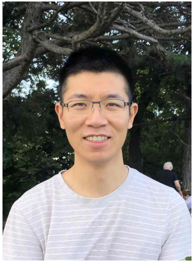
Are all models wrong? Fundamental limits in distribution-free empirical model falsification

Manuel M. Müller
University of Cambridge

Conference on Learning Theory
1st July 2025
Lyon, France

Collaborators



Yuetian Luo

Rina Foygel Barber

- M., M. M., Luo, Y. and Barber, R. F. (2025). Are all models wrong? Fundamental limits in distribution-free empirical model falsification. arXiv:2502.06765.

Setup

Standard setup:

- $\mathcal{Z}$ ...data space
- $\mathcal{F}$ ...family of functions
- $\ell : \mathcal{F} \times \mathcal{Z} \to [0, \infty)$ ...loss

Setup

Standard setup:

- $\mathcal{Z}$ ...data space
- $\mathcal{F}$ ...family of functions
- $\ell : \mathcal{F} \times \mathcal{Z} \to [0, \infty)$ ...loss

E.g. in regression could have:

- $\mathcal{Z} = \mathbb{R}^d \times \mathbb{R}$
- $\mathcal{F}$ could be given by:
  - $\{x \mapsto x^\top \beta : \beta \in \mathbb{R}^d\}$
  - a smoothness class
  - a neural network architecture
  - ...
- for $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ and $f \in \mathcal{F}$, e.g. $\ell\big(f, (x, y)\big) := \big(f(x) - y\big)^2$.

Setup

Standard setup:

- $\mathcal{Z}$ ...data space
- $\mathcal{F}$ ...family of functions
- $\ell : \mathcal{F} \times \mathcal{Z} \to [0, \infty)$ ...loss

Define the risk

$$R_P(f) := \mathbb{E}_{Z\sim P}\{\ell(f, Z)\},$$

for some distribution P on $\mathcal{Z}$.

E.g. in regression could have:

- $\mathcal{Z} = \mathbb{R}^d \times \mathbb{R}$
- $\mathcal{F}$ could be given by:
  - $\{x \mapsto x^\top \beta : \beta \in \mathbb{R}^d\}$
  - a smoothness class
  - a neural network architecture
  - ...
- for $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ and $f \in \mathcal{F}$, e.g. $\ell\big(f, (x, y)\big) := \big(f(x) - y\big)^2$.

Setup

Standard setup:

- $\mathcal{Z}$ ...data space
- $\mathcal{F}$ ...family of functions
- $\ell : \mathcal{F} \times \mathcal{Z} \to [0, \infty)$ ...loss

Define the risk

$$R_P(f) := \mathbb{E}_{Z \sim P}\{\ell(f, Z)\},$$

for some distribution P on $\mathcal{Z}$.

Define the **model class risk** as:

$$R_P(\mathcal{F}) := \inf_{f \in \mathcal{F}} R_P(f).$$

E.g. in regression could have:

- $\mathcal{Z} = \mathbb{R}^d \times \mathbb{R}$
- $\mathcal{F}$ could be given by:
  - $\{x \mapsto x^\top \beta : \beta \in \mathbb{R}^d\}$
  - a smoothness class
  - a neural network architecture
  - ...
- for $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ and $f \in \mathcal{F}$, e.g. $\ell\big(f, (x, y)\big) := \big(f(x) - y\big)^2$.

Main question

Define the **model class risk** as:

$$R_P(\mathcal{F}) := \inf_{f \in \mathcal{F}} R_P(f).$$

Main question

Define the **model class risk** as:

$$R_P(\mathcal{F}) := \inf_{f \in \mathcal{F}} R_P(f).$$

Overall goal: *Given an i.i.d. data set* $\mathcal{D}_n := (Z_1, \ldots, Z_n) \sim P^n$, *can we conduct meaningful inference on* $R_P(\mathcal{F})$ *without making assumptions on* P*?*

Main question

Define the **model class risk** as:

$$R_P(\mathcal{F}) := \inf_{f \in \mathcal{F}} R_P(f).$$

Overall goal: *Given an i.i.d. data set $\mathcal{D}_n := (Z_1, \ldots, Z_n) \sim P^n$, can we conduct meaningful inference on $R_P(\mathcal{F})$ without making assumptions on P?*

Definition. Fix $\alpha \in (0, 1)$, $n \geq 1$ and $\mathcal{F}$. $\hat{L}_\alpha(\mathcal{F}, \cdot) : \mathcal{Z}^n \to [0, \infty]$ is a distribution-free (DF) lower bound on $R_P(\mathcal{F})$ if

$$\inf_P \mathbb{P}_{\mathcal{D}_n \sim P^n}\Big(R_P(\mathcal{F}) \geq \hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n)\Big) \geq 1 - \alpha.$$

Main question

Define the **model class risk** as:

$$R_P(\mathcal{F}) := \inf_{f \in \mathcal{F}} R_P(f).$$

Overall goal: *Given an i.i.d. data set $\mathcal{D}_n := (Z_1, \ldots, Z_n) \sim P^n$, can we conduct meaningful inference on $R_P(\mathcal{F})$ without making assumptions on P?*

Definition. Fix $\alpha \in (0,1)$, $n \geq 1$ and $\mathcal{F}$. $\hat{L}_\alpha(\mathcal{F}, \cdot) : \mathcal{Z}^n \to [0, \infty]$ is a distribution-free (DF) lower bound on $R_P(\mathcal{F})$ if

$$\inf_P \mathbb{P}_{\mathcal{D}_n \sim P^n}\left(R_P(\mathcal{F}) \geq \hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n)\right) \geq 1 - \alpha.$$

Remark: Upper bounds are simpler, as $R_P(\mathcal{F}) \leq R_P(f_0)$ for all $f_0 \in \mathcal{F}$ $\longrightarrow$ **Asymmetry**.

Main question

Define the **model class risk** as:

$$R_P(\mathcal{F}) := \inf_{f \in \mathcal{F}} R_P(f).$$

Overall goal: *Given an i.i.d. data set $\mathcal{D}_n := (Z_1, \ldots, Z_n) \sim P^n$, can we conduct meaningful inference on $R_P(\mathcal{F})$ without making assumptions on P?*

Definition. Fix $\alpha \in (0,1)$, $n \geq 1$ and $\mathcal{F}$. $\hat{L}_\alpha(\mathcal{F}, \cdot) : \mathcal{Z}^n \to [0, \infty]$ is a distribution-free (DF) lower bound on $R_P(\mathcal{F})$ if

$$\inf_P \mathbb{P}_{\mathcal{D}_n \sim P^n}\Big(R_P(\mathcal{F}) \geq \hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n)\Big) \geq 1 - \alpha.$$

Remark: Upper bounds are simpler, as $R_P(\mathcal{F}) \leq R_P(f_0)$ for all $f_0 \in \mathcal{F} \longrightarrow$ **Asymmetry**.
Note: Can use lower bound on $R_P(\mathcal{F})$ to upper bound the excess risk $R_P(f_0) - \inf_{f \in \mathcal{F}} R_P(f)$.

A trivial solution

Definition. Fix $\alpha \in (0,1)$, $n \geq 1$ and $\mathcal{F}$. $\hat{L}_\alpha(\mathcal{F}, \cdot) : \mathcal{Z}^n \to [0, \infty]$ is a distribution-free (DF) lower bound on $R_P(\mathcal{F})$ if

$$\inf_P \mathbb{P}_{\mathcal{D}_n \sim P^n}\left(R_P(\mathcal{F}) \geq \hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n)\right) \geq 1 - \alpha.$$

Trivial solution: Take

$$\hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n) := \begin{cases} 0 & \text{with probability } 1 - \alpha, \\ \infty & \text{with probability } \alpha. \end{cases}$$

A trivial solution

Definition. Fix $\alpha \in (0,1)$, $n \geq 1$ and $\mathcal{F}$. $\hat{L}_\alpha(\mathcal{F}, \cdot) : \mathcal{Z}^n \to [0, \infty]$ is a distribution-free (DF) lower bound on $R_P(\mathcal{F})$ if

$$\inf_P \mathbb{P}_{\mathcal{D}_n \sim P^n}\left(R_P(\mathcal{F}) \geq \hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n)\right) \geq 1 - \alpha.$$

Trivial solution: Take

$$\hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n) := \begin{cases} 0 & \text{with probability } 1-\alpha, \\ \infty & \text{with probability } \alpha. \end{cases}$$

Let's call any lower bound **trivial** (for $\mathcal{F}$ and P) that satisfies

$$\mathbb{P}\left(\hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n) > 0\right) \leq \alpha + o(1).$$

A trivial solution

Definition. Fix $\alpha \in (0,1)$, $n \geq 1$ and $\mathcal{F}$. $\hat{L}_\alpha(\mathcal{F}, \cdot) : \mathcal{Z}^n \to [0, \infty]$ is a distribution-free (DF) lower bound on $R_P(\mathcal{F})$ if

$$\inf_P \mathbb{P}_{\mathcal{D}_n \sim P^n}\Big(R_P(\mathcal{F}) \geq \hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n)\Big) \geq 1 - \alpha.$$

Trivial solution: Take

$$\hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n) := \begin{cases} 0 & \text{with probability } 1-\alpha, \\ \infty & \text{with probability } \alpha. \end{cases}$$

Let's call any lower bound **trivial** (for $\mathcal{F}$ and P) that satisfies

$$\mathbb{P}\Big(\hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n) > 0\Big) \leq \alpha + o(1).$$

(When) Can we get nontrivial lower bounds?

A first lower bound

Define the **empirical (model class) risk**:

$$\hat{R}(f, \mathcal{D}_n) := \frac{1}{n} \sum_{i=1}^{n} \ell(f, Z_i) \quad \text{and} \quad \hat{R}(\mathcal{F}, \mathcal{D}_n) := \inf_{f \in \mathcal{F}} \hat{R}(f, \mathcal{D}_n).$$

A first lower bound

Define the **empirical (model class) risk**:

$$\hat{R}(f, \mathcal{D}_n) := \frac{1}{n} \sum_{i=1}^{n} \ell(f, Z_i) \quad \text{and} \quad \hat{R}(\mathcal{F}, \mathcal{D}_n) := \inf_{f \in \mathcal{F}} \hat{R}(f, \mathcal{D}_n).$$

Theorem. The following gives a valid lower bound:

$$\hat{L}_{\alpha}^{\mathrm{ERM}}(\mathcal{F}, \mathcal{D}_n) := \alpha \cdot \hat{R}(\mathcal{F}, \mathcal{D}_n).$$

(Proof idea: Combine Markov's inequality with $\mathbb{E}_P[\hat{R}(\mathcal{F}, \mathcal{D}_n)] \leq \inf_{f \in \mathcal{F}} R_P(f) = R_P(\mathcal{F})$.)

A first lower bound

Define the **empirical (model class) risk**:

$$\hat{R}(f, \mathcal{D}_n) := \frac{1}{n} \sum_{i=1}^{n} \ell(f, Z_i) \quad \text{and} \quad \hat{R}(\mathcal{F}, \mathcal{D}_n) := \inf_{f \in \mathcal{F}} \hat{R}(f, \mathcal{D}_n).$$

Theorem. The following gives a valid lower bound:

$$\hat{L}_{\alpha}^{\mathrm{ERM}}(\mathcal{F}, \mathcal{D}_n) := \alpha \cdot \hat{R}(\mathcal{F}, \mathcal{D}_n).$$

(Proof idea: Combine Markov's inequality with $\mathbb{E}_P[\hat{R}(\mathcal{F}, \mathcal{D}_n)] \leq \inf_{f \in \mathcal{F}} R_P(f) = R_P(\mathcal{F})$.)

Remark 1: If the loss is bounded, can improve this to $\big(1 - o(1)\big) \cdot \hat{R}(\mathcal{F}, \mathcal{D}_n)$.

A first lower bound

Define the **empirical (model class) risk**:

$$\hat{R}(f, \mathcal{D}_n) := \frac{1}{n} \sum_{i=1}^{n} \ell(f, Z_i) \quad \text{and} \quad \hat{R}(\mathcal{F}, \mathcal{D}_n) := \inf_{f \in \mathcal{F}} \hat{R}(f, \mathcal{D}_n).$$

Theorem. The following gives a valid lower bound:

$$\hat{L}_{\alpha}^{\mathrm{ERM}}(\mathcal{F}, \mathcal{D}_n) := \alpha \cdot \hat{R}(\mathcal{F}, \mathcal{D}_n).$$

(Proof idea: Combine Markov's inequality with $\mathbb{E}_P[\hat{R}(\mathcal{F}, \mathcal{D}_n)] \leq \inf_{f \in \mathcal{F}} R_P(f) = R_P(\mathcal{F})$.)

Remark 1: If the loss is bounded, can improve this to $\big(1 - o(1)\big) \cdot \hat{R}(\mathcal{F}, \mathcal{D}_n)$.
Remark 2: This is non-zero if and only if $\hat{R}(\mathcal{F}, \mathcal{D}_n) > 0$.

A first lower bound

Define the **empirical (model class) risk**:

$$\hat{R}(f, \mathcal{D}_n) := \frac{1}{n} \sum_{i=1}^{n} \ell(f, Z_i) \quad \text{and} \quad \hat{R}(\mathcal{F}, \mathcal{D}_n) := \inf_{f \in \mathcal{F}} \hat{R}(f, \mathcal{D}_n).$$

Theorem. The following gives a valid lower bound:

$$\hat{L}_{\alpha}^{\mathrm{ERM}}(\mathcal{F}, \mathcal{D}_n) := \alpha \cdot \hat{R}(\mathcal{F}, \mathcal{D}_n).$$

(Proof idea: Combine Markov's inequality with $\mathbb{E}_P[\hat{R}(\mathcal{F}, \mathcal{D}_n)] \leq \inf_{f \in \mathcal{F}} R_P(f) = R_P(\mathcal{F})$.)

Remark 1: If the loss is bounded, can improve this to $\big(1 - o(1)\big) \cdot \hat{R}(\mathcal{F}, \mathcal{D}_n)$.
Remark 2: This is non-zero if and only if $\hat{R}(\mathcal{F}, \mathcal{D}_n) > 0$.

> **Question:** *In modern settings, we often have* $\hat{R}(\mathcal{F}, \mathcal{D}_n) = 0$. *Can we still have* $\hat{L}_{\alpha}(\mathcal{F}, \mathcal{D}_n) > 0$ *then?*

Fundamental hardness

Question: *In modern settings, we often have $\hat{R}(\mathcal{F}, \mathcal{D}_n) = 0$. Can we still have $\hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n) > 0$ then?*

Fundamental hardness

Question: *In modern settings, we often have $\hat{R}(\mathcal{F}, \mathcal{D}_n) = 0$. Can we still have $\hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n) > 0$ then?*

Theorem. Any DF lower bound $\hat{L}_\alpha(\mathcal{F}, \cdot)$ must satisfy for all $N \geq n$ and P that

$$\mathbb{P}_P\left\{\hat{L}(\mathcal{F}, \mathcal{D}_n) \leq \hat{R}(\mathcal{F}, \mathcal{D}_N)\right\} \geq 1 - \alpha - \frac{n^2}{2N}.$$

Fundamental hardness

Question: *In modern settings, we often have* $\hat{R}(\mathcal{F}, \mathcal{D}_n) = 0$*. Can we still have* $\hat{L}_\alpha(\mathcal{F}, \mathcal{D}_n) > 0$ *then?*

Theorem. Any DF lower bound $\hat{L}_\alpha(\mathcal{F}, \cdot)$ must satisfy for all $N \geq n$ and P that

$$\mathbb{P}_P\left\{\hat{L}(\mathcal{F}, \mathcal{D}_n) \leq \hat{R}(\mathcal{F}, \mathcal{D}_N)\right\} \geq 1 - \alpha - \frac{n^2}{2N}.$$

Suppose $\mathcal{F}$ and P are such that $\hat{R}(\mathcal{F}, \mathcal{D}_N) = 0$ for $N \gg n^2$. Then:

$$\mathbb{P}_P\left\{\hat{L}(\mathcal{F}, \mathcal{D}_n) > 0\right\} \leq \alpha + o(1).$$

What have we learned so far?

Definition. (Interpolation capacity of $\mathcal{F}$ under P)

$$N(\mathcal{F}, P) := \sup\Big\{k \in \mathbb{N} \,:\, \mathbb{P}_{\mathcal{D}_k \sim P^k}\big(\hat{R}(\mathcal{F}, \mathcal{D}_k) = 0\big) = 1\Big\}$$

What have we learned so far?

Definition. (Interpolation capacity of $\mathcal{F}$ under P)

$$N(\mathcal{F}, P) := \sup\Big\{k \in \mathbb{N} \,:\, \mathbb{P}_{\mathcal{D}_k \sim P^k}\big(\hat{R}(\mathcal{F}, \mathcal{D}_k) = 0\big) = 1\Big\}$$
$$N_+(\mathcal{F}, P) := \sup\Big\{k \in \mathbb{N} \,:\, \mathbb{P}_{\mathcal{D}_k \sim P^k}\big(\hat{R}(\mathcal{F}, \mathcal{D}_k) = 0\big) > 0\Big\}.$$

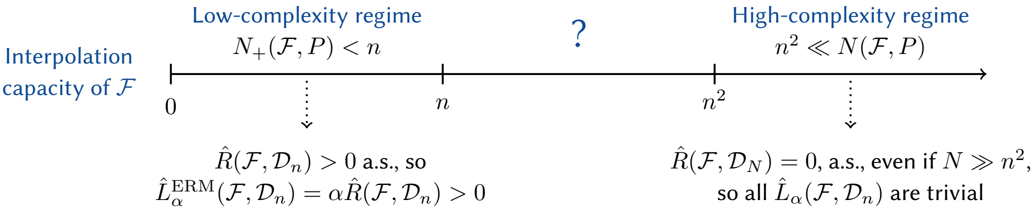
Of course, $N(\mathcal{F}, P) \leq N_+(\mathcal{F}, P)$.

Definition. (Interpolation capacity of $\mathcal{F}$ under P)

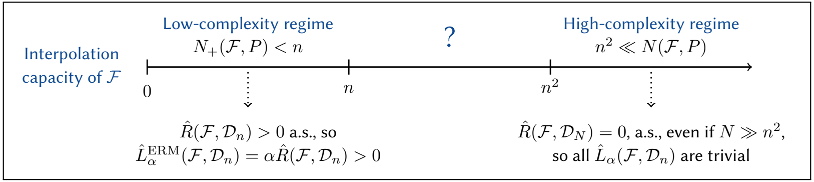
$$N(\mathcal{F},P) := \sup\Big\{k \in \mathbb{N} \,:\, \mathbb{P}_{\mathcal{D}_k\sim P^k}\big(\hat{R}(\mathcal{F},\mathcal{D}_k)=0\big)=1\Big\}$$

$$N_+(\mathcal{F},P) := \sup\Big\{k \in \mathbb{N} \,:\, \mathbb{P}_{\mathcal{D}_k\sim P^k}\big(\hat{R}(\mathcal{F},\mathcal{D}_k)=0\big)>0\Big\}.$$

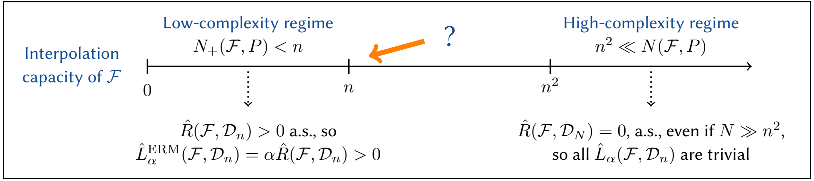
Of course, $N(\mathcal{F},P) \le N_+(\mathcal{F},P)$.



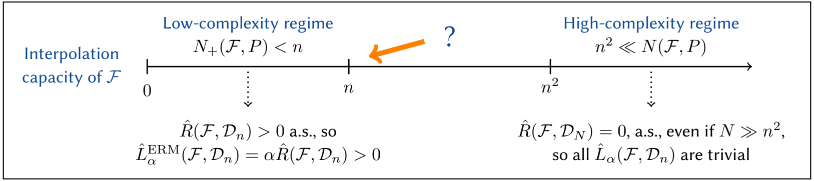
What about the gap?



What about the gap?



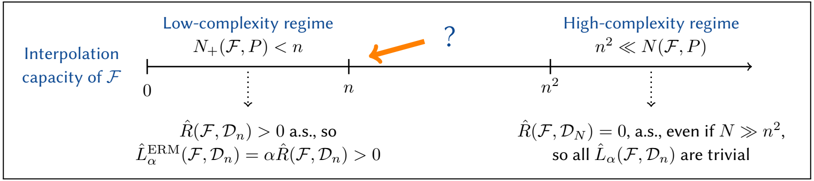
What about the gap?



Example A: Linear Regression.

Take $\mathcal{F} = \mathcal{F}_{\mathrm{lin}}^{(d)} = \{x \mapsto x^\top \beta \,:\, \beta \in \mathbb{R}^d\}$ and $\ell\big(f, (x, y)\big) = \big(y - f(x)\big)^2$.

What about the gap?

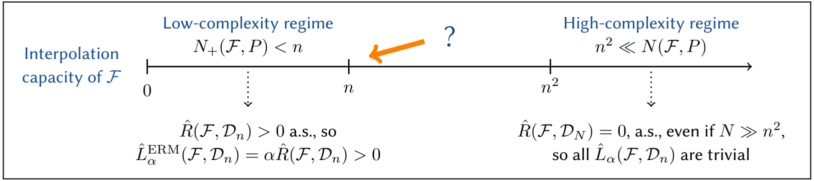


Example A: Linear Regression.

Take $\mathcal{F} = \mathcal{F}_{\mathrm{lin}}^{(d)} = \{x \mapsto x^\top \beta \,:\, \beta \in \mathbb{R}^d\}$ and $\ell\big(f, (x, y)\big) = \big(y - f(x)\big)^2$.
For P with a density on $\mathbb{R}^d \times \mathbb{R}$:

$$N(\mathcal{F}_{\mathrm{lin}}^{(d)}, P) = N_+(\mathcal{F}_{\mathrm{lin}}^{(d)}, P) = d.$$

What about the gap?



Example A: Linear Regression.

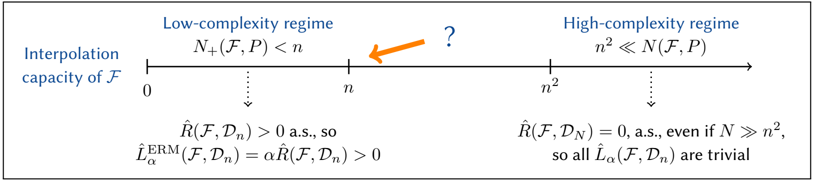
Take $\mathcal{F} = \mathcal{F}_{\mathrm{lin}}^{(d)} = \{x \mapsto x^\top \beta \,:\, \beta \in \mathbb{R}^d\}$ and $\ell\big(f, (x, y)\big) = \big(y - f(x)\big)^2$.
For P with a density on $\mathbb{R}^d \times \mathbb{R}$:

$$N(\mathcal{F}_{\mathrm{lin}}^{(d)}, P) = N_+(\mathcal{F}_{\mathrm{lin}}^{(d)}, P) = d.$$

Proposition. Let P have a density, and $P_X = \mathcal{N}_d(0, \Sigma)$ for $\Sigma \succ 0$. Then, for $d \gg n$ and any $\hat{L}_\alpha$:

$$\mathbb{P}_{\mathcal{D}_n \sim P^n}\Big\{\hat{L}_\alpha(\mathcal{F}_{\mathrm{lin}}^{(d)}, \mathcal{D}_n) > 0\Big\} \leq \alpha + o(1).$$

What about the gap?



Example A: Linear Regression.

Take $\mathcal{F} = \mathcal{F}_{\mathrm{lin}}^{(d)} = \{x \mapsto x^\top \beta \,:\, \beta \in \mathbb{R}^d\}$ and $\ell\big(f, (x, y)\big) = \big(y - f(x)\big)^2$.
For P with a density on $\mathbb{R}^d \times \mathbb{R}$:

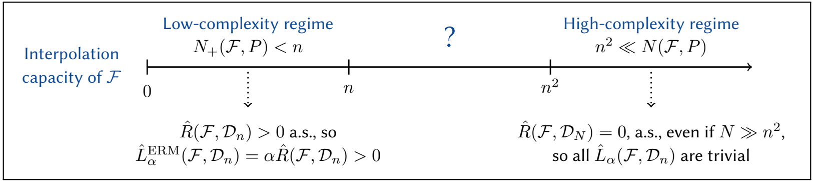
$$N(\mathcal{F}_{\mathrm{lin}}^{(d)}, P) = N_+(\mathcal{F}_{\mathrm{lin}}^{(d)}, P) = d.$$

Proposition. Let P have a density, and $P_X = \mathcal{N}_d(0, \Sigma)$ for $\Sigma \succ 0$. Then, for $d \gg n$ and any $\hat{L}_\alpha$:

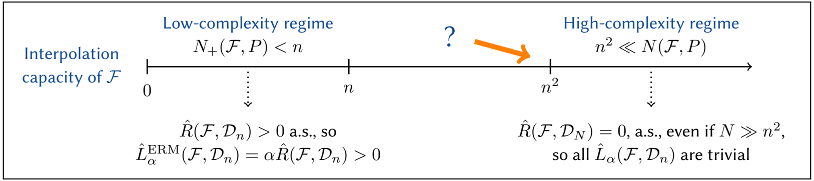
$$\mathbb{P}_{\mathcal{D}_n \sim P^n}\left\{\hat{L}_\alpha(\mathcal{F}_{\mathrm{lin}}^{(d)}, \mathcal{D}_n) > 0\right\} \le \alpha + o(1).$$

$\Longrightarrow$ **See the paper for far more general results.**

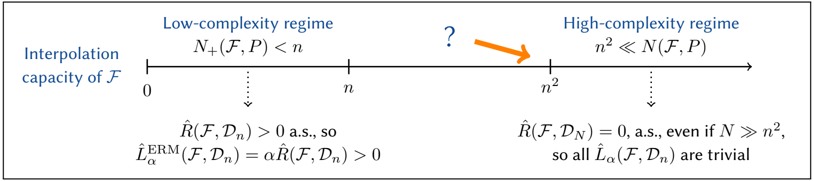
What about the gap?



What about the gap?



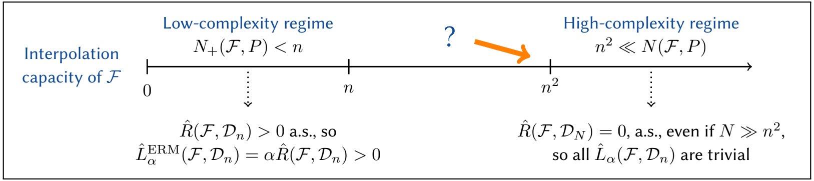
What about the gap?



Example B: Piecewise constant regression (e.g. random forests).

Take $\mathcal{F} = \mathcal{F}_{\mathrm{pwc}}^{(m)} := \{f : \mathbb{R}^d \to \mathbb{R} \,:\, |\{f(x) \,:\, x \in \mathbb{R}^d\}| \leq m\}$ and $\ell\big(f, (x, y)\big) = \big(y - f(x)\big)^2$.

What about the gap?



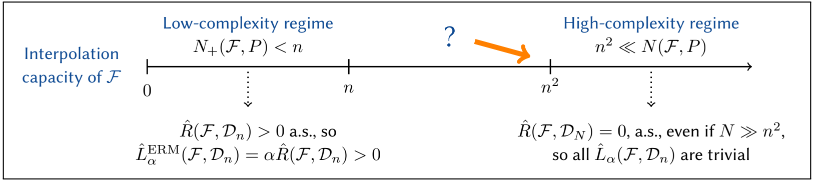
Example B: Piecewise constant regression (e.g. random forests).

Take $\mathcal{F} = \mathcal{F}_{\mathrm{pwc}}^{(m)} := \{f : \mathbb{R}^d \to \mathbb{R} \,:\, |\{f(x) \,:\, x \in \mathbb{R}^d\}| \leq m\}$ and $\ell\big(f, (x, y)\big) = \big(y - f(x)\big)^2$.

Theorem. There exists $m_n \propto n^2$, such that the following is a DF lower bound for $\mathcal{F}_{\mathrm{pwc}}^{(m_n)}$:

$$\hat{L}_\alpha^{\mathrm{pwc}}(\mathcal{F}_{\mathrm{pwc}}^{(m_n)}, \mathcal{D}_n) := \frac{\alpha}{2} \hat{R}(\mathcal{F}_{\mathrm{pwc}}^{(n-1)}, \mathcal{D}_n).$$

What about the gap?



Example B: Piecewise constant regression (e.g. random forests).

Take $\mathcal{F} = \mathcal{F}_{\mathrm{pwc}}^{(m)} := \{f : \mathbb{R}^d \to \mathbb{R} : |\{f(x) : x \in \mathbb{R}^d\}| \leq m\}$ and $\ell\big(f, (x, y)\big) = \big(y - f(x)\big)^2$.

Theorem. There exists $m_n \propto n^2$, such that the following is a DF lower bound for $\mathcal{F}_{\mathrm{pwc}}^{(m_n)}$:

$$\hat{L}_\alpha^{\mathrm{pwc}}(\mathcal{F}_{\mathrm{pwc}}^{(m_n)}, \mathcal{D}_n) := \frac{\alpha}{2} \hat{R}(\mathcal{F}_{\mathrm{pwc}}^{(n-1)}, \mathcal{D}_n).$$

Note: Let P be atom-free. Then $N(\mathcal{F}_{\mathrm{pwc}}^{(m_n)}, P) = m_n \propto n^2$ and at the same time almost surely $\hat{R}(\mathcal{F}_{\mathrm{pwc}}^{(n-1)}, \mathcal{D}_n) > 0$ and hence:

$$\hat{L}_\alpha^{\mathrm{pwc}}(\mathcal{F}_{\mathrm{pwc}}^{(m_n)}, \mathcal{D}_n) > 0.$$

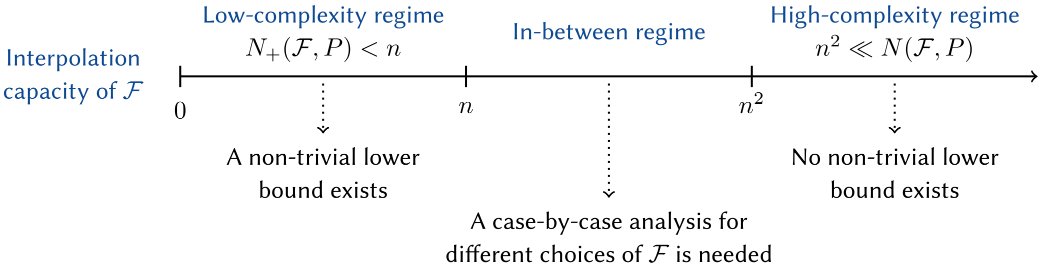
Summary: our findings

Question: When can we construct empirical, non-zero, distribution-free lower bounds on the model class risk $R_P(\mathcal{F}) := \inf_{f \in \mathcal{F}} \mathbb{E}_{Z \sim P}\{\ell(f, Z)\}$?

Summary: our findings

Question: When can we construct empirical, non-zero, distribution-free lower bounds on the model class risk $R_P(\mathcal{F}) := \inf_{f \in \mathcal{F}} \mathbb{E}_{Z \sim P}\{\ell(f, Z)\}$?

Answer: This is driven by two phase-transitions in the interpolation capacity of $\mathcal{F}$ under P:



Thank you!

Reference:

- M., M. M., Luo, Y. and Barber, R. F. (2025). Are all models wrong? Fundamental limits in distribution-free empirical model falsification. arXiv:2502.06765.

Find the slides on manuelmmueller.github.io.