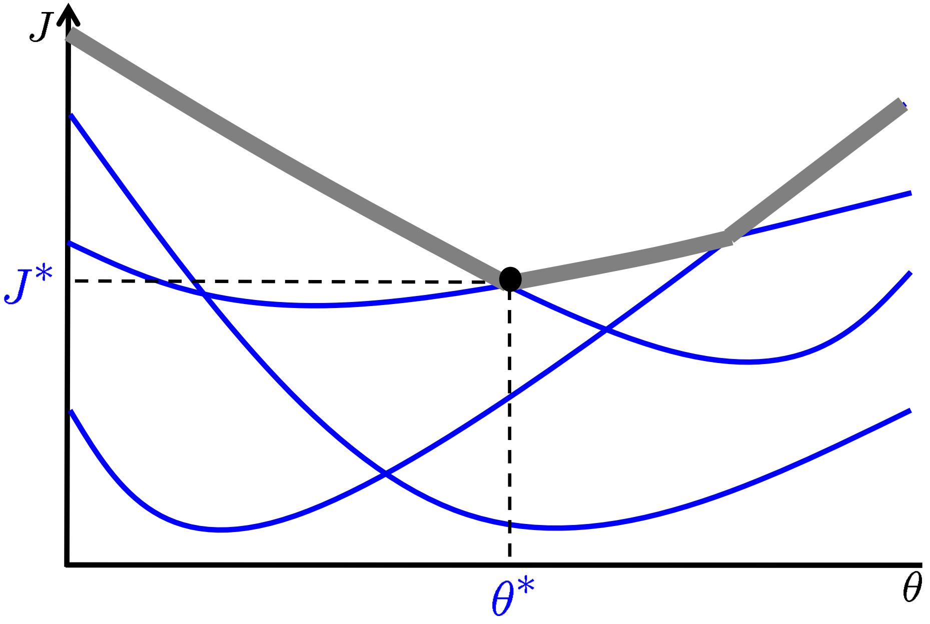
$$\delta_1, \delta_2, \ldots, \delta_N \longrightarrow J(\theta,\delta_1), J(\theta,\delta_2), \ldots, J(\theta,\delta_N)$$



$$\min_{\theta} \max_{\delta_i, i=1,\ldots,N} J(\theta,\delta_i)$$

example: minimize $J(\theta, \delta)$

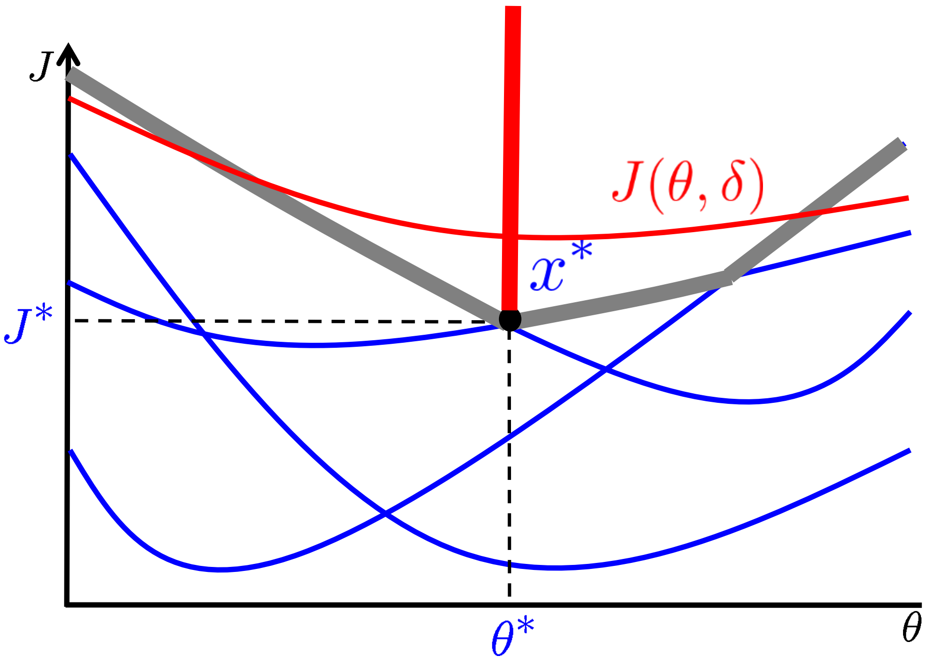
$$\delta_1, \delta_2, \ldots, \delta_N \quad \longrightarrow \quad J(\theta, \delta_1), J(\theta, \delta_2), \ldots, J(\theta, \delta_N)$$



$$\min_{\theta} \max_{\delta_i, i=1,\ldots,N} J(\theta, \delta_i) \quad \Longrightarrow \quad \begin{array}{ll} \displaystyle\min_{x=(\theta,J)} & J \\ \text{subject to:} & J \geq J(\theta, \delta_i), \quad i = 1, \ldots, N \end{array}$$

example: minimize $J(\theta,\delta)$

$$\delta_1, \delta_2, \ldots, \delta_N \quad \longrightarrow \quad J(\theta,\delta_1), J(\theta,\delta_2), \ldots, J(\theta,\delta_N)$$



$$\text{Risk} = \mathbb{P}\{\delta : J^* \not\geq J(\theta^*,\delta)\}$$

$$\Uparrow$$

$$\min_{\theta} \max_{\delta_i, i=1,\ldots,N} J(\theta,\delta_i) \quad \Rightarrow \quad \begin{array}{l} \min\limits_{x=(\theta,J)} \quad J \\ \text{subject to: } \; J \geq J(\theta,\delta_i), \quad i = 1,\ldots,N \end{array}$$

robust optimization

$$\begin{aligned} \min_{x} \quad & c(x) \\ \text{subject to:} \quad & x \in \mathcal{X}_{\delta_i}, \quad i = 1, \ldots, N \end{aligned}$$

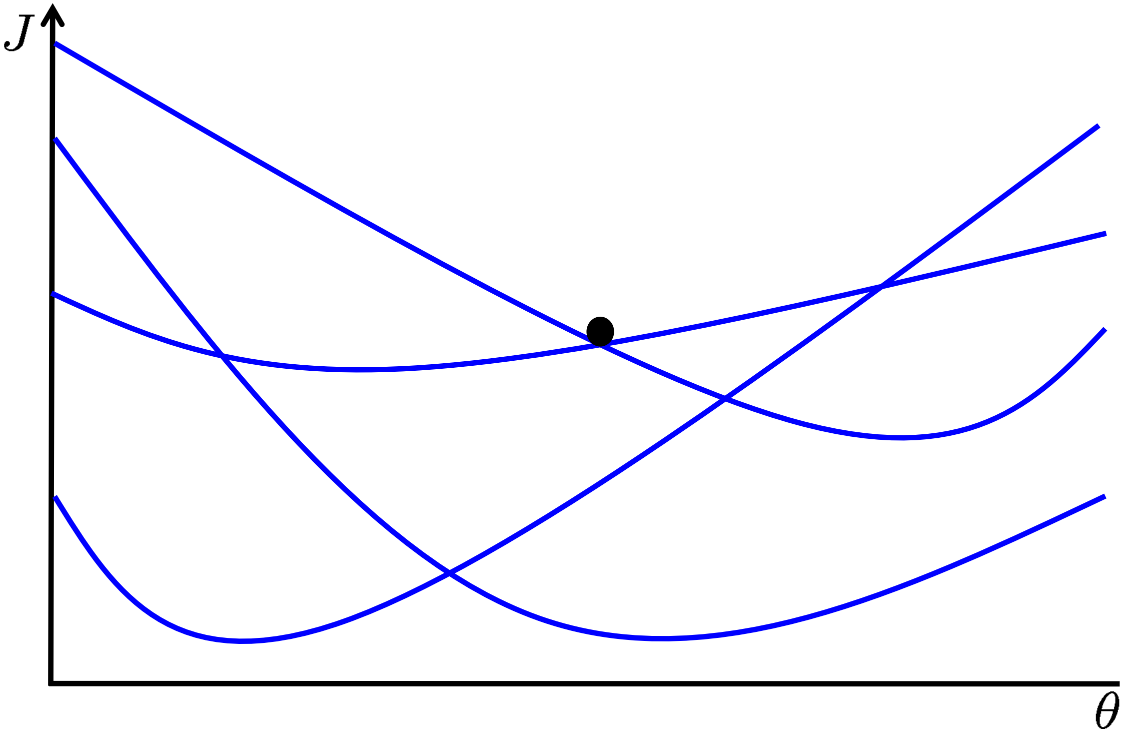
$$\longrightarrow \quad \mathrm{Risk} = \mathbb{P}\{\delta : x^* \notin \mathcal{X}_\delta\}$$

⟶ *can we certify the Risk?*

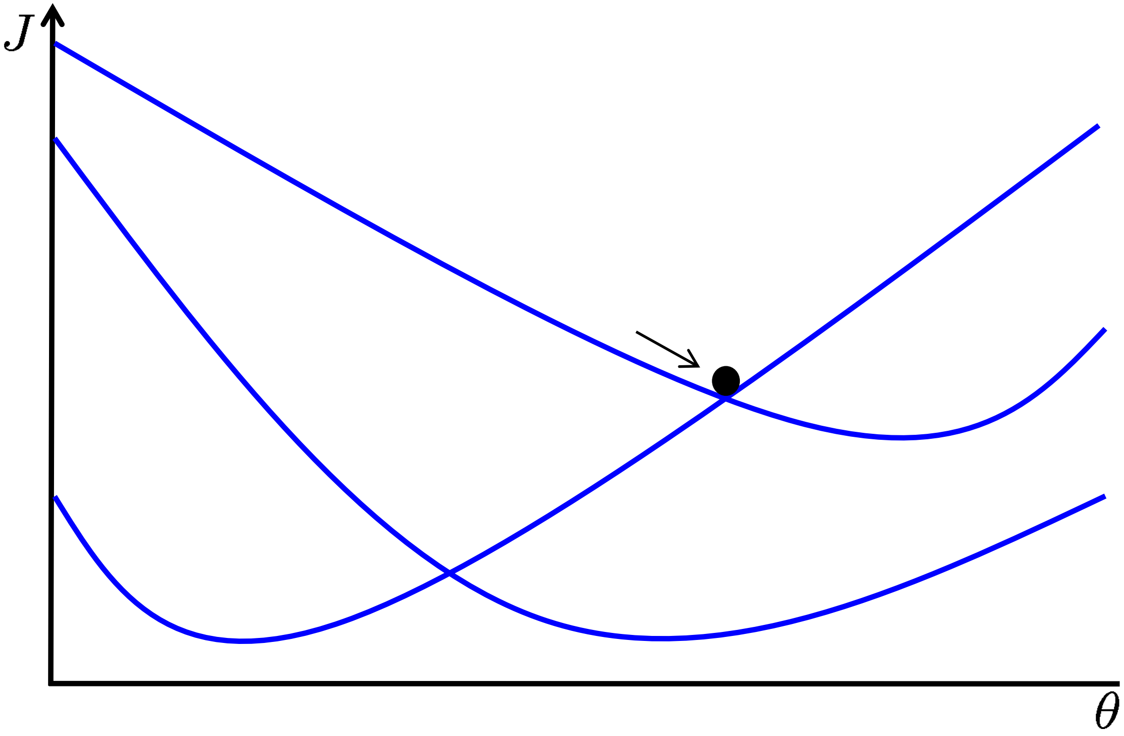
⟶ *can we certify the Risk?*

a beautiful marriage: Risk and Complexity

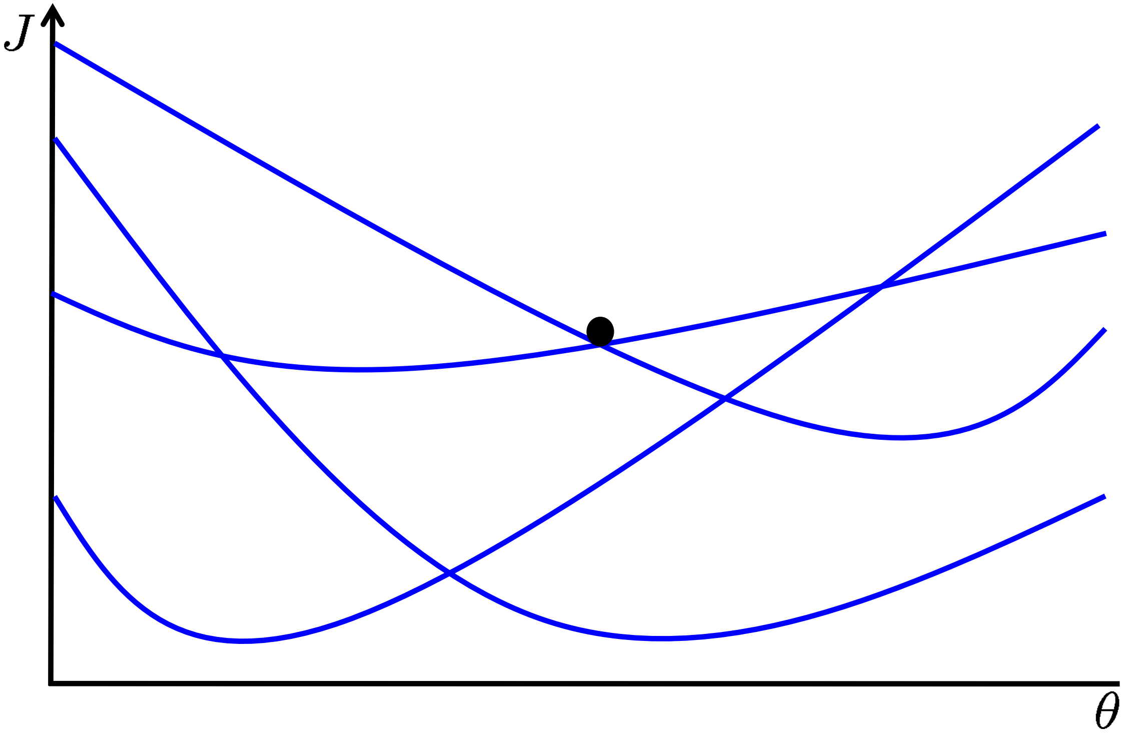
complexity



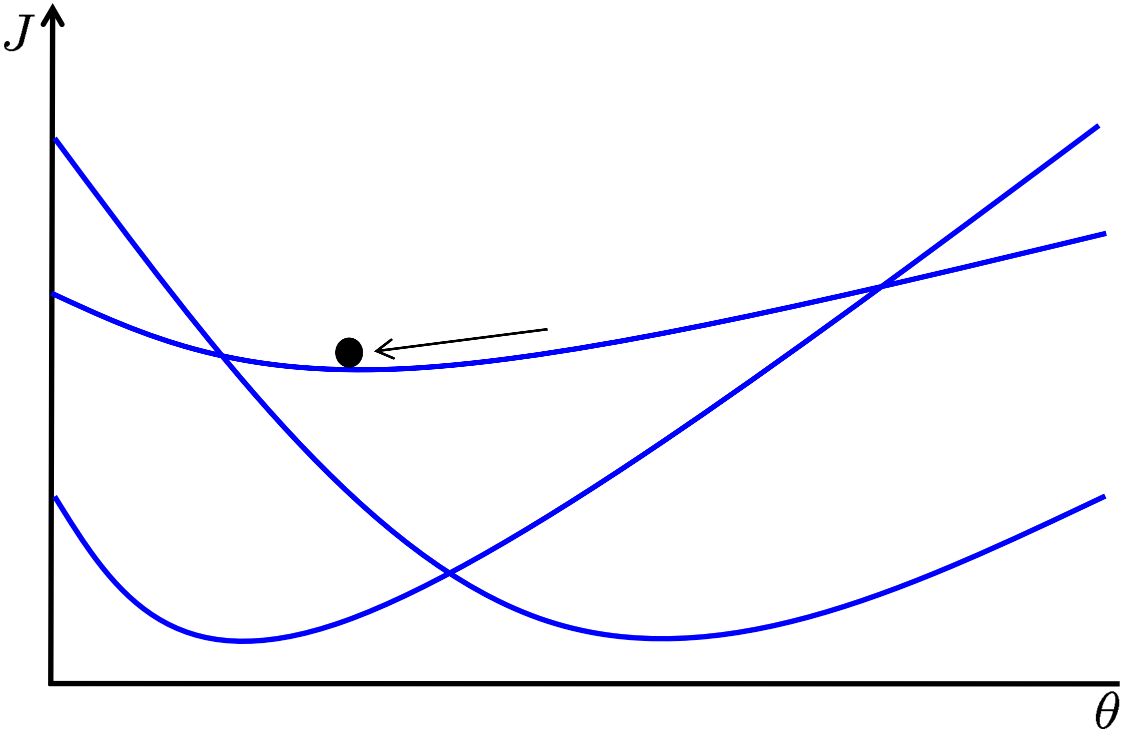
complexity



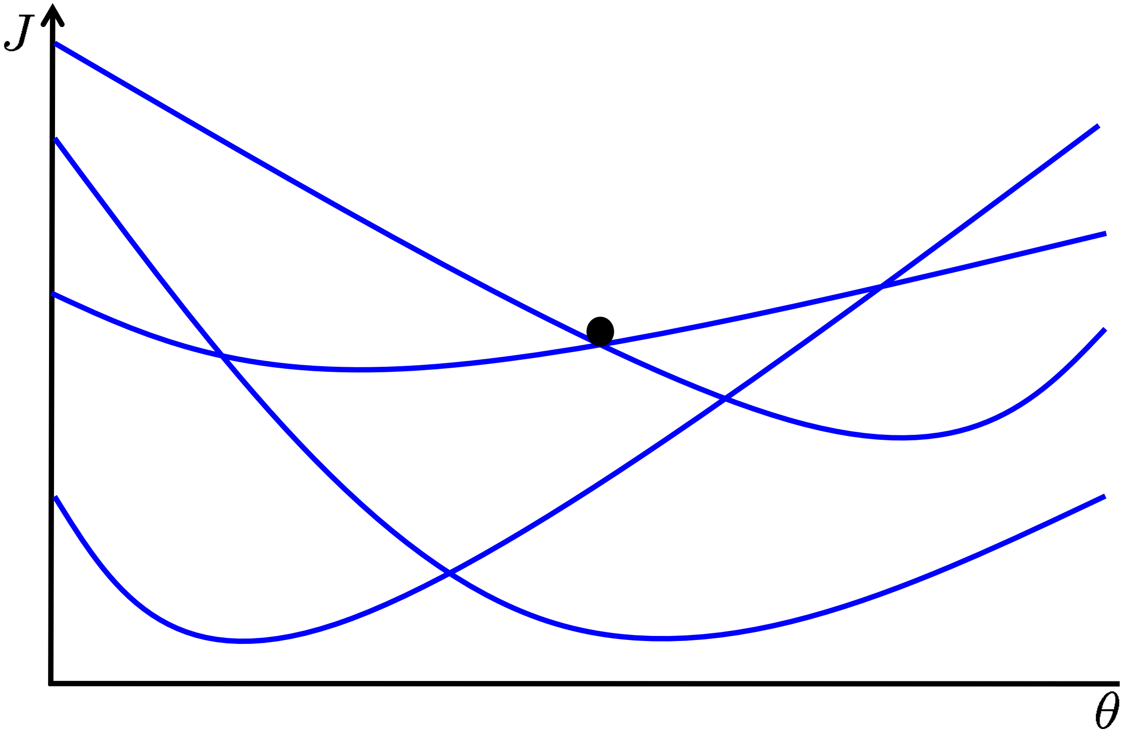
complexity



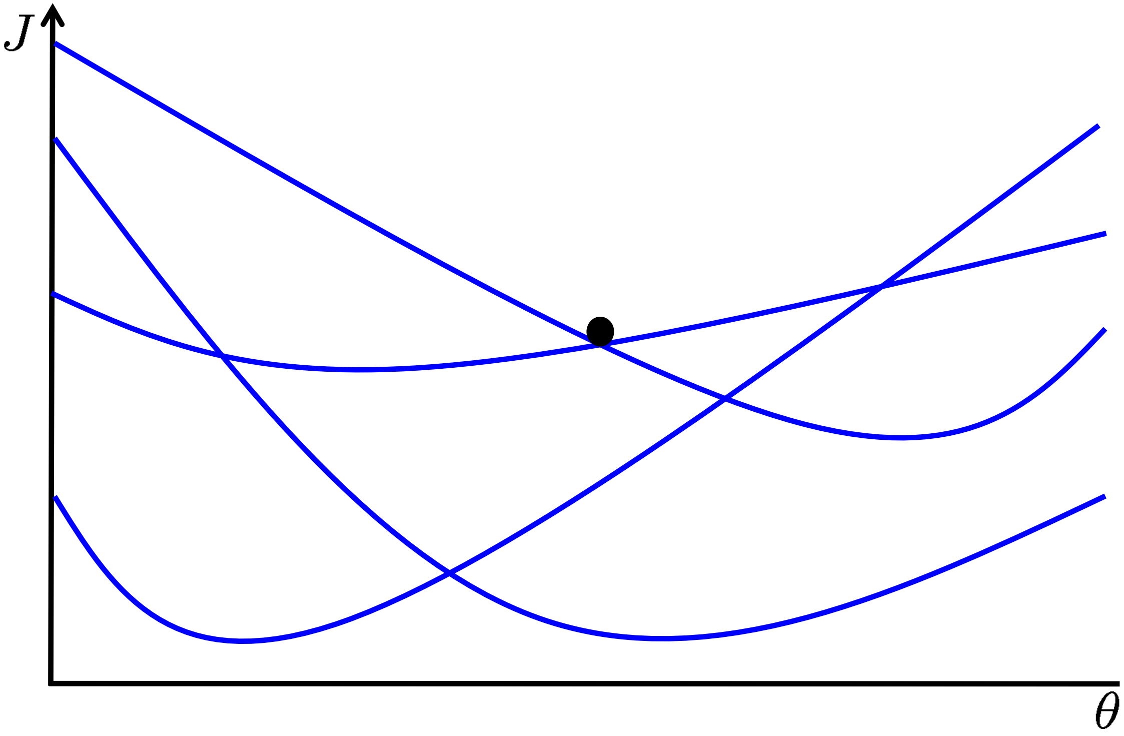
complexity



complexity



complexity

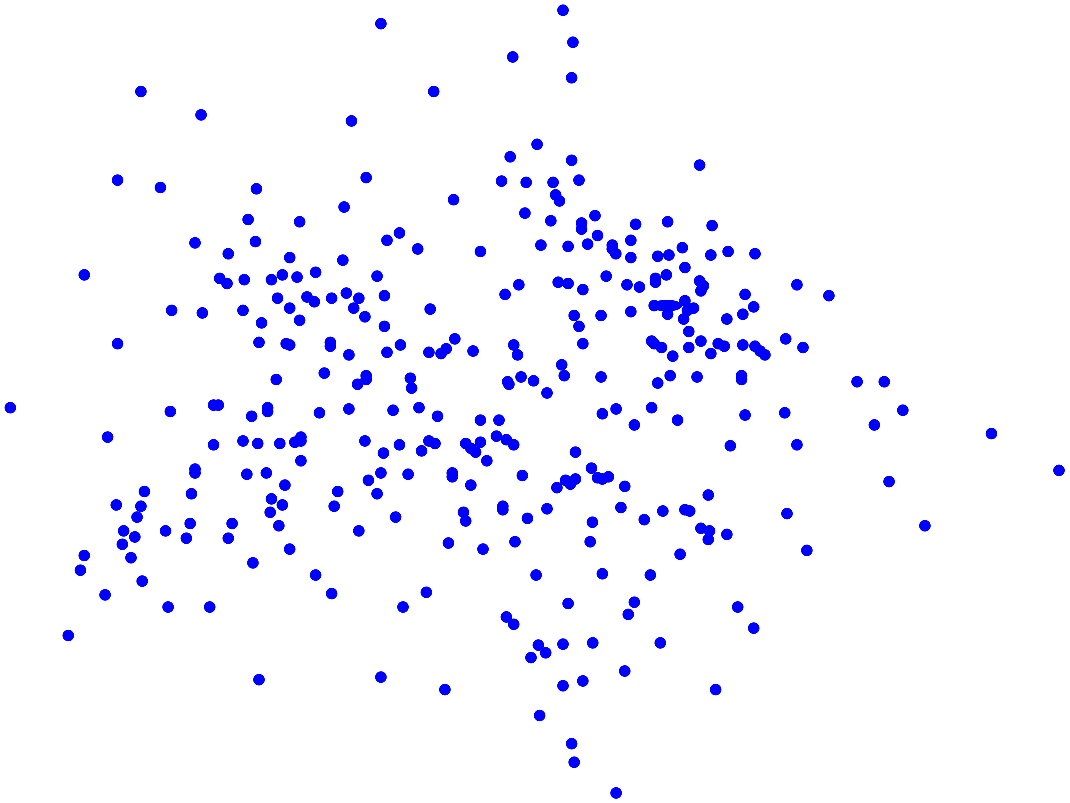


complexity = 2

⟶ if only two δ_i are maintained, solution doesn't change

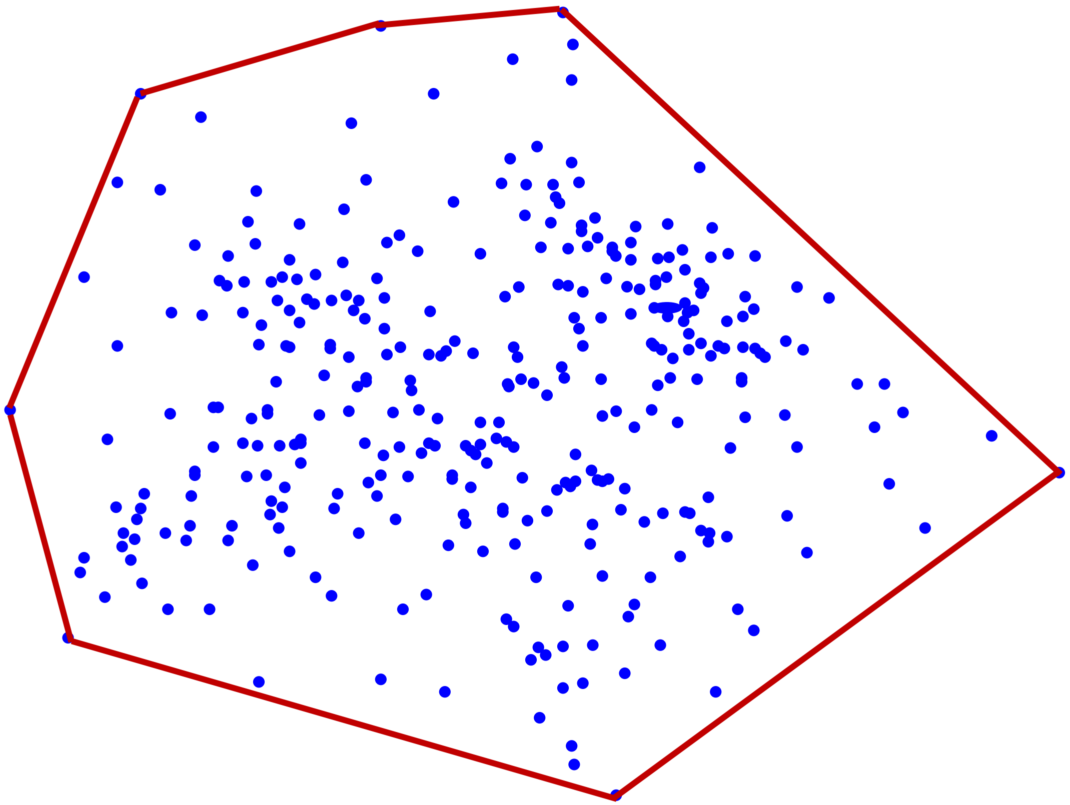
complexity - another example

500 points



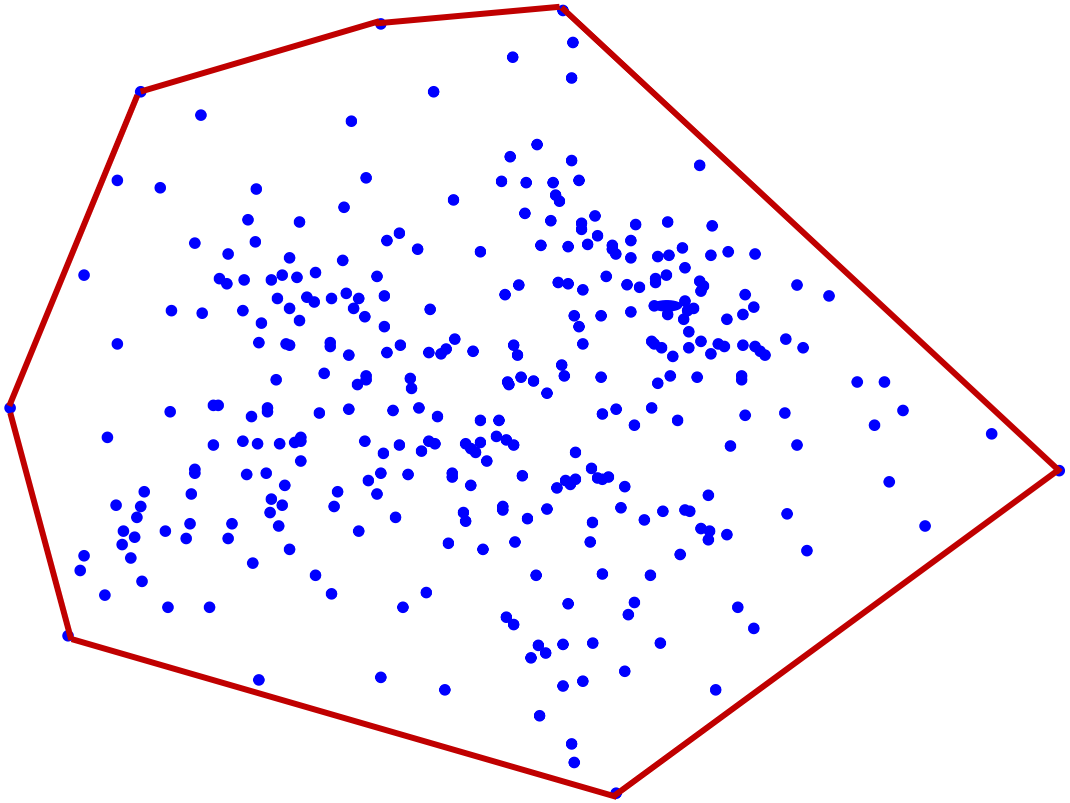
complexity - another example

500 points



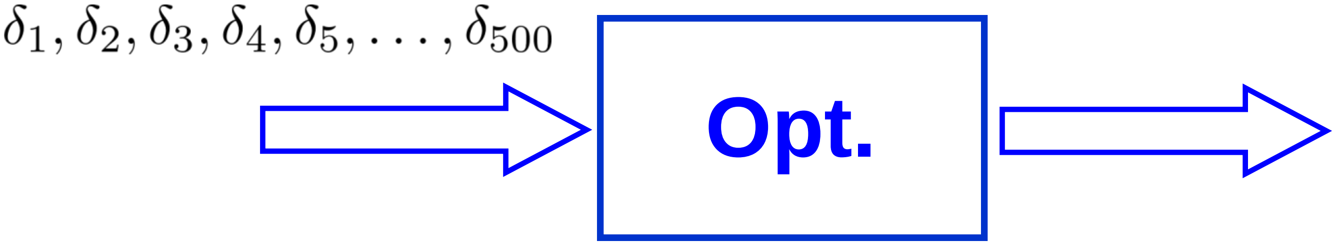
complexity - another example

500 points



7 points at the boundary

complexity - another example

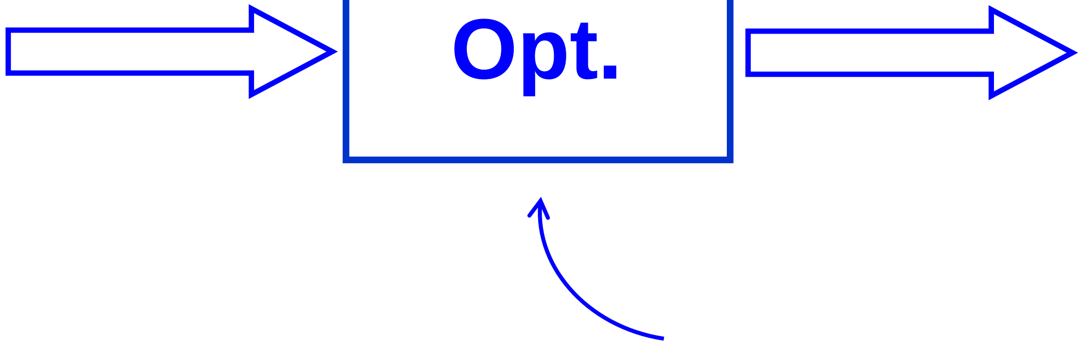


$\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \ldots, \delta_{500}$

Opt.

complexity - another example

$\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \ldots, \delta_{500}$



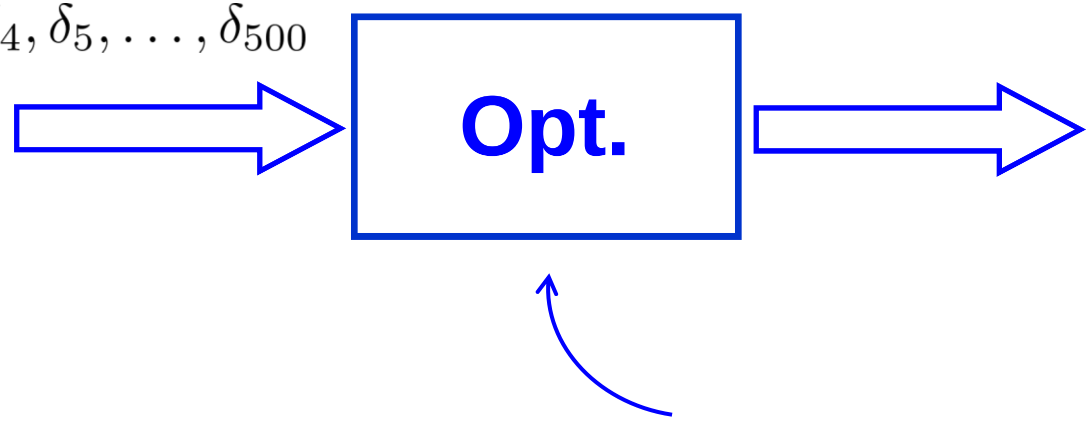
Opt.

$$\min_{S \in \{\text{convex sets}\}} \quad Vol(S)$$

$$\text{subject to:} \quad \delta_i \in S$$

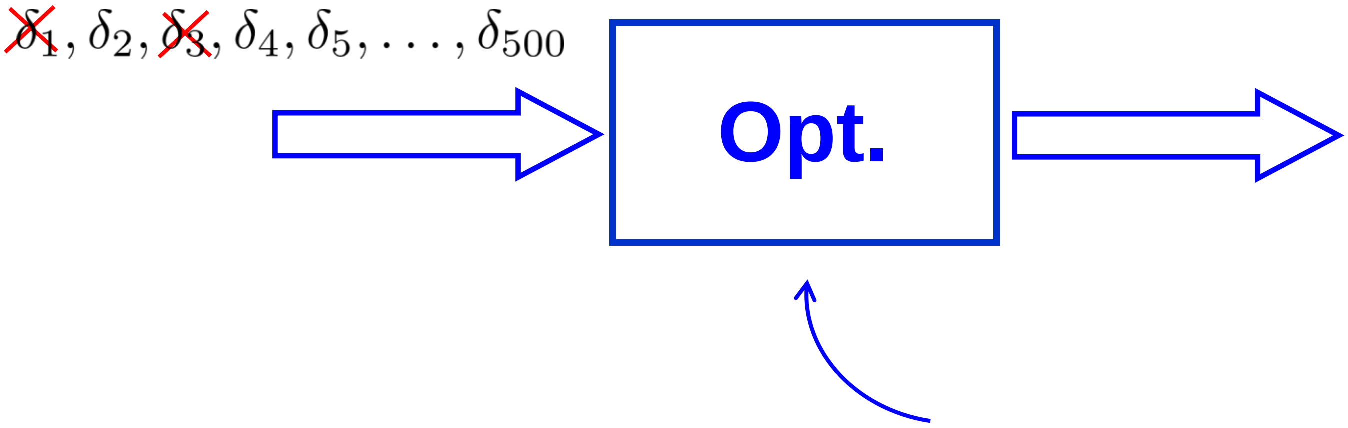
complexity - another example

$\delta_1, \delta_2, \xcancel{\delta_3}, \delta_4, \delta_5, \ldots, \delta_{500}$



$$\begin{aligned} \min_{S\in\{\text{convex sets}\}} \quad & Vol(S) \\ \text{subject to:} \quad & \delta_i \in S \end{aligned}$$

complexity - another example



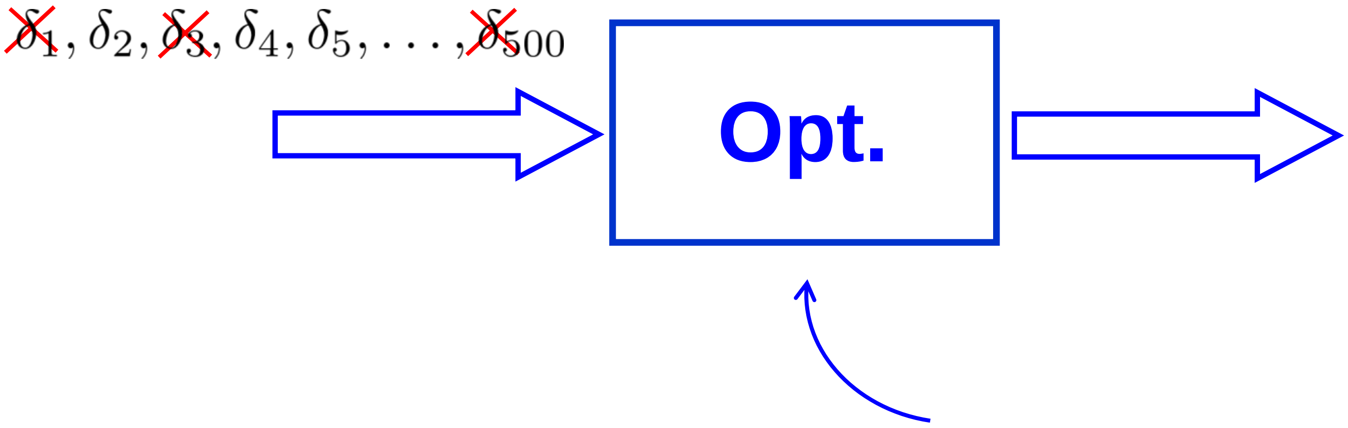
$\cancel{\delta_1}, \delta_2, \cancel{\delta_3}, \delta_4, \delta_5, \ldots, \delta_{500}$

Opt.

$$\min_{S \in \{\text{convex sets}\}} \quad Vol(S)$$

$$\text{subject to:} \quad \delta_i \in S$$

complexity - another example

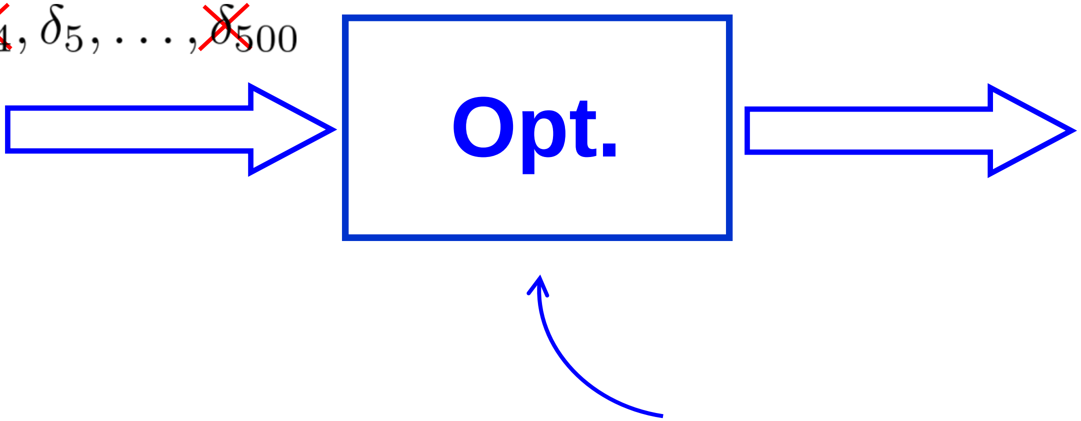


$\cancel{\delta_1}, \delta_2, \cancel{\delta_3}, \delta_4, \delta_5, \ldots, \cancel{\delta_{500}}$

Opt.

$$\begin{aligned} \min_{S\in\{\text{convex sets}\}} \quad & Vol(S) \\ \text{subject to:} \quad & \delta_i \in S \end{aligned}$$

complexity - another example

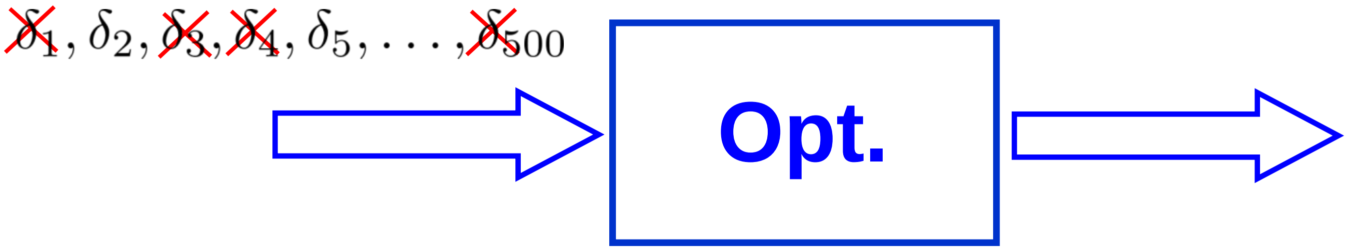


$\cancel{\delta_1}, \delta_2, \cancel{\delta_3}, \cancel{\delta_4}, \delta_5, \ldots, \cancel{\delta_{500}}$

Opt.

$$\min_{S \in \{\text{convex sets}\}} \quad Vol(S)$$
$$\text{subject to:} \quad \delta_i \in S$$

complexity - another example



$\cancel{\delta_1}, \delta_2, \cancel{\delta_3}, \cancel{\delta_4}, \delta_5, \ldots, \cancel{\delta_{500}}$

Opt.

7 points at the boundary $\Rightarrow$ if only these 7 points are maintained,
"Opt" gives the same solution

complexity = 7

complexity - another example

$\cancel{\delta_1}, \delta_2, \cancel{\delta_3}, \cancel{\delta_4}, \delta_5, \ldots, \cancel{\delta_{500}}$

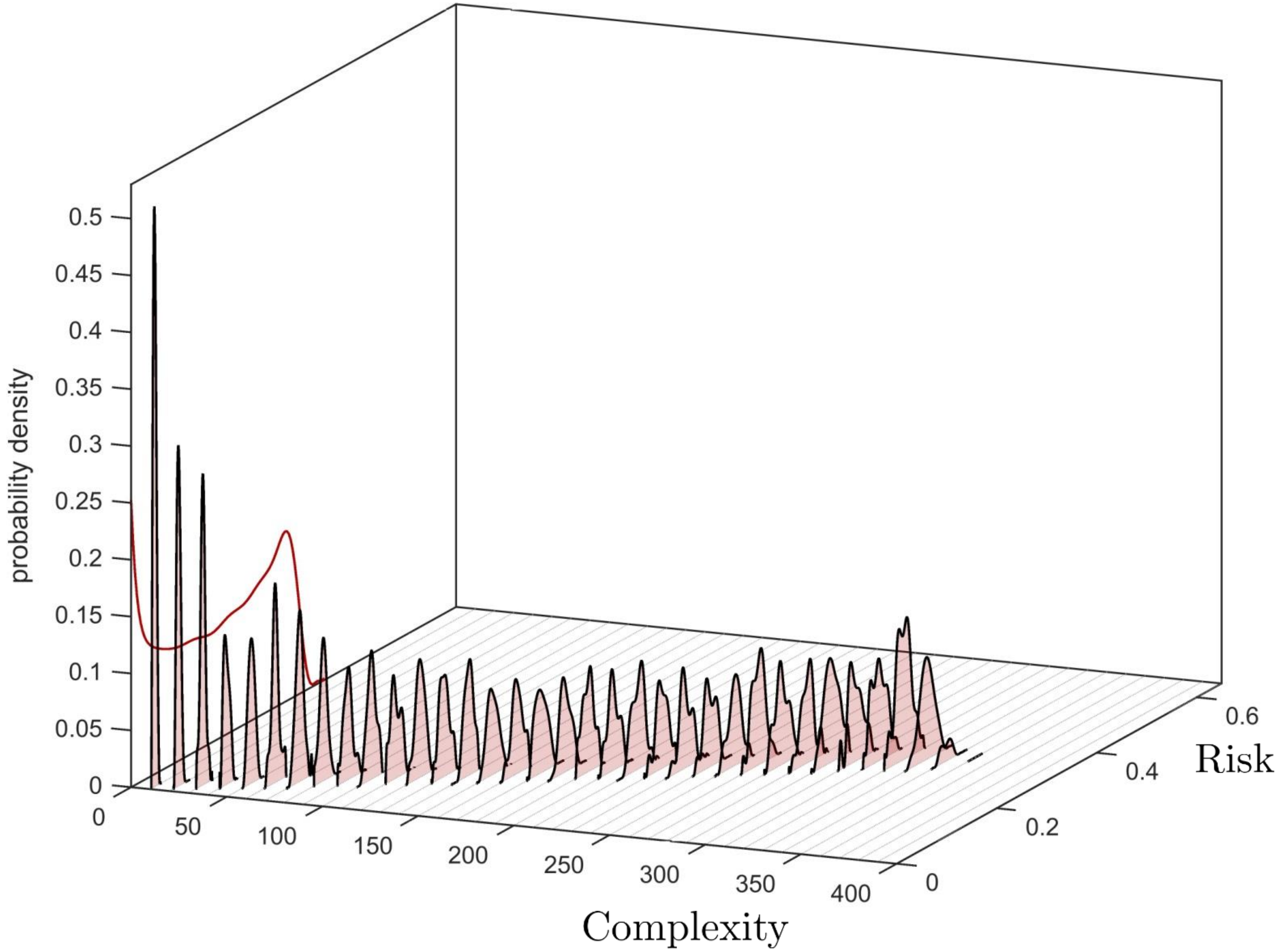
Opt.

7 points at the boundary $\Longrightarrow$ if only these 7 points are maintained, “Opt” gives the same solution

complexity = 7

intuitively: low Complexity $\Longrightarrow$ low Risk

risk-complexity distribution



robust optimization

$$\min_{x \in \mathcal{X}} \quad c(x)$$

$$\text{subject to:} \quad x \in \mathcal{X}_{\delta_i}, \quad i = 1, \ldots, N \qquad \delta_i \quad i.i.d.$$

[M.C. Campi, S. Garatti, Mathematical Programming, 2016]

robust optimization

$$\min_{x \in \mathcal{X}} \quad c(x)$$

$$\text{subject to:} \quad x \in \mathcal{X}_{\delta_i}, \quad i = 1, \ldots, N \qquad \delta_i \quad i.i.d.$$

Theorem

Suppose only one irreducible support set.

$\mathbb{P}\{\underline{\varepsilon}(\text{Complexity}) \leq \text{Risk} \leq \overline{\varepsilon}(\text{Complexity})\} \geq 1 - \beta$

where, with the position $t = 1 - \varepsilon$, $\underline{\varepsilon}$ and $\overline{\varepsilon}$ are given by:

$$\binom{N}{k} t^{N-k} - \frac{\beta}{2N} \sum_{i=k}^{N-1} \binom{i}{k} t^{i-k} - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{k} t^{i-k} = 0.$$

[M.C. Campi, S. Garatti, Mathematical Programming, 2019]

robust optimization

confidence = 10^{-4} ; 10^{-6}; 10^{-8}

[S. Garatti, M.C. Campi, Mathematical Programming, 2019]

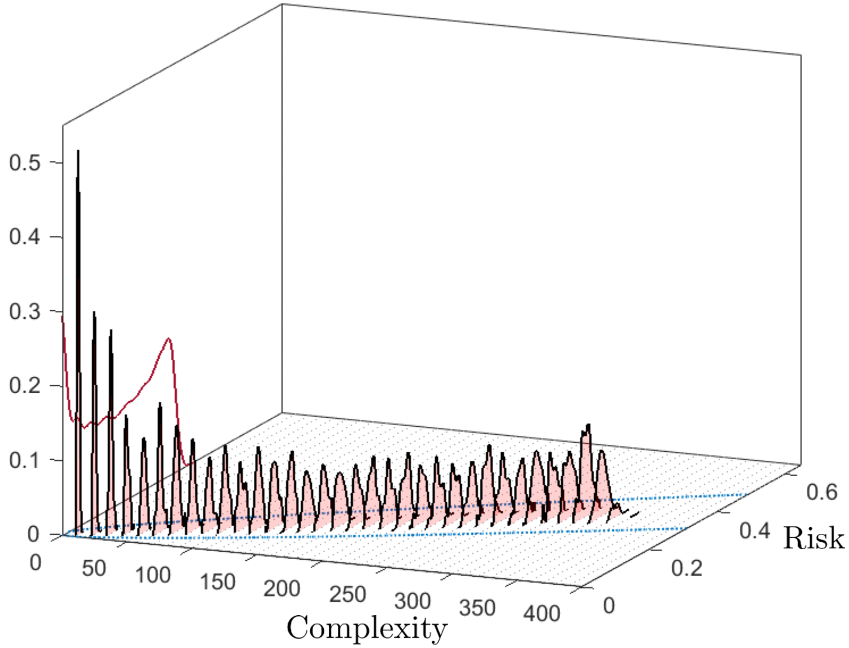
robust optimization

confidence = 10^{-4} ; 10^{-6}; 10^{-8}

holds distribution-free!

[S. Garatti, M.C. Campi, Mathematical Programming, 2019]

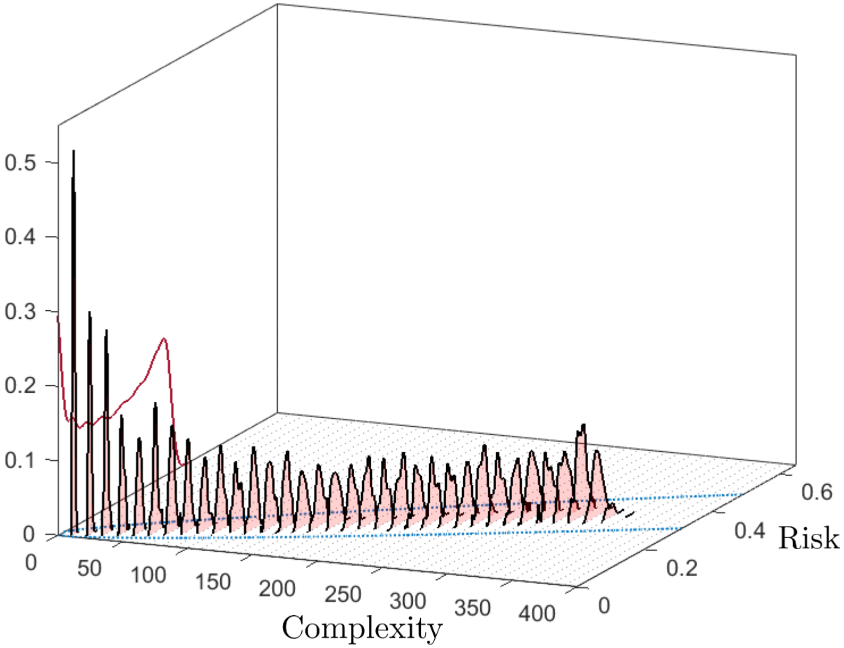
robust optimization



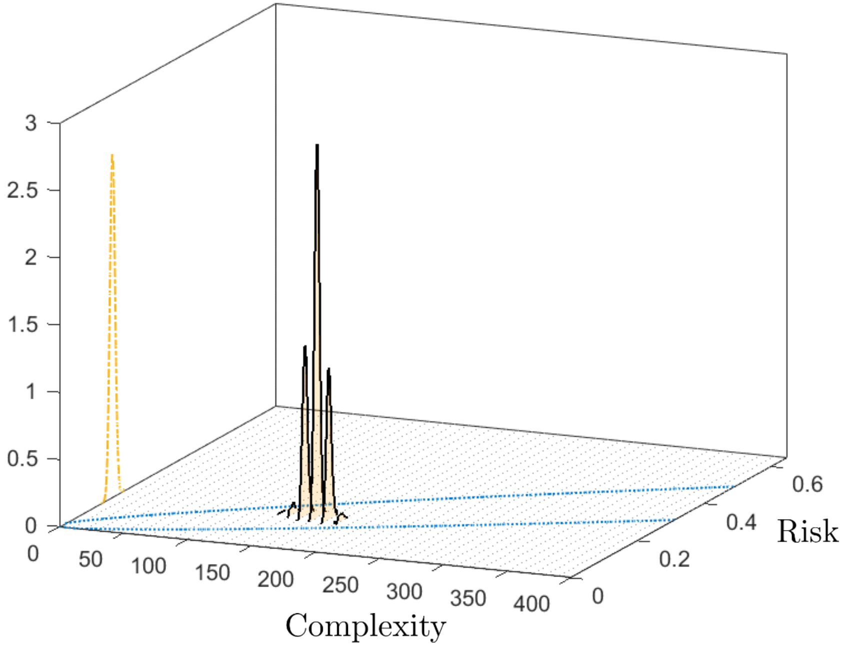
a distribution in R^{400}

[S. Garatti, M.C. Campi, Mathematical Programming, 2019]

robust optimization



a distribution in R^{400}



another distribution in R^{400}

[S. Garatti, M.C. Campi, Mathematical Programming, 2019]

relaxed schemes

$$\begin{aligned} &\min_{x \in \mathcal{X}} \quad c(x) \\ &\text{subject to:} \quad x \in \mathcal{X}_{\delta_i}\,, \qquad i = 1, \ldots, N \end{aligned}$$

relaxed schemes

$$
\begin{aligned}
&\min_{x \in \mathcal{X}} \quad c(x) \\
&\text{subject to:} \quad f(x, \delta_i) \leq 0, \quad i = 1, \ldots, N
\end{aligned}
$$

relaxed schemes

$$\min_{x \in \mathcal{X}, \xi_i \geq 0} \quad c(x) + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{subject to:} \quad f(x, \delta_i) \leq \xi_i, \quad i = 1, \ldots, N$$

relaxed schemes

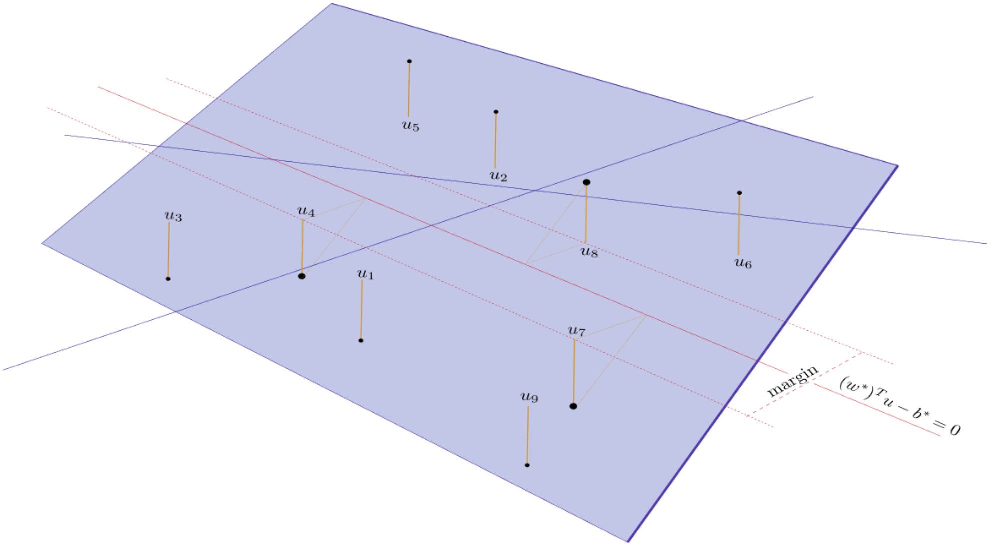
$$\min_{x\in\mathcal{X},\xi_i\geq 0} \quad c(x)+\rho\sum_{i=1}^{N}\xi_i$$

$$\text{subject to:} \quad f(x,\delta_i)\leq\xi_i, \quad i=1,\ldots,N$$

for example, SVM:

$$\min_{w\in\mathbb{R}^d,b\in\mathbb{R},\xi_i\geq 0} \|w\|^2+\rho\sum_{i=1}^{N}\xi_i$$

$$\text{subject to: } 1-y_i(\langle w,u_i\rangle-b)\leq\xi_i$$



relaxed schemes

$$\min_{x \in \mathcal{X}, \xi_i \geq 0} \quad c(x) + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{subject to:} \quad f(x, \delta_i) \leq \xi_i, \quad i = 1, \ldots, N$$

relaxed schemes

$$\min_{x \in \mathcal{X}, \xi_i \geq 0} \quad c(x) + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{subject to: } \quad f(x, \delta_i) \leq \xi_i, \quad i = 1, \ldots, N$$

$$\text{Complexity} \rightarrow \#(\{\xi_i^* > 0\} \cup \{\delta_i \text{ giving the solution}\})$$

[S. Garatti, M.C. Campi, Mathematical Programming, 2019]

relaxed schemes

$$\min_{x \in \mathcal{X}, \xi_i \geq 0} \quad c(x) + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{subject to:} \quad f(x, \delta_i) \leq \xi_i, \quad i = 1, \ldots, N$$

$\longrightarrow$ *same theorem as before applies*

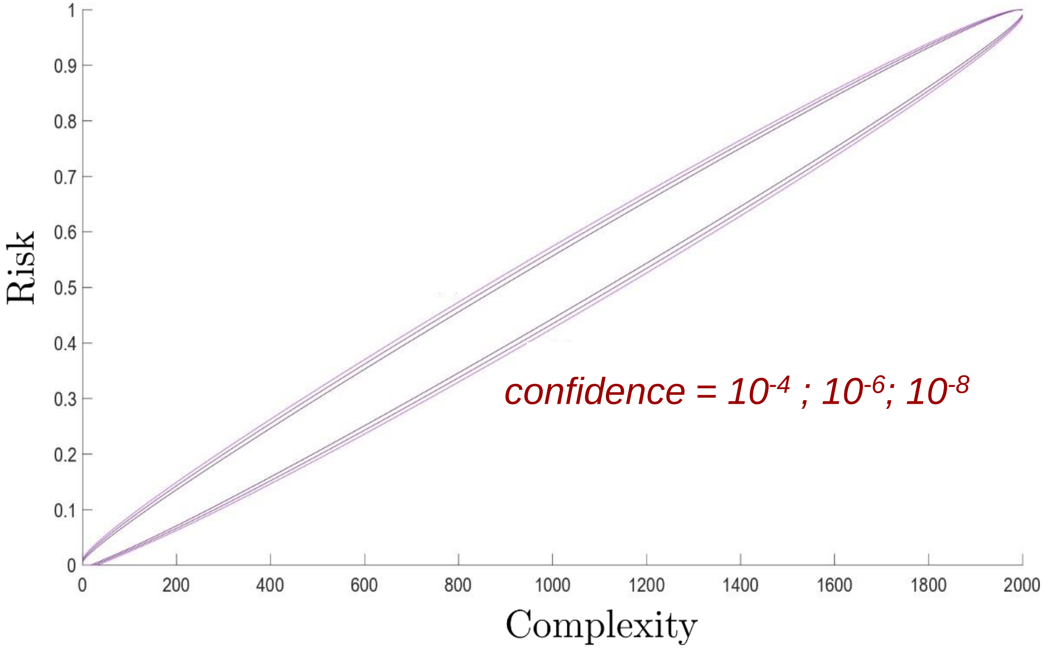
Theorem

Suppose only one irreducible support set.

$\mathbb{P}\{\underline{\varepsilon}(\text{Complexity}) \leq \text{Risk} \leq \overline{\varepsilon}(\text{Complexity})\} \geq 1 - \beta$

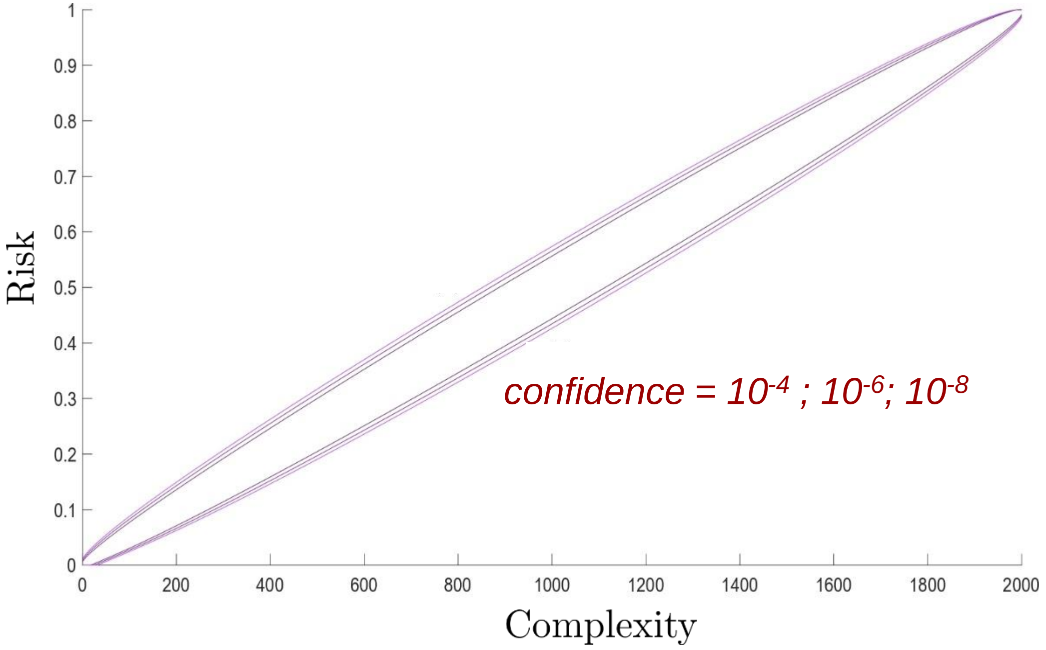
where, with the position $t = 1 - \varepsilon$, $\underline{\varepsilon}$ and $\overline{\varepsilon}$ are given by:

$$\binom{N}{k} t^{N-k} - \frac{\beta}{2N} \sum_{i=k}^{N-1} \binom{i}{k} t^{i-k} - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{k} t^{i-k} = 0.$$



confidence = 10^{-4} ; 10^{-6}; 10^{-8}

holds distribution-free!



confidence = 10^{-4} ; 10^{-6}; 10^{-8}

holds distribution-free!

should we aim higher?

prior + data → design

~~prior~~ + data → guarantees

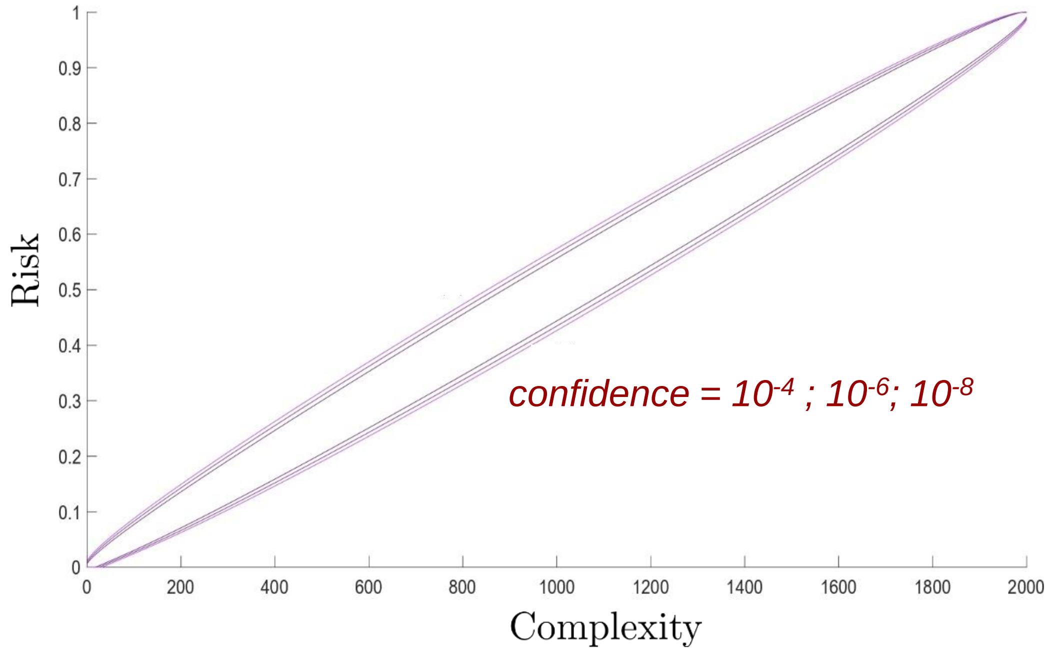
confidence = 10^{-4} ; 10^{-6}; 10^{-8}

holds distribution-free!

should we aim higher?

prior + data → design → check for features of interest

~~prior~~ + data → guarantees → always remain intact



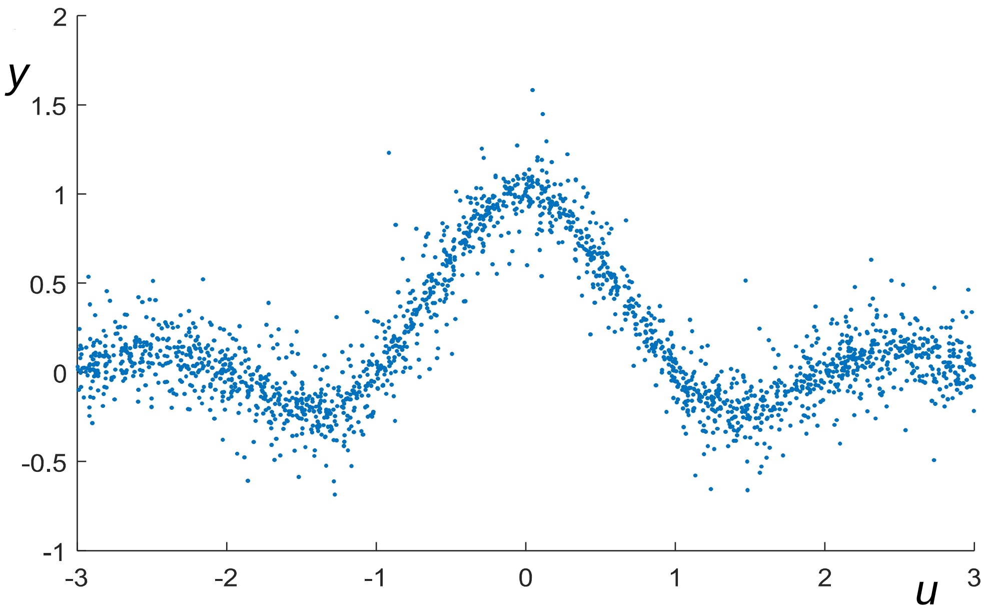
confidence = 10^{-4} ; 10^{-6}; 10^{-8}

holds distribution-free!

→ *develop trust in the decision*

→ *tune hyper-parameters*

Example: SVR

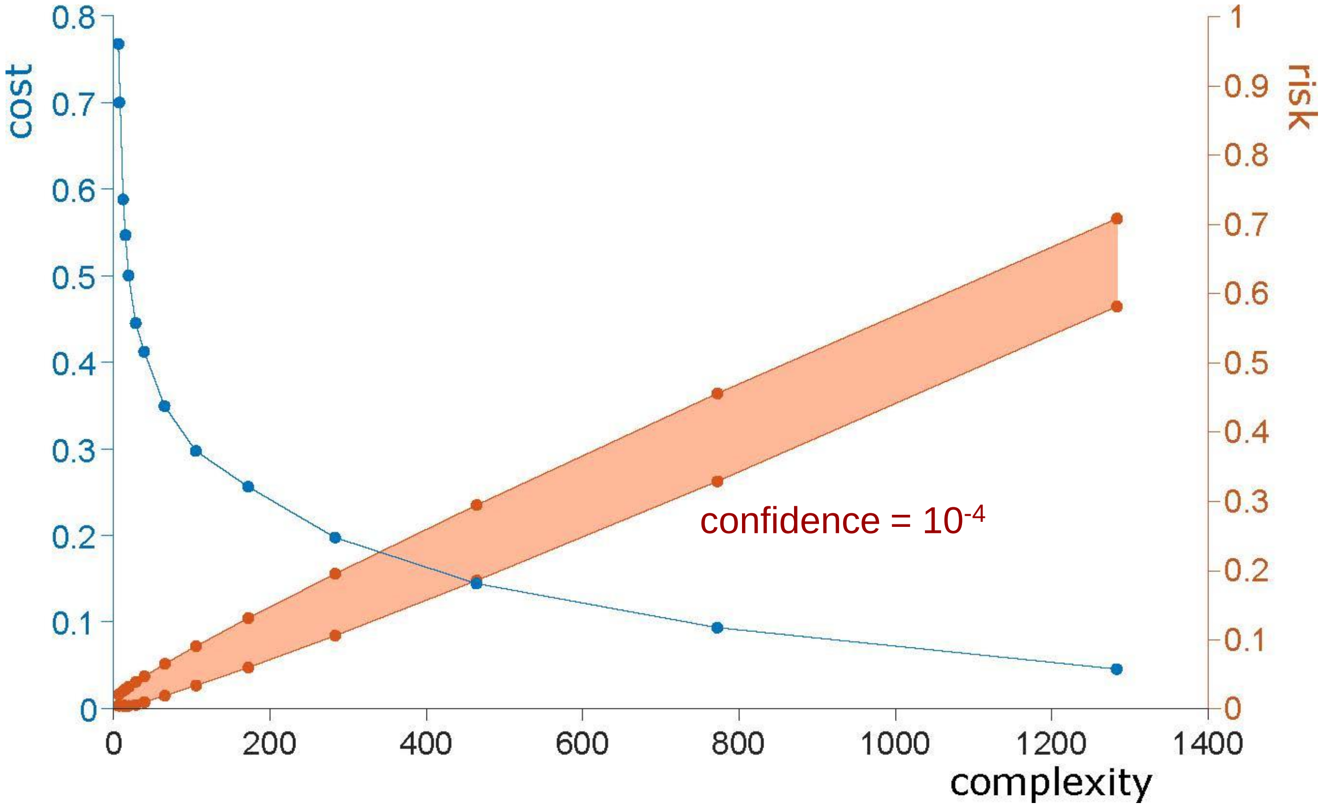


$$\min_{w,\gamma,b,\xi_i\geq 0} \gamma + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{subject to: } |y_i - \langle w, \phi_i \rangle - b| - \gamma \leq \xi_i, \quad i = 1, \ldots, 2000$$

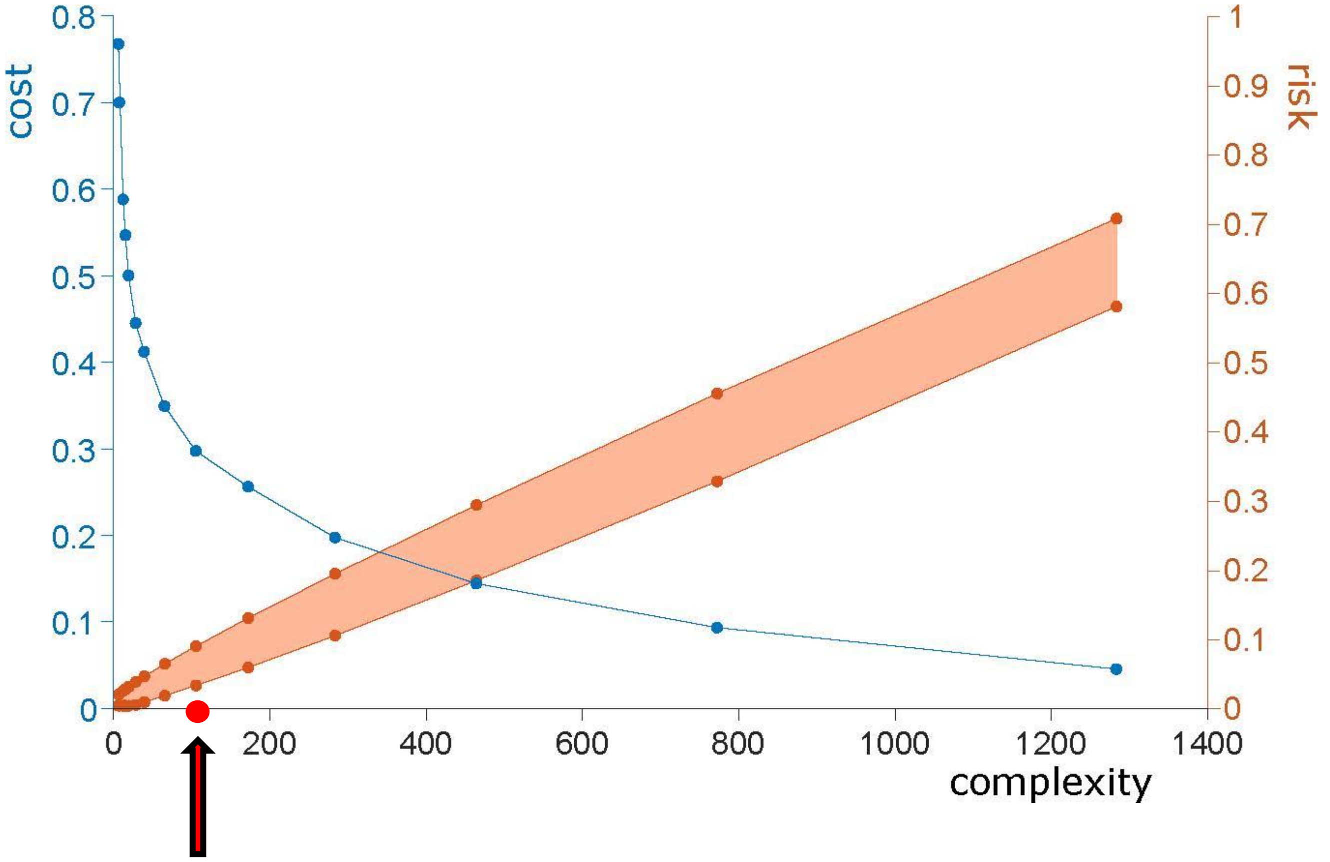
$$\langle \phi_i, \phi_j \rangle = \exp(-(u_i - u_j))^2 \qquad \text{(Gaussian kernel)}$$

Example: SVR



$$\rho = \left(\frac{3}{5}\right)^{\ell}, \quad \ell = 0, \ldots, 14$$

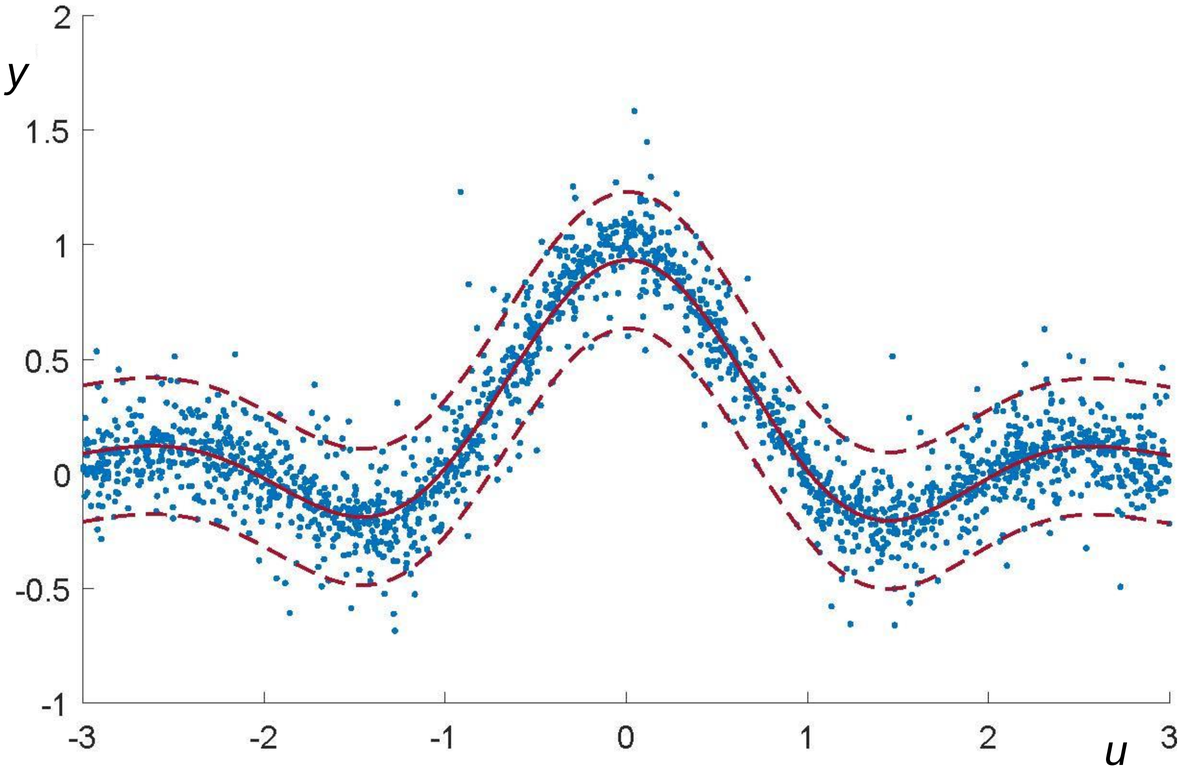
Example: SVR



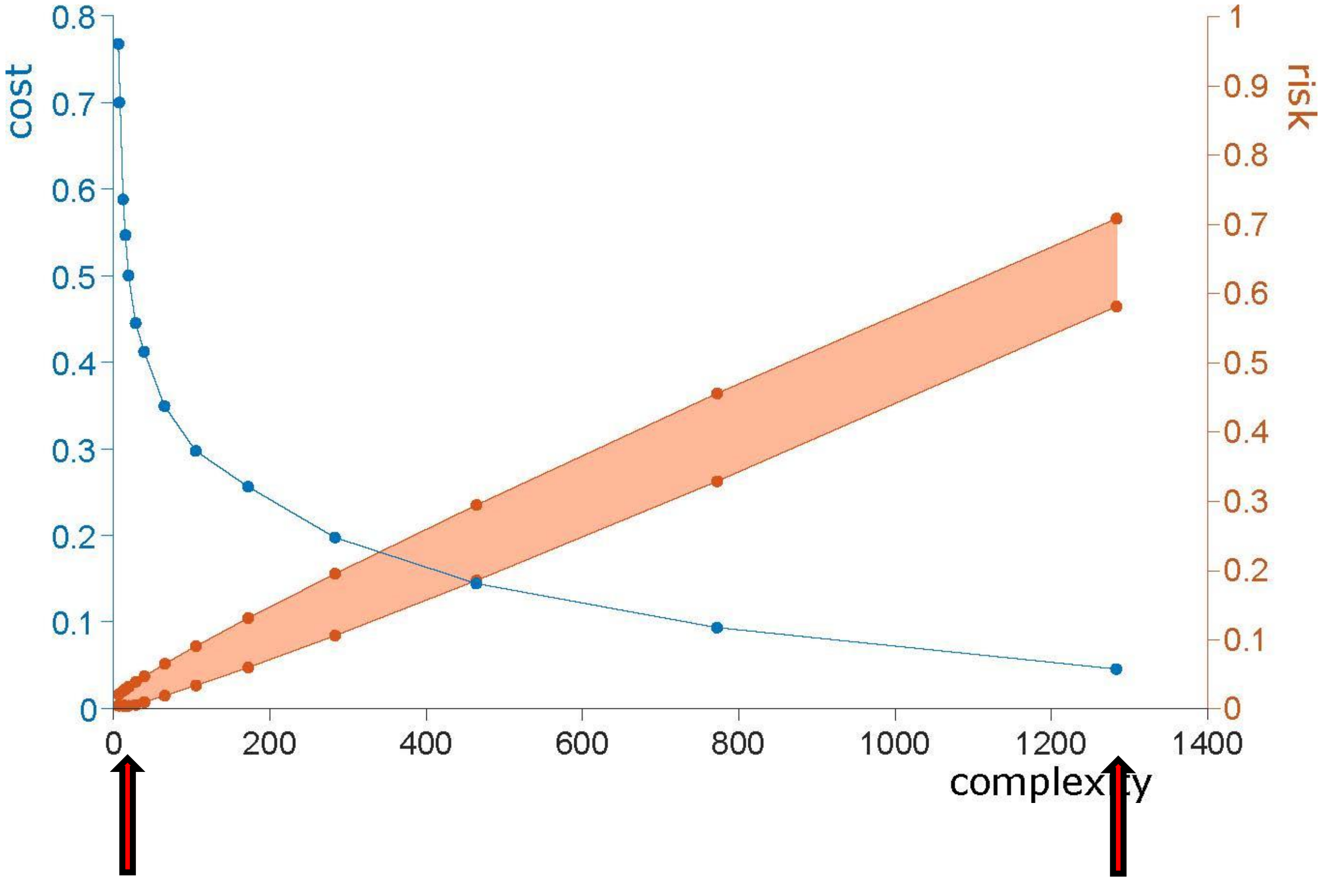
$\rho = (3/5)^9$

$\text{risk} \in [0.032, 0.08]$ $\gamma^* = 0.3$

Example: SVR



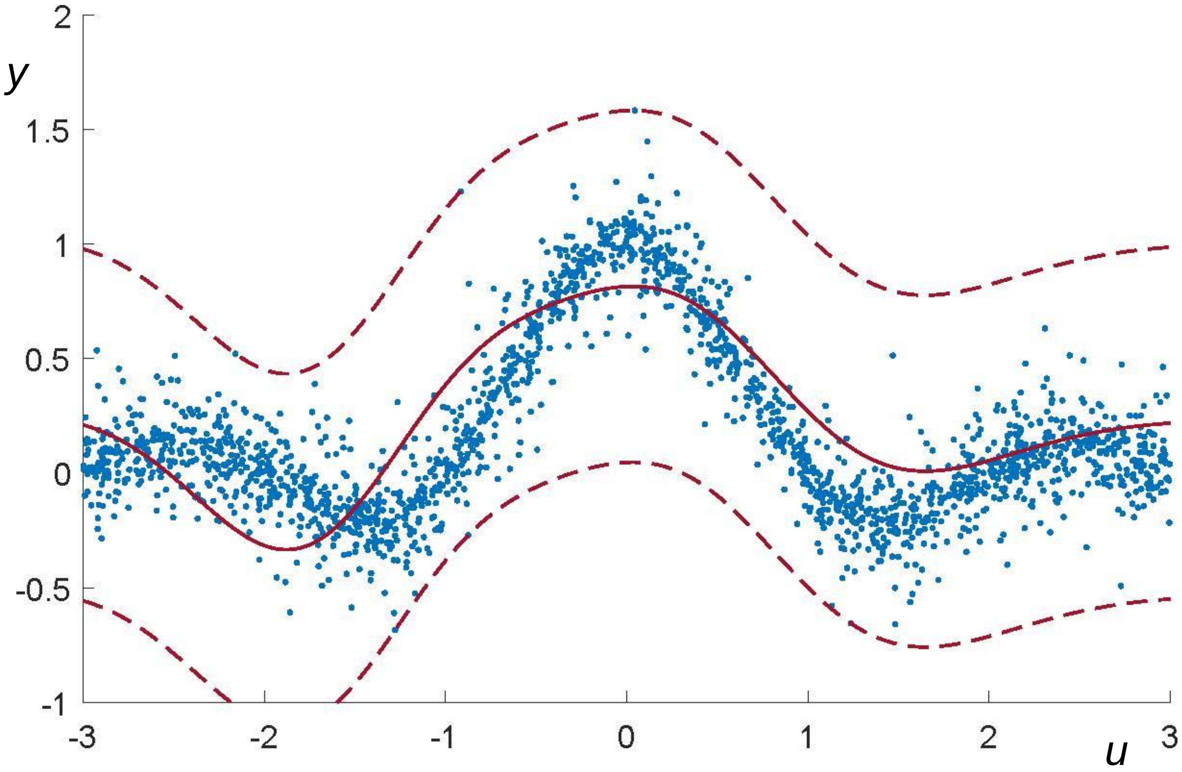
Example: SVR



$\rho = (3/5)^0$

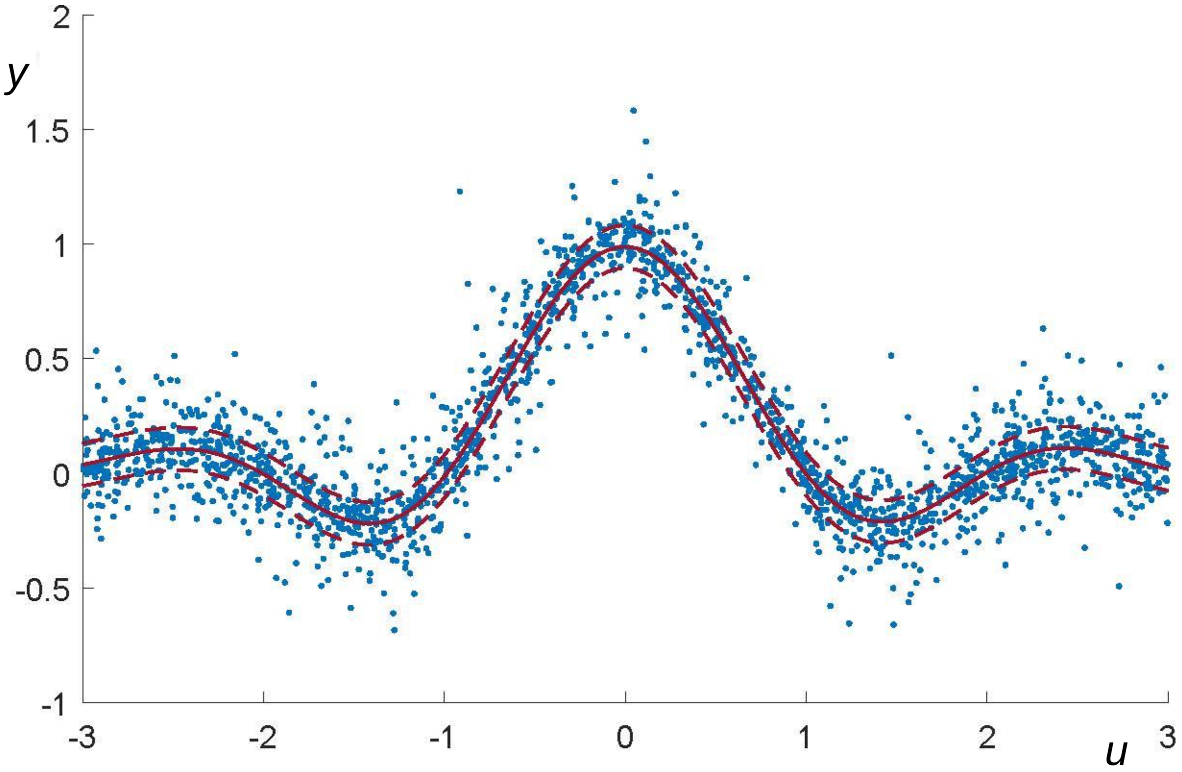
$\rho = (3/5)^{14}$

Example: SVR



$\rho = (3/5)^0$

Example: SVR



$\rho = (3/5)^{14}$

conclusions

conclusions

→ we aim to develop a science for data-driven optimization in the presence of uncertainty

conclusions

→ we aim to develop a science for data-driven optimization in the presence of uncertainty

→ that is, we want to optimize based on data while also providing certificates that maintain intact their validity when the hypothesized prior fails to be fully correct

conclusions

→ we aim to develop a science for data-driven optimization in the presence of uncertainty

→ that is, we want to optimize based on data while also providing certificates that maintain intact their validity when the hypothesized prior fails to be fully correct

→ Scenario Optimization is a body of direct data-driven methodologies with distribution-free certificates

conclusions

→ we aim to develop a science for data-driven optimization in the presence of uncertainty

→ that is, we want to optimize based on data while also providing certificates that maintain intact their validity when the hypothesized prior fails to be fully correct

→ Scenario Optimization is a body of direct data-driven methodologies with distribution-free certificates

→ The results show that data can well stand the dual role of (i) providing information to optimize; (ii) assessing the quality of the ensuing result

Thank you all!!!