Quantitative Input Feature Usage

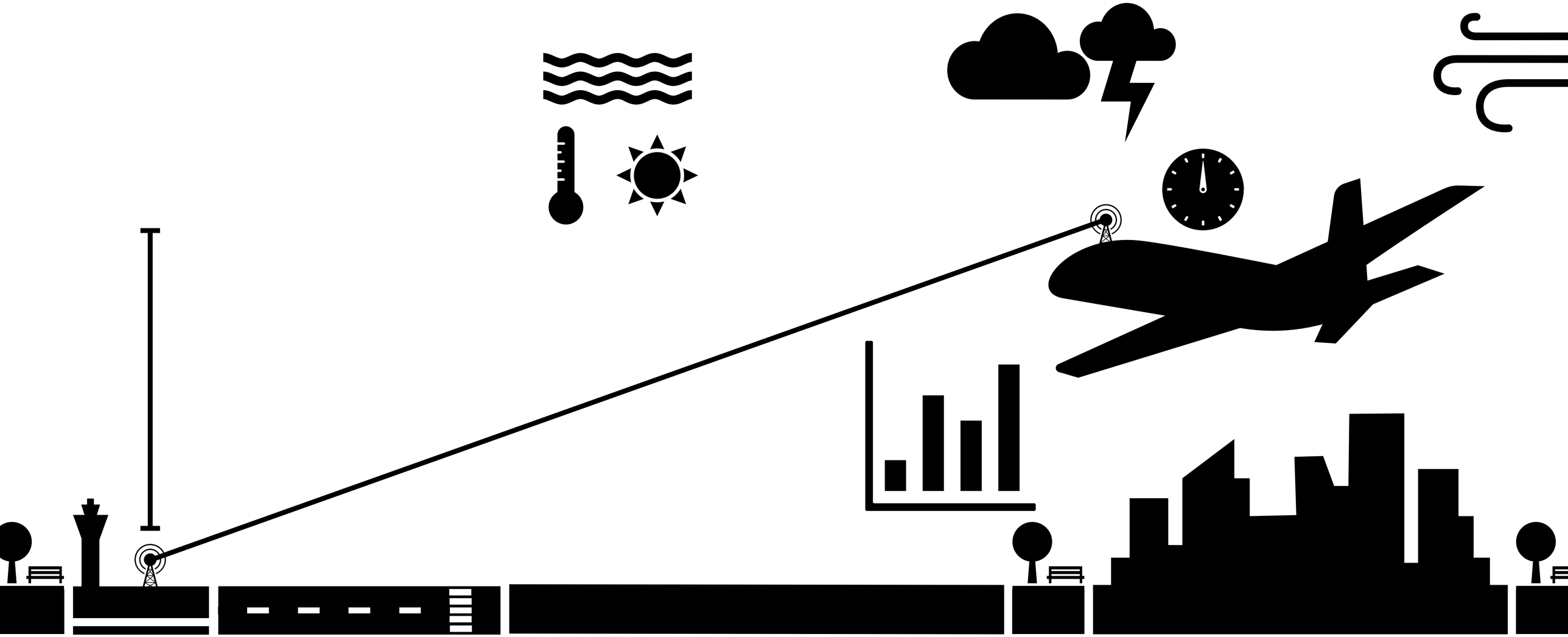
Denis Mazzucato, Marco Campion, Caterina Urban

École Normale Supérieure, Inria

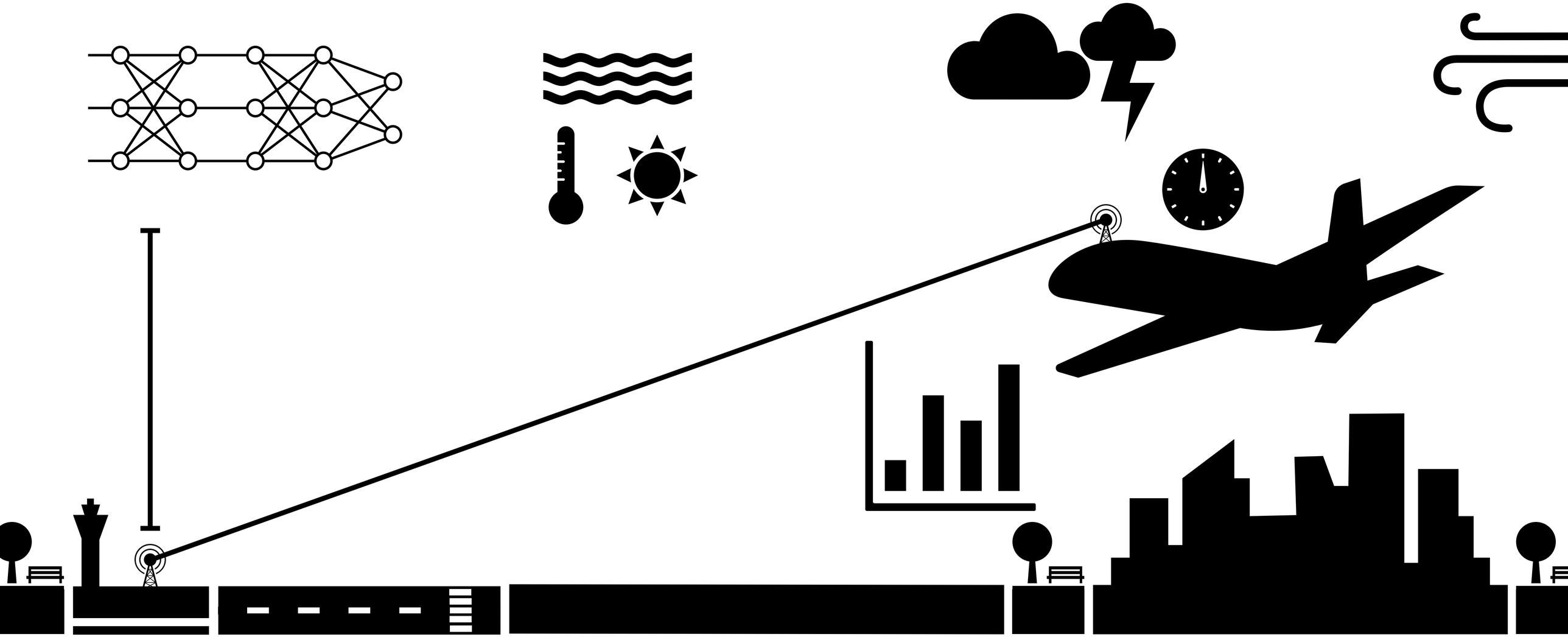
25th May 2023 — Challenges of Software Verification, Venice

Landing alarm system 🚨

Landing alarm system

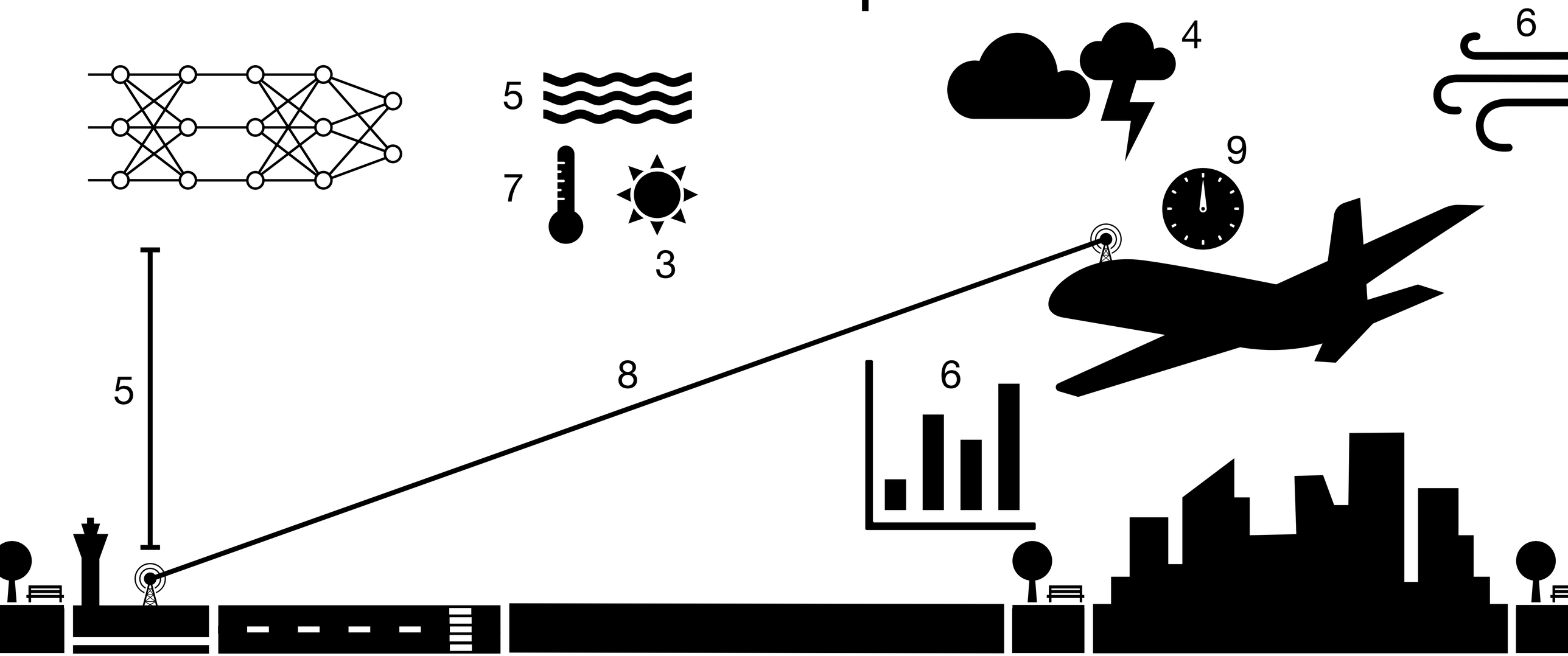


Landing alarm system

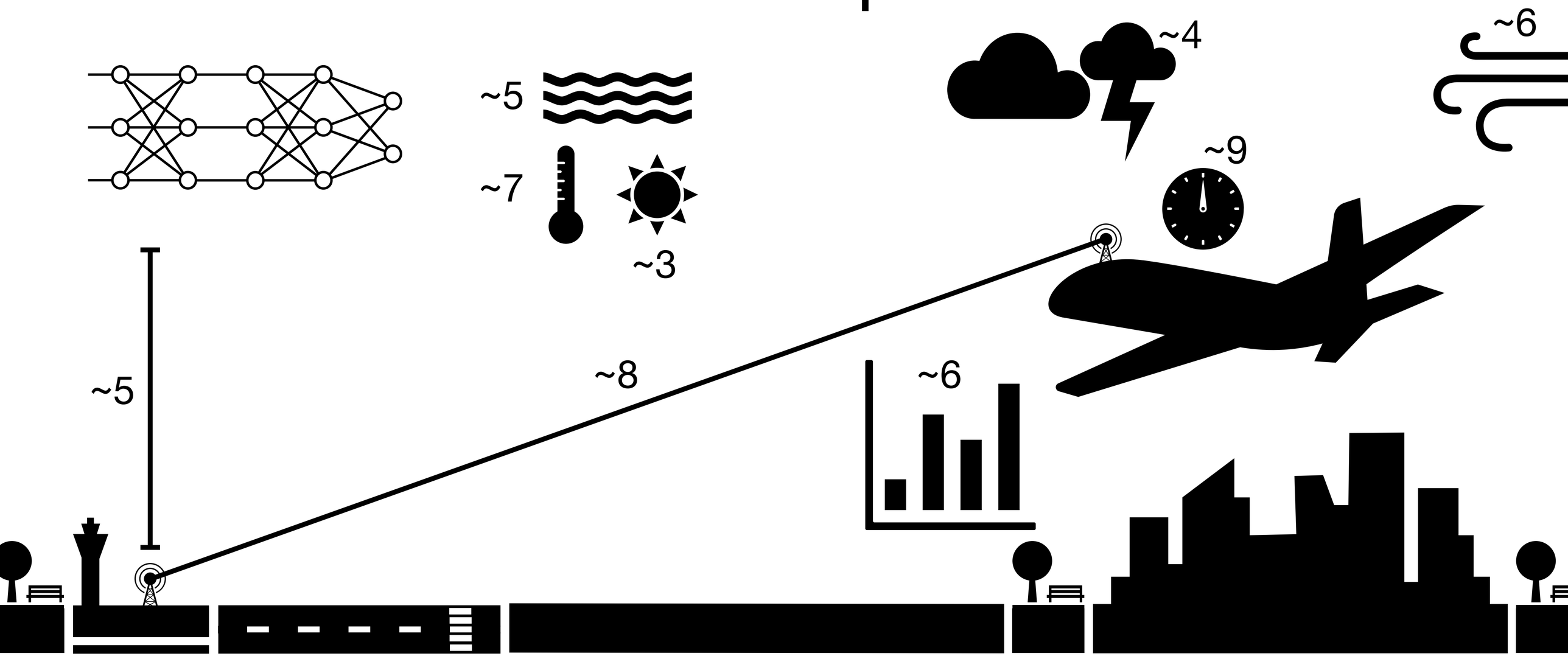


Landing alarm system

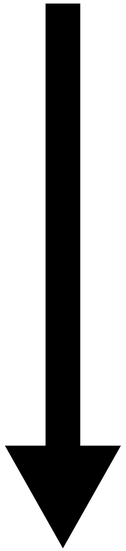
Stochastic methods: feature importance



Stochastic methods: feature importance



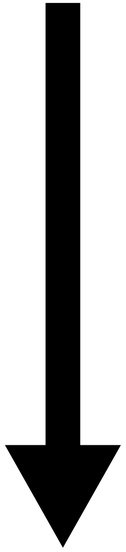
Formal methods: quantitative analysis



Qualitative

Is the i-th input feature of P unused?

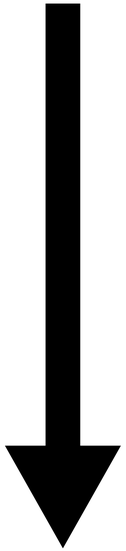
[Urban18]



Quantitative

How much is the i-th input feature of P used?

Caterina Urban, Peter Müller. An Abstract Interpretation Framework for Input Data Usage. ESOP 2018



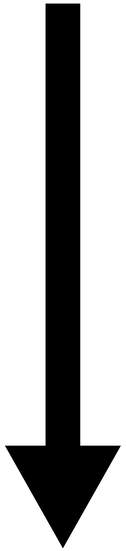
Qualitative

Is the i-th input feature of P unused?

[Urban18]

$$P \models \mathbb{U}_i$$

Caterina Urban, Peter Müller. An Abstract Interpretation Framework for Input Data Usage. ESOP 2018



Qualitative

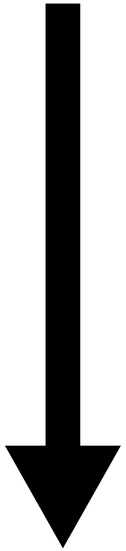
Is the i-th input feature of P unused?

[Urban18]

$$P \models \mathbb{U}_i$$

$$\mathbb{U}_i = \{\text{program } S \mid S \text{ does not use the input feature } i\}$$

Caterina Urban, Peter Müller. An Abstract Interpretation Framework for Input Data Usage. ESOP 2018



Qualitative

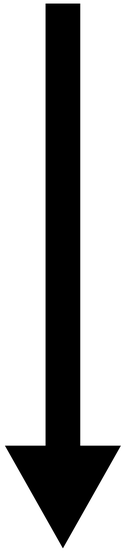
Is the i-th input feature of P unused?

[Urban18]

$$P \models \mathbb{U}_i$$

$$\mathbb{U}_i = \{ \quad [\![S]\!] \quad \mid S \text{ does not use the input feature } i \}$$

Caterina Urban, Peter Müller. An Abstract Interpretation Framework for Input Data Usage. ESOP 2018



Qualitative

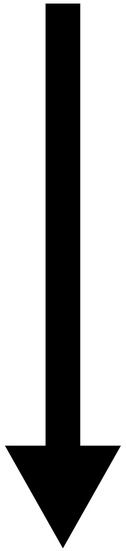
Is the i-th input feature of P unused?

[Urban18]

$$P \models \mathbb{U}_i$$

$$\mathbb{U}_i = \{ \quad [\![S]\!] \quad | \quad \mathsf{unused}_i([\![S]\!]) \quad \}$$

Caterina Urban, Peter Müller. An Abstract Interpretation Framework for Input Data Usage. ESOP 2018



Qualitative

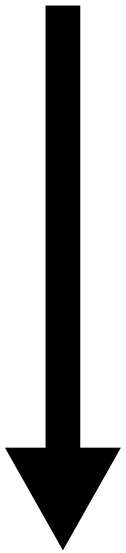
Is the i-th input feature of P unused?

[Urban18]

$$P \models \mathbb{U}_i$$

$$\mathbb{U}_i = \{ \ [\![S]\!] \mid \mathsf{unused}_i([\![S]\!]) \ \}$$

Caterina Urban, Peter Müller. An Abstract Interpretation Framework for Input Data Usage. ESOP 2018



Qualitative

Is the i-th input feature of P unused?

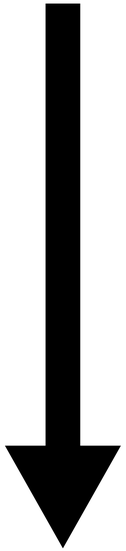
[Urban18]

$$P \models \mathbb{U}_i$$

$$\mathbb{U}_i = \{\ [\![S]\!] \mid \mathsf{unused}_i([\![S]\!])\ \}$$

$$P \models \mathbb{U}_i \iff \{[\![P]\!]\} \subseteq \mathbb{U}_i$$

Caterina Urban, Peter Müller. An Abstract Interpretation Framework for Input Data Usage. ESOP 2018



Qualitative

Is the i-th input feature of P unused?

[Urban18]

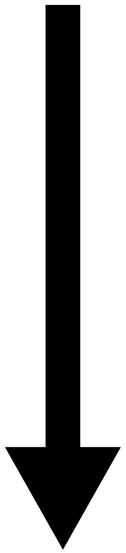
$$P \models \mathbb{U}_i$$

$$\mathbb{U}_i = \{\ [\![S]\!] \mid \mathsf{unused}_i([\![S]\!])\ \}$$

$$P \models \mathbb{U}_i \iff \{[\![P]\!]\} \subseteq \mathbb{U}_i \quad \Leftarrow \quad \gamma([\![P]\!]^{\natural}) \subseteq \mathbb{U}_i$$

Caterina Urban, Peter Müller. An Abstract Interpretation Framework for Input Data Usage. ESOP 2018

$$P \models \mathbb{Q}_i$$

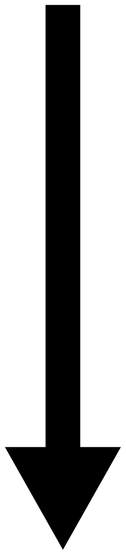


Quantitative

How much is the i-th input feature of P used?

$$P \models \mathbb{Q}_i$$

$$\mathbb{Q}_i = \{\ [\![S]\!] \mid \mathsf{impact}_i([\![S]\!])\ \}$$

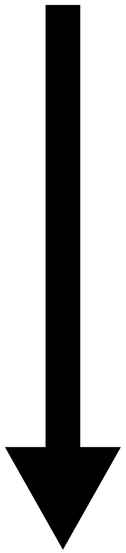


Quantitative

How much is the i-th input feature of P used?

$$P \models \mathbb{Q}_i \qquad \text{impact}_i \in \text{Traces} \to \mathbb{R}$$

$$\mathbb{Q}_i = \{\ [\![S]\!] \mid \text{impact}_i([\![S]\!])\ \}$$

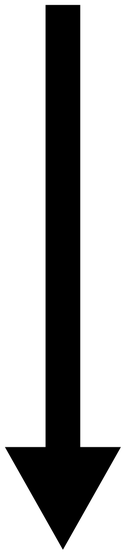


Quantitative

How much is the i-th input feature of P used?

$$P \models \mathbb{Q}_i^k \qquad \text{impact}_i \in \text{Traces} \to \mathbb{R}$$

$$\mathbb{Q}_i^k = \{\ [\![S]\!] \mid \text{impact}_i([\![S]\!]) \leq k\ \}$$



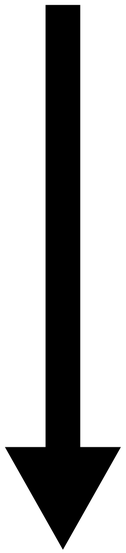
Quantitative

How much is the i-th input feature of P used?

$$P \models \mathbb{Q}_i^k \qquad \mathsf{impact}_i \in \mathsf{Traces} \to \mathbb{R}$$

$$\mathbb{Q}_i^k = \{\ [\![S]\!] \mid \mathsf{impact}_i([\![S]\!]) \leq k\ \}$$

$$P \models \mathbb{Q}_i^k \iff \{[\![P]\!]\} \subseteq \mathbb{Q}_i^k \quad \Leftarrow \quad \gamma([\![P]\!]^\natural) \subseteq \mathbb{Q}_i^k$$



Quantitative

How much is the i-th input feature of P used?

$$P \vDash \mathbb{U}_i$$

$$\mathbb{U}_i = \{\ [\![S]\!] \mid \mathsf{unused}_i([\![S]\!])\ \}$$

$$P \vDash \mathbb{U}_i \iff \{[\![P]\!]\} \subseteq \mathbb{U}_i \quad \Leftarrow \quad \gamma([\![P]\!]^\natural) \subseteq \mathbb{U}_i$$

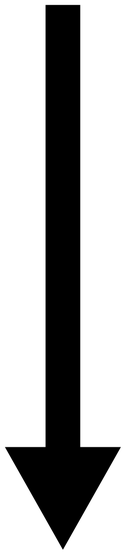
cato

ENS | PSL Inria

Traces $\to \mathbb{R}$

$S]\!]) \leq k\ \}$

$$P \vDash \mathbb{Q}_i^k \iff \{[\![P]\!]\} \subseteq \mathbb{Q}_i^k \quad \Leftarrow \quad \gamma([\![P]\!]^\natural) \subseteq \mathbb{Q}_i^k$$



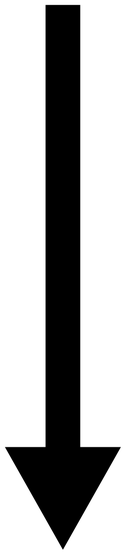
Quantitative

How much is the i-th input feature of P used?

$$P \models \mathbb{Q}_i^k \qquad \mathsf{impact}_i \in \mathsf{Traces} \to \mathbb{R}$$

$$\mathbb{Q}_i^k = \{\ [\![S]\!] \mid \mathsf{impact}_i([\![S]\!]) \leq k\ \}$$

$$P \models \mathbb{Q}_i^k \iff \{[\![P]\!]\} \subseteq \mathbb{Q}_i^k \quad \Leftarrow \quad \gamma([\![P]\!]^\natural) \subseteq \mathbb{Q}_i^k$$



Quantitative

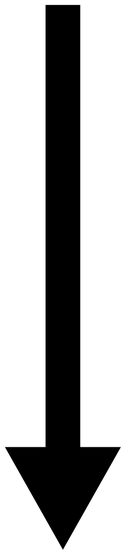
How much is the i-th input feature of P used?

$$P \models \mathbb{Q}_i^k \qquad \text{impact}_i \in \text{Traces} \to \mathbb{D}$$

Quantities Domain

$$\mathbb{Q}_i^k = \{\ [\![S]\!] \mid \text{impact}_i([\![S]\!]) \leq k\ \}$$

$$P \models \mathbb{Q}_i^k \iff \{[\![P]\!]\} \subseteq \mathbb{Q}_i^k \quad \Leftarrow \quad \gamma([\![P]\!]^\natural) \subseteq \mathbb{Q}_i^k$$



Quantitative

How much is the i-th input feature of P used?

How to define "$\text{impact}_i \in \text{Traces} \to \mathbb{D}$"?

How to define "$\text{impact}_i \in \text{Traces} \to \mathbb{D}$"?

The number of output changes

(with repetitions)

result from perturbations on the i-th input feature

for any input X

How to define "$\text{impact}_i \in \text{Traces} \to \mathbb{D}$"?

The number of output changes

(with repetitions)

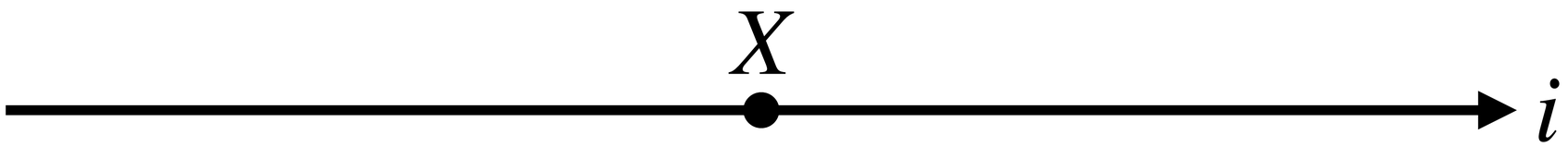
result from perturbations on the i-th input feature

for any input X

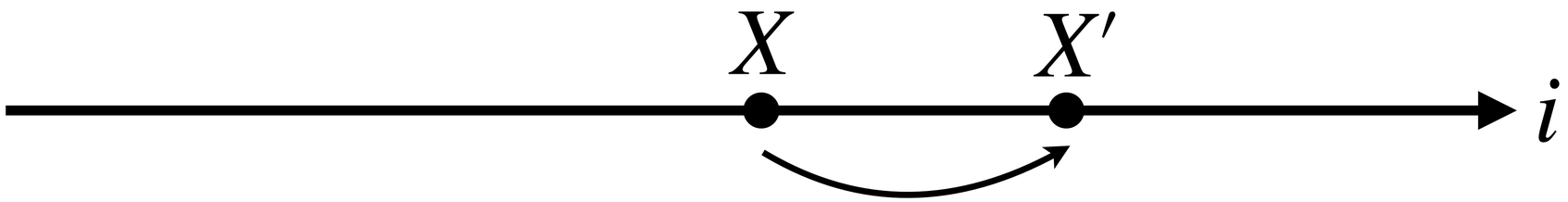
$$\text{CountChanges}_i \in \text{Traces} \to \mathbb{N}^\infty$$

$$\mathrm{CountChanges}_i \in \mathrm{Traces} \to \mathbb{N}^\infty$$

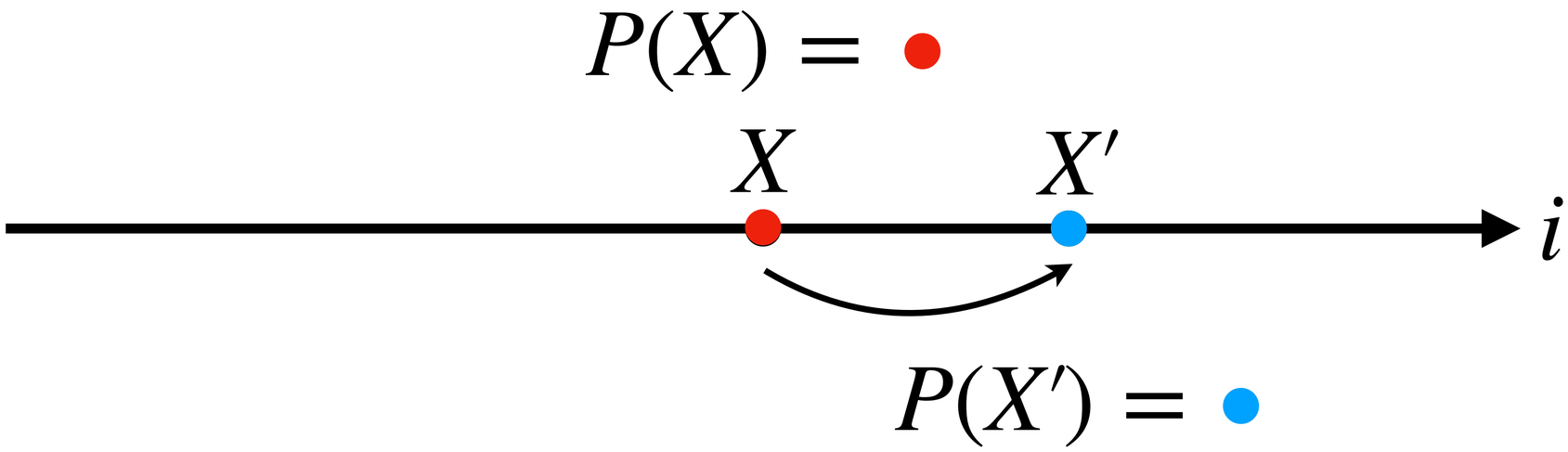
$\mathsf{CountChanges}_i \in \mathsf{Traces} \to \mathbb{N}^\infty$



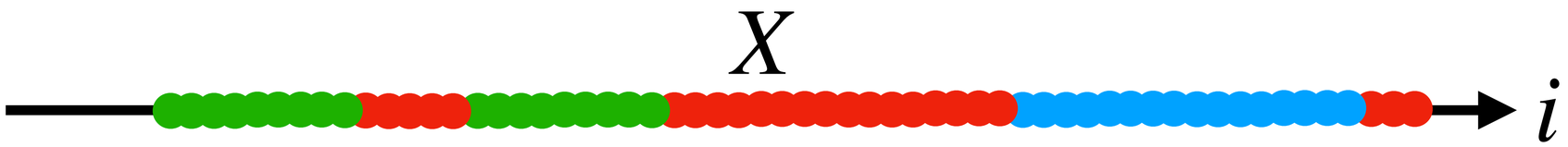
$\mathrm{CountChanges}_i \in \mathrm{Traces} \to \mathbb{N}^\infty$



$\mathsf{CountChanges}_i \in \mathsf{Traces} \to \mathbb{N}^\infty$

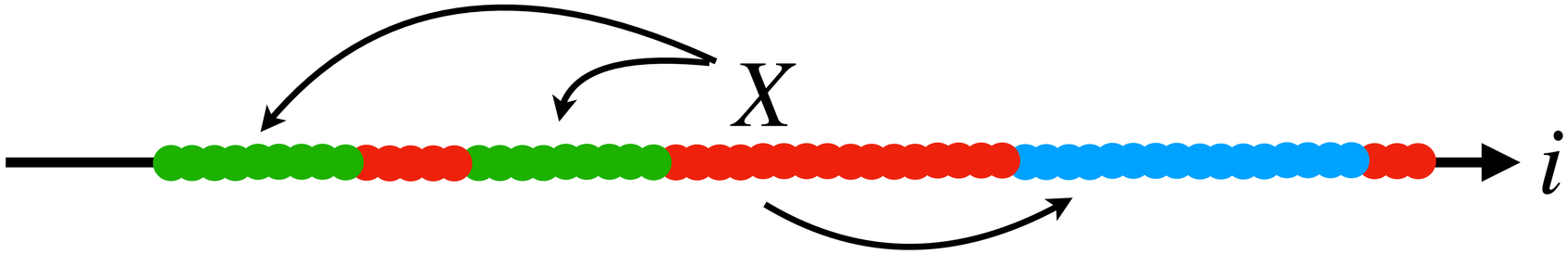


$\mathrm{CountChanges}_i \in \mathrm{Traces} \to \mathbb{N}^\infty$



$$\mathsf{CountChanges}_i \in \mathsf{Traces} \to \mathbb{N}^\infty$$

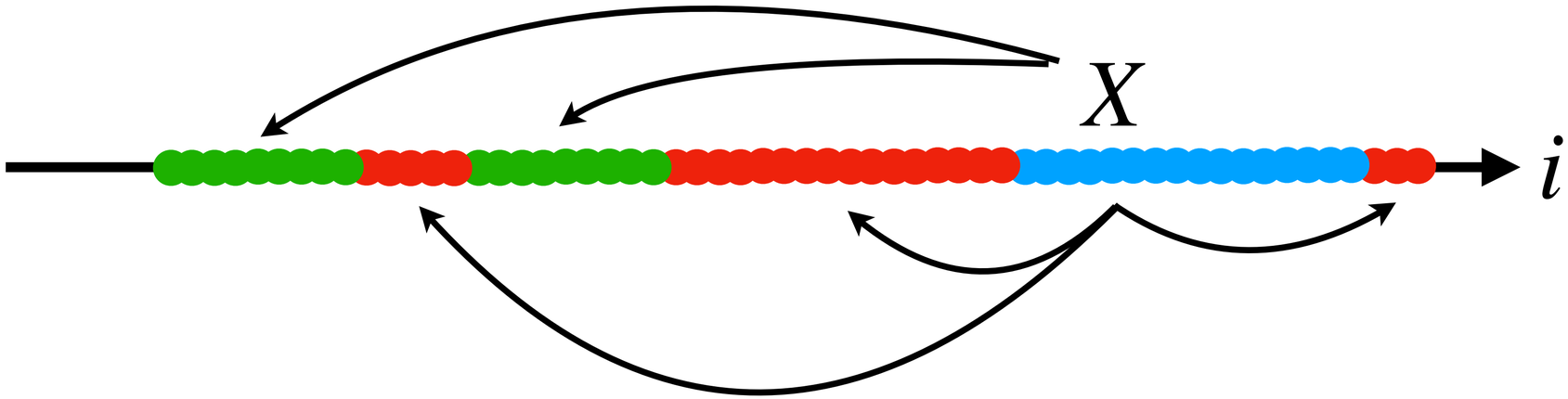
starting from X leading to ● $\Longrightarrow$ 3 changes



$\mathsf{CountChanges}_i \in \mathsf{Traces} \to \mathbb{N}^\infty$

starting from X leading to ● $\Longrightarrow$ 3 changes

starting from X leading to ● $\Longrightarrow$ 5 changes

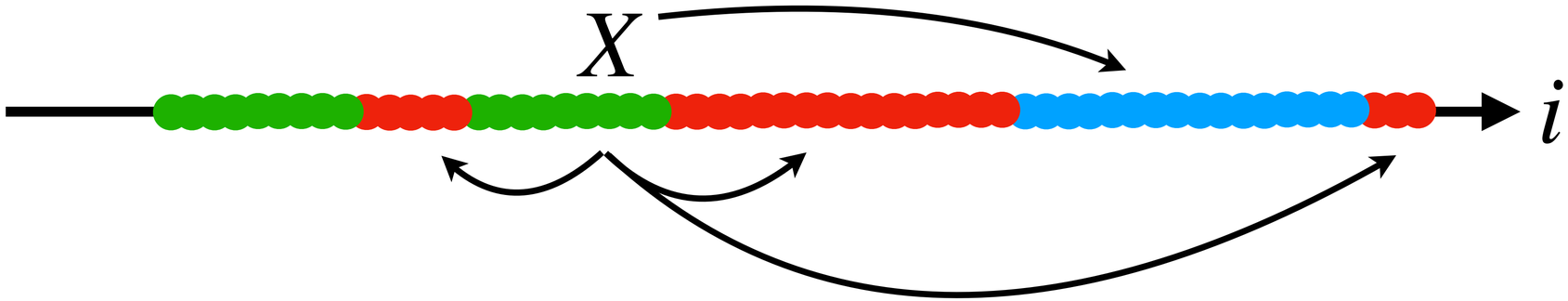


$\mathsf{CountChanges}_i \in \mathsf{Traces} \to \mathbb{N}^\infty$

starting from X leading to ● $\Longrightarrow$ 3 changes

starting from X leading to ● $\Longrightarrow$ 5 changes

starting from X leading to ● $\Longrightarrow$ 4 changes



$$\mathsf{CountChanges}_i \in \mathsf{Traces} \to \mathbb{N}^\infty$$

starting from X leading to ● $\Longrightarrow$ 3 changes

starting from X leading to ● $\Longrightarrow$ 5 changes

starting from X leading to ● $\Longrightarrow$ 4 changes

$$\mathsf{CountChanges}_i \in \mathsf{Traces} \to \mathbb{N}^\infty$$

$$\mathsf{CountChanges}_i(P) = 5$$

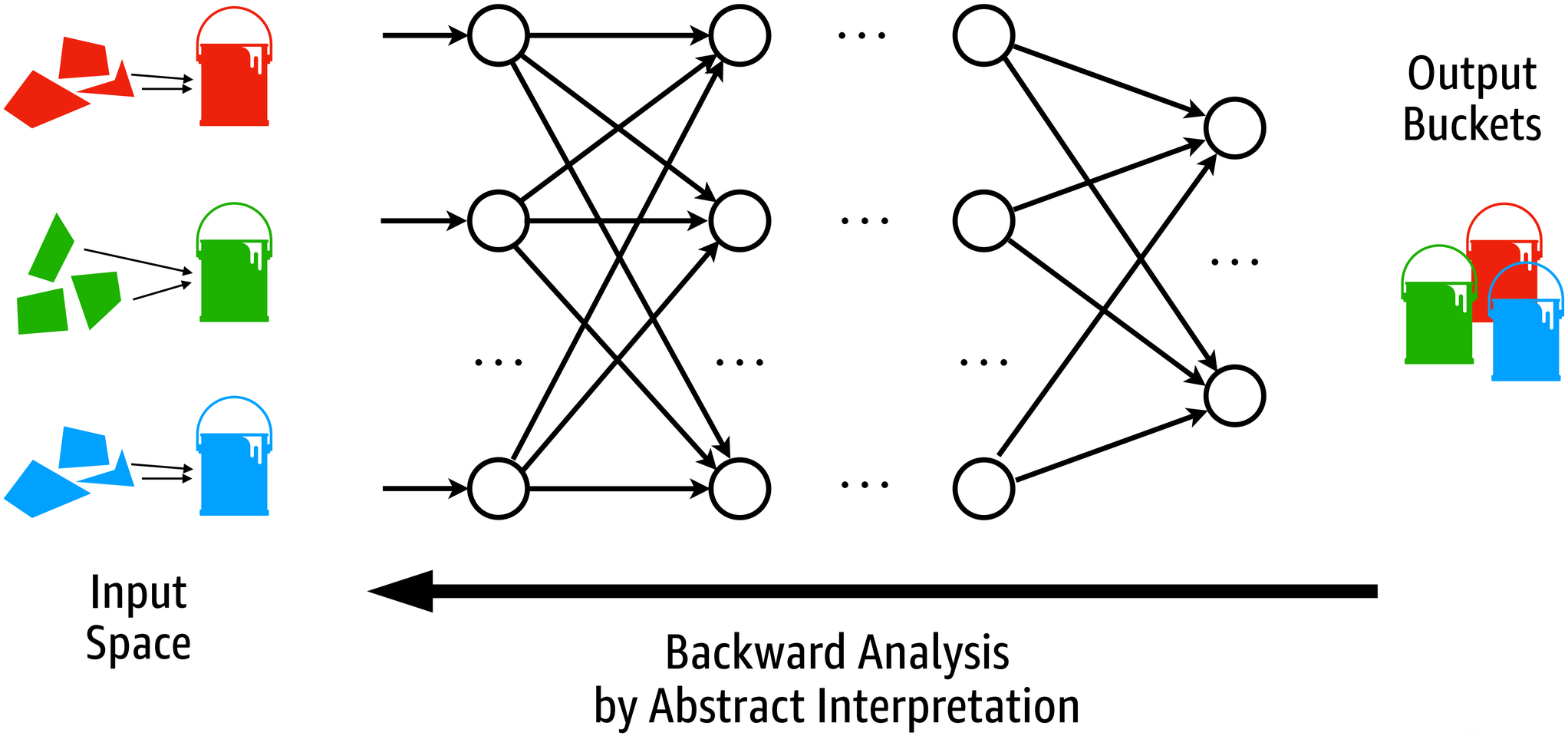
$$\mathsf{CountChanges}_i \in \mathsf{Traces} \to \mathbb{N}^\infty$$

$$\mathsf{CountChanges}_i(P) = 5$$

Continuous input space ✓

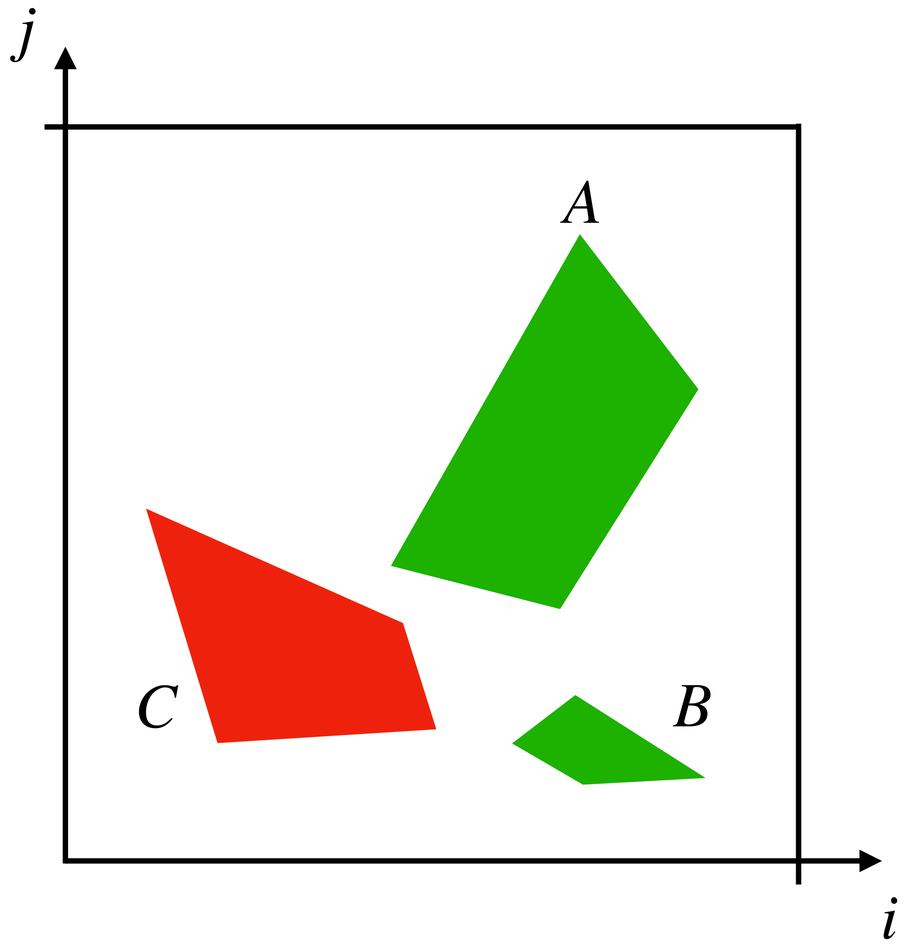
Discrete output space ✓

ImpactAnalysis$_i$

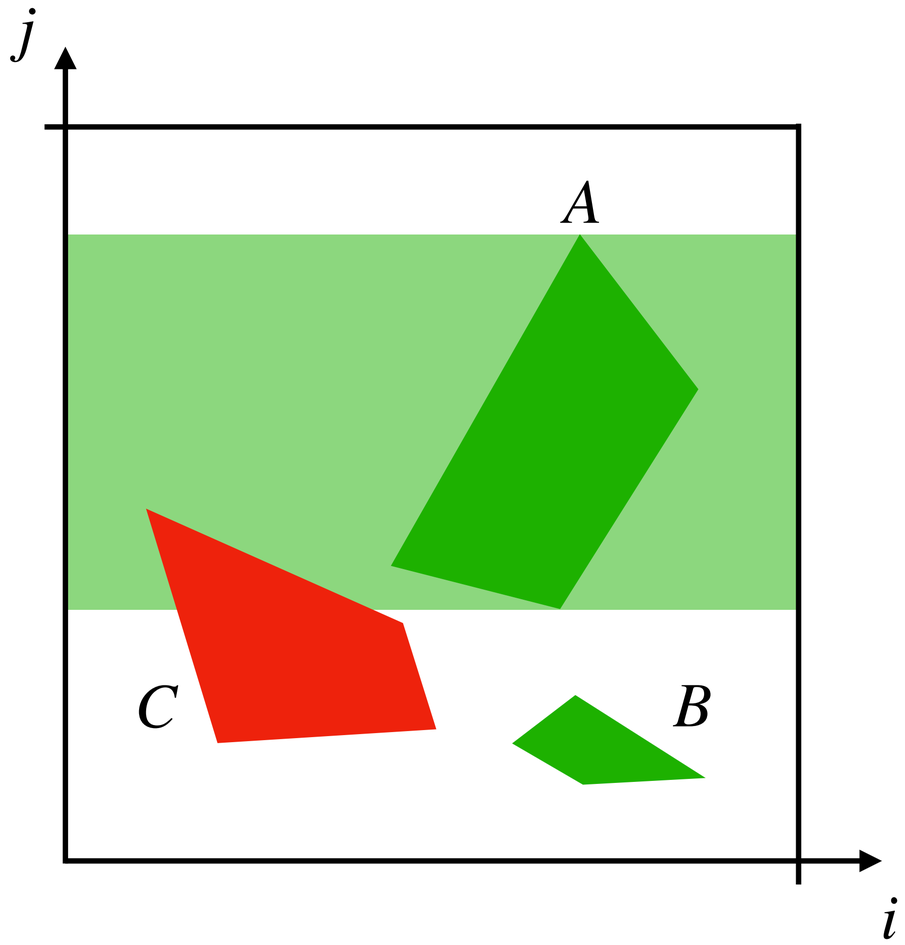


ImpactAnalysis$^{\natural}_{i}$

2 input features

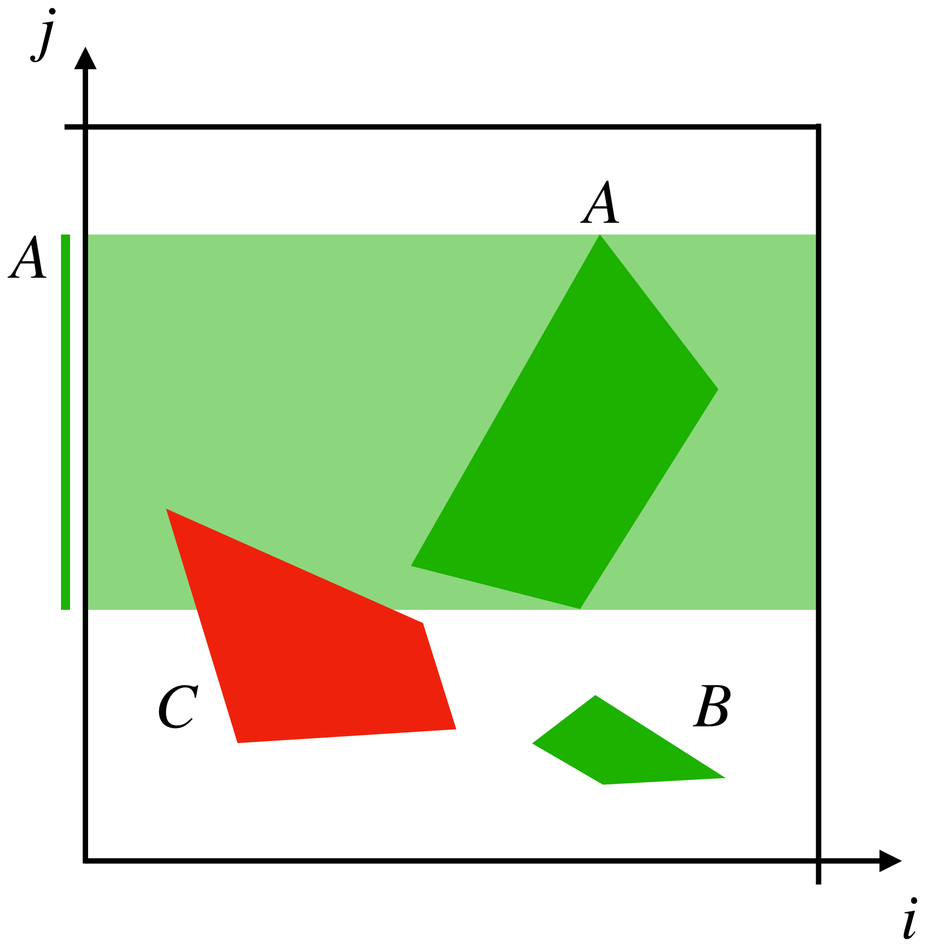


Perturbations of i



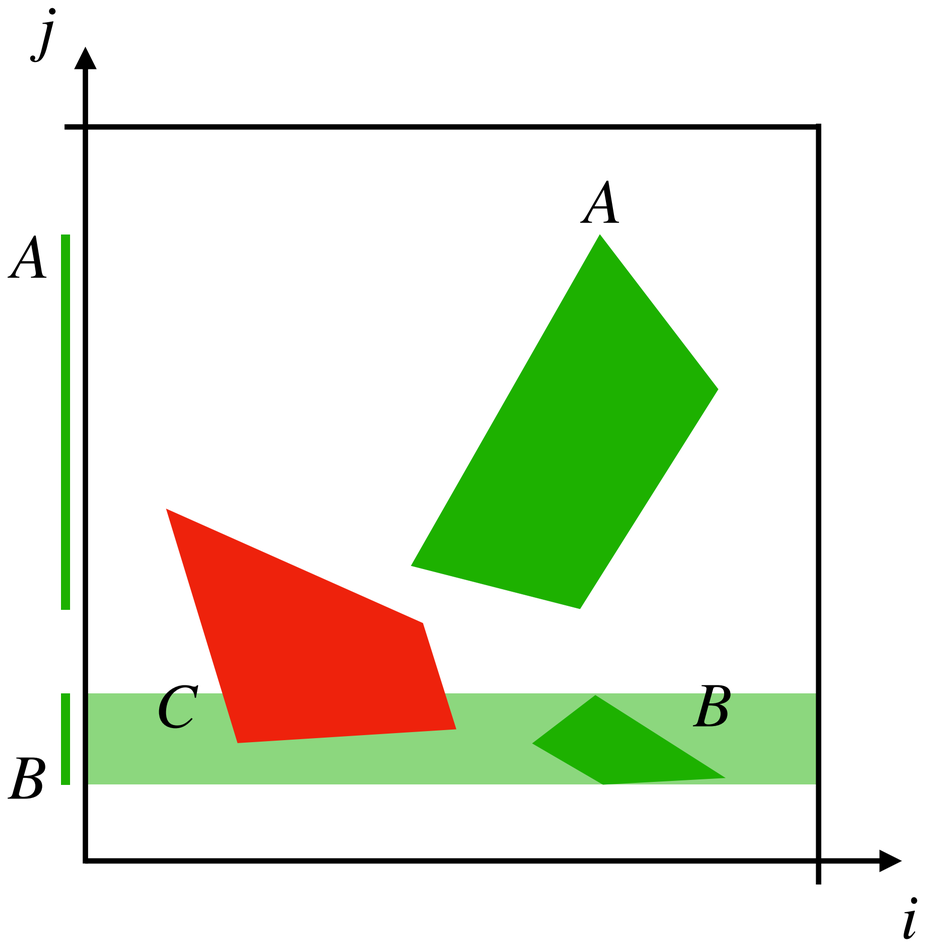
Projecting away the feature i

Perturbations of i



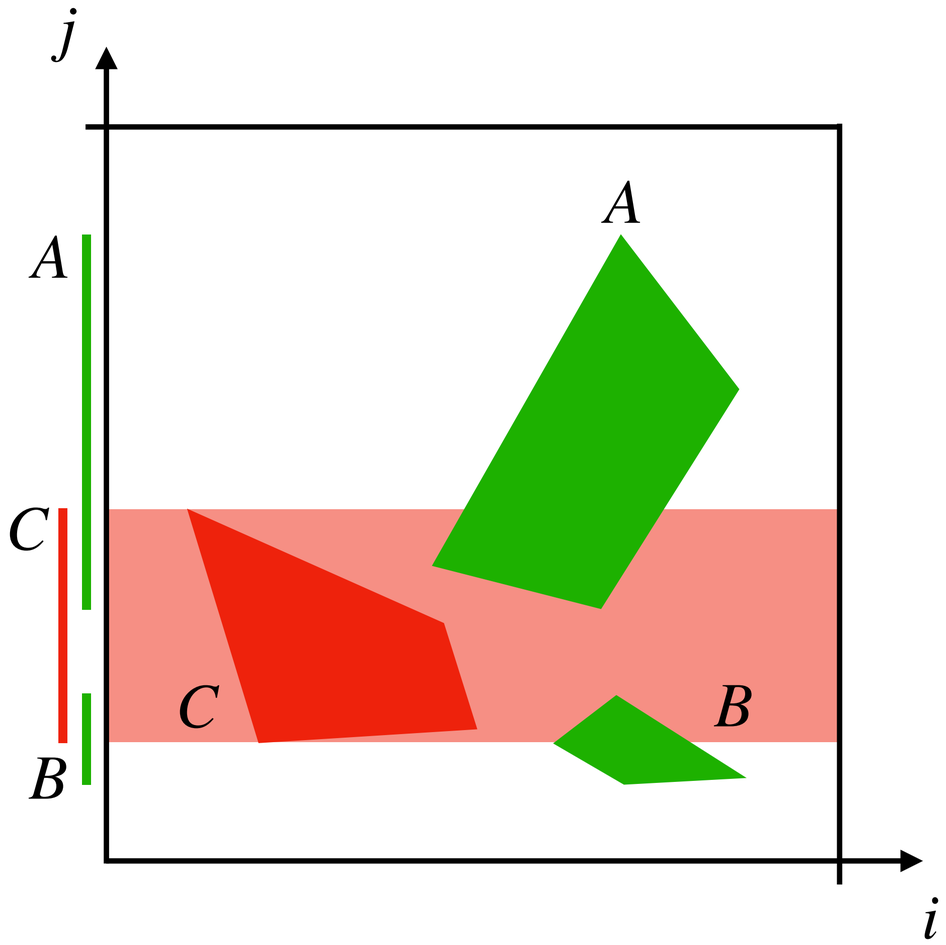
Projecting away the feature i

Perturbations of i



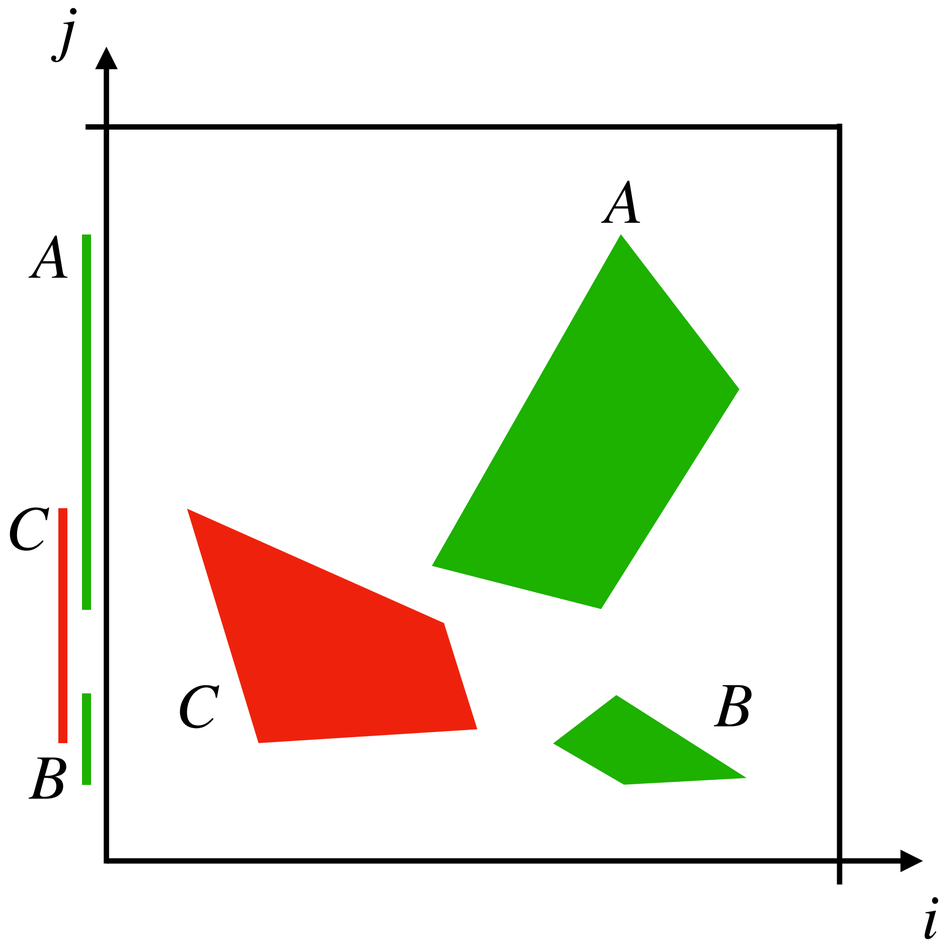
Projecting away the feature i

Perturbations of i



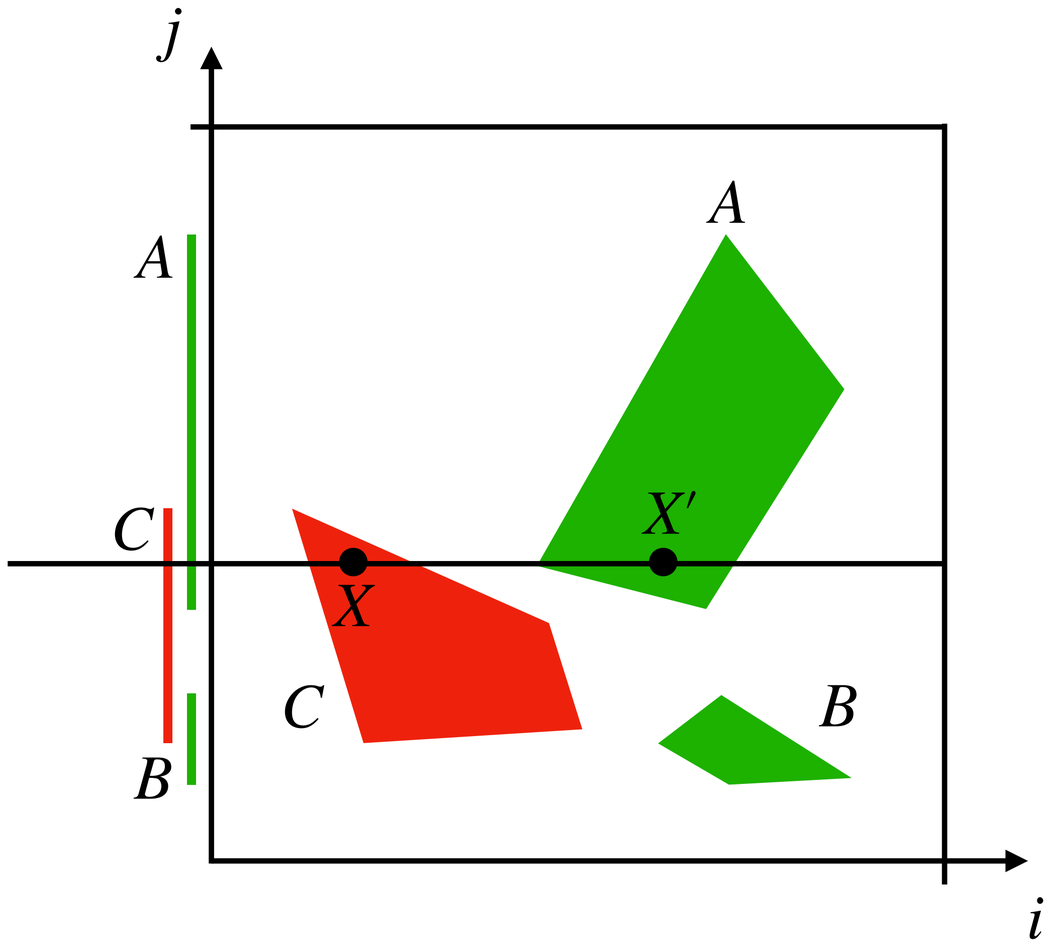
Projecting away the feature i

Meaning of intersections

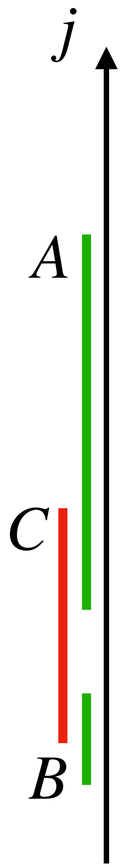


Meaning of intersections

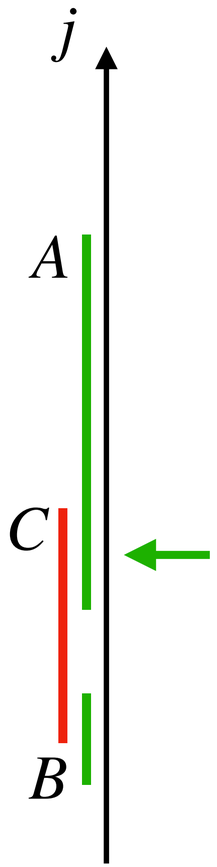
Intersection between A and C



Count the intersections

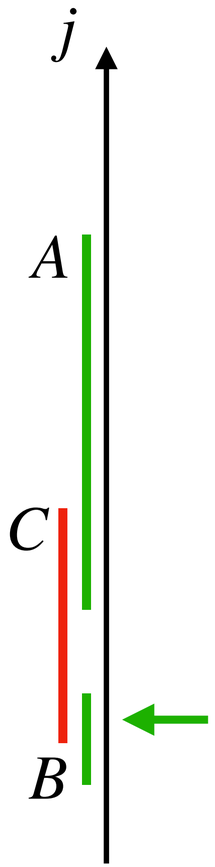


Count the intersections



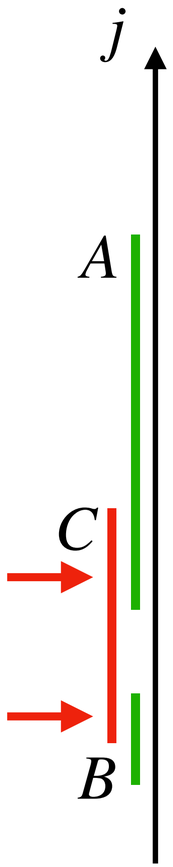
starting from ● $\Longrightarrow$ 1 intersection

Count the intersections



starting from ● $\Longrightarrow$ 1 intersection

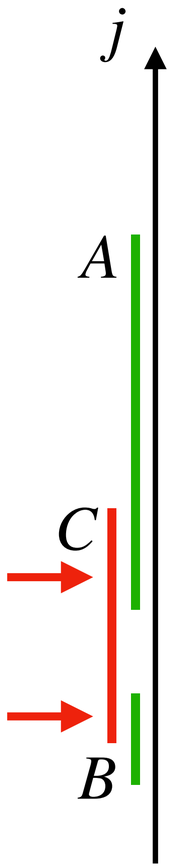
Count the intersections



starting from ● $\Longrightarrow$ 1 intersection

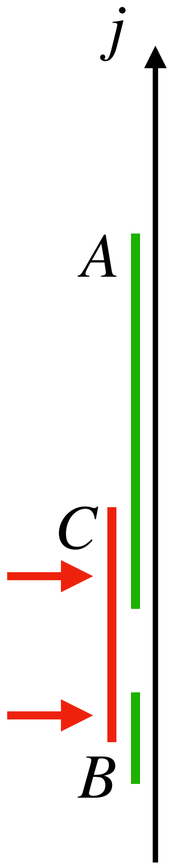
starting from ● $\Longrightarrow$ 2 intersections

Count the intersections



$$\mathsf{ImpactAnalysis}_i^{\natural}(P) = 2$$

Count the intersections



concrete

$$\mathsf{CountChanges}_i(P)$$

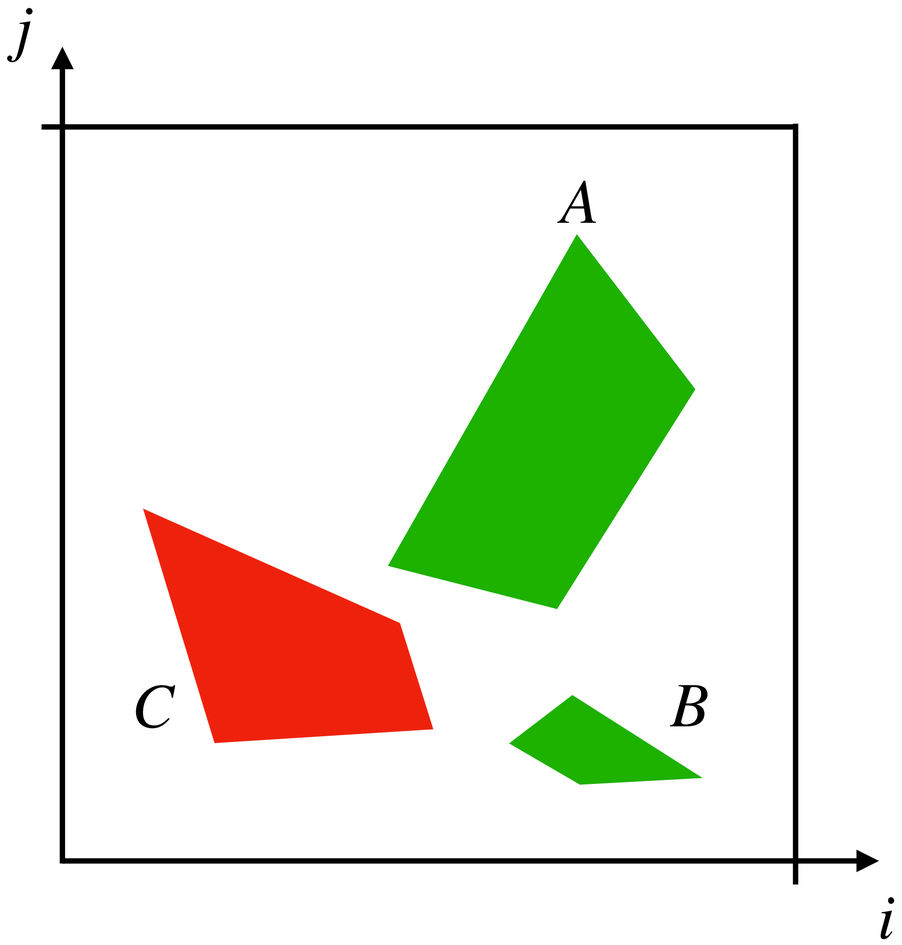
$$\leq$$

$$\mathsf{ImpactAnalysis}_i^{\natural}(P) = 2$$

abstract

Sound?

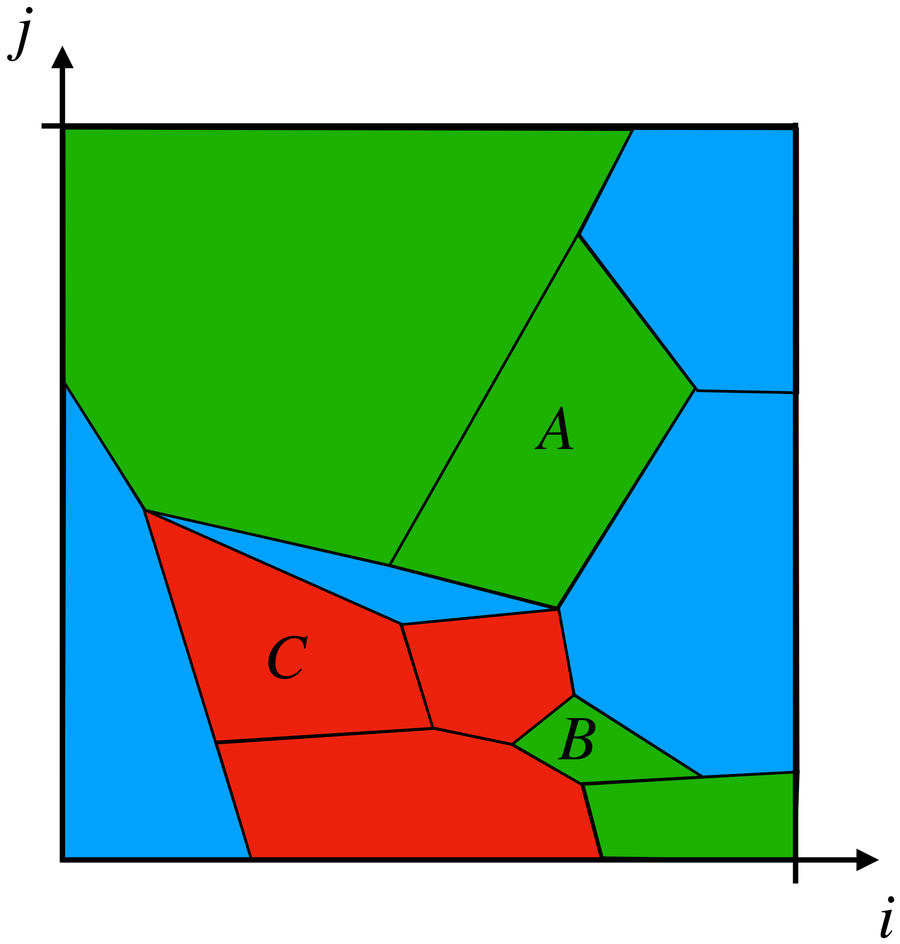
2 input features



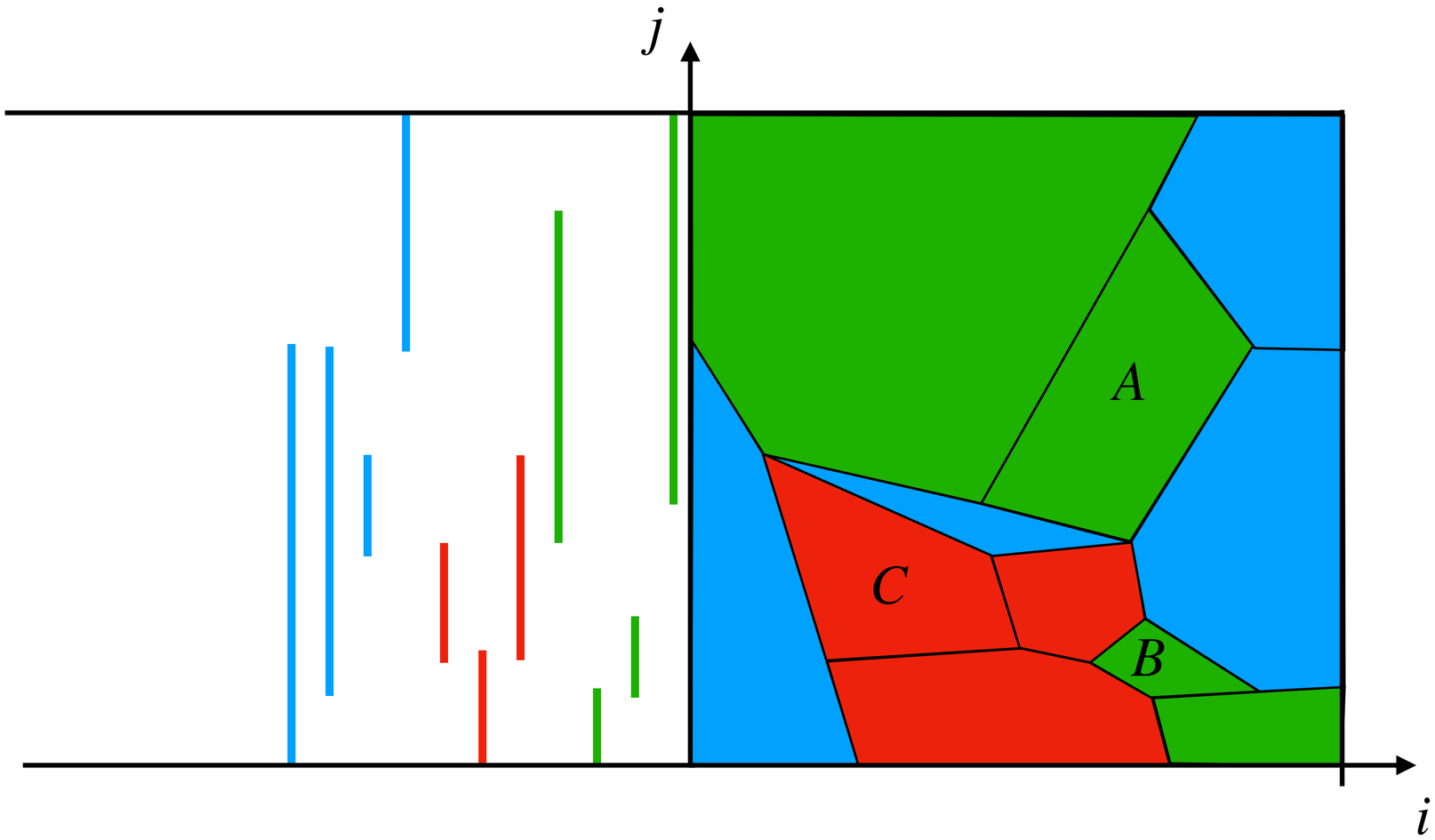
Sound?

2 input features

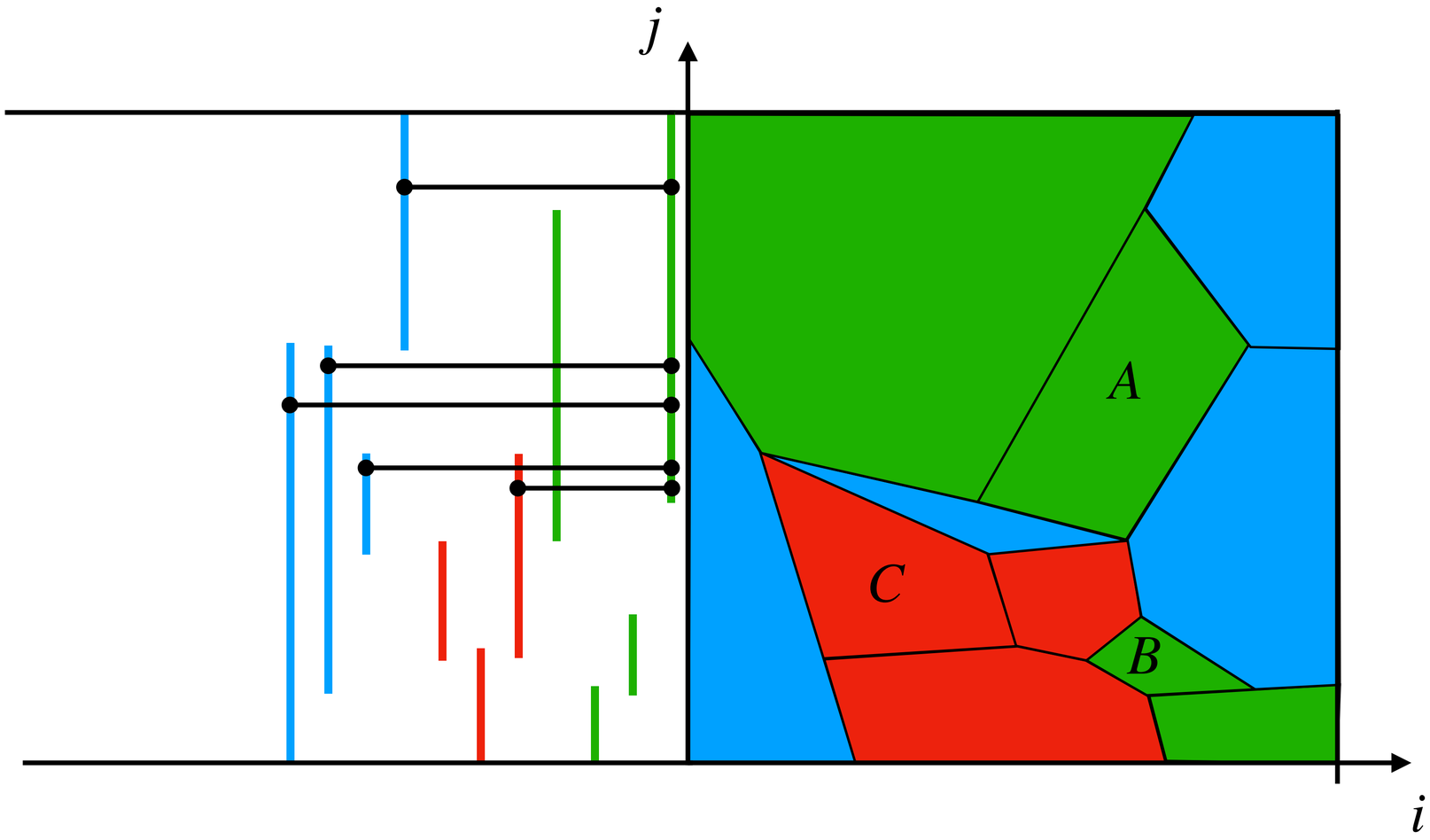
3 output buckets



Sound?

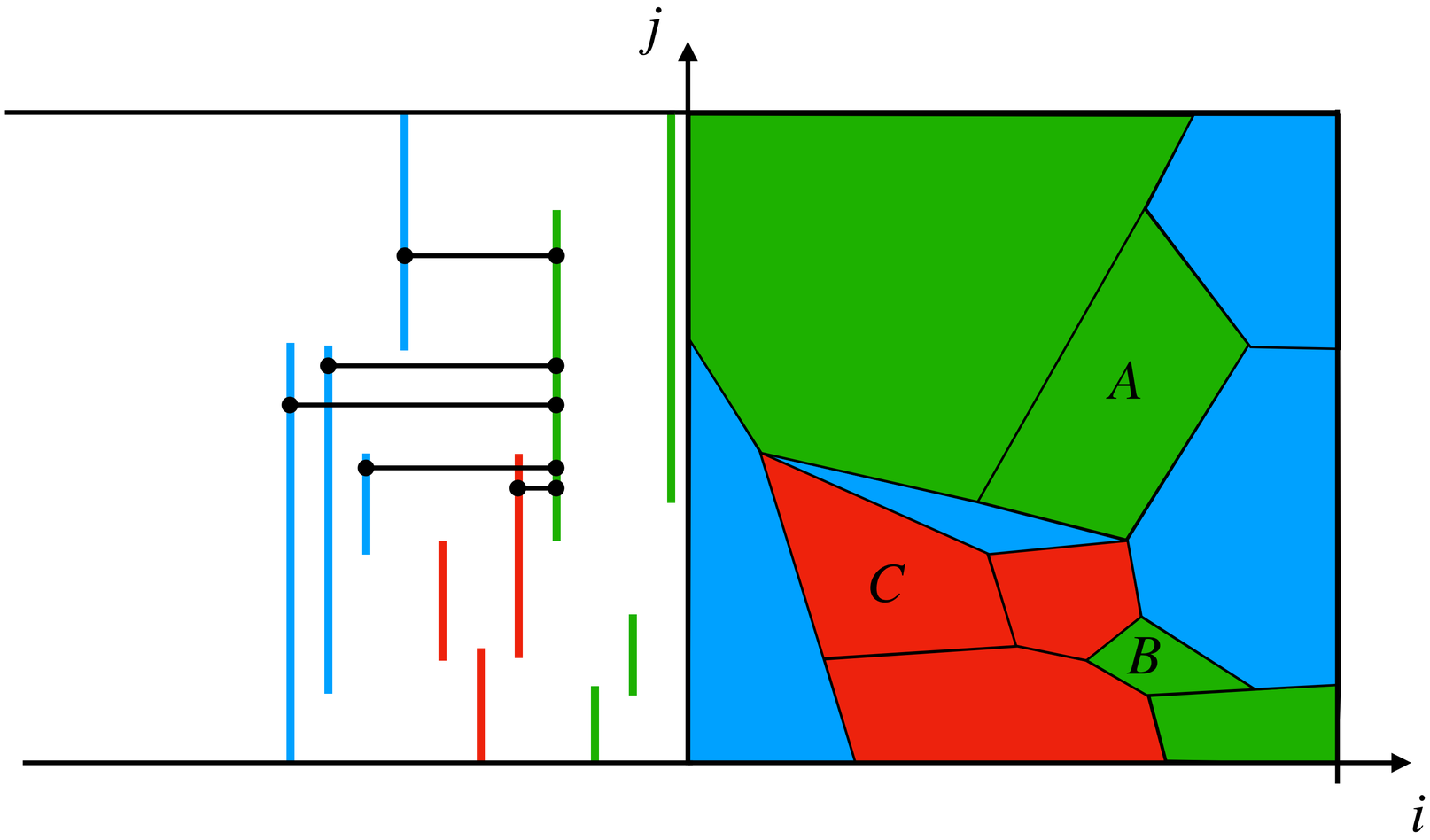


Sound?



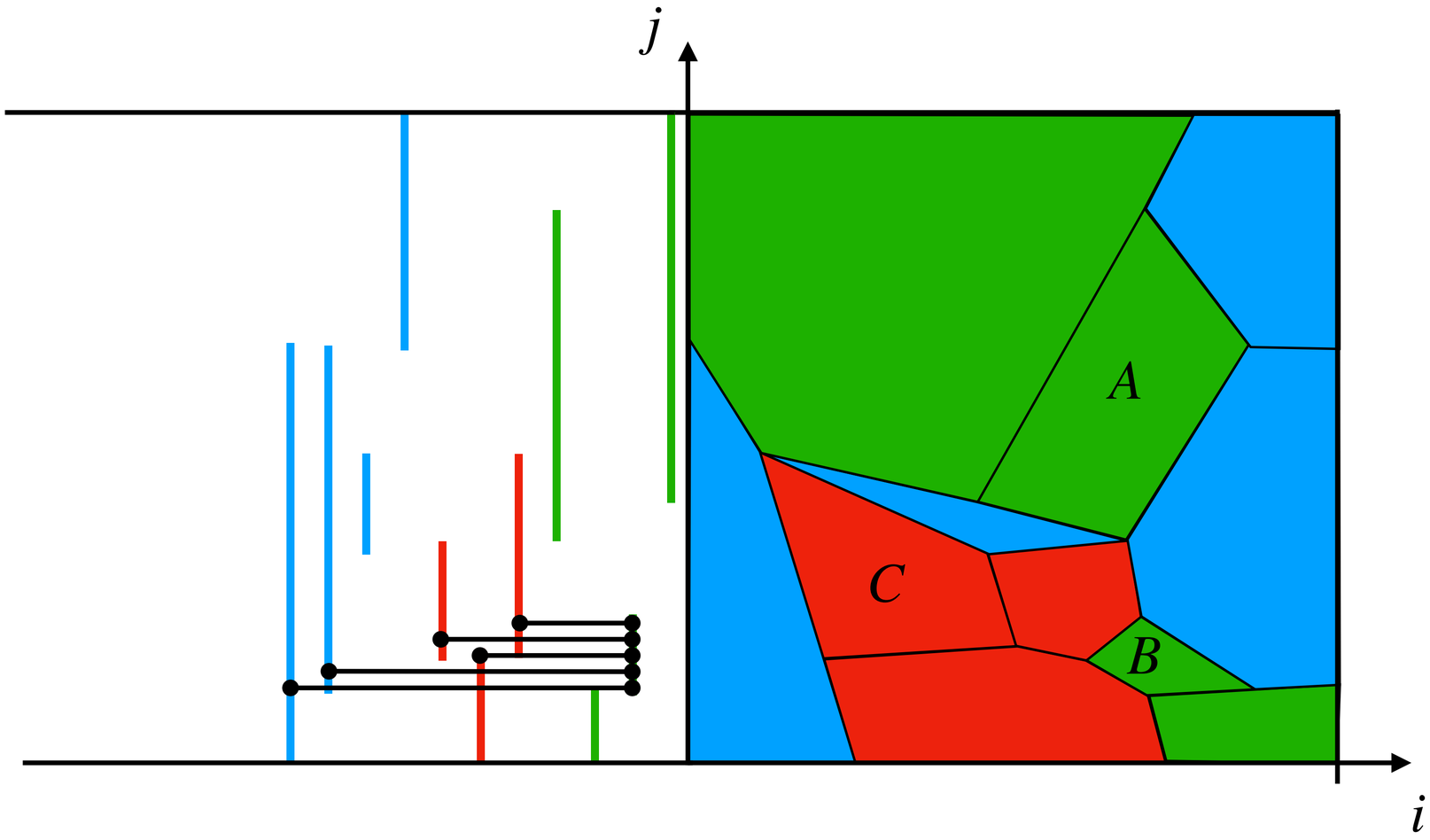
$\bullet \Longrightarrow \{5,$

Sound?



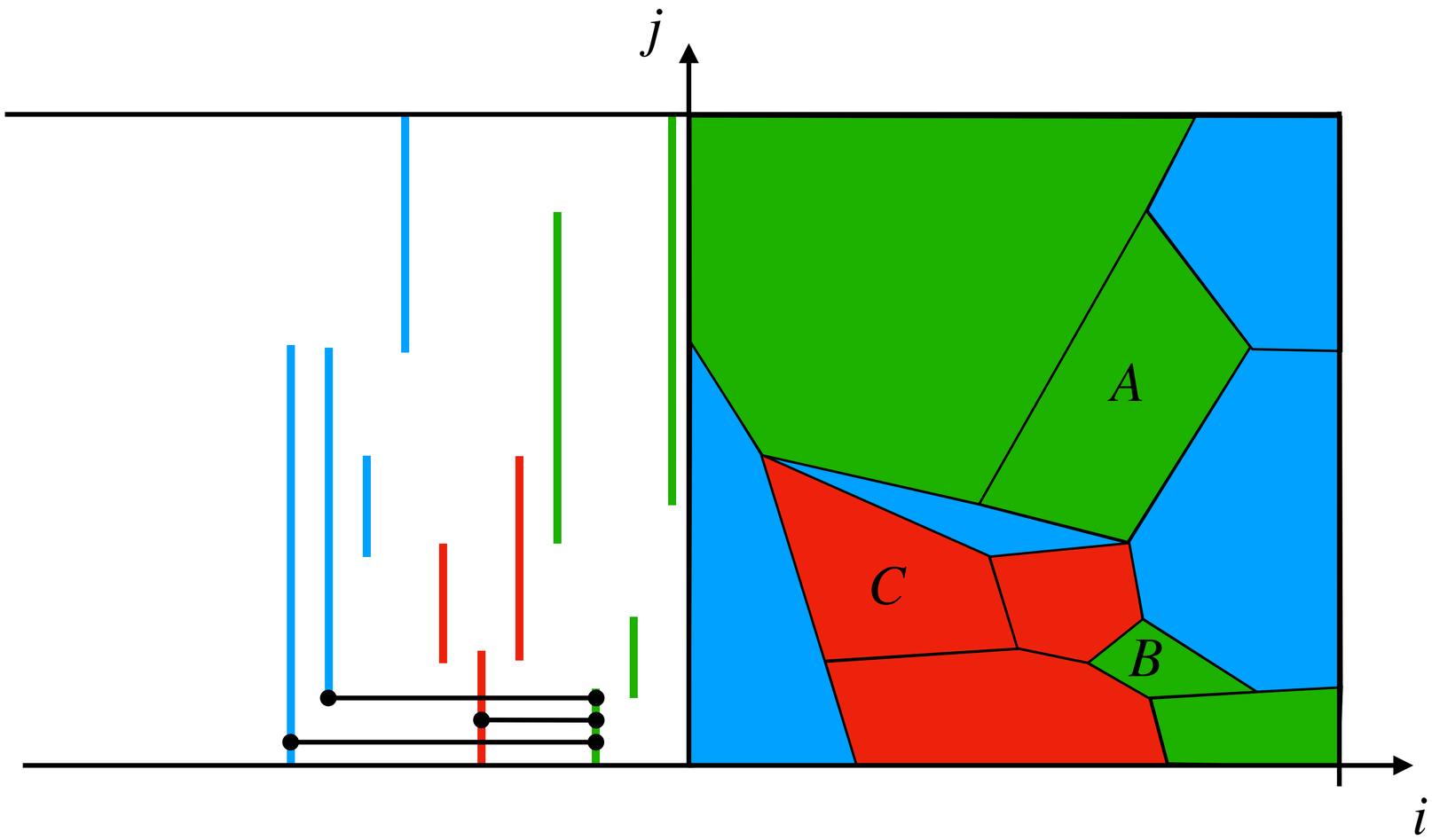
● $\Longrightarrow$ {5,5

Sound?



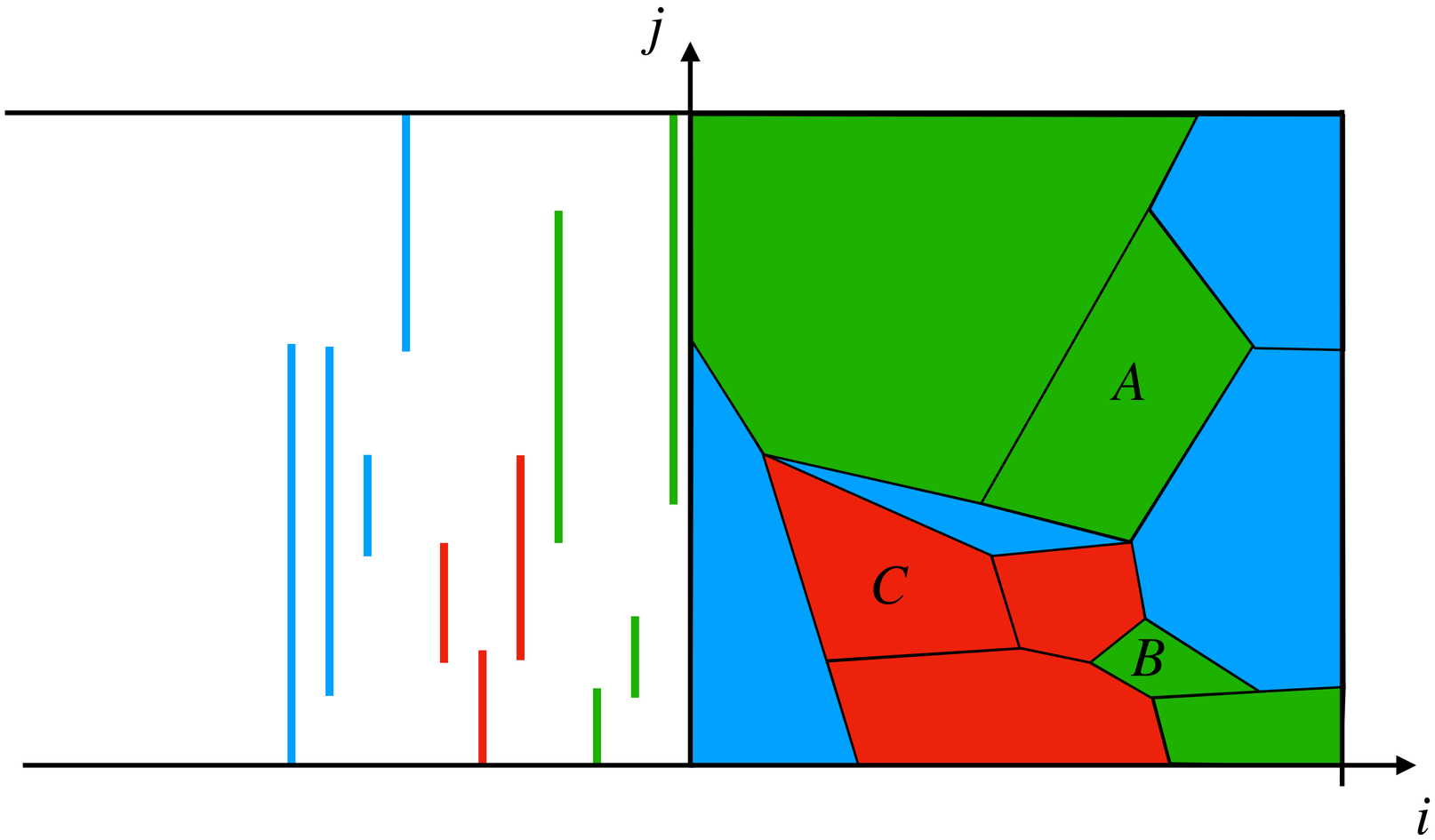
● $\Longrightarrow$ {5,5,5

Sound?



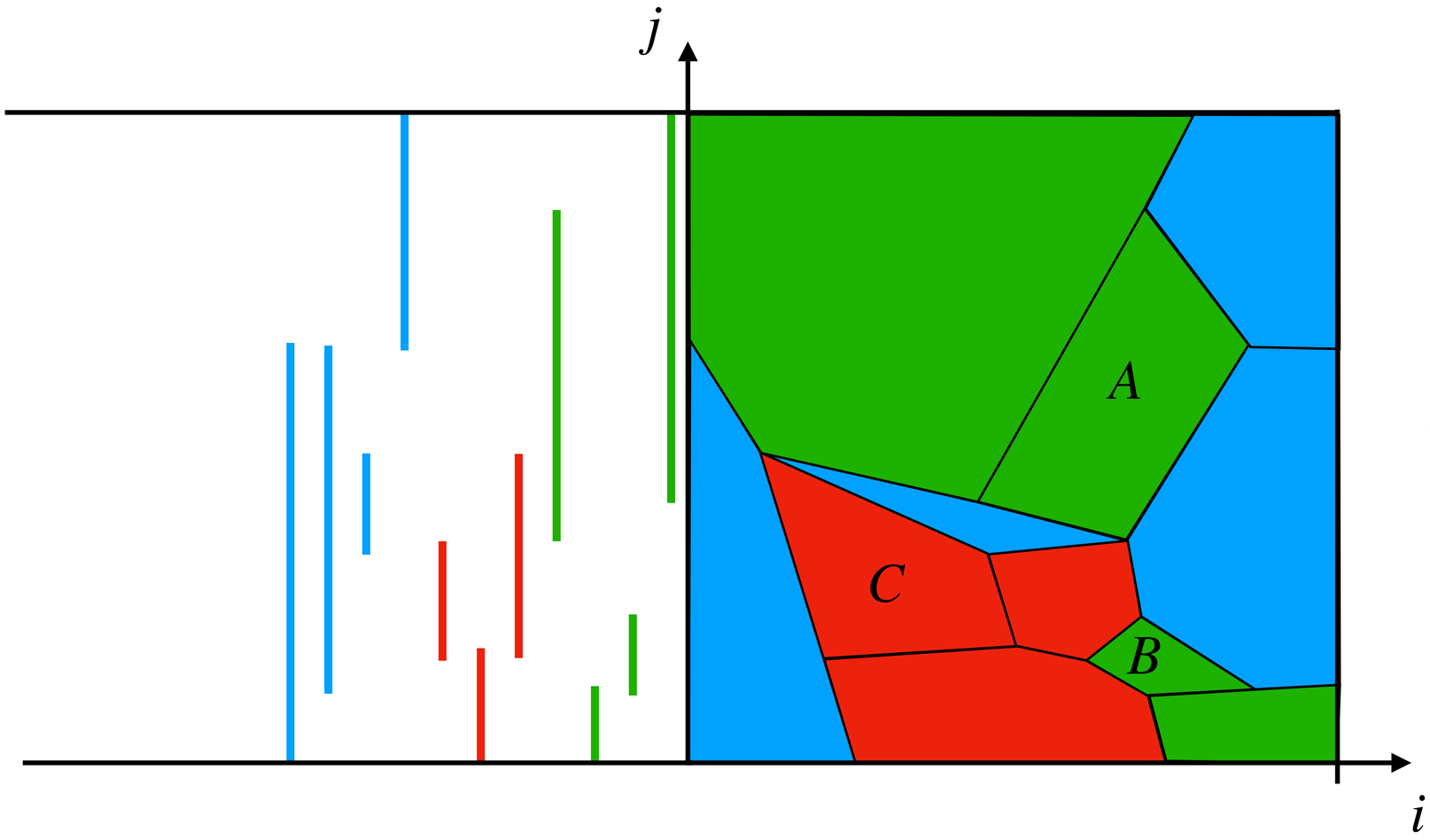
$\bullet \Longrightarrow \{5,5,5,3\}$

Sound?



- $\Longrightarrow$ 5 intersections

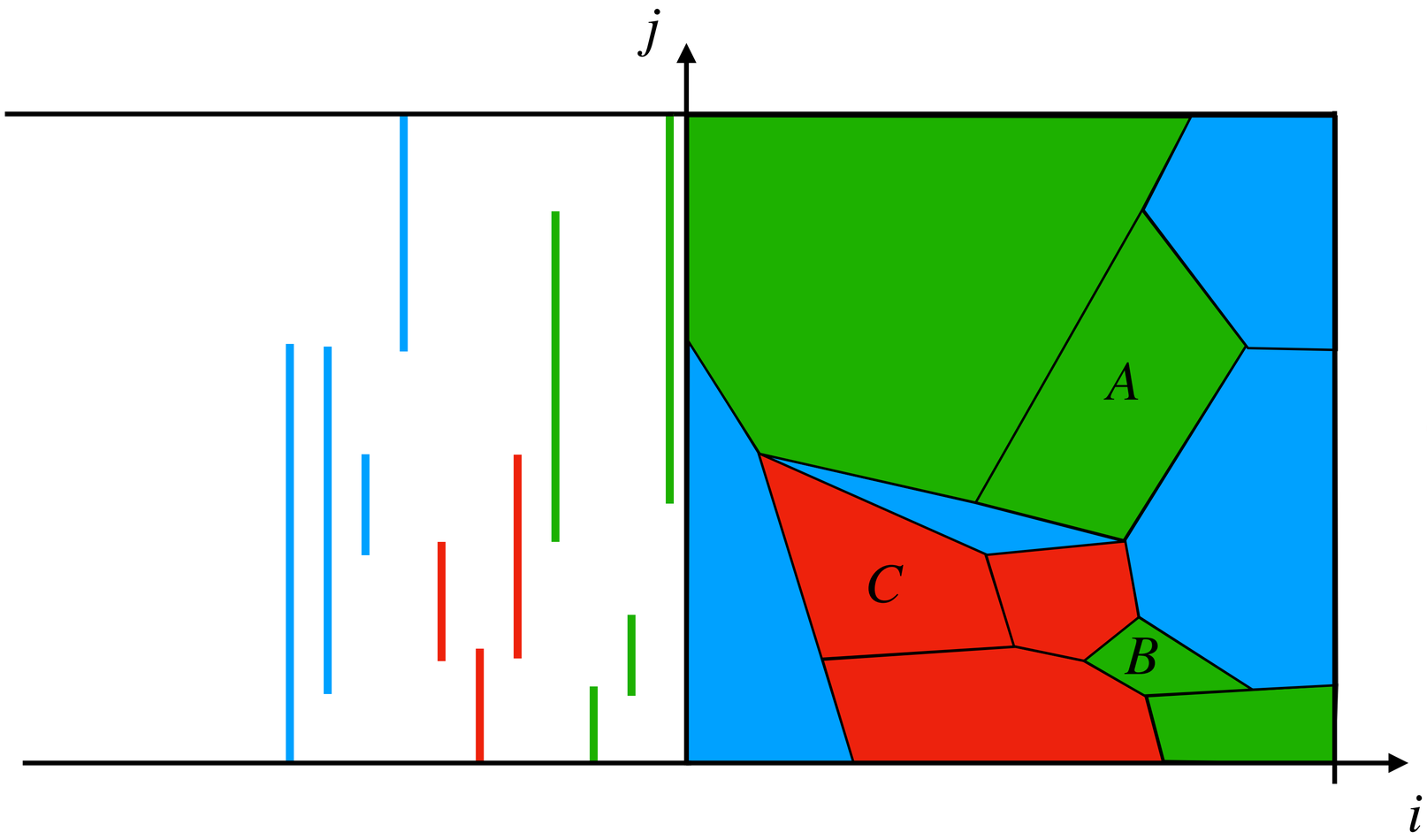
Sound?



● $\Longrightarrow$ 5 intersections

● $\Longrightarrow$ 6 intersections

Sound?

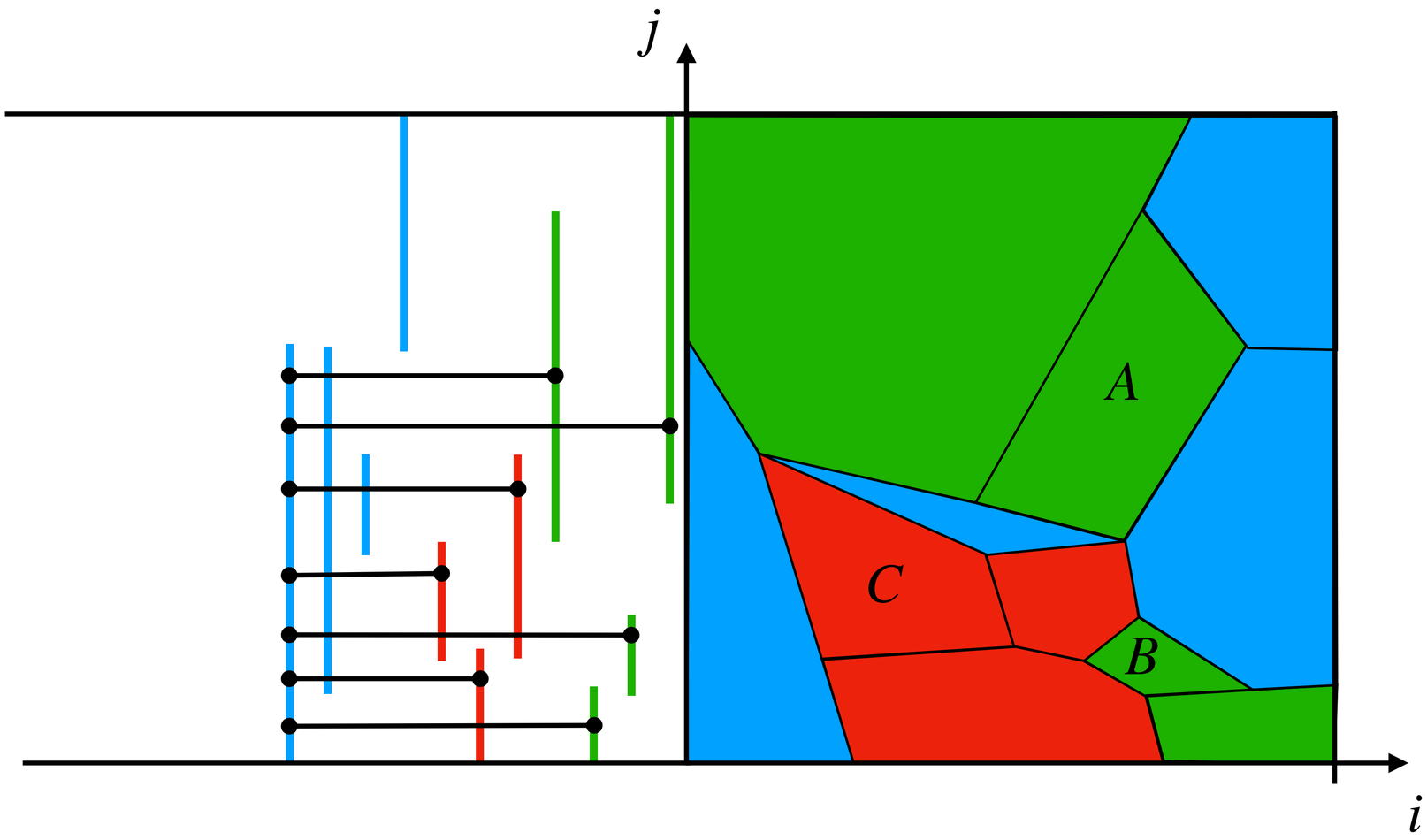


● $\Longrightarrow$ 5 intersections

● $\Longrightarrow$ 6 intersections

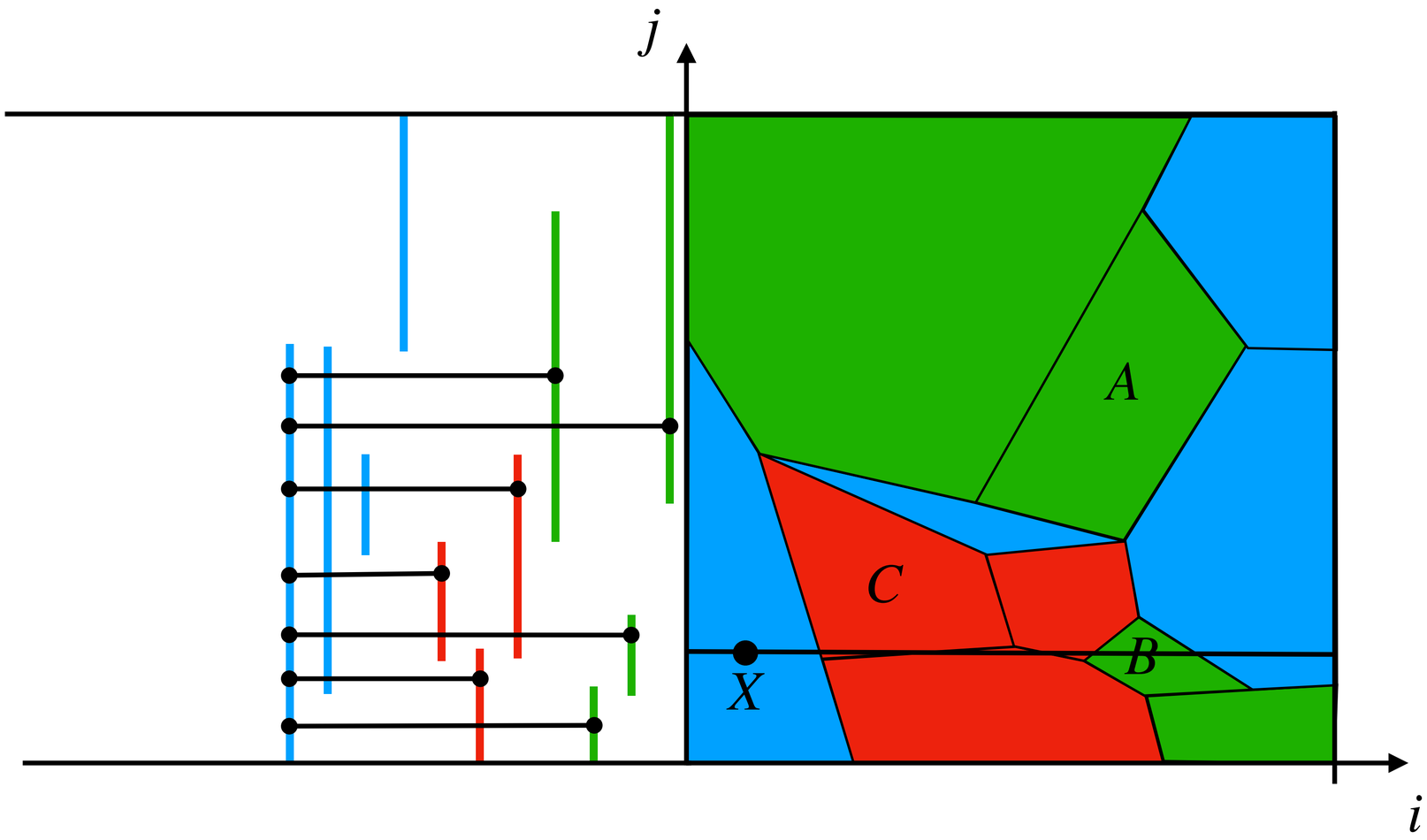
● $\Longrightarrow$ 7 intersections

Sound?



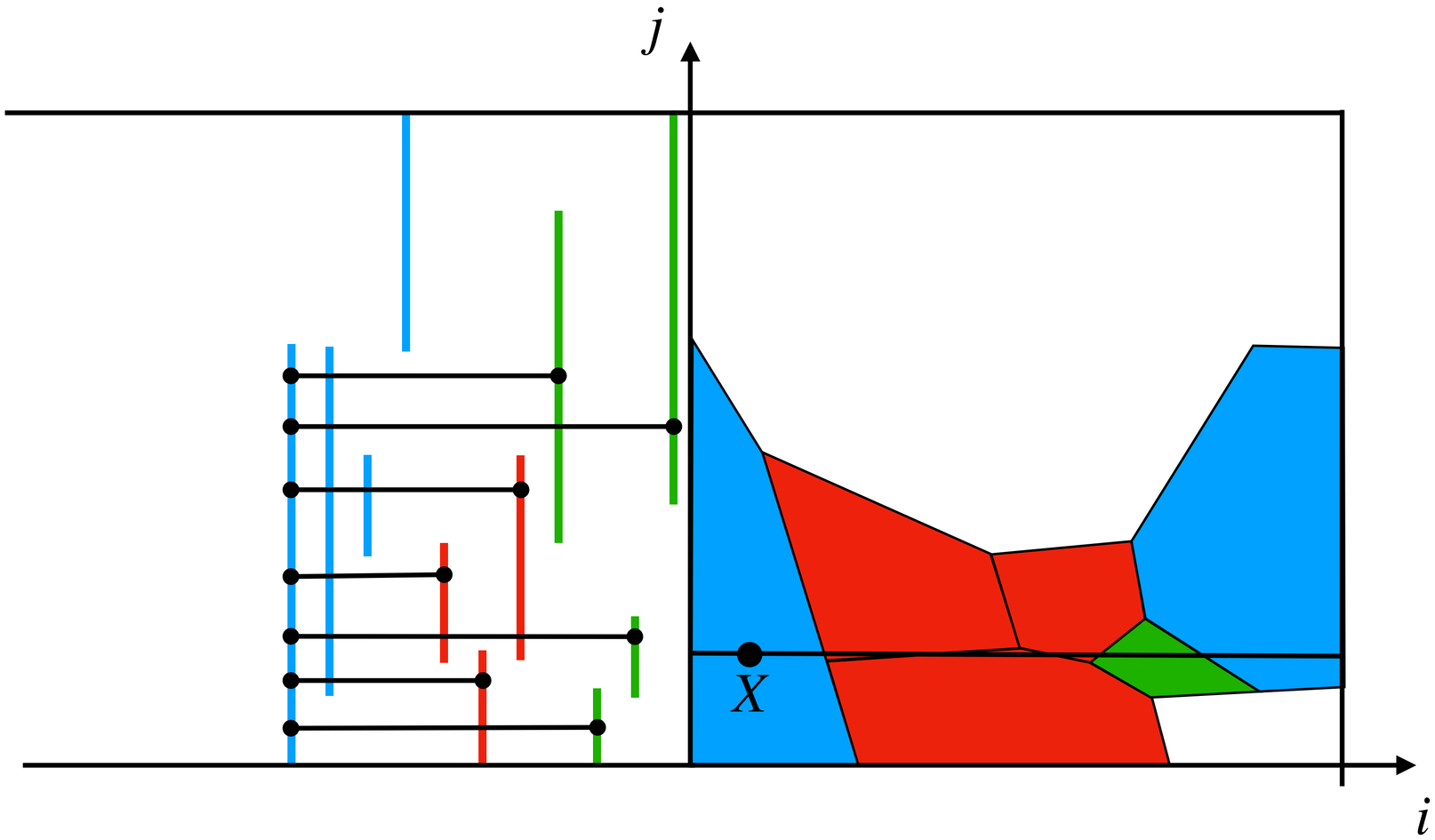
- ● $\Longrightarrow$ 5 intersections
- ● $\Longrightarrow$ 6 intersections
- ● $\Longrightarrow$ 7 intersections

Sound?



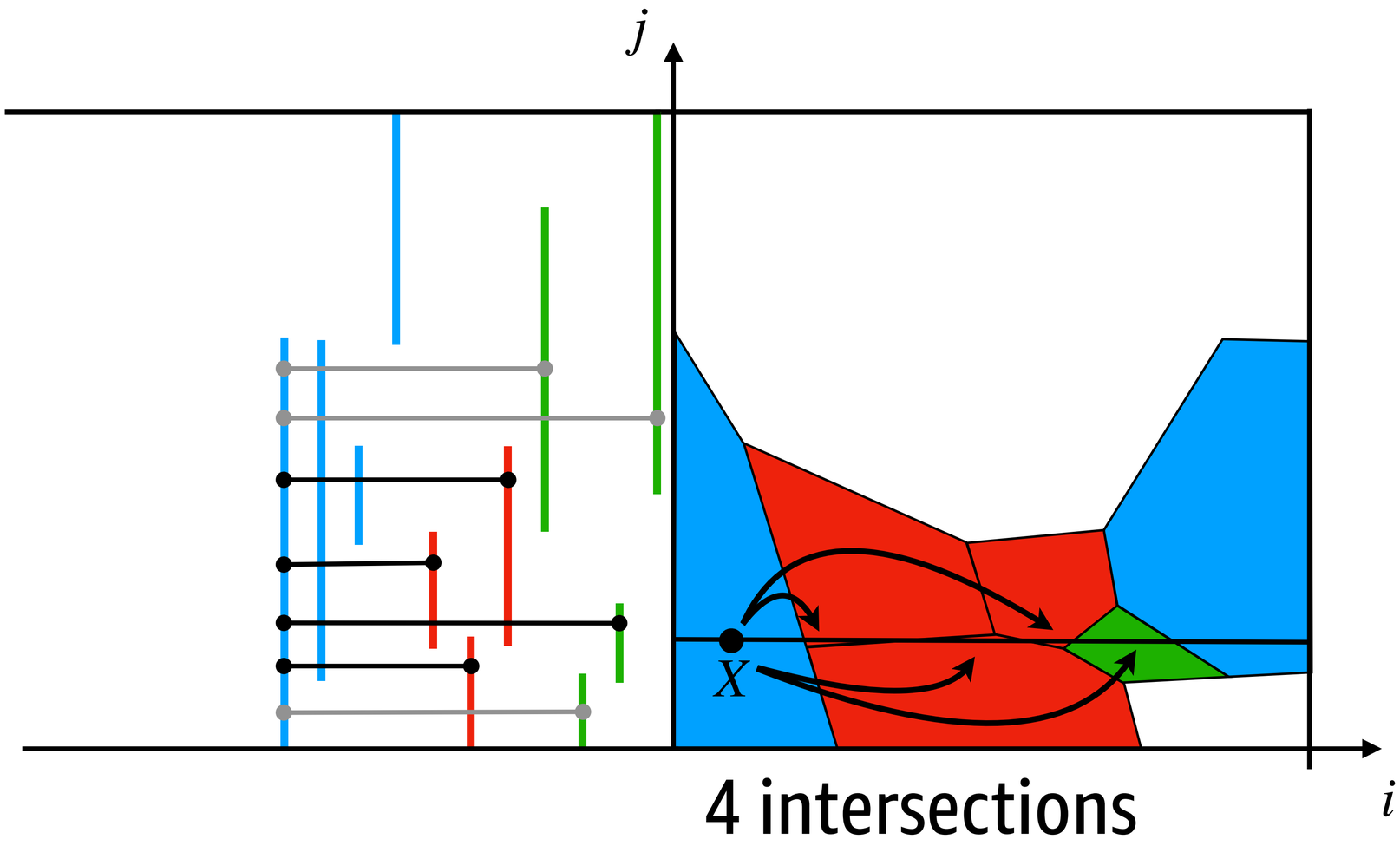
7 intersections

Sound?



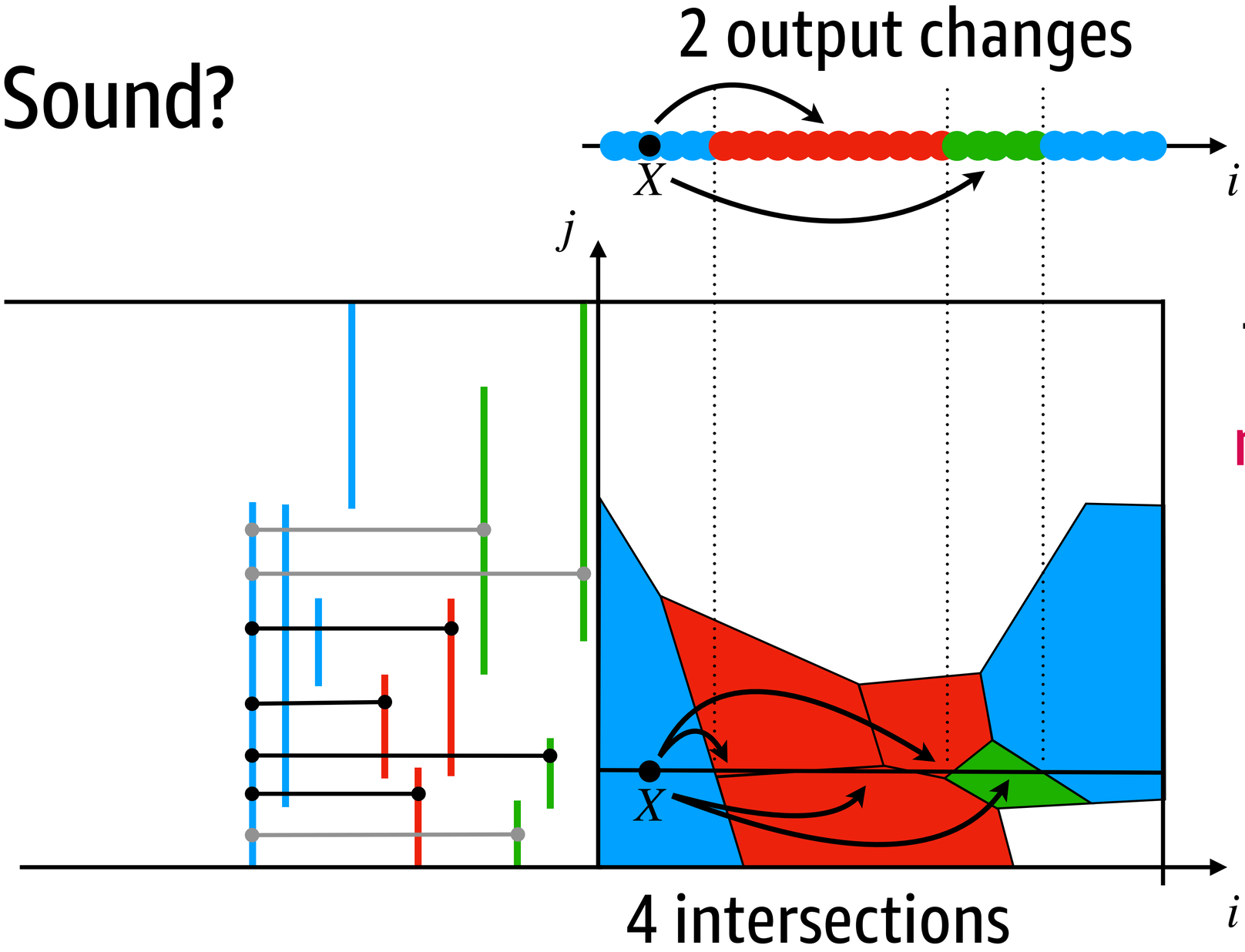
7 intersections

Sound?



7 intersections

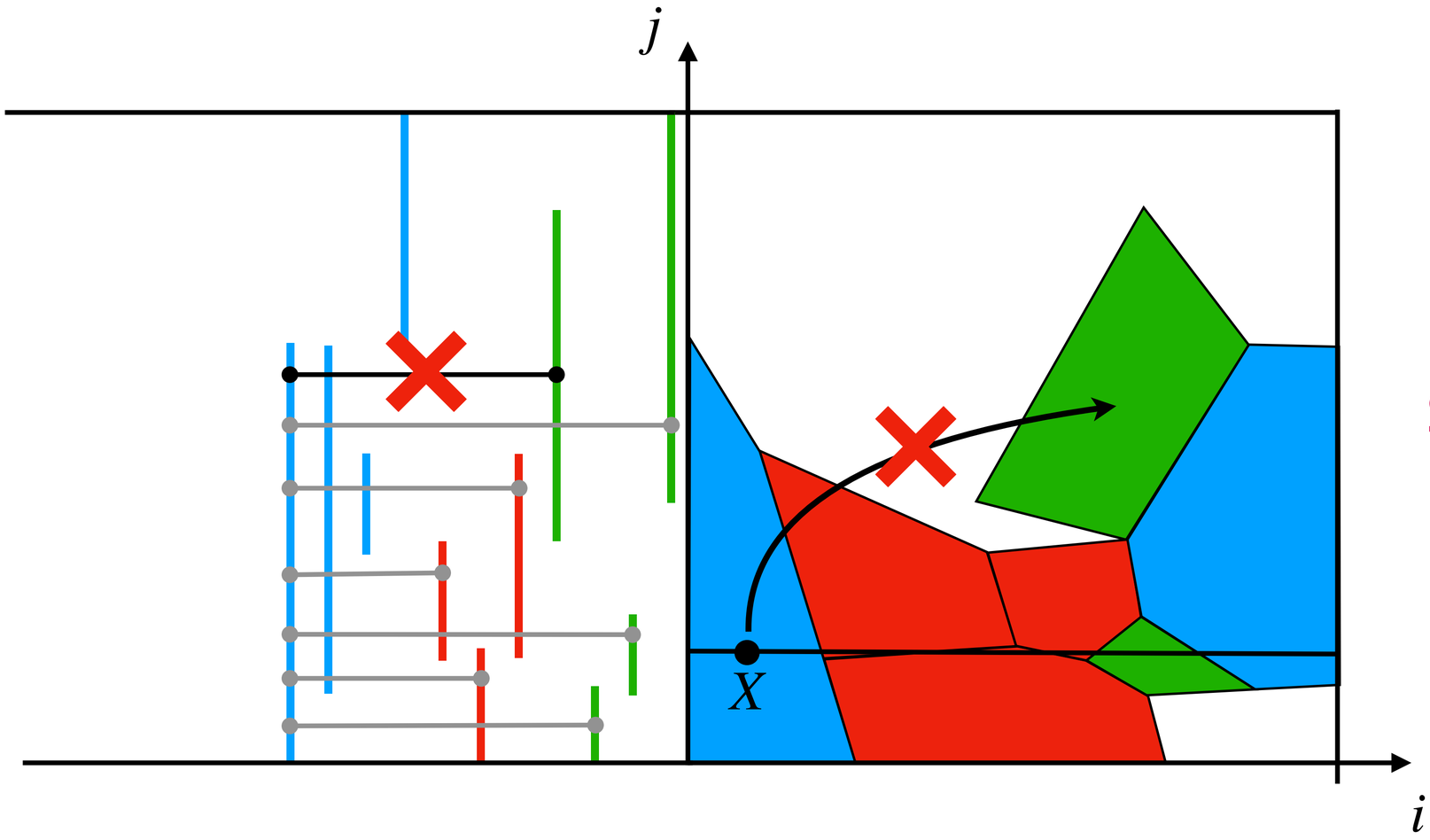
Sound?



The abstraction counts more changes than real

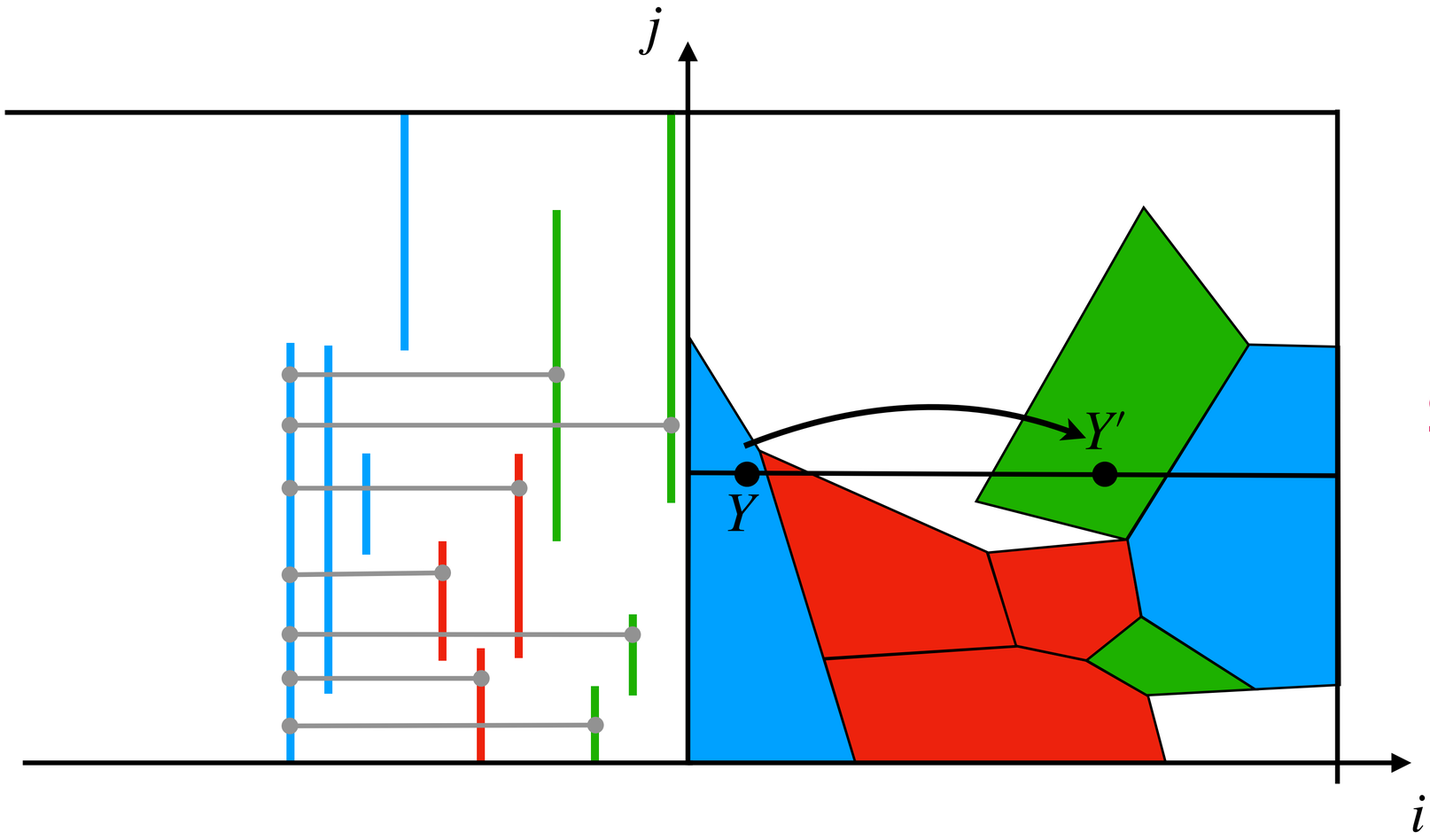
7 intersections

Sound?



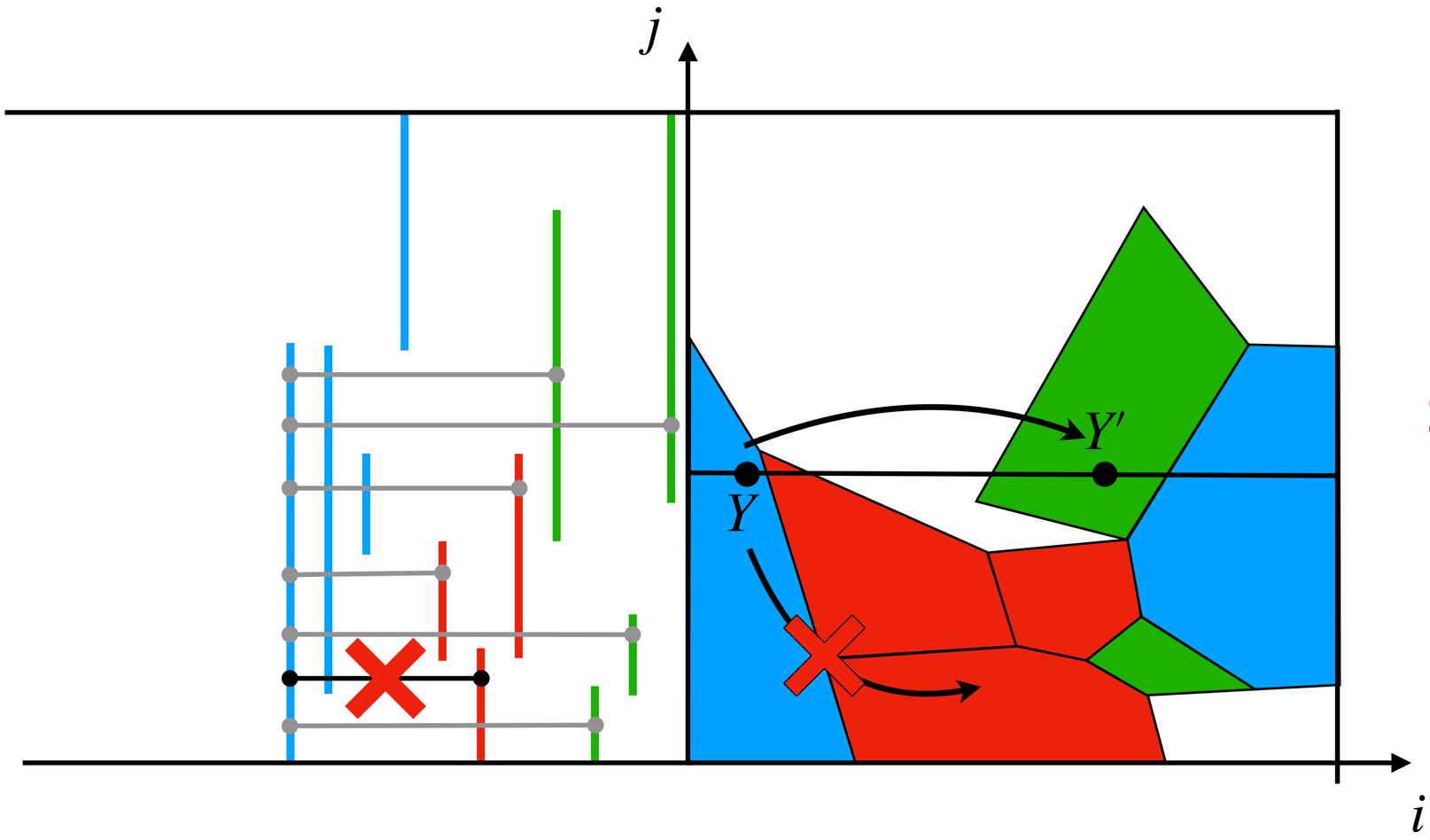
The abstraction adds spurious perturbations

Sound?



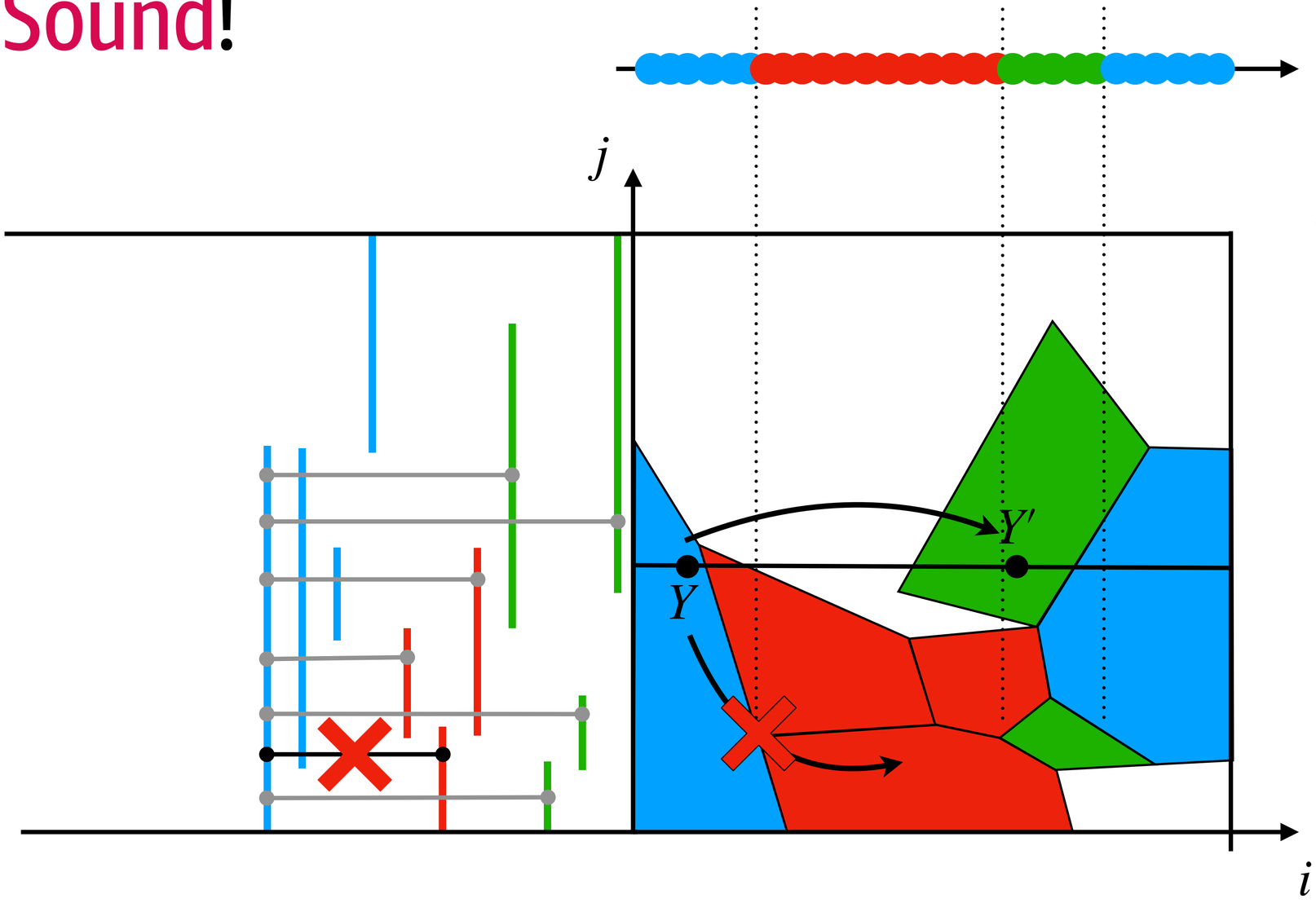
The abstraction adds
spurious perturbations

Sound?



The abstraction adds
spurious perturbations

Sound!



The abstraction adds
spurious perturbations
and
counts more changes
than real

↓

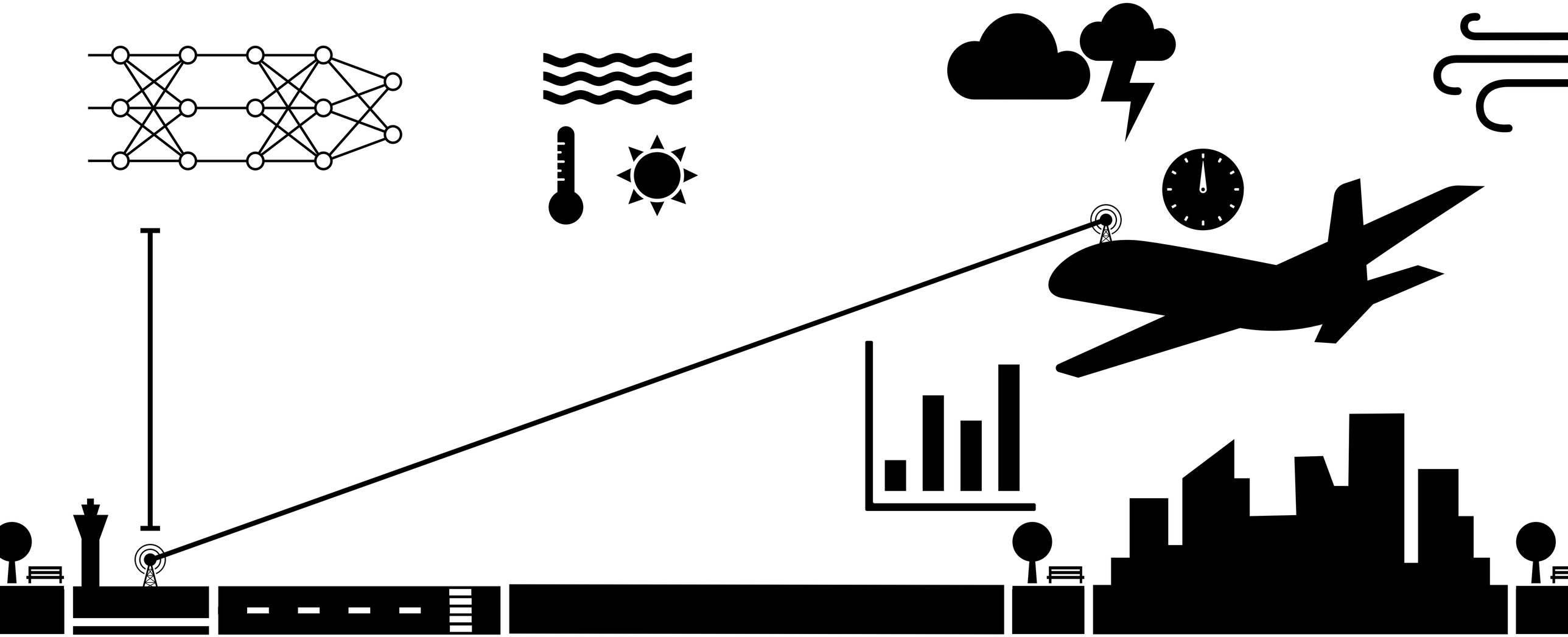
Sound!

Sound!

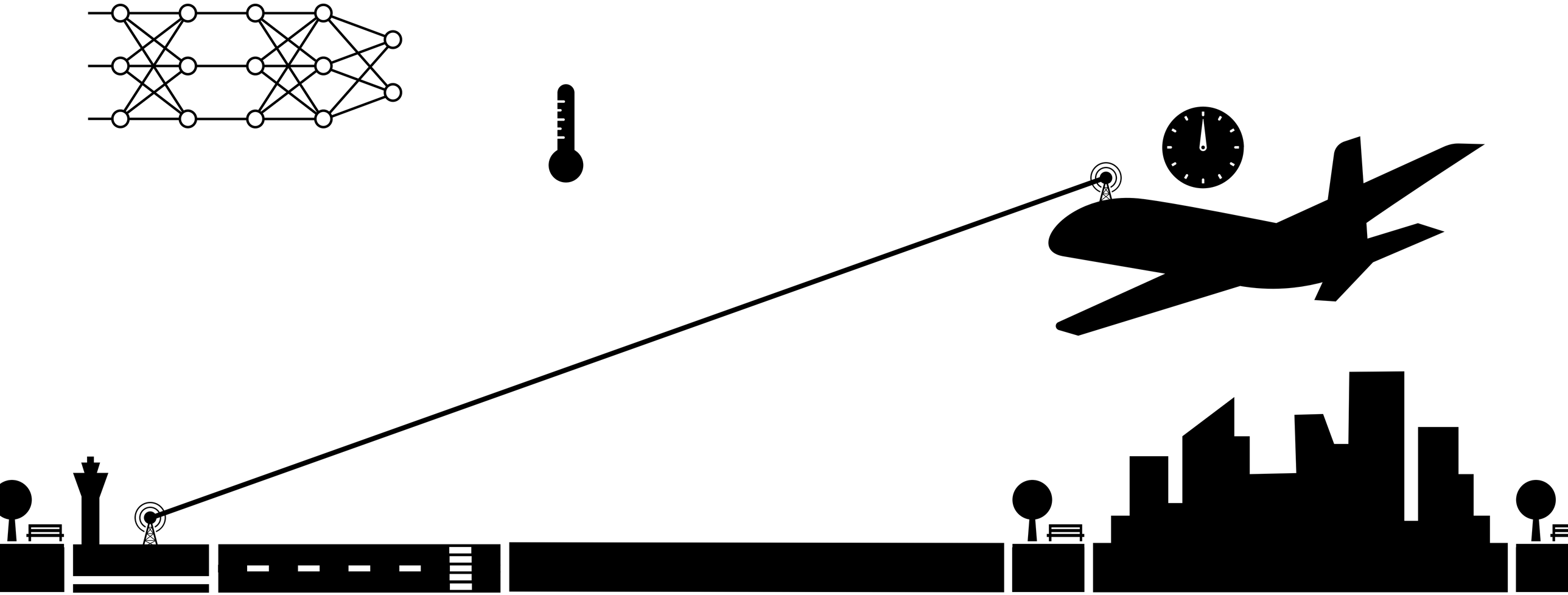
concrete abstract

$$\text{CountChanges}_i(P) \leq \text{ImpactAnalysis}_i^{\lightning}(P)$$

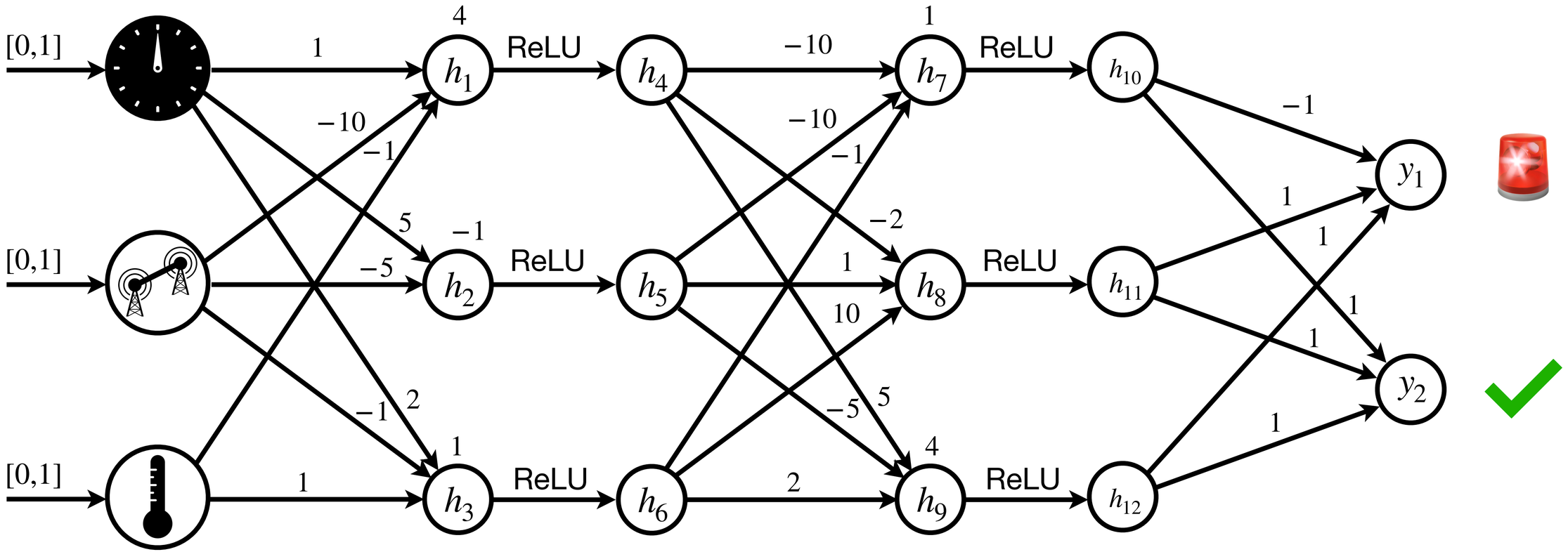
Landing alarm system



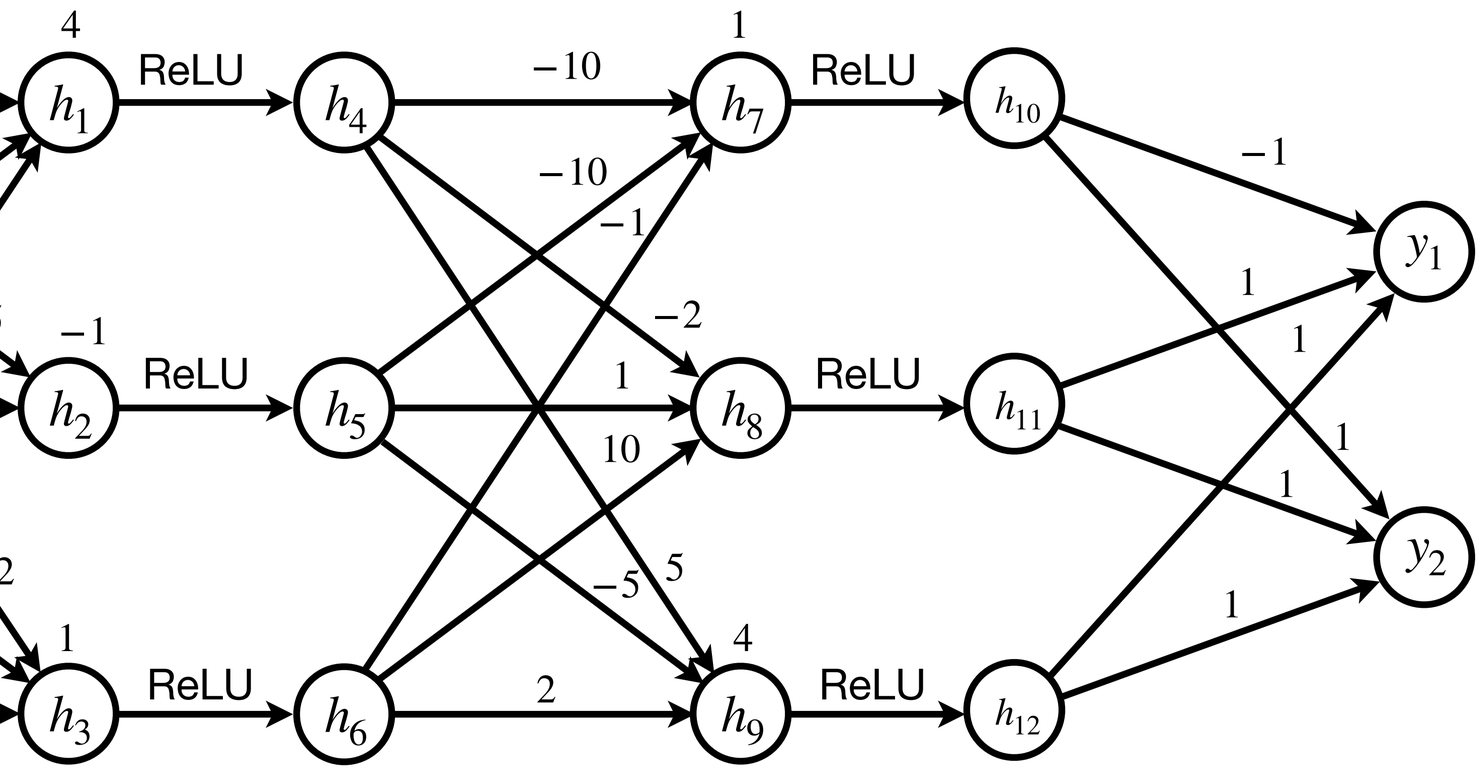
Simplified landing alarm system



Feed-forward neural network



Bucket abstraction



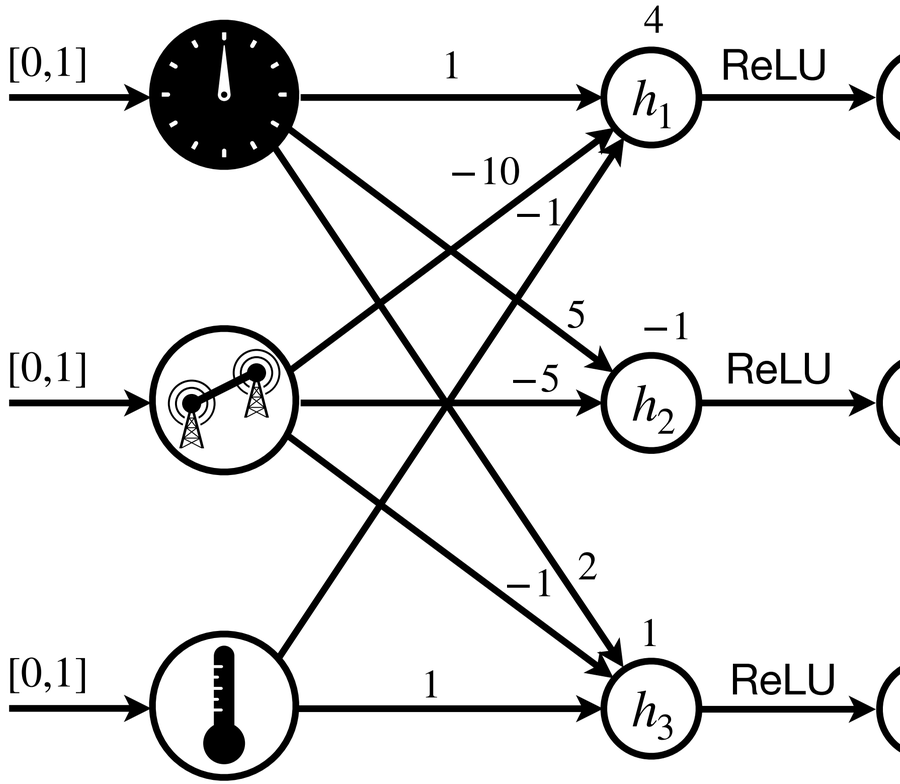
$y_2 \leq y_1$

$y_1 \leq y_2$

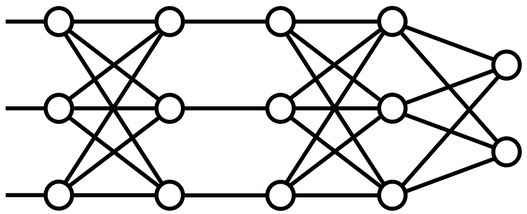
$$\text{ImpactAnalysis}_{\text{clock}}(P) = 75$$

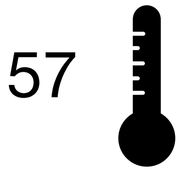
$$\text{ImpactAnalysis}_{\text{antenna}}(P) = 83$$

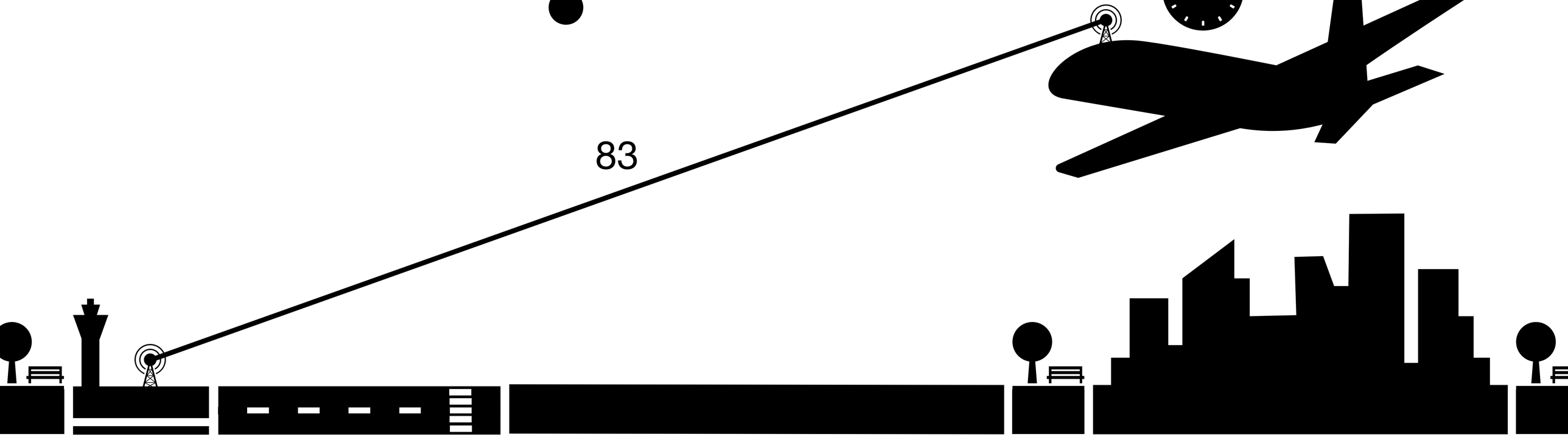
$$\text{ImpactAnalysis}_{\text{thermometer}}(P) = 57$$



Simplified landing alarm system







Experiments setup

- Diabetes
- Wine quality
- RPG Videogame
- Rain Sidney

- 4 Databases

Experiments setup

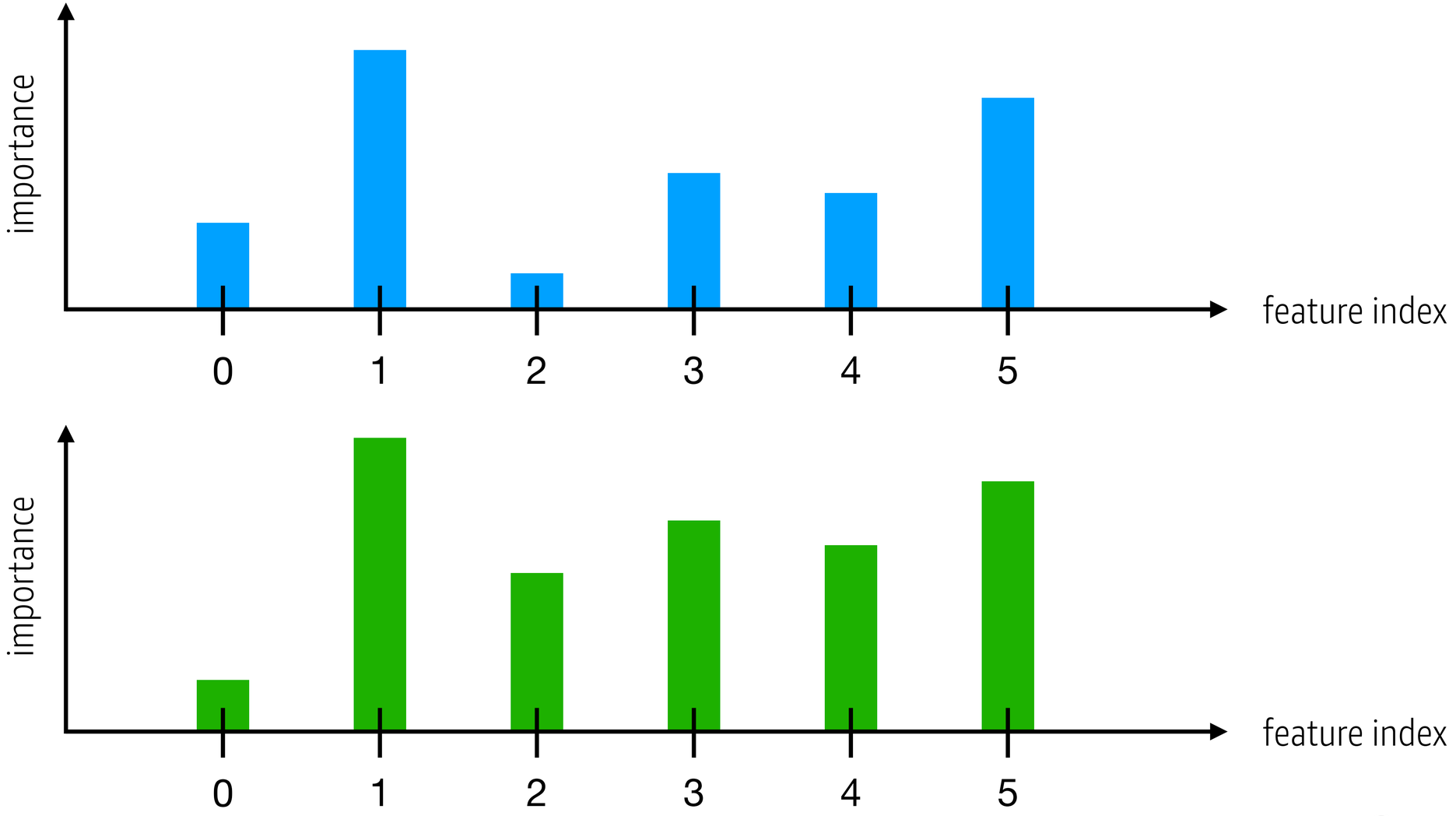
- Diabetes
- Wine quality
- RPG Videogame
- Rain Sidney

→ 4 Databases

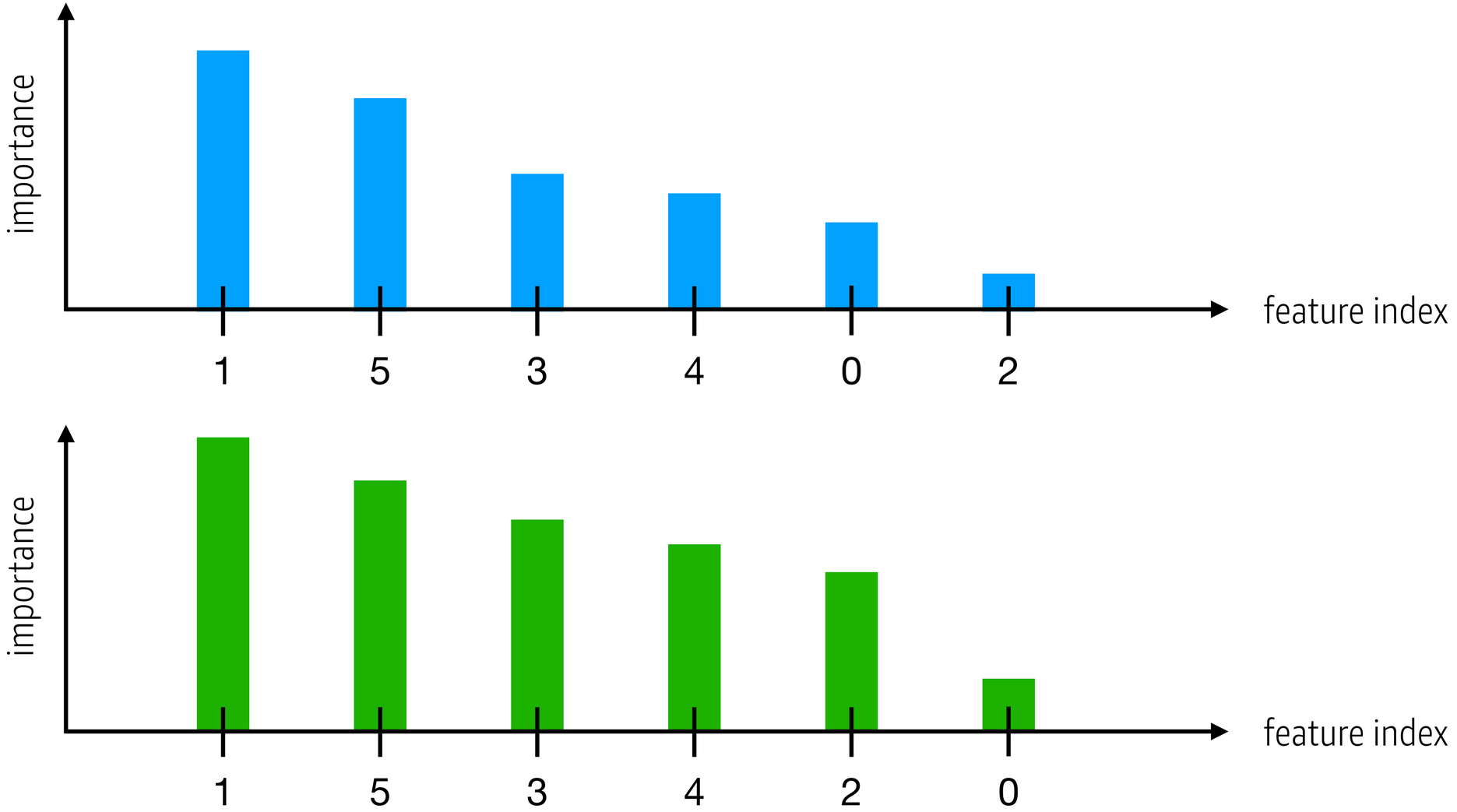
Common prefix between ordering of input features

- Baseline vs Permutation Feature Importance
- Baseline vs Retraining
- Baseline vs ImpactAnalysis

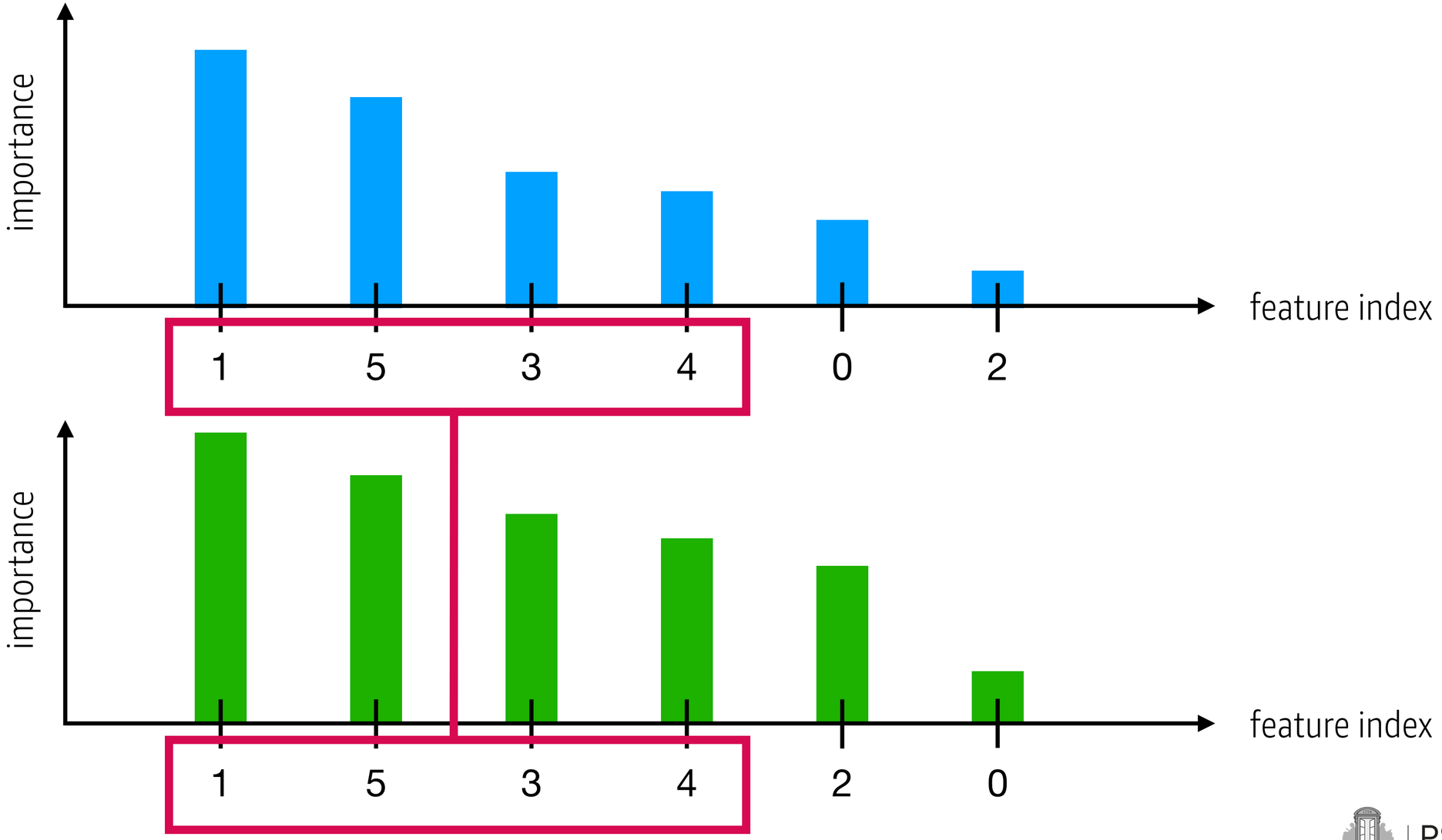
Length of the common prefix



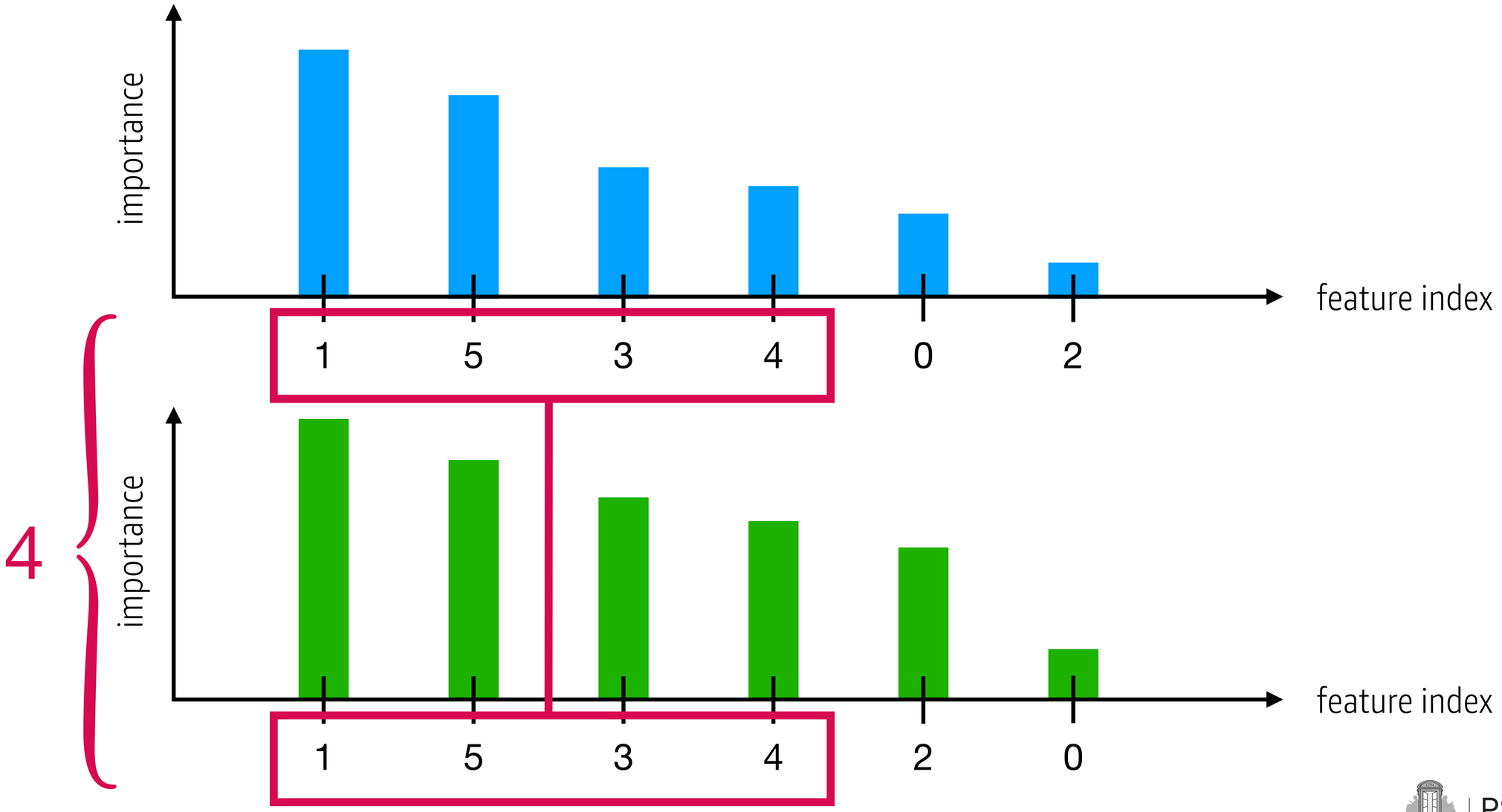
Length of the common prefix



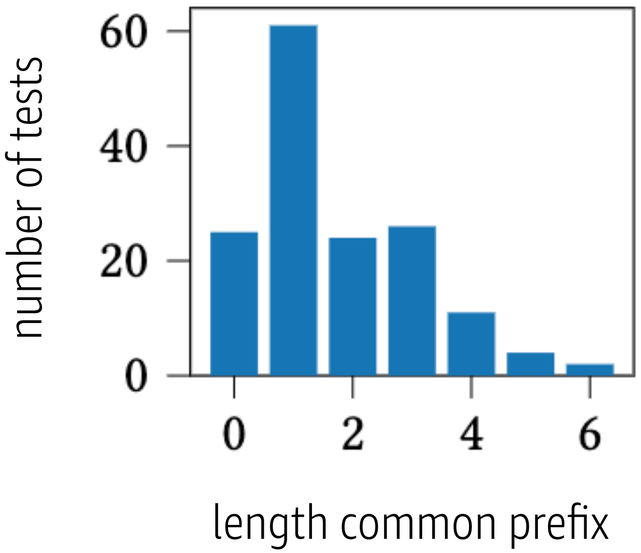
Length of the common prefix

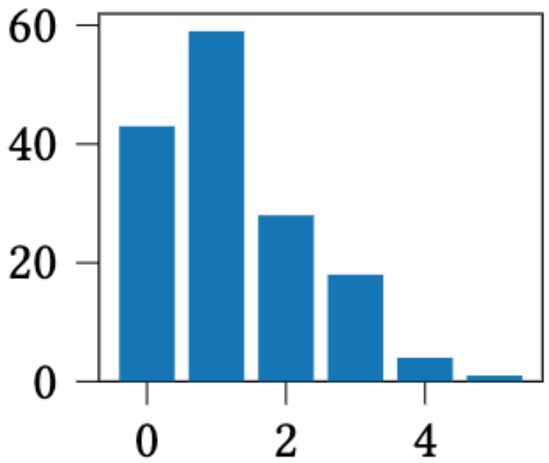


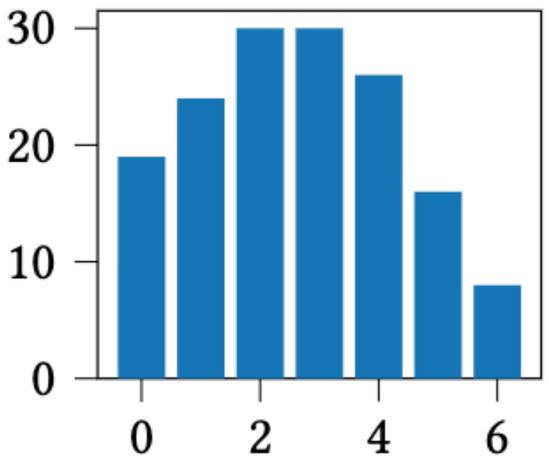
Length of the common prefix



Baseline vs **PFI** **Retraining** **ImpactAnalysis**

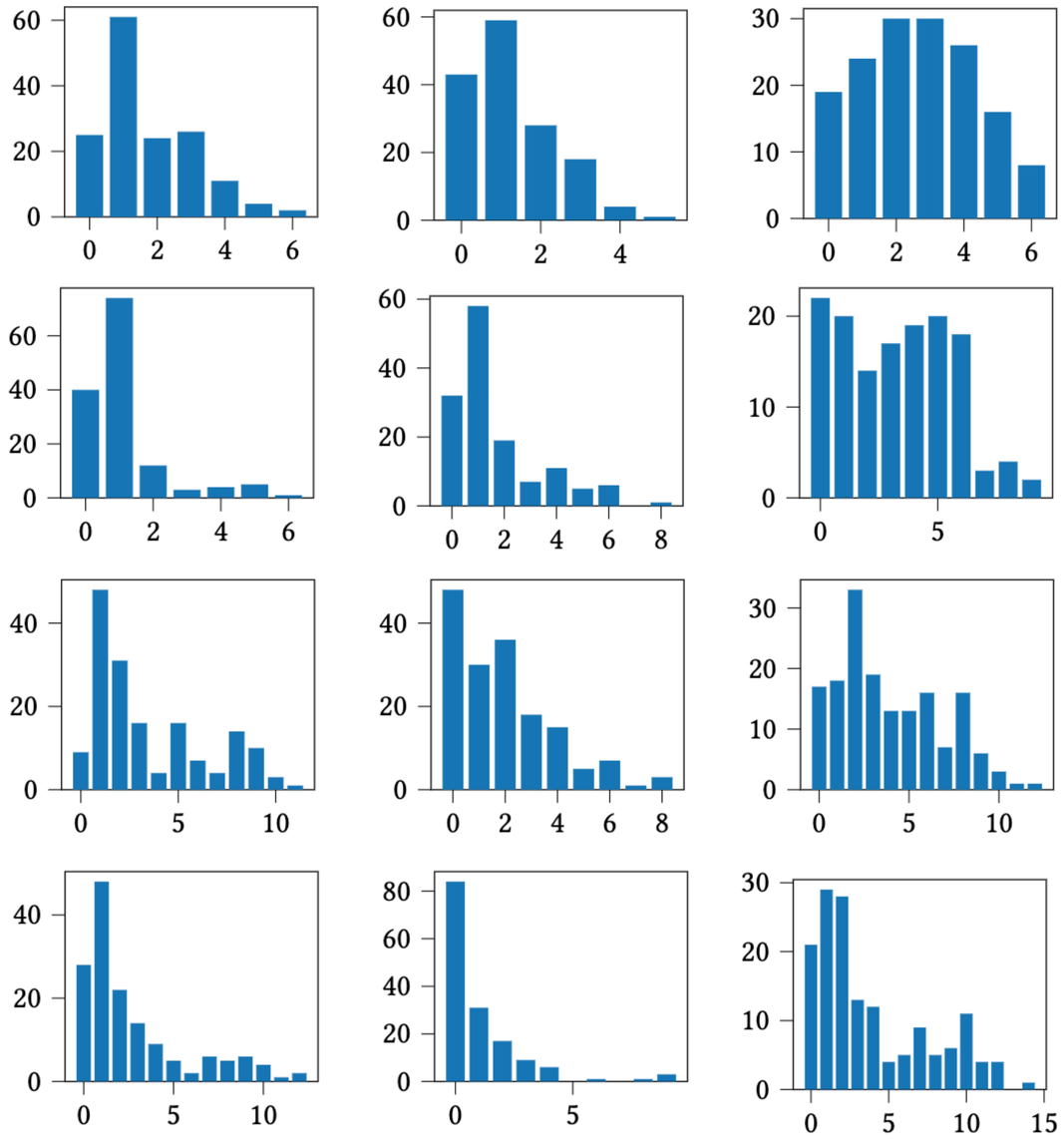




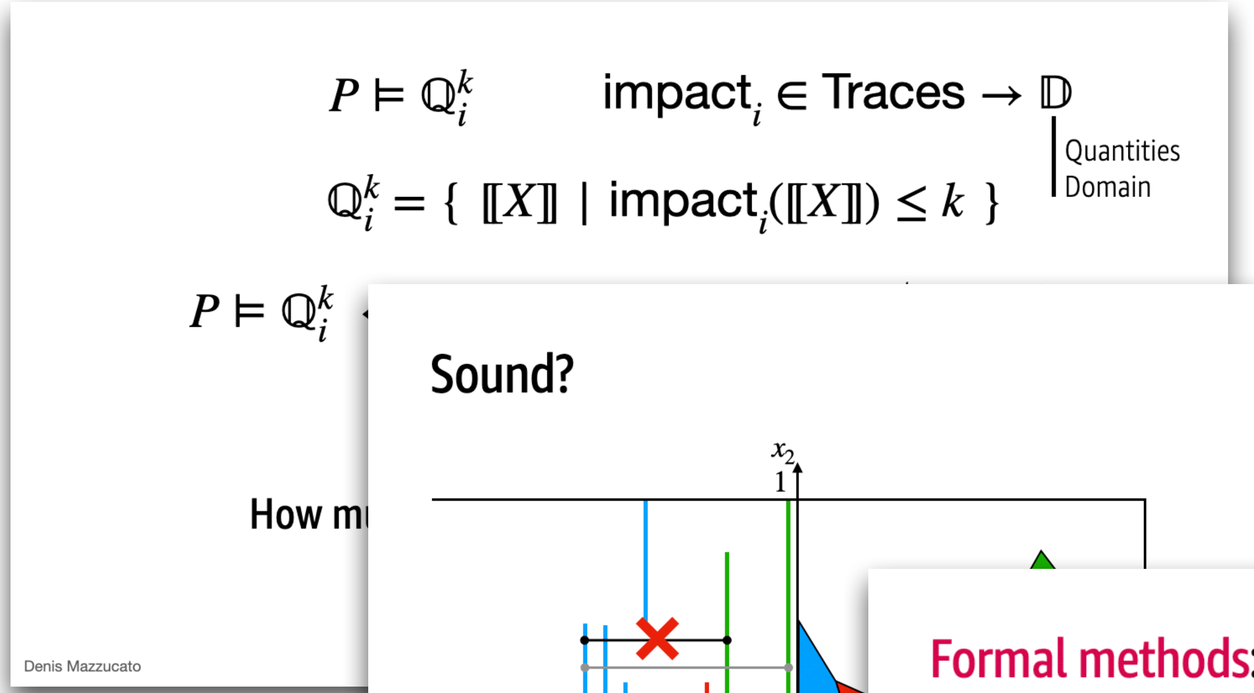


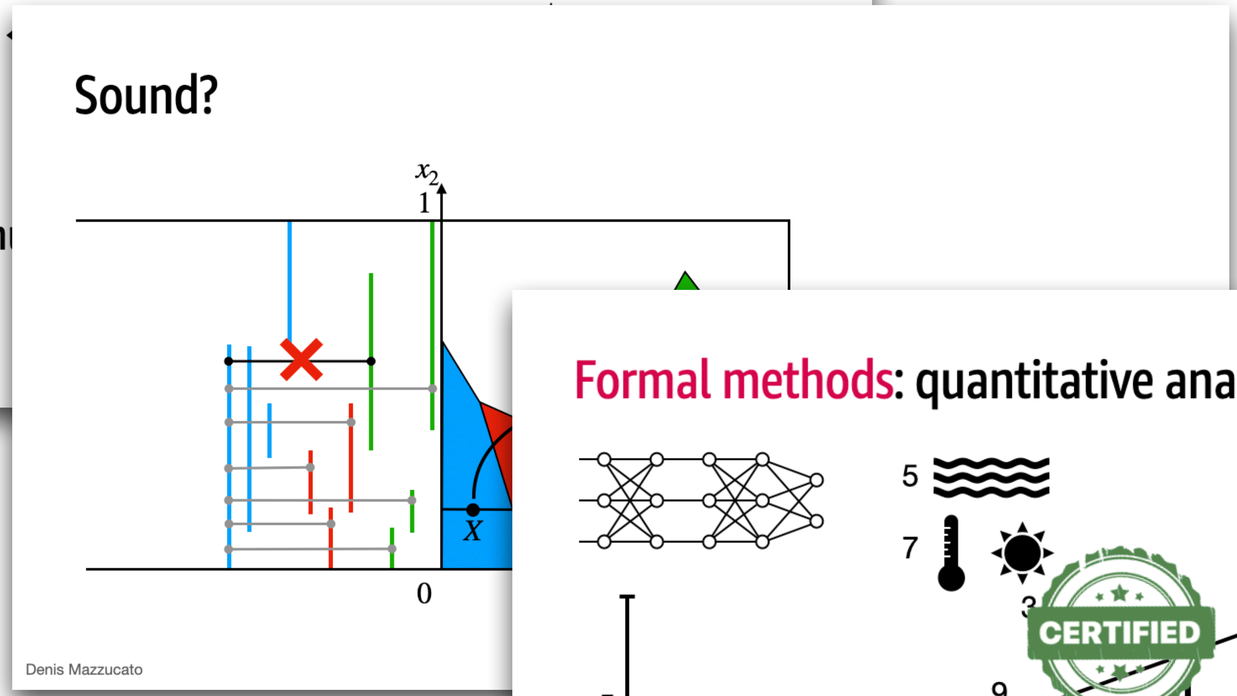
Diabetes

Baseline vs **PFI** **Retraining** **ImpactAnalysis**



Conclusion



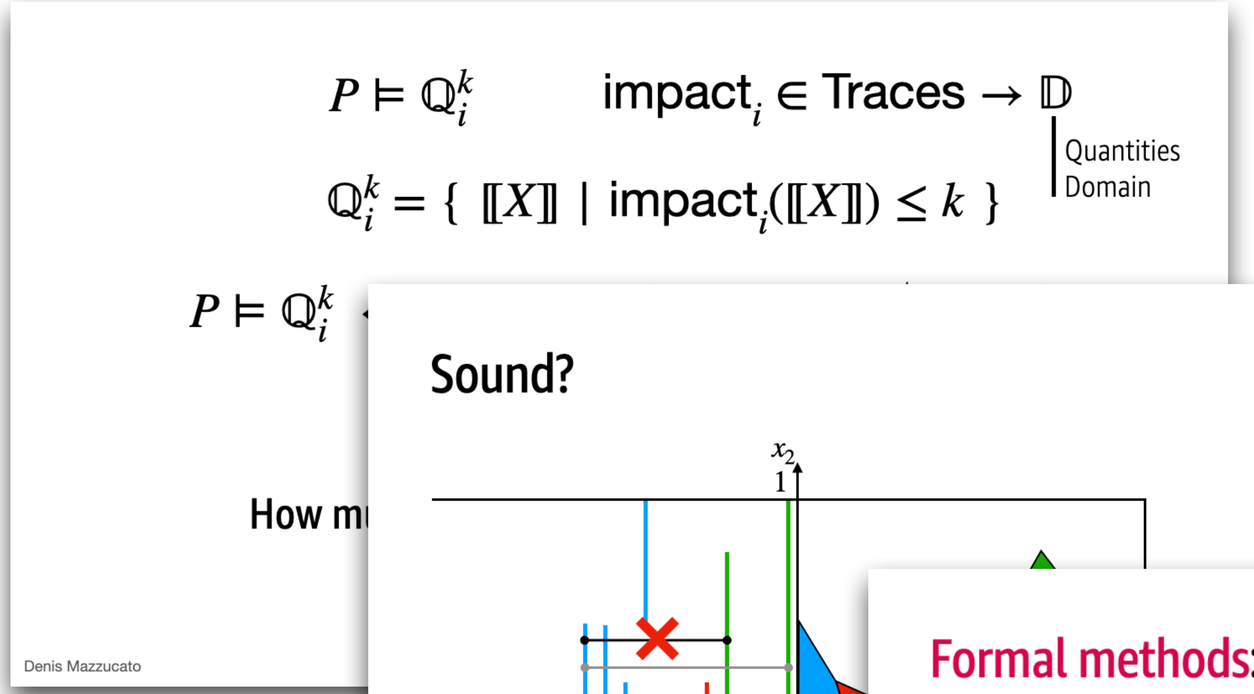


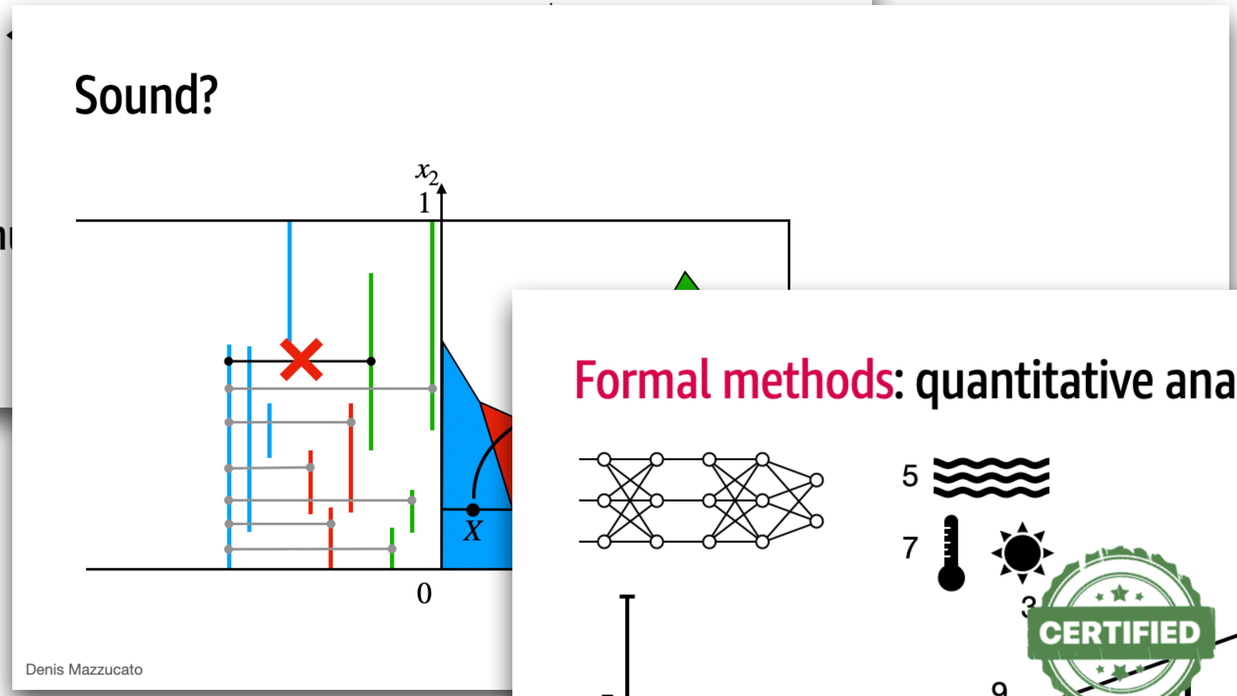
- Impact formal framework
- Concrete CountChanges_i
- Sound abstraction ImpactAnalysis_i

Conclusion

Sound by construction!

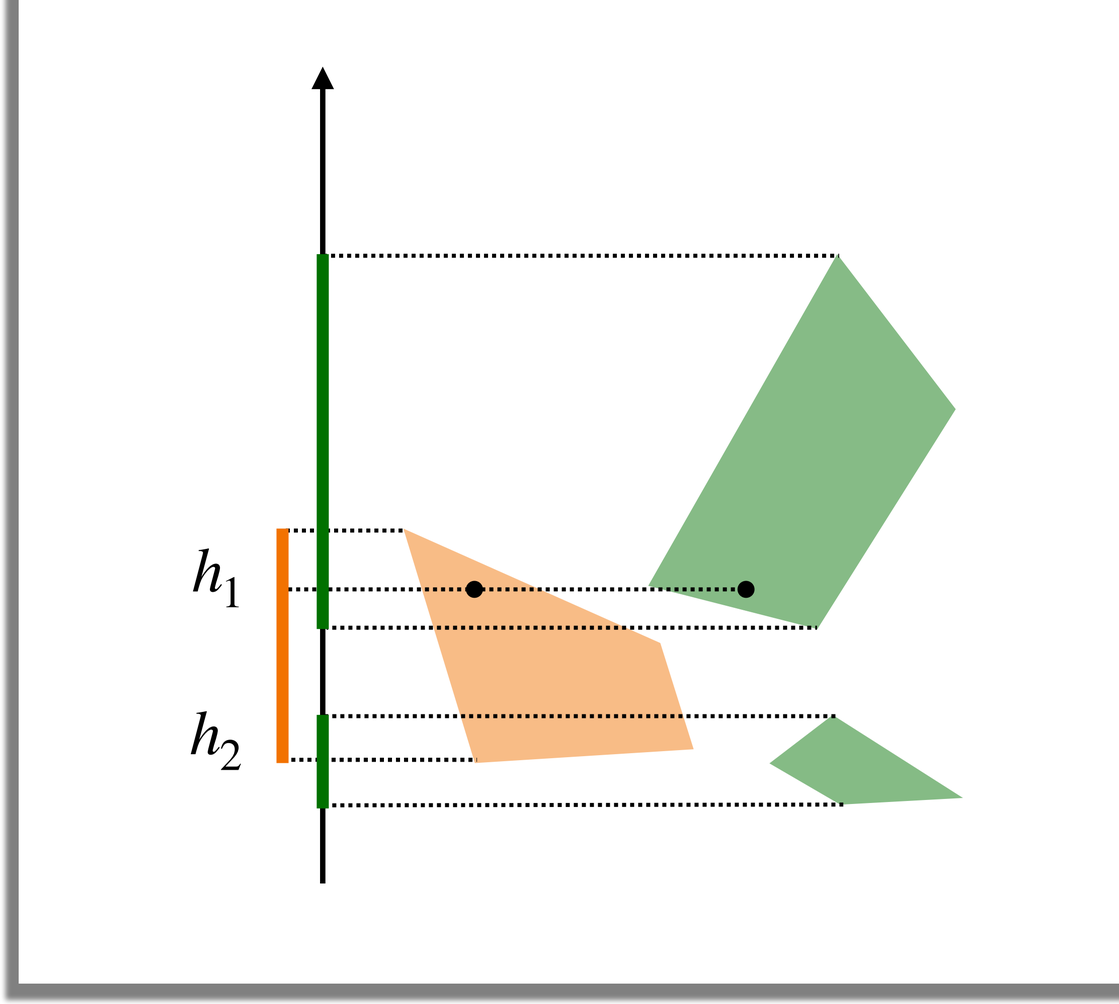
- Impact formal framework
- Concrete CountChanges_i
- Sound abstraction ImpactAnalysis_i





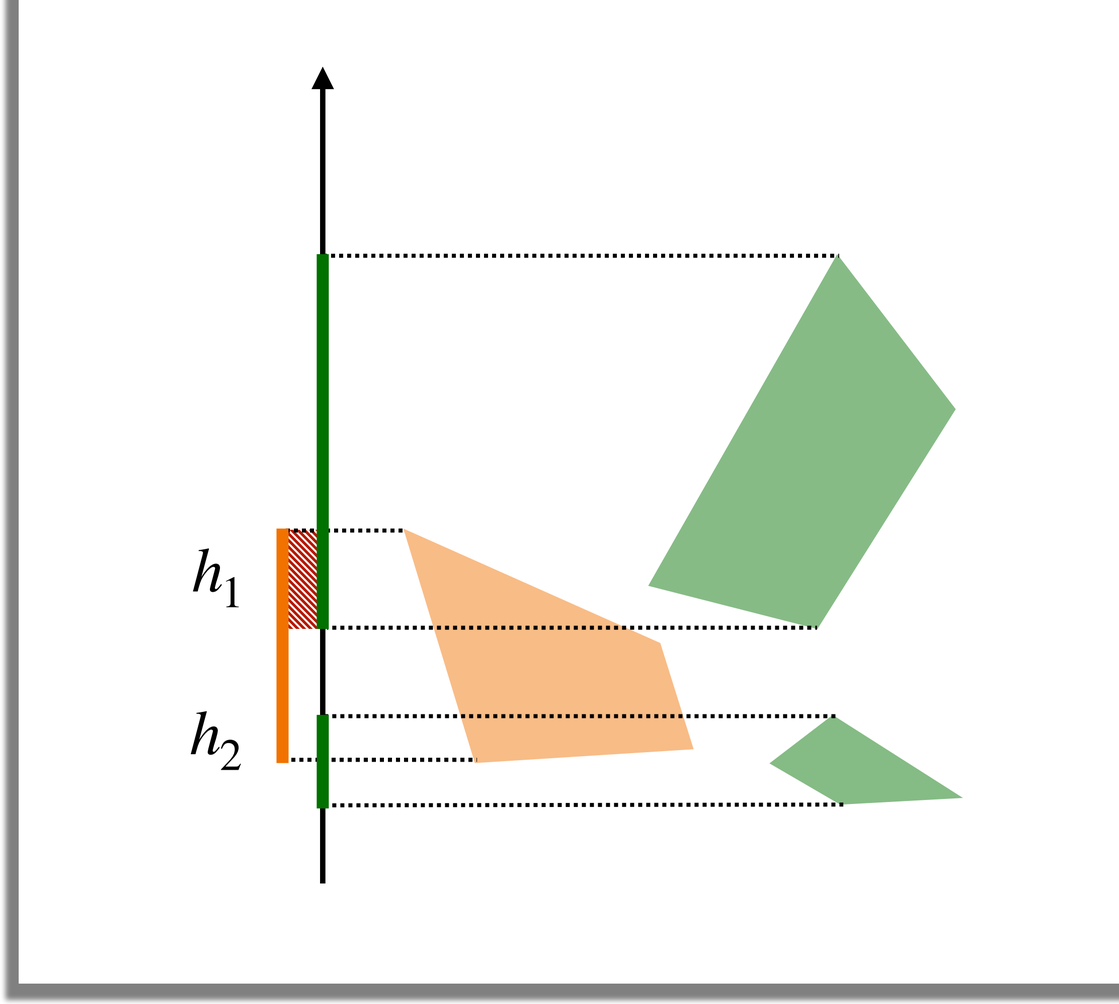
Future work

- Other impact instances
  - Checking Intersections



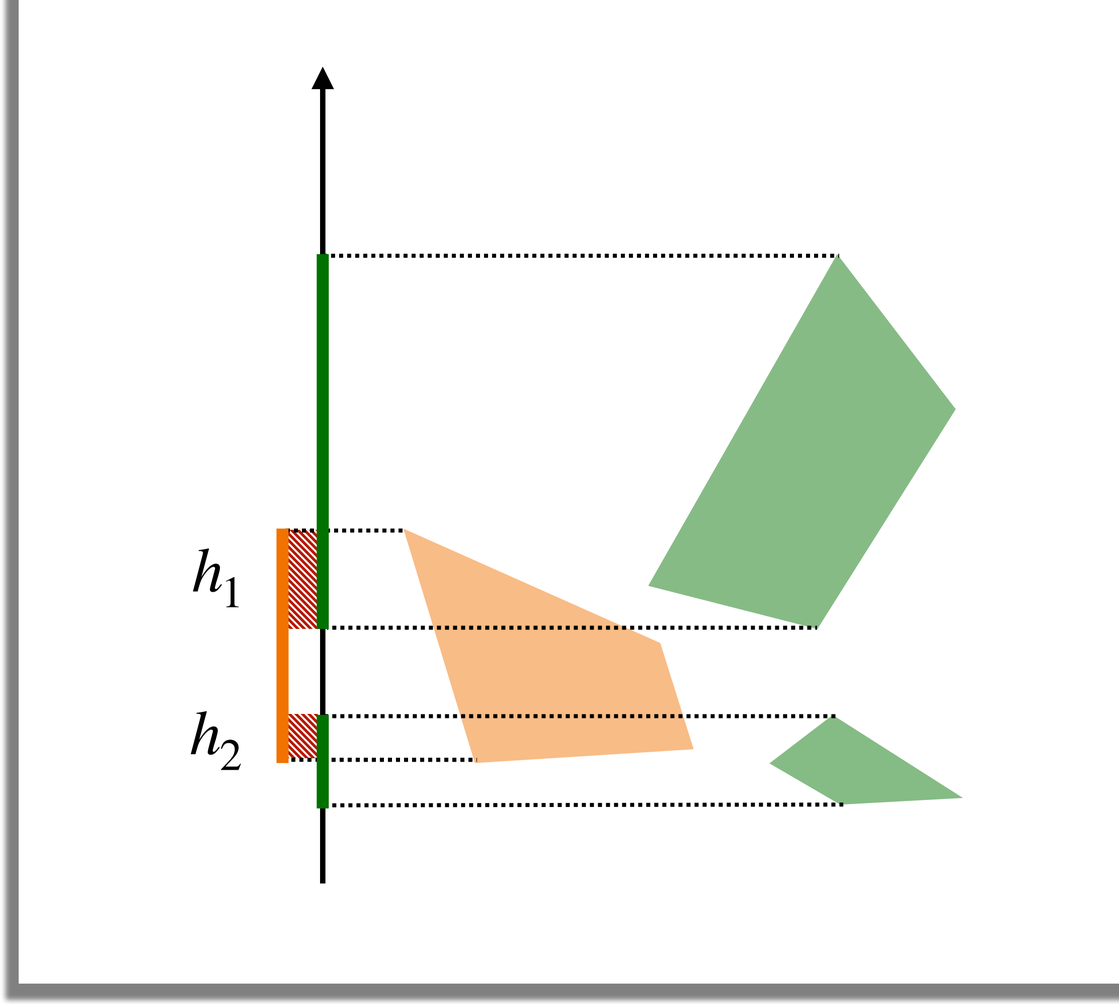
Future work

- Other impact instances
  - Checking Intersections
  - Maximum Volume



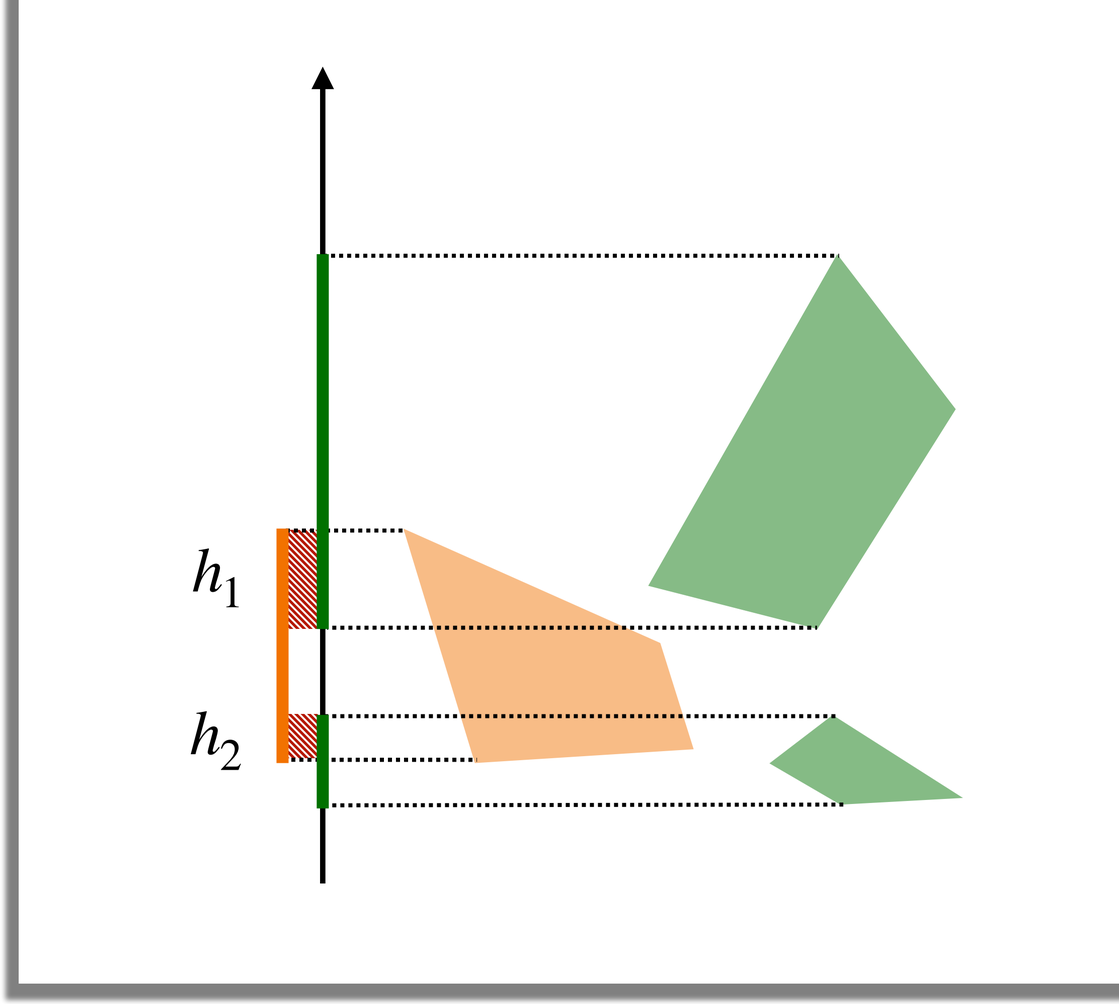
Future work

- Other impact instances
  - Checking Intersections
  - Maximum Volume
  - Disjoint Sum of Volumes



Future work

- Other impact instances
  - Checking Intersections
  - Maximum Volume
  - Disjoint Sum of Volumes
- Other application contexts
  - Not only neural networks
  - Jupyter notebooks



Future work

- Other impact instances
  - Checking Intersections
  - Maximum Volume
  - Disjoint Sum of Volumes
- Other application contexts
  - Not only neural networks
  - Jupyter notebooks

Our work

- Impact formal framework
- Concrete CountChanges.
- Abstract ImpactAnalysis.

Sound by construction!