On Continuous Monitoring of Risk Violations under Unknown Shift

Rajeev Verma / University of Amsterdam

UvA - BOSCH DELTA LAB

AMLAB Amsterdam Machine Learning Lab

On Continuous Monitoring of Risk Violations under Unknown Shift

Alexander Timans

Rajeev Verma / University of Amsterdam

UvA - BOSCH DELTA LAB

AMLAB Amsterdam Machine Learning Lab

The role of statistical inference in machine learning

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

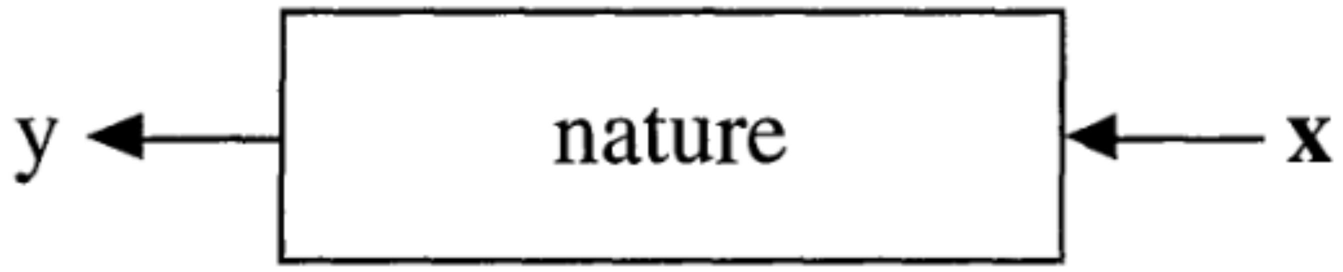
Abstract. There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.

The role of statistical inference in machine learning

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

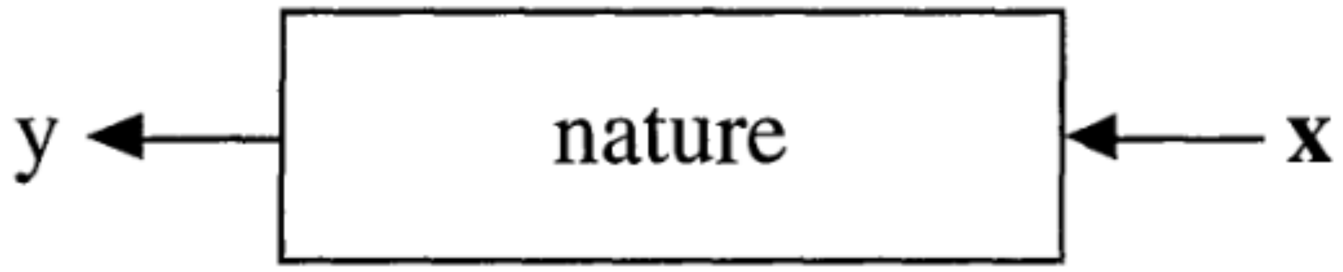


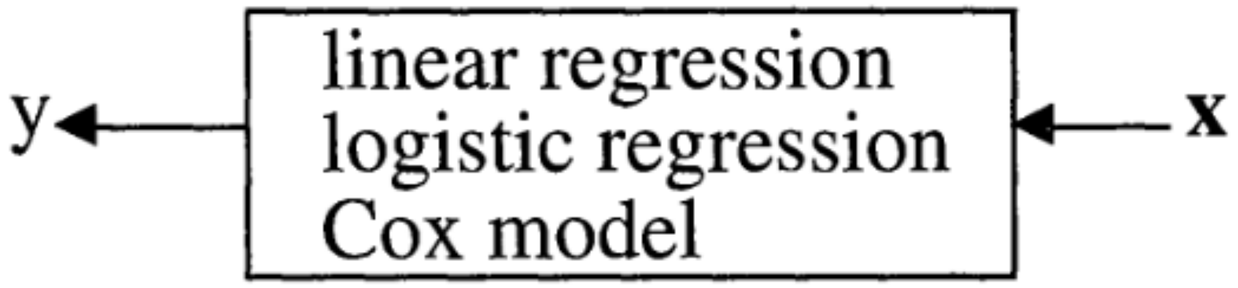
The role of statistical inference in machine learning

Statistical Science
2001, Vol. 16, No. 3, 199–231

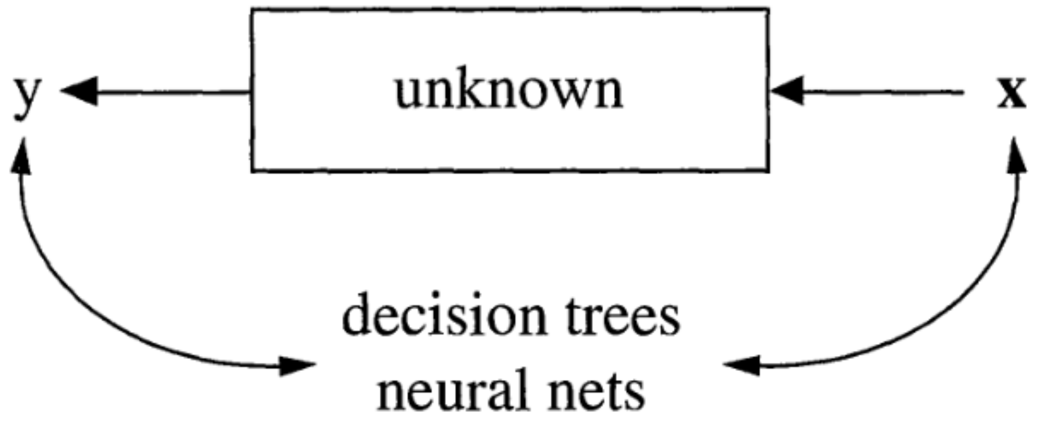
Statistical Modeling: The Two Cultures

Leo Breiman



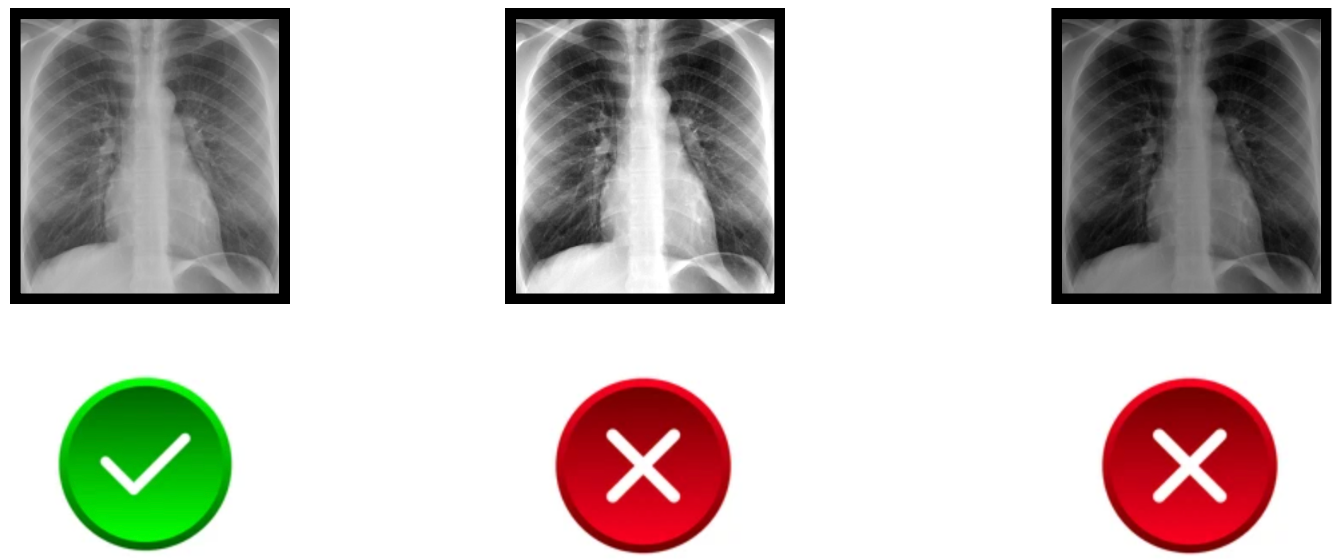


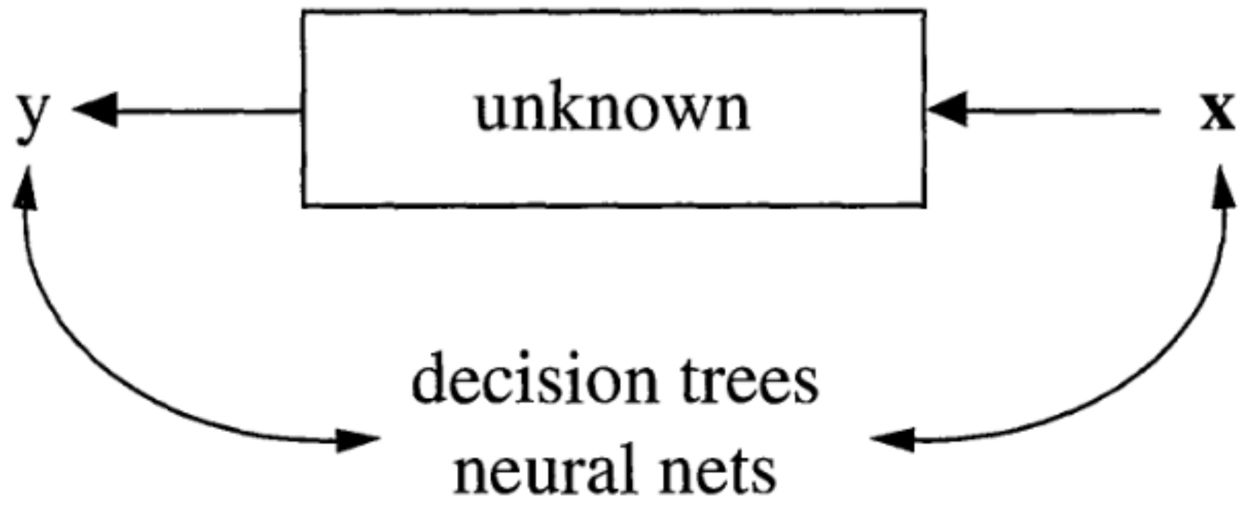
The data-modelling culture



The algorithmic modelling culture

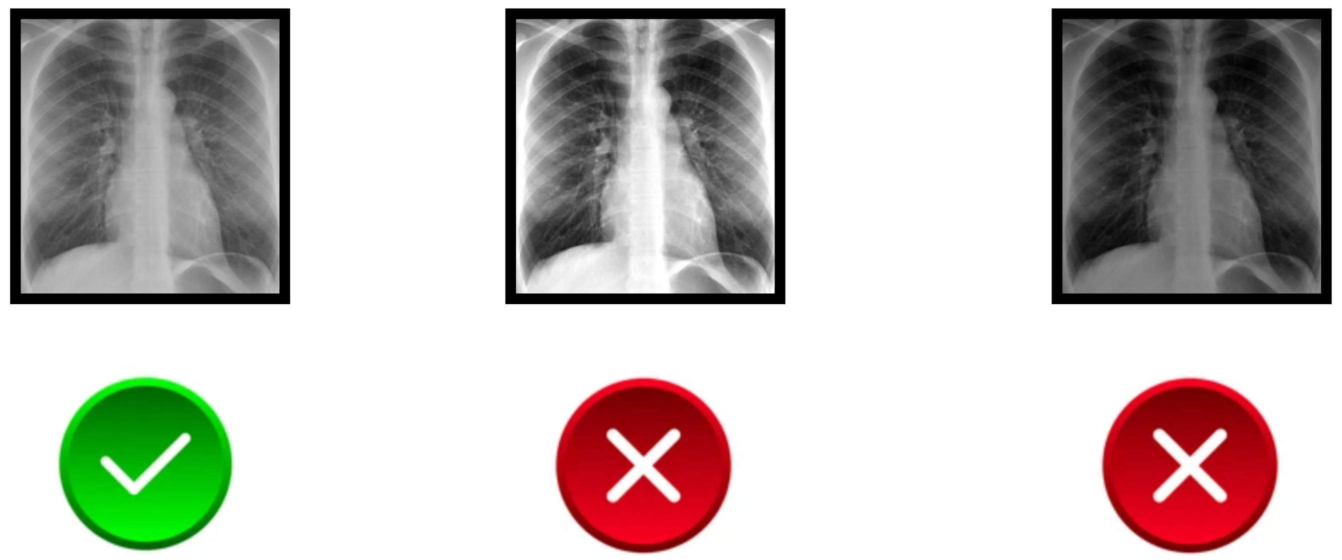
The role of statistical inference in machine learning

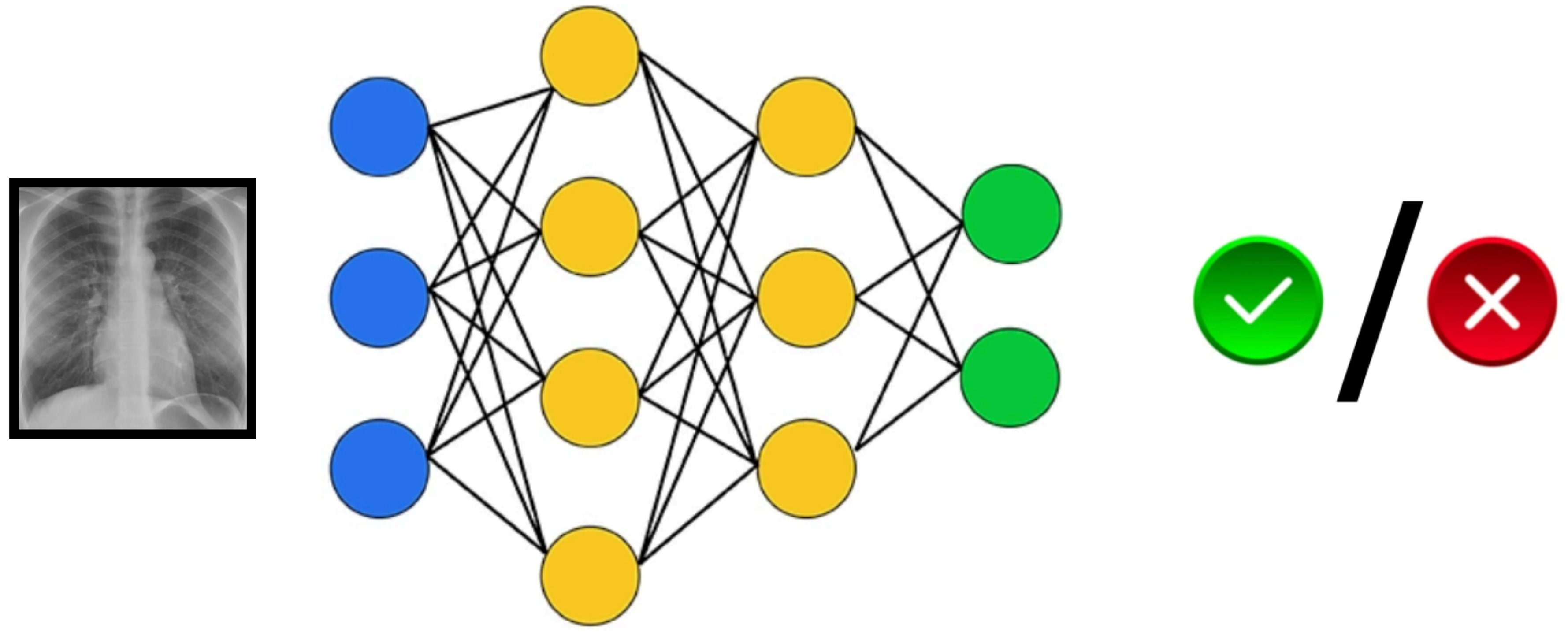




The algorithmic modeling culture

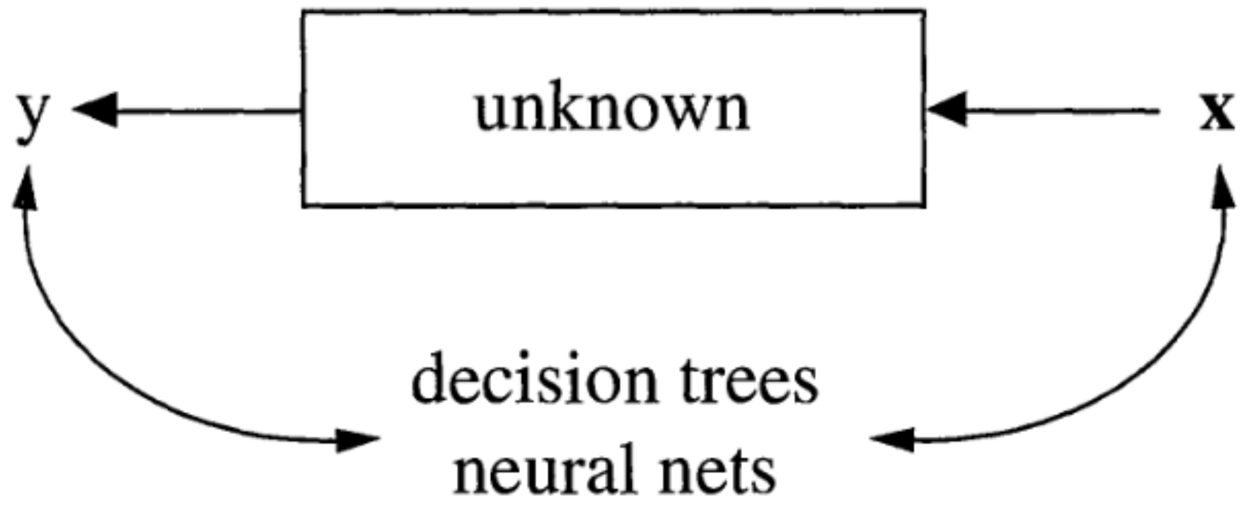
Machine Learning



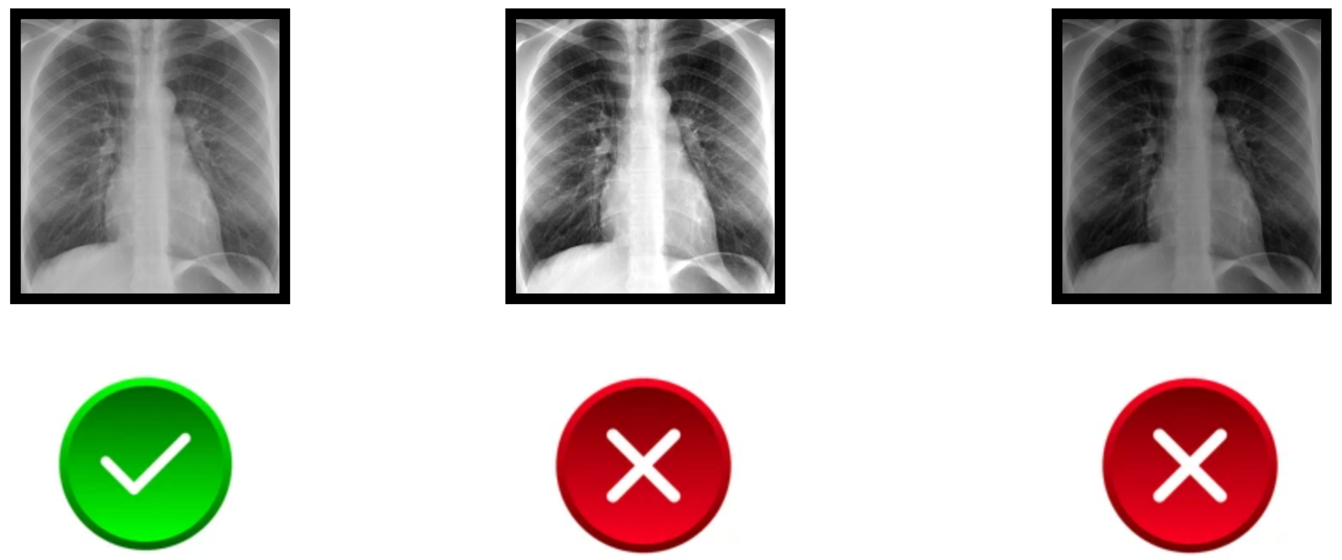


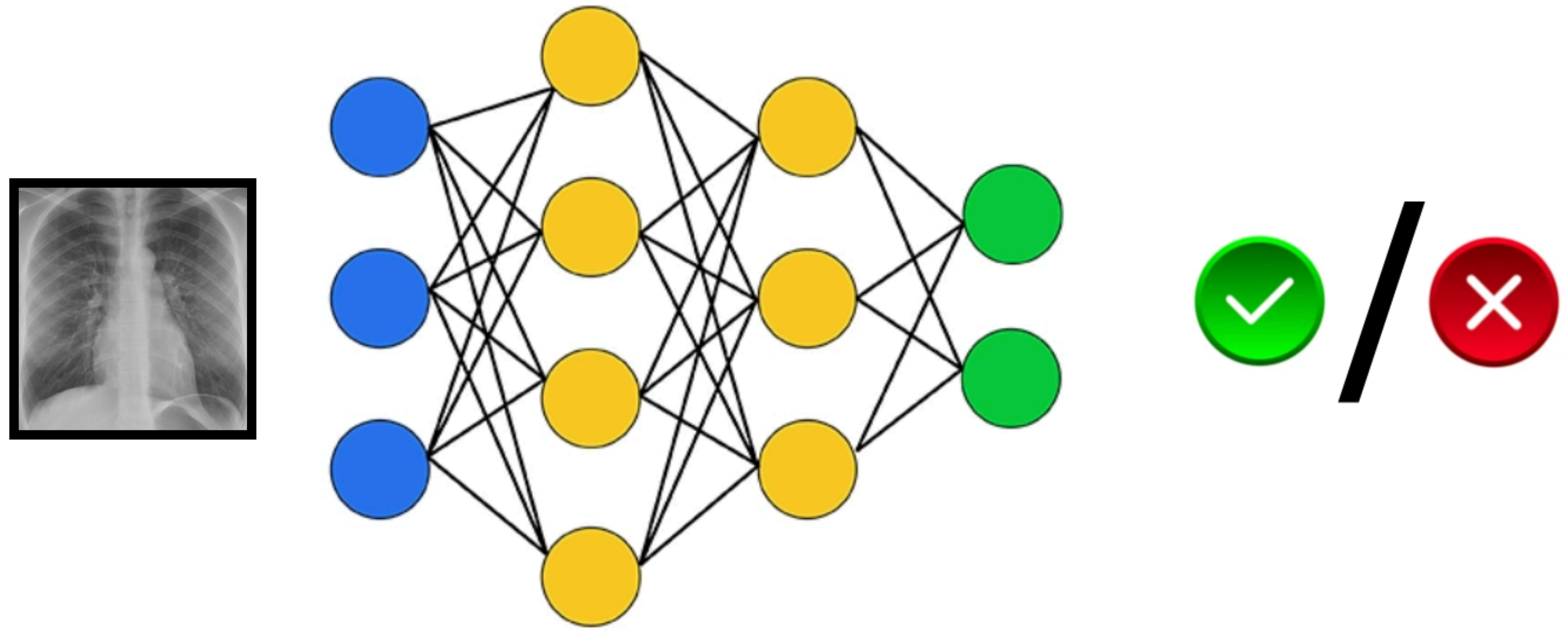
Model validation. Measured by predictive accuracy.

Machine Learning

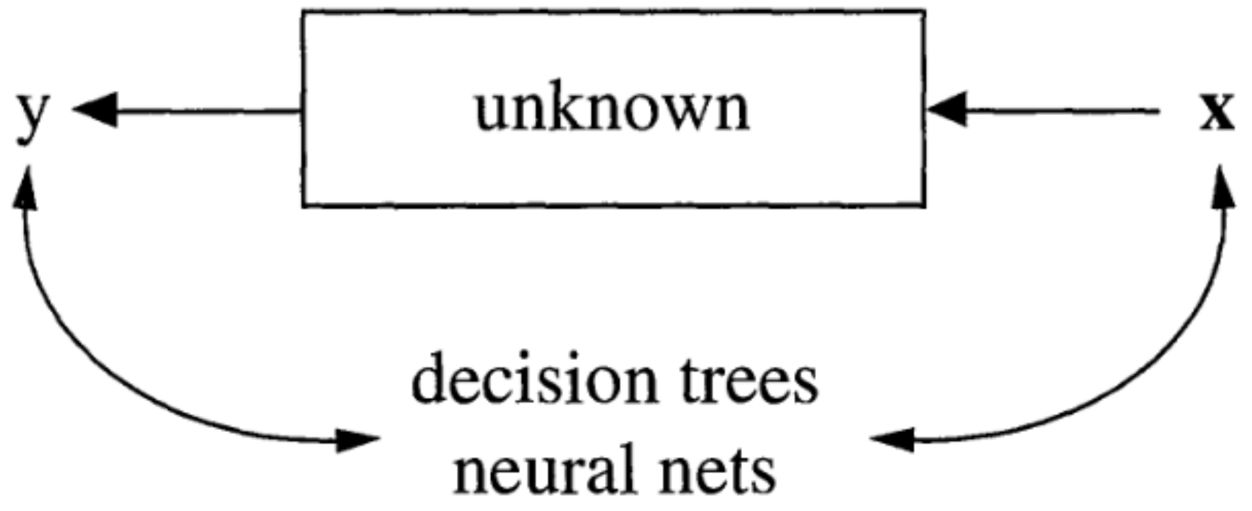


The algorithmic modeling culture



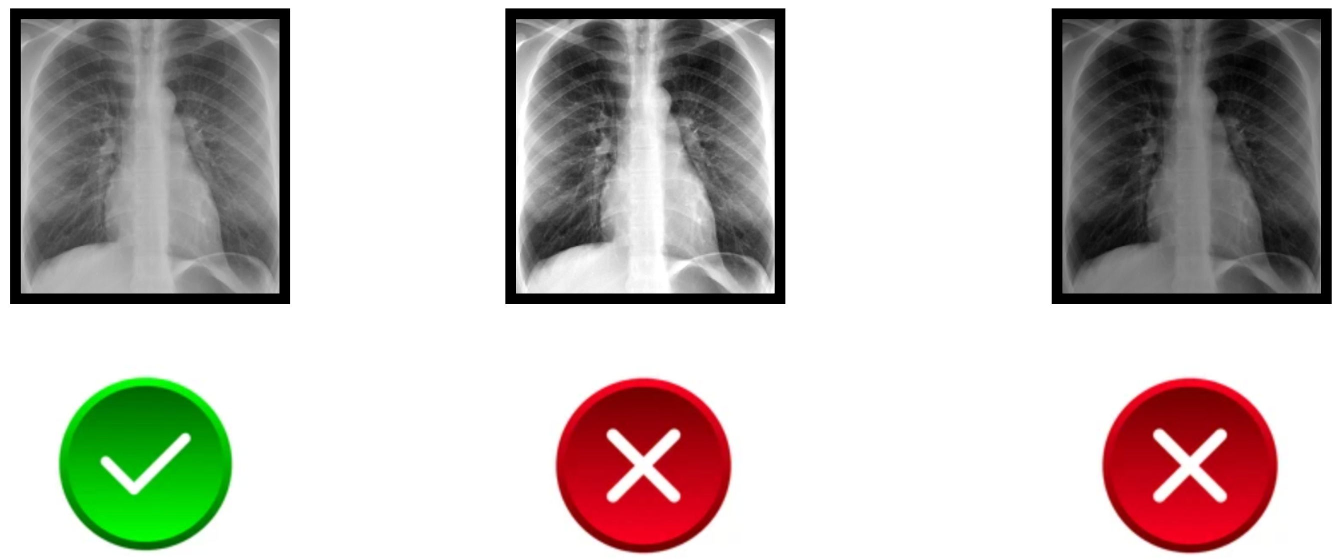


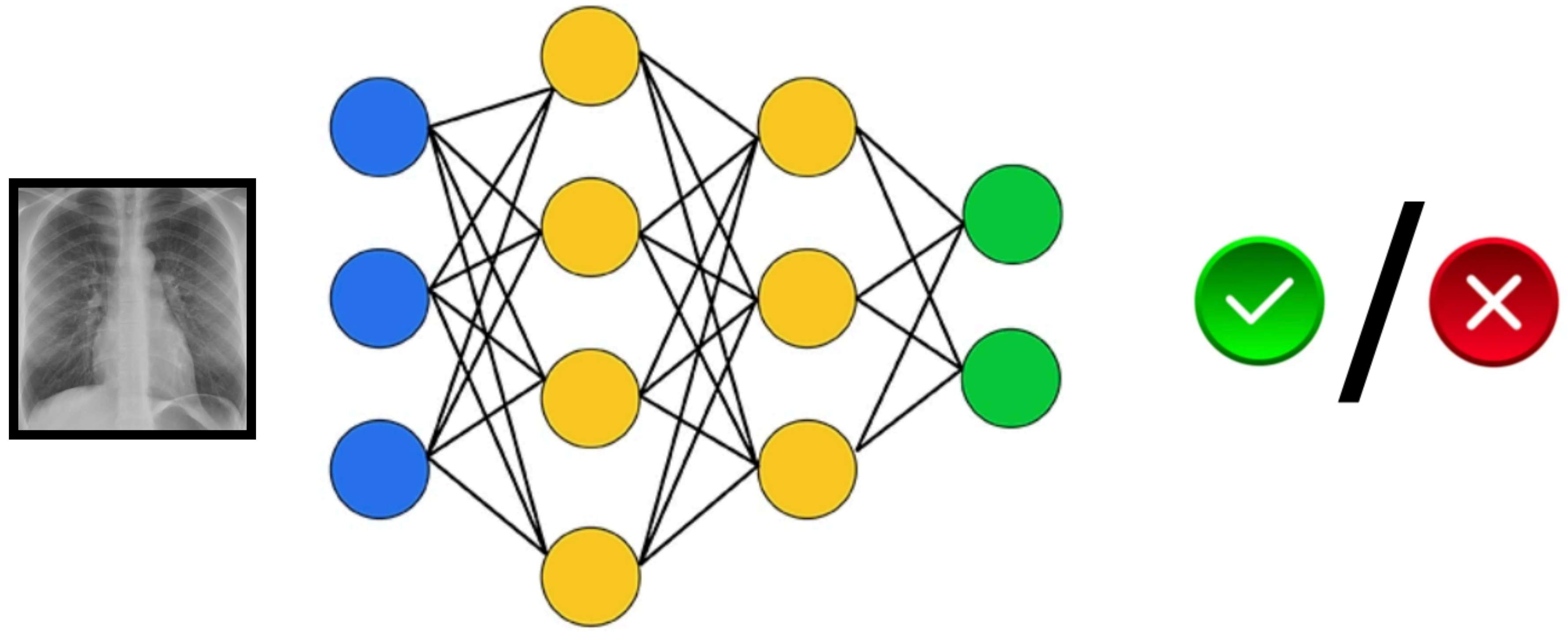
Model validation. Measured by predictive accuracy.

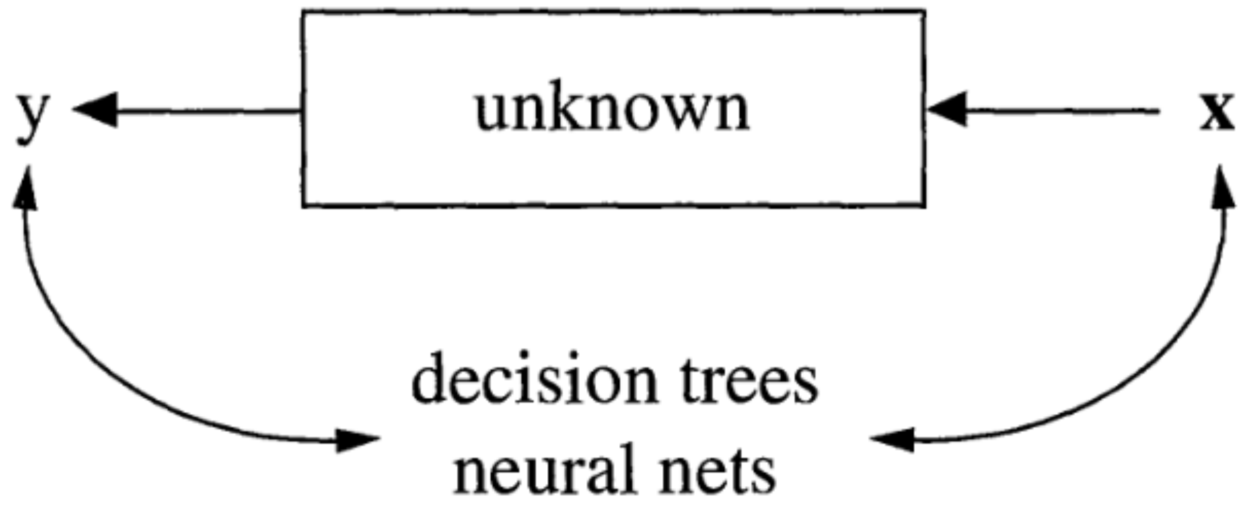


The algorithmic modeling culture

Machine Learning: is this enough?



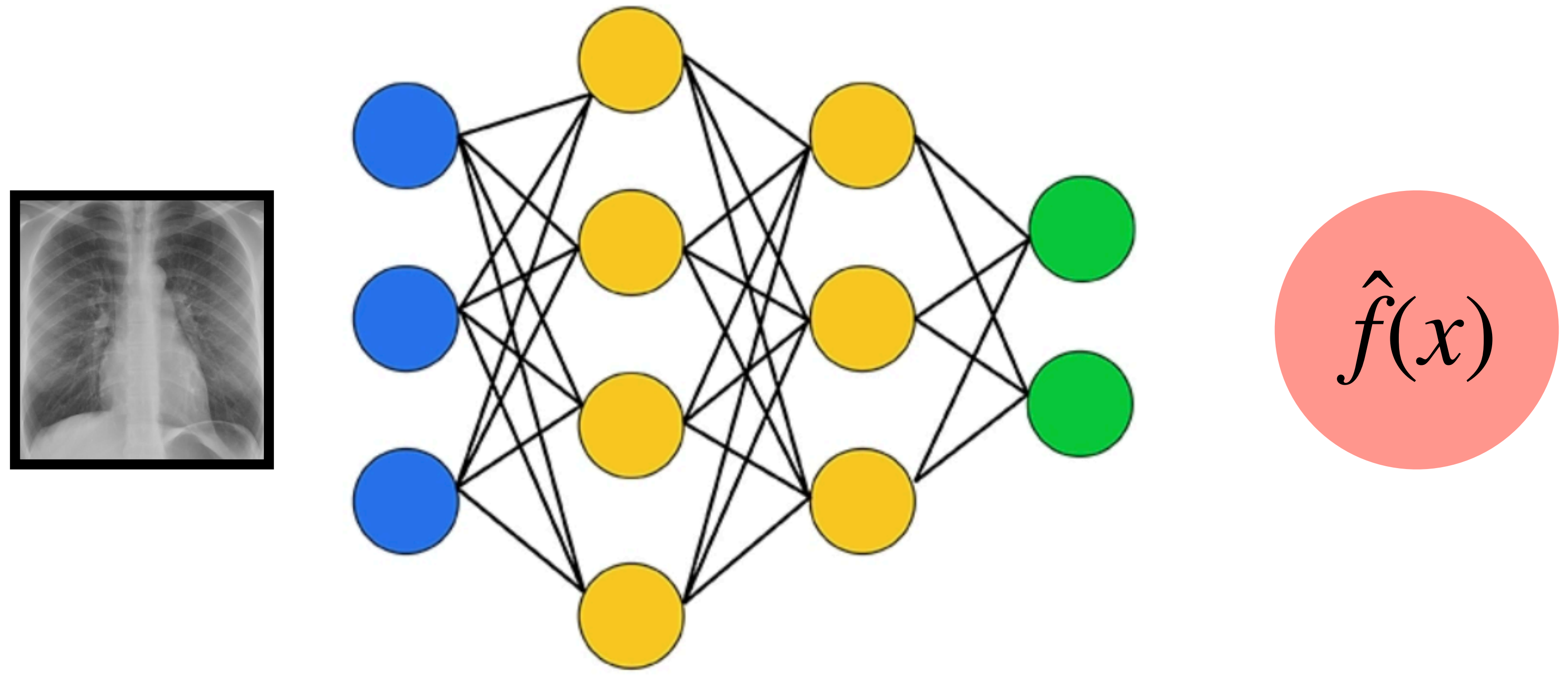




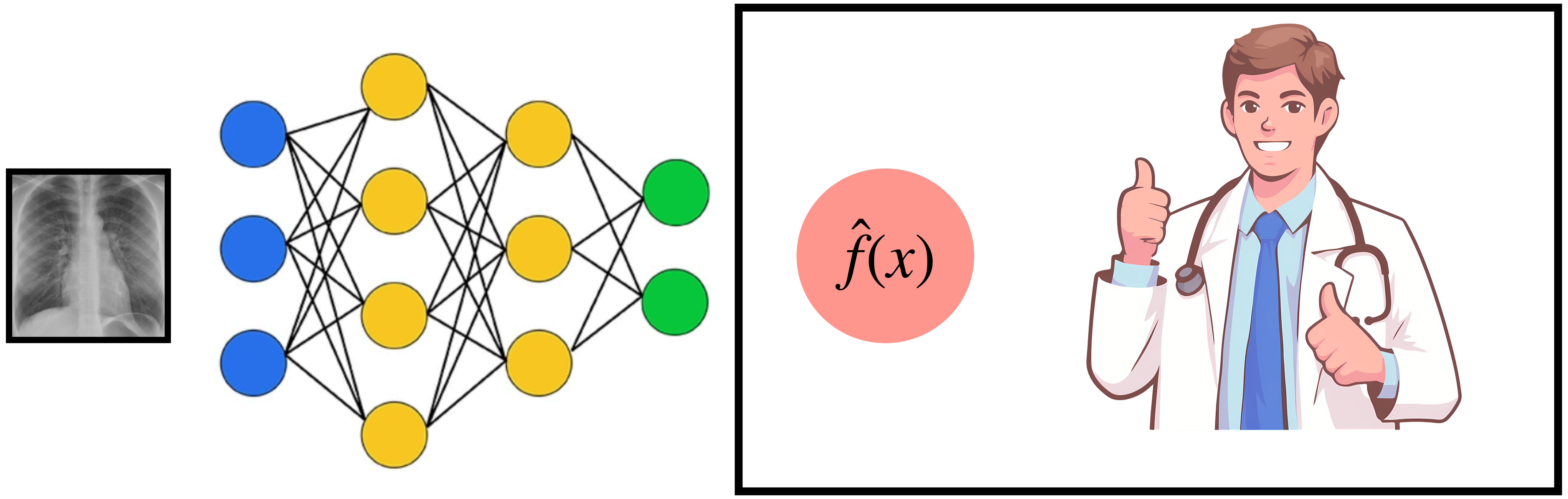
The algorithmic modeling culture

Model validation. Measured by predictive accuracy.

Machine Learning: is this enough?

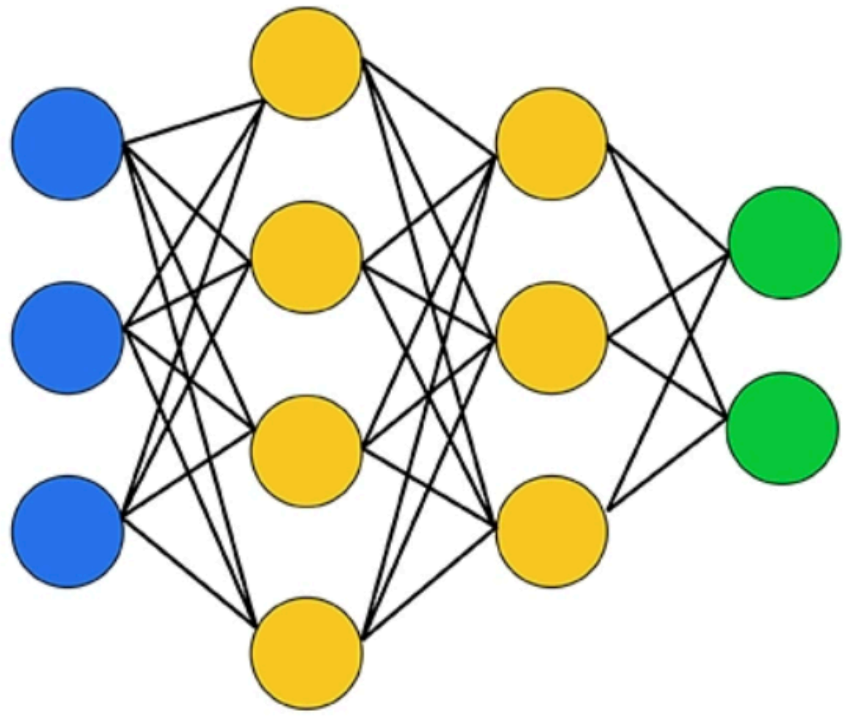


Machine Learning: Risk Controlling frameworks



$$g(\hat{f}(x), \cdot\ )$$

Machine Learning: Risk Controlling frameworks



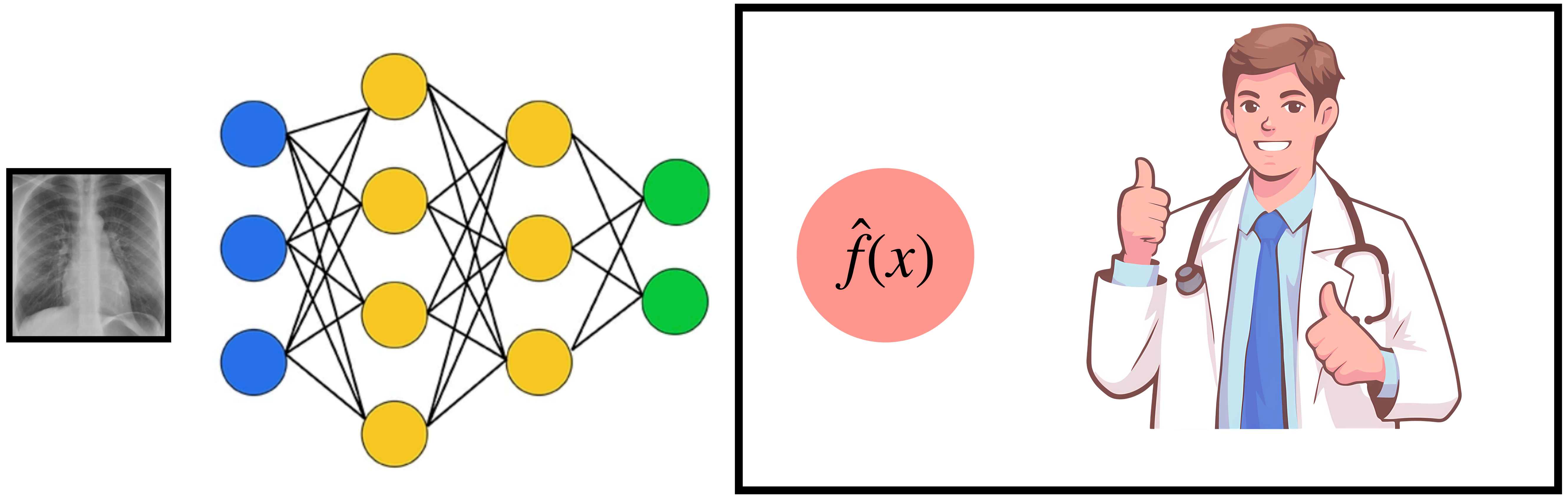
$\hat{f}(x)$

$$\hat{f}(x) \geq \psi$$

treat the patient

$$g(\hat{f}(x), \cdot\,)$$

Machine Learning: Risk Controlling frameworks

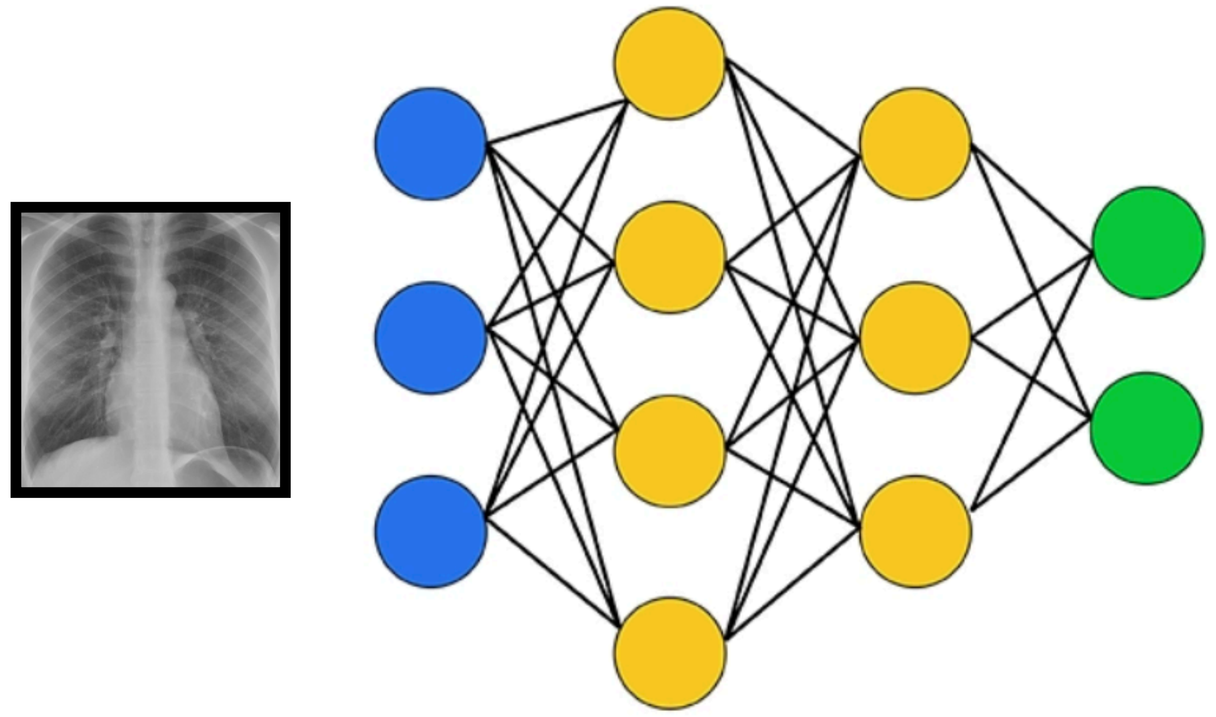


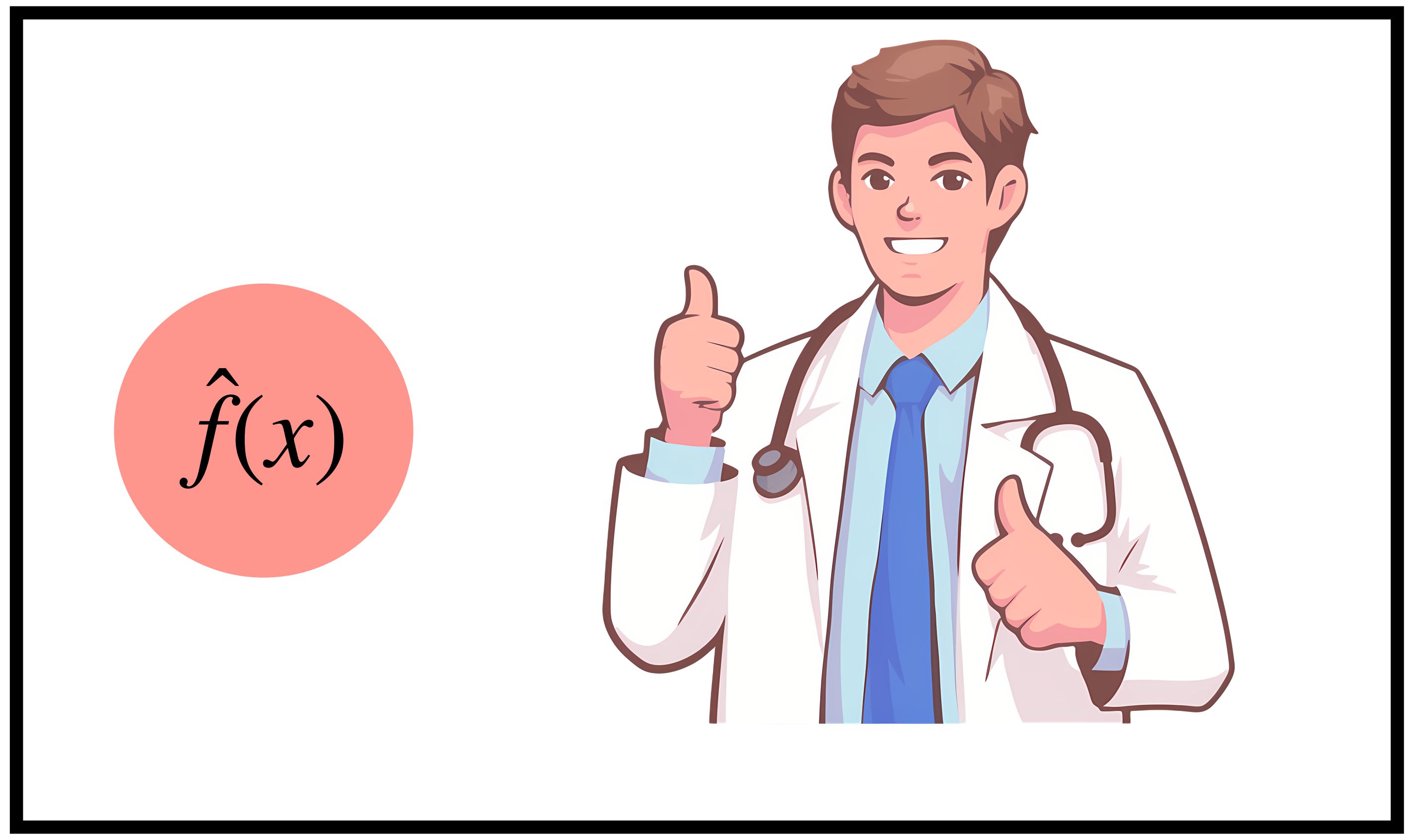
$$g(\hat{f}(x), \psi)$$

$$\hat{f}(x) \geq \psi$$

treat the patient

Machine Learning: Risk Controlling frameworks





$$\hat{f}(x) \geq \psi$$

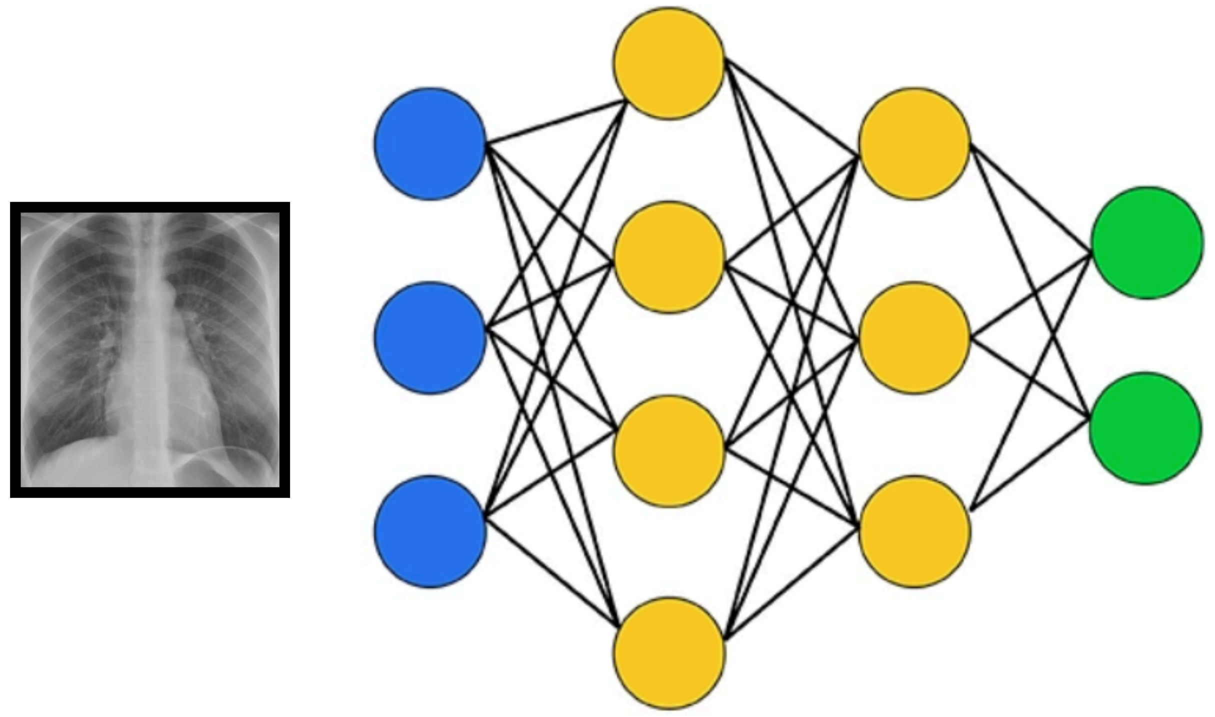
treat the patient

$$\hat{f}_{\psi}(x) := g(\hat{f}(x), \psi)$$

Machine Learning: Risk Controlling frameworks

$$\ell$$

e.g. false positive rate



$\hat{f}(x)$

$$\hat{f}(x) \geq \psi$$

treat the patient

$$\hat{f}_{\psi}(x) := g(\hat{f}(x), \psi)$$

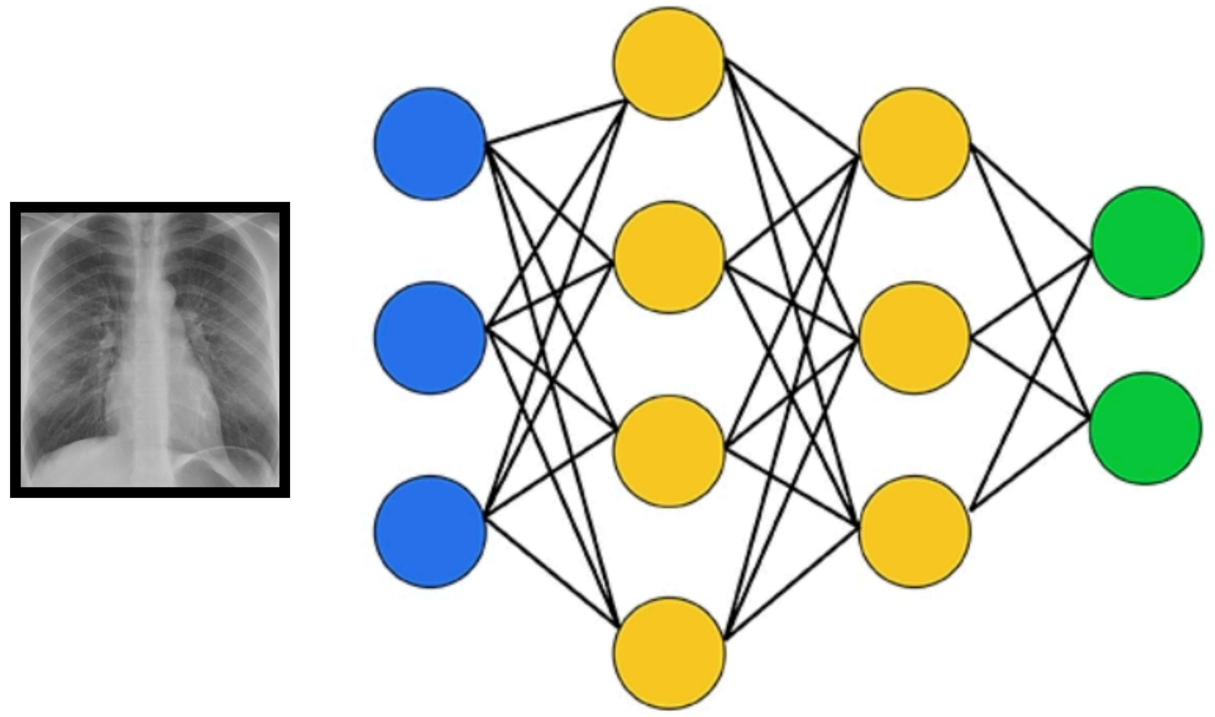
Machine Learning: Risk Controlling frameworks

ℓ

e.g. false positive rate

$$\mathcal{R}(\psi) = \mathbb{E}_P\left[\ell\left(\hat{f}_\psi(X), Y\right)\right]$$

control this risk



$\hat{f}(x)$

$\hat{f}(x) \geq \psi$

treat the patient

$$\hat{f}_\psi(x) := g(\hat{f}(x), \psi)$$

Machine Learning: Risk Controlling frameworks

Conformal Risk Control

Anastasios N. Angelopoulos[1], Stephen Bates[1], Adam Fisch[2], Lihua Lei[3], and Tal Schuste

[1]University of California, Berkeley
[2]Massachusetts Institute of Technology
[3]Stanford University
[4]Google Research

Abstract

We extend conformal prediction to control the expected value of any monotone loss function. The algorithm generalizes split conformal prediction together with its coverage guarantee. Like conformal

Distribution-Free, Risk-Controlling Prediction Sets

Stephen Bates*, Anastasios Angelopoulos*, Lihua Lei*, Jitendra Malik, Michael I. Jordar

August 6, 2021

Abstract

While improving prediction accuracy has been the focus of machine learning in recent years, this alone does not suffice for reliable decision-making. Deploying learning systems in consequential settings

Learn then Test:
Calibrating Predictive Algorithms to Achieve Risk Control

Anastasios N. Angelopoulos, Stephen Bates, Emmanuel J. Candès, Michael I. Jordan, Lihua Lei

October 3, 2022

Abstract

We introduce a framework for calibrating machine learning models so that their predictions satisfy explicit,

Machine Learning: Risk Controlling frameworks

Conformal Risk Control

Anastasios N. Angelopoulos[1], Stephen Bates[1], Adam Fisch[2], Lihua Lei[3], and Tal Schuste

[1]University of California, Berkeley
[2]Massachusetts Institute of Technology
[3]Stanford University
[4]Google Research

Abstract

We extend conformal prediction to control the expected value of any monotone loss function. The algorithm generalizes split conformal prediction together with its coverage guarantee. Like conformal

Distribution-Free, Risk-Controlling Prediction Sets

Stephen Bates*, Anastasios Angelopoulos*, Lihua Lei*, Jitendra Malik, Michael I. Jordar

August 6, 2021

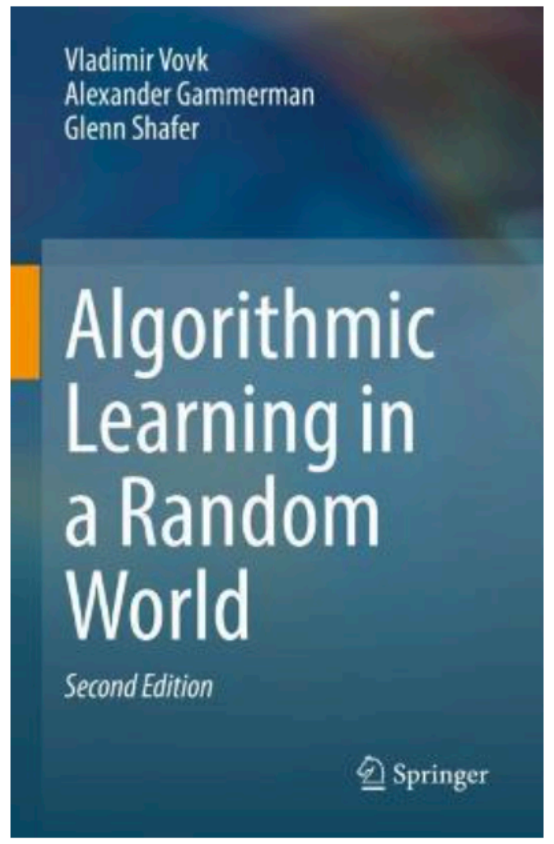
Abstract

While improving prediction accuracy has been the focus of machine learning in recent years, this alone does r ... ettings

Algorithmic learning in a random world

by Vladimir Vovk, Alex Gammerman, and Glenn Shafer

Springer, 2005 (first edition), 2022 (second edition)



The main topic of this book is conformal prediction, a method
machine learning, and unlike other state-of-the-art methods, th

The book integrates mathematical theory and revealing experir
are applied to independent and identically distributed data, and
results to models called repetitive structures, which originate ir
existing methods of machine learning, including newer methoc

Topics and Features:

- Describes how conformal predictors yield accurate and reliable
- Handles both classification and regression problems
- Explains how to apply the new algorithms to real-world data se
- Demonstrates the infeasibility of some standard prediction task
- Explains connections with Kolmogorov's algorithmic randomn
- Develops new methods of probability forecasting and shows h

Researchers in computer science, statistics, and artificial intelli
machine learning. Practitioners and students in all areas of rese

Learn then Test: Calibrating Predictive Algorithms to Achieve Risk Control

Anastasios N. Angelopoulos, Stephen Bates, Emmanuel J. Candès, Michael I. Jordan, Lihua Lei

October 3, 2022

Abstract

We introduce a framework for calibrating machine learning models so that their predictions satisfy explicit,

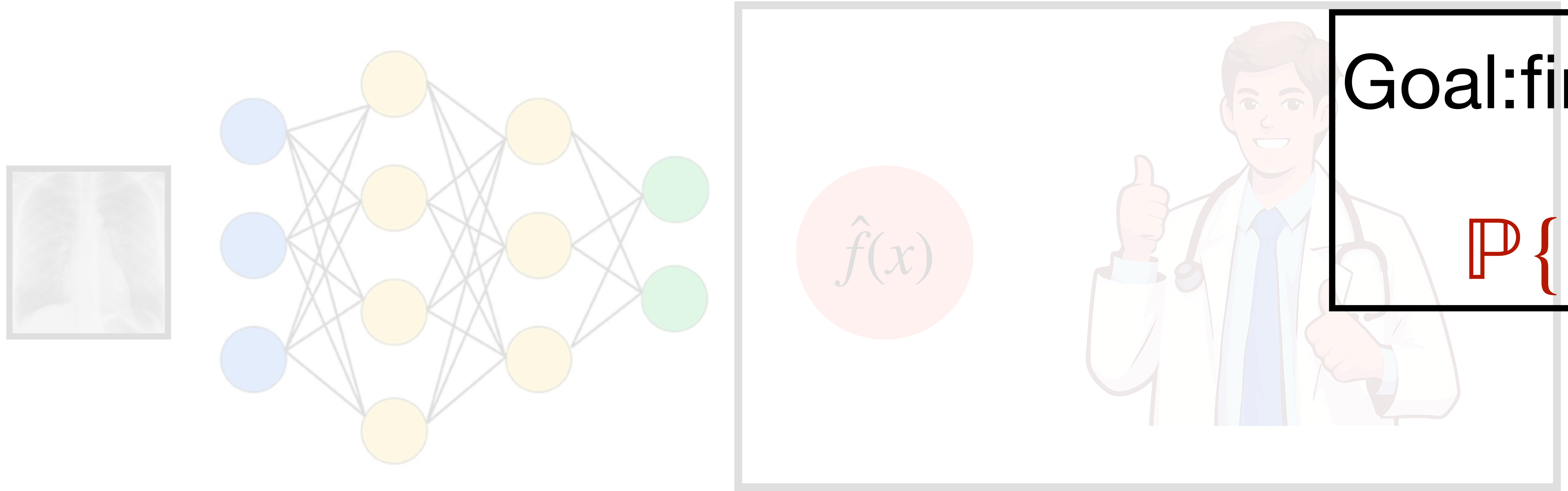
Machine Learning: Risk Controlling frameworks

ℓ

e.g. false positive rate

$$\mathcal{R}(\psi) = \mathbb{E}_P\left[\ell\left(\hat{f}_\psi(X), Y\right)\right]$$

control this risk



Goal:find threshold(s) $\hat{\psi}$ such that

$$\mathbb{P}\{\mathcal{R}(\hat{\psi}) \leq \epsilon\} \geq 1 - \delta$$

$\hat{f}(x) \geq \psi$

treat the patient

$$\hat{f}_\psi(x) := g(\hat{f}(x), \psi)$$

$\{(x_i, y_i)\}_{i=1}^N$ $(x, y) \sim P_0$

calibration set

$$\mathcal{R}(\psi) = \mathbb{E}_P\left[\ell\left(\hat{f}_\psi(X), Y\right)\right]$$

control this risk

Goal:find threshold(s) $\hat{\psi}$ such that

$$\mathbb{P}\{\mathcal{R}(\hat{\psi}) \leq \epsilon\} \geq 1 - \delta$$

Risk Controlling frameworks: conformal prediction

$\{(x_i, y_i)\}_{i=1}^N$ $(x, y) \sim P_0$

calibration set

$s : X \times Y \to [0,B]$

compatibility score

$$\mathcal{R}(\psi) = \mathbb{E}_P \left[\ell \left(\hat{f}_\psi(X), Y \right) \right]$$

control this risk

Goal:find threshold(s) $\hat{\psi}$ such that

$$\mathbb{P}\{\mathcal{R}(\hat{\psi}) \leq \epsilon\} \geq 1 - \delta$$

Risk Controlling frameworks: conformal prediction

$\{(x_i, y_i)\}_{i=1}^{N}$ $(x, y) \sim P_0$

calibration set

$s : X \times Y \to [0,B]$

compatibility score

$\ell : \mathbb{I}\{s(x, y) \leq \psi\}$

loss function

$$\mathscr{R}(\psi) = \mathbb{E}_P \left[\ell \left(\hat{f}_\psi(X), Y \right) \right]$$

control this risk

Goal:find threshold(s) $\hat{\psi}$ such that

$$\mathbb{P}\{\mathscr{R}(\hat{\psi}) \leq \epsilon\} \geq 1 - \delta$$

Risk Controlling frameworks: conformal prediction

$\{(x_i, y_i)\}_{i=1}^N \qquad (x, y) \sim P_0$

calibration set

$s : X \times Y \to [0,B]$

compatibility score

$\ell : \mathbb{I}\{s(x, y) \leq \psi\}$

loss function

$\mathscr{R}(\psi) = \mathbb{P}\{s(X, Y) \leq \psi\}$

Risk function

$$\mathscr{R}(\psi) = \mathbb{E}_P\left[\ell\left(\hat{f}_\psi(X), Y\right)\right]$$

control this risk

Goal:find threshold(s) $\hat{\psi}$ such that

$$\mathbb{P}\{\mathscr{R}(\hat{\psi}) \leq \epsilon\} \geq 1 - \delta$$

Risk Controlling frameworks: conformal prediction

$\{(x_i, y_i)\}_{i=1}^N \qquad (x, y) \sim P_0$

calibration set

$s : X \times Y \to [0,B]$

compatibility score

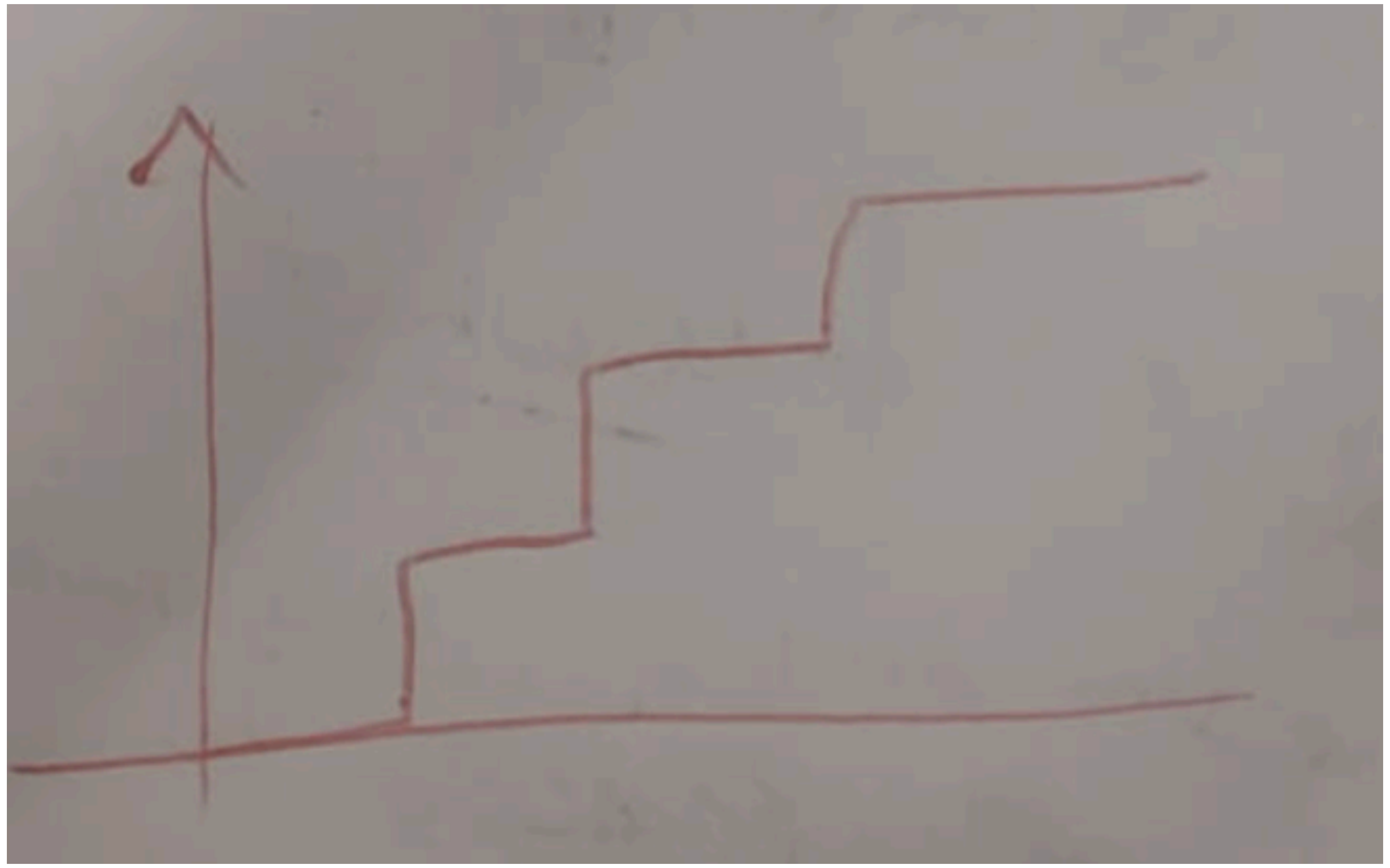
$\ell : \mathbb{I}\{s(x, y) \leq \psi\}$

loss function

$\mathscr{R}(\psi) = \mathbb{P}\{s(X, Y) \leq \psi\}$

Risk function

empirical CDF function



$s(X, Y) \longrightarrow$

Goal:find threshold(s) $\hat{\psi}$ such that

$$\mathbb{P}\{\mathscr{R}(\hat{\psi}) \leq \epsilon\} \geq 1 - \delta$$

Risk Controlling frameworks: conformal prediction

$\{(x_i, y_i)\}_{i=1}^{N}$ $(x, y) \sim P_0$

calibration set

$s : X \times Y \to [0,B]$

compatibility score

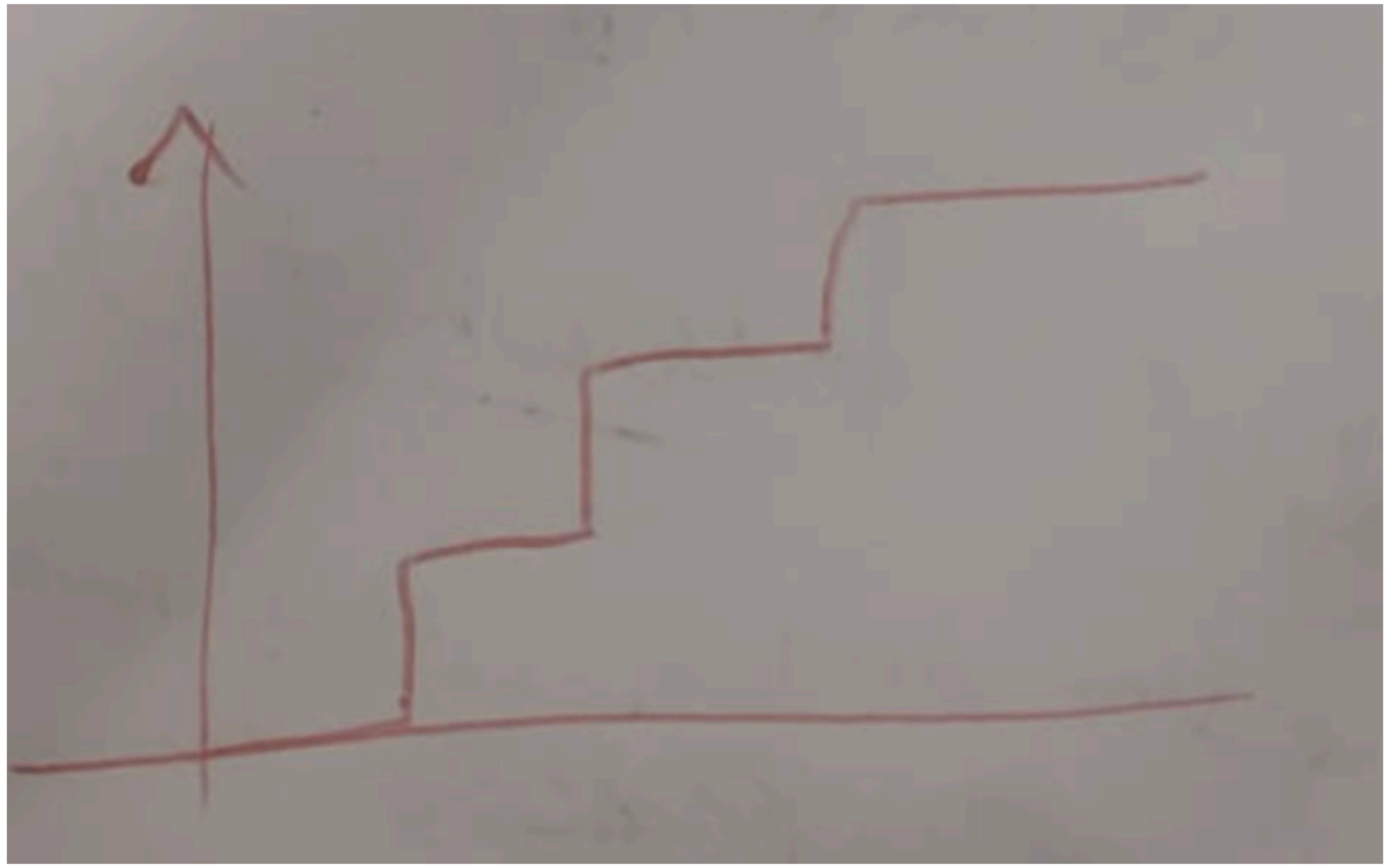
$\ell : \mathbb{I}\{s(x,y) \le \psi\}$

loss function

$\mathscr{R}(\psi) = \mathbb{P}\{s(X,Y) \le \psi\}$

Risk function

empirical CDF function



$s(X, Y) \longrightarrow$

Goal:find threshold(s) $\hat{\psi}$ such that

$\mathbb{P}\{\mathscr{R}(\hat{\psi}) \le \epsilon\} \ge 1 - \delta$

Risk Controlling frameworks: conformal prediction

$$\mathscr{R}(\psi) = \mathbb{E}_P\left[\ell\left(\hat{f}_\psi(X), Y\right)\right]$$

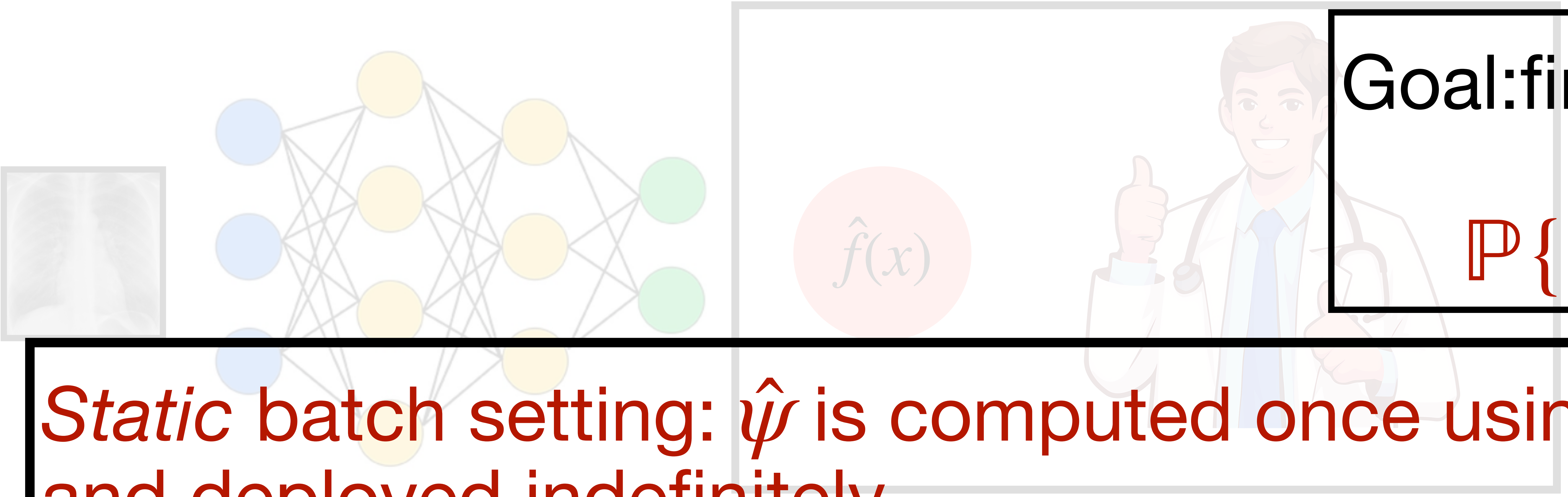
control this risk

Goal: find threshold(s) $\hat{\psi}$ such that

$$\mathbb{P}\{\mathscr{R}(\hat{\psi}) \leq \epsilon\} \geq 1 - \delta$$

Static batch setting: $\hat{\psi}$ is computed once using hold-out calibration set, and deployed indefinitely.

The risk control guarantee holds on the static distribution P_0 over time.



When accurate prediction models yield harmful self-fulfilling prophecies

Wouter A.C. van Amsterdam, MD, PhD* W.A.C.VANAMSTERDAM-3@UMCUTRECHT.NL
Department of Data Science and Biostatistics
Julius Center of Health Sciences and Primary Care
University Medical Center Utrecht, Utrecht, the Netherlands
University of Utrecht, Utrecht, the Netherlands
Heidelberglaan 100, 3584 CX Utrecht, the Netherlands
corresponding author

Nan van Geloven, PhD
Department of Biomedical Data Sciences
Leiden University Medical Center, Leiden, the Netherlands

Jesse H. Krijthe, PhD
Pattern Recognition & Bioinformatics
Delft University of Technology, Delft, the Netherlands

Rajesh Ranganath, PhD
Courant Institute of Mathematical Science, Department of Computer Science
Center for Data Science
New York University, New York City, USA

Giovanni Cinà*, PhD G.CINA@AMSTERDAMUMC.NL
Department of Medical Informatics
Amsterdam University Medical Center, Amsterdam, the Netherlands
Institute for Logic, Language and Computation
University of Amsterdam, Amsterdam, the Netherlands
Pacmed, Amsterdam, the Netherlands

On Continuous Monitoring of Risk Violations under Unknown Shift

[1]UvA-Bosch Delta Lab, University of Amsterdam
[2]Department of Computer Science, Johns Hopkins University

Abstract

Machine learning systems deployed in the real world must operate under dynamic and often unpredictable distribution shifts. This challenges the validity of statistical safety assurances on the system's risk established beforehand. Common risk control frameworks rely on fixed assumptions and lack mechanisms to continuously monitor deployment reliability. In this work, we propose a general framework for the real-time monitoring of

systems has the potential to thwart any 'quality assurance' stamp these methods derive from their static inference. Challenges like outliers, distribution shifts and feedback loops are commonplace [Koh et al., 2021]. In fact, van Amsterdam et al. [2025] argue that an effective machine learning model should *actively* affect the real-world—distribution shift is then not merely an artifact or deployment challenge, but rather a manifestation of a successfully operating system. Hence, any decision-making parameters necessitate *continuous monitoring* during deployment, and the user should be notified when statistical reliability is faltering.

On Continuous Monitoring of Risk Violations under Unknown Shift

There is a need to continuously monitor the risk control guarantees on the predictive systems.

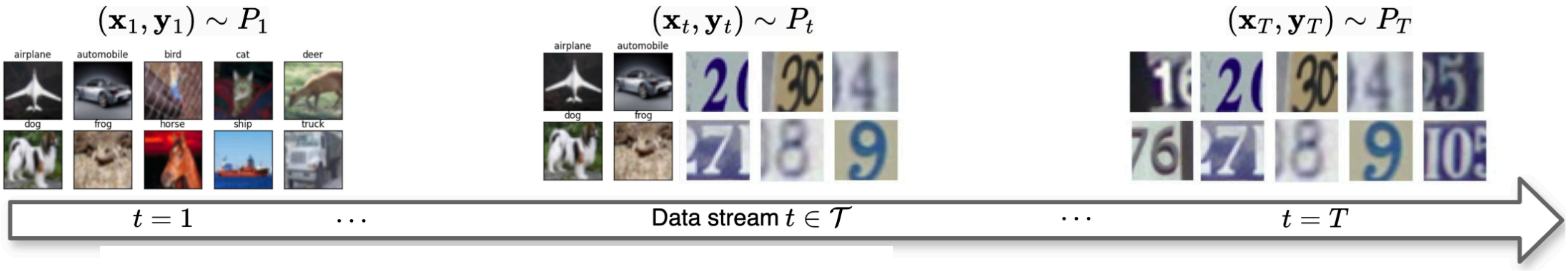
[1]UvA-Bosch Delta Lab, University of Amsterdam
[2]Department of Computer Science, Johns Hopkins University

Abstract

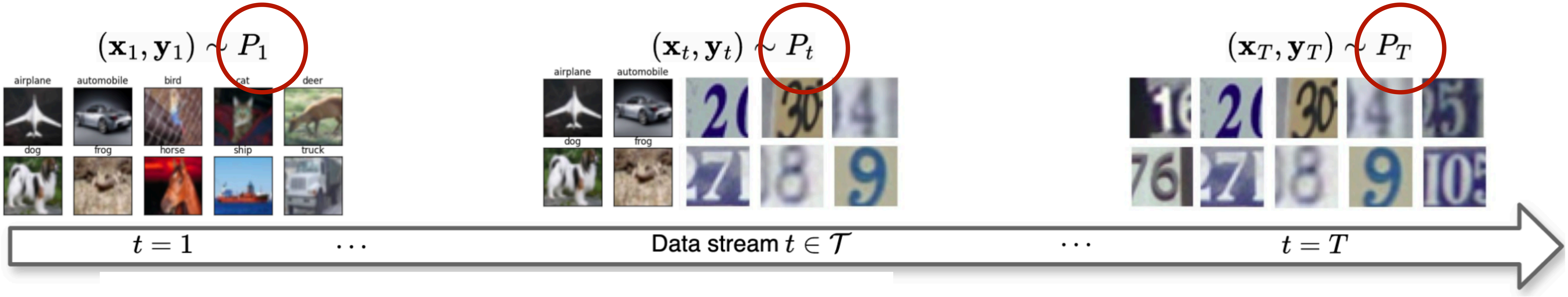
Machine learning systems deployed in the real world must operate under dynamic and often unpredictable distribution shifts. This challenges the validity of statistical safety assurances on the system's risk established beforehand. Common risk control frameworks rely on fixed assumptions and lack mechanisms to continuously monitor deployment reliability. In this work, we propose a general framework for the real-time monitoring of

systems has the potential to thwart any 'quality assurance' stamp these methods derive from their static inference. Challenges like outliers, distribution shifts and feedback loops are commonplace [Koh et al., 2021]. In fact, van Amsterdam et al. [2025] argue that an effective machine learning model should *actively* affect the real-world—distribution shift is then not merely an artifact or deployment challenge, but rather a manifestation of a successfully operating system. Hence, any decision-making parameters necessitate *continuous monitoring* during deployment, and the user should be notified when statistical reliability is faltering.

On Continuous Monitoring of Risk Violations under Unknown Shift

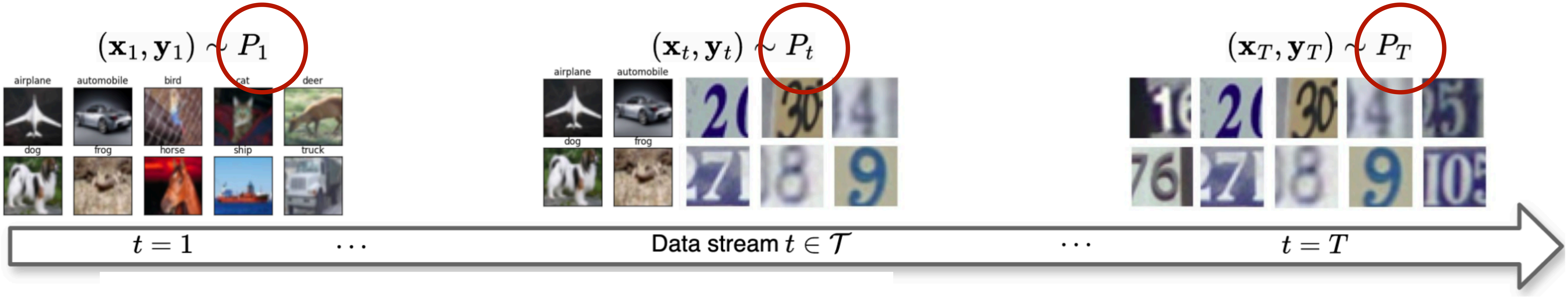


On Continuous Monitoring of Risk Violations under Unknown Shift



On Continuous Monitoring of Risk Violations under Unknown Shift

$$Z = \ell(\hat{\psi}, \hat{f}, X, Y)$$



$$\mathbb{E}_{P_1}[Z_1] \leq \epsilon \qquad \mathbb{E}_{P_t}[Z_t] \leq \epsilon \qquad \mathbb{E}_{P_T}[Z_T] \leq \epsilon$$

On Continuous Monitoring of Risk Violations under Unknown Shift

$$Z = \ell(\hat{\psi}, \hat{f}, X, Y)$$

$(\mathbf{x}_1, \mathbf{y}_1) \sim P_1$ $(\mathbf{x}_t, \mathbf{y}_t) \sim P_t$ $(\mathbf{x}_T, \mathbf{y}_T) \sim P_T$

$t = 1$ $\cdots$ Data stream $t \in \mathcal{T}$ $\cdots$ $t = T$

$\mathbb{E}_{P_1}[Z_1] \leq \epsilon$ $\mathbb{E}_{P_t}[Z_t] \leq \epsilon$ $\mathbb{E}_{P_T}[Z_T] \leq \epsilon$

Goal: for the considered threshold $\hat{\psi}$, decide whether it controls the instantaneous risk by ideally sample access at each time step.

On Continuous Monitoring of Risk Violations under Unknown Shift

$$Z = \ell(\hat{\psi}, \hat{f}, X, Y)$$

$(\mathbf{x}_1, \mathbf{y}_1) \sim P_1$ $(\mathbf{x}_t, \mathbf{y}_t) \sim P_t$ $(\mathbf{x}_T, \mathbf{y}_T) \sim P_T$

$t = 1$ $\cdots$ Data stream $t \in \mathcal{T}$ $\cdots$ $t = T$

$\mathbb{E}_{P_1}[Z_1] \leq \epsilon$ $\mathbb{E}_{P_t}[Z_t] \leq \epsilon$ $\mathbb{E}_{P_T}[Z_T] \leq \epsilon$

Approach: deploy a risk tracker $M_t(\psi)$ to monitor risk violations.

On Continuous Monitoring of Risk Violations under Unknown Shift

Properties of risk tracker: if the risk violations happen, the tracker should grow.

If the tracker does grow, it should signal risk violations with high probability.

Approach: deploy a risk tracker $M_t(\psi)$ to monitor risk violations.

On Continuous Monitoring of Risk Violations under Unknown Shift

$$H_0(\psi) : \mathbb{E}_{P_t}[Z_t \mid \mathcal{F}_{t-1}] \leq \epsilon \quad \forall t \in \mathcal{T} \qquad \text{(risk controlled)}$$

$$H_1(\psi) : \exists t \in \mathcal{T} \quad \mathbb{E}_{P_t}[Z_t \mid \mathcal{F}_{t-1}] > \epsilon \qquad \text{(risk violated)}$$

On Continuous Monitoring of Risk Violations under Unknown Shift

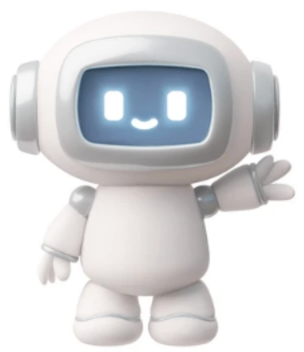
$$H_0(\psi) : \mathbb{E}_{P_t}[Z_t \,|\, \mathcal{F}_{t-1}] \leq \epsilon \quad \forall t \in \mathcal{T}$$ (risk controlled)

$$H_1(\psi) : \exists t \in \mathcal{T} \quad \mathbb{E}_{P_t}[Z_t \,|\, \mathcal{F}_{t-1}] > \epsilon$$ (risk violated)

How to design a tracker $M_t(\psi)$?

How to design a tracker $M_t(\psi)$?

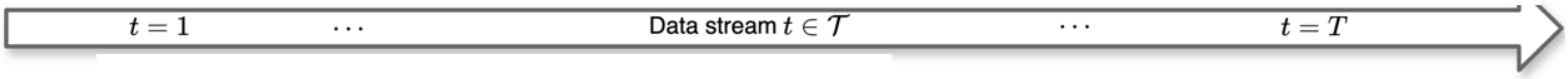
Nature



Forecaster

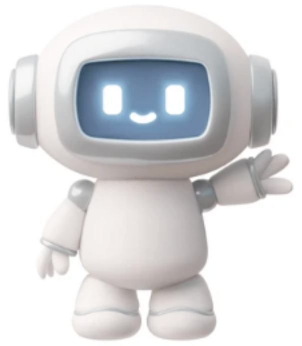
z_t

π_t estimate of the risk at the next time step: t



A sequential forecasting game.

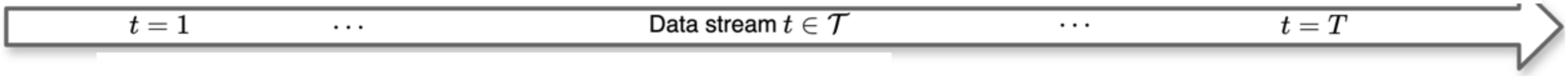
Nature



Forecaster

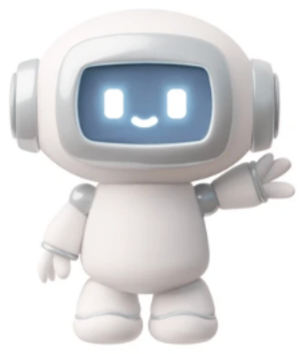
z_t realised loss value

π_t estimate of the risk at the next time step: t



A sequential forecasting game.

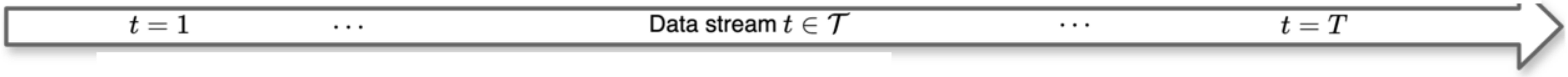
Nature



Forecaster

z_t realised loss value

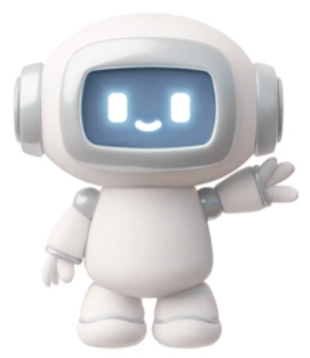
π_t estimate of the risk at the next time step: t



Forecaster incurs error: $\delta_t = z_t - \pi_t$

A sequential forecasting game.

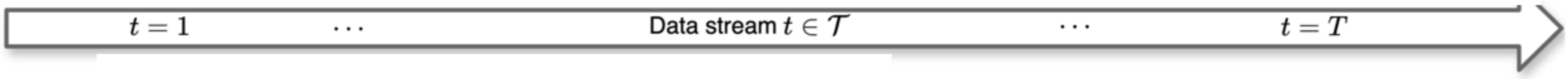
Nature



Forecaster

Z_t realised loss value

π_t estimate of the risk at the next time step: t

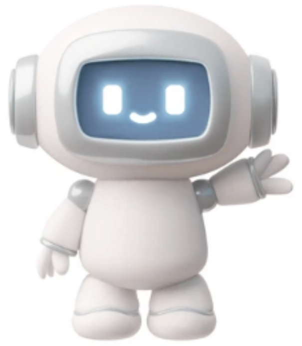


Forecaster incurs error: $\delta_t = z_t - \pi_t$

If the forecaster is playing their best move: $\pi_t = \mathbb{E}_{P_t}[Z_t \mid \mathscr{F}_{t-1}]$, then the forecaster would incur diminishing errors.

A sequential forecasting game.

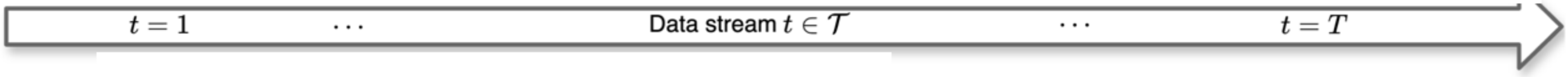
Nature



Forecaster

z_t realised loss value

π_t estimate of the risk at the next time step: t

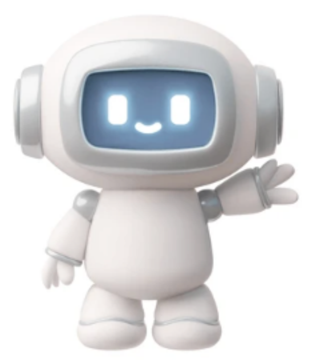


Forecaster incurs error: $\delta_t = z_t - \pi_t$

If the forecaster is playing their best move: $\pi_t = \mathbb{E}_{P_t}[Z_t \,|\, \mathcal{F}_{t-1}]$, then the forecaster won't incur error, as $\mathbb{E}_{P_t}[Z_t - \pi_t \,|\, \mathcal{F}_{t-1}] = 0$.

A sequential forecasting game.

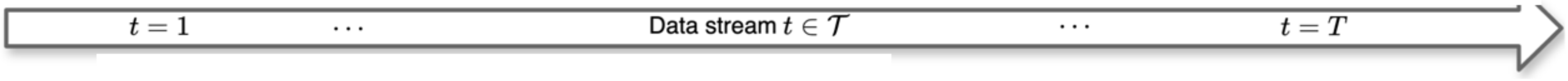
Nature



Forecaster

z_t realised loss value

π_t estimate of the risk at the next time step: t



Forecaster incurs error: $\delta_t = z_t - \pi_t$

The error process $(\delta_t)_{t \in \mathcal{T}}$ can be used to construct the tracker.

A sequential forecasting game.

$$M_t(\psi) = \prod_{i=1}^{t} (1 + \lambda_i \cdot \delta_i) = \prod_{i=1}^{t} (1 + \lambda_i(z_i - \epsilon))$$

If the risk is controlled, then the tracker will not grow.

$$\mathbb{E}[M_t(\psi) \,|\, \mathcal{F}_{t-1}] = M_{t-1} \cdot \mathbb{E}_{P_t}[\lambda_t(z_t - \epsilon) \,|\, \mathcal{F}_{t-1}]$$

$$M_t(\psi) = \prod_{i=1}^{t} (1 + \lambda_i \cdot \delta_i) = \prod_{i=1}^{t} (1 + \lambda_i(z_i - \epsilon))$$

If the risk is controlled, then the tracker will not grow.

$$\mathbb{E}[M_t(\psi) \mid \mathscr{F}_{t-1}] = M_{t-1} \cdot \mathbb{E}_{P_t}[\lambda_t(z_t - \epsilon) \mid \mathscr{F}_{t-1}]$$

Test supermartingale

Submitted to Statistical Science

Game-Theoretic Statistics and Safe Anytime-Valid Inference

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk and Glenn Shafer

Abstract. Safe anytime-valid inference (SAVI) provides measures of statistical evidence and certainty—e-processes for testing and confidence sequences for estimation—that remain valid at all stopping times, accommodating con-

Testing by betting:

$M_t(\psi)$ is the wealth process of an agent actively betting against the null.

Submitted to Statistical Science

Game-Theoretic Statistics and Safe Anytime-Valid Inference

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk and Glenn Shafer

Abstract. Safe anytime-valid inference (SAVI) provides measures of statistical evidence and certainty—e-processes for testing and confidence sequences for estimation—that remain valid at all stopping times, accommodating continuous monitoring and analysis of accumulating data and optional stopping or continuation for any reason. These measures crucially rely on test martin-

On Continuous Monitoring of Risk Violations under Unknown Shift

Properties of risk tracker: if the risk violations happen, the tracker should grow.

If the tracker does grow, it should signal risk violations with high probability.

Approach: deploy a risk tracker $M_t(\psi)$ to monitor risk violations.

Lemma 4.2 (False alarm guarantee). *For any $\psi \in \Psi$ such that $\mathbb{E}_{P_t}[\mathbf{z}_t \mid \mathcal{F}_{t-1}] \leq \epsilon \; \forall t \in \mathcal{T}$ satisfies the null, it holds that $\mathbb{P}\left(\exists t \in \mathcal{T} : M_t(\psi) \geq 1/\delta\right) \leq \delta$.*

If the tracker does grow, it should signal risk violations with high probability.

Definition 4.4 (Growth rate optimality (GRO)). *The betting rate λ_t is growth rate optimal if it satisfies the condition $\lambda_t = \arg\max_{\lambda \in [0, 1/\epsilon)} \mathbb{E}_{H_1}[\log M_t(\psi)]$.*

If the risk violations happen, the tracker should grow.

Empirical demonstration

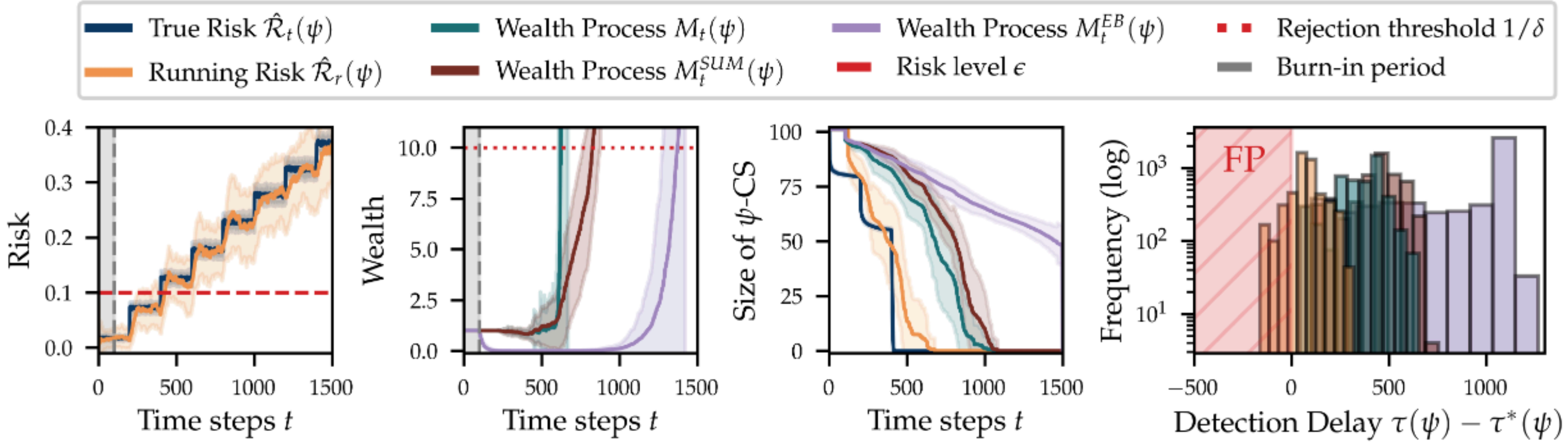


Figure 2: Results for **outlier detection with a stepwise shift** (§ 6.1). *From left to right:* Visuals of the growing risk and wealth process behaviour with respective rejection thresholds ϵ and $1/\delta$, for a single threshold candidate (here $\psi = 0.50$); the behaviour of the valid threshold set ψ-CS (Eq. 5), which eventually shrinks to zero signalling a model update; and the empirical distributions of detection delays $\tau(\psi) - \tau_*(\psi)$ across all $\psi \in \Psi$, including the false alarm region (FP). We also have $B = 1, S = 50$ and $t_{out} = 200$, with results evaluated over $R = 50$ trials (mean and std. deviation).

Empirical demonstration

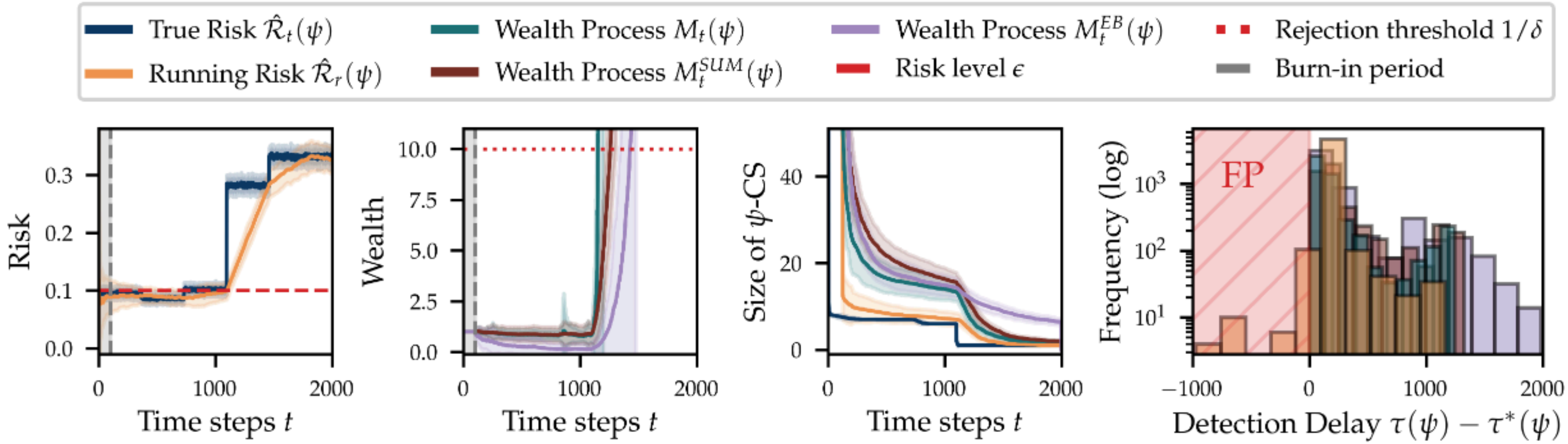


Figure 3: Results for **set prediction with a temporal shift on FMoW** (§ 6.2). *From left to right:* Visuals of the growing risk and wealth process behaviour with respective rejection thresholds ϵ and $1/\delta$, for a single threshold candidate (here $\psi = 0.08$); the behaviour of the valid threshold set ψ-CS (Eq. 5), which eventually tends to zero signalling a model update; and the empirical distributions of detection delays $\tau(\psi) - \tau_*(\psi)$ across all $\psi \in \Psi$, including the false alarm region (FP). We also have $B = 1$ and $S = 365$ (one year), with results evaluated over $R = 50$ trials (mean and std. deviation).

More results in the paper:

Confidence sets, asymptotic consistency and detection delay bound, betting strategie

The role of statistical inference in machine learning

Questions?

The role of statistical inference in machine learning