# Algorithm design for improved decision-making

Jessie Finocchiaro

# Algorithmic predictions are used to make decisions


https://armman.org/

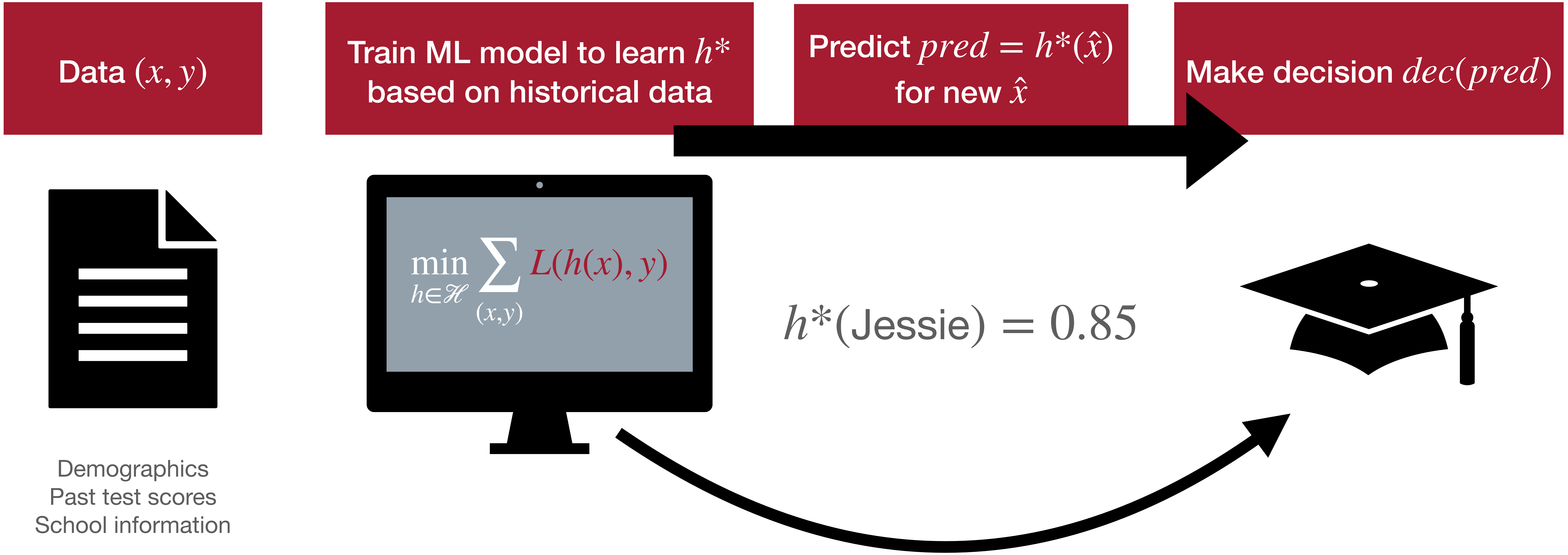
armman
Delivering India's Future

Wealthfront

https://www.wealthfront.com/


https://www.theguardian.com/commentisfree/2020/aug/19/its-not-just-a-levels-algorithms-have-a-nightmarish-new-power-over-our-lives

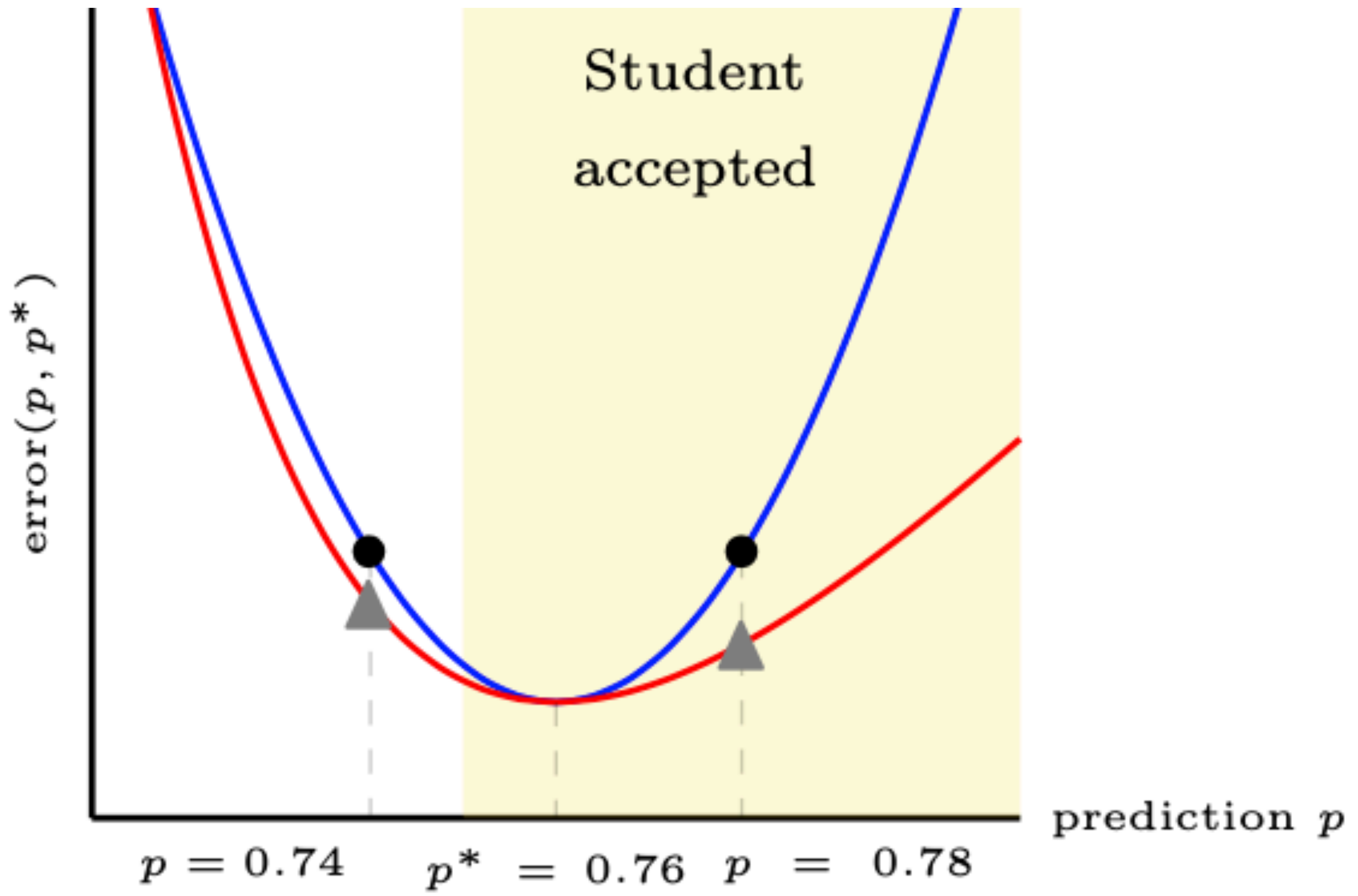
Folk wisdom: we can make better decisions with "less" if we just predict $dec(pred)$




Goal: incorporate the structure of $dec$ into $L(\cdot, \cdot)$

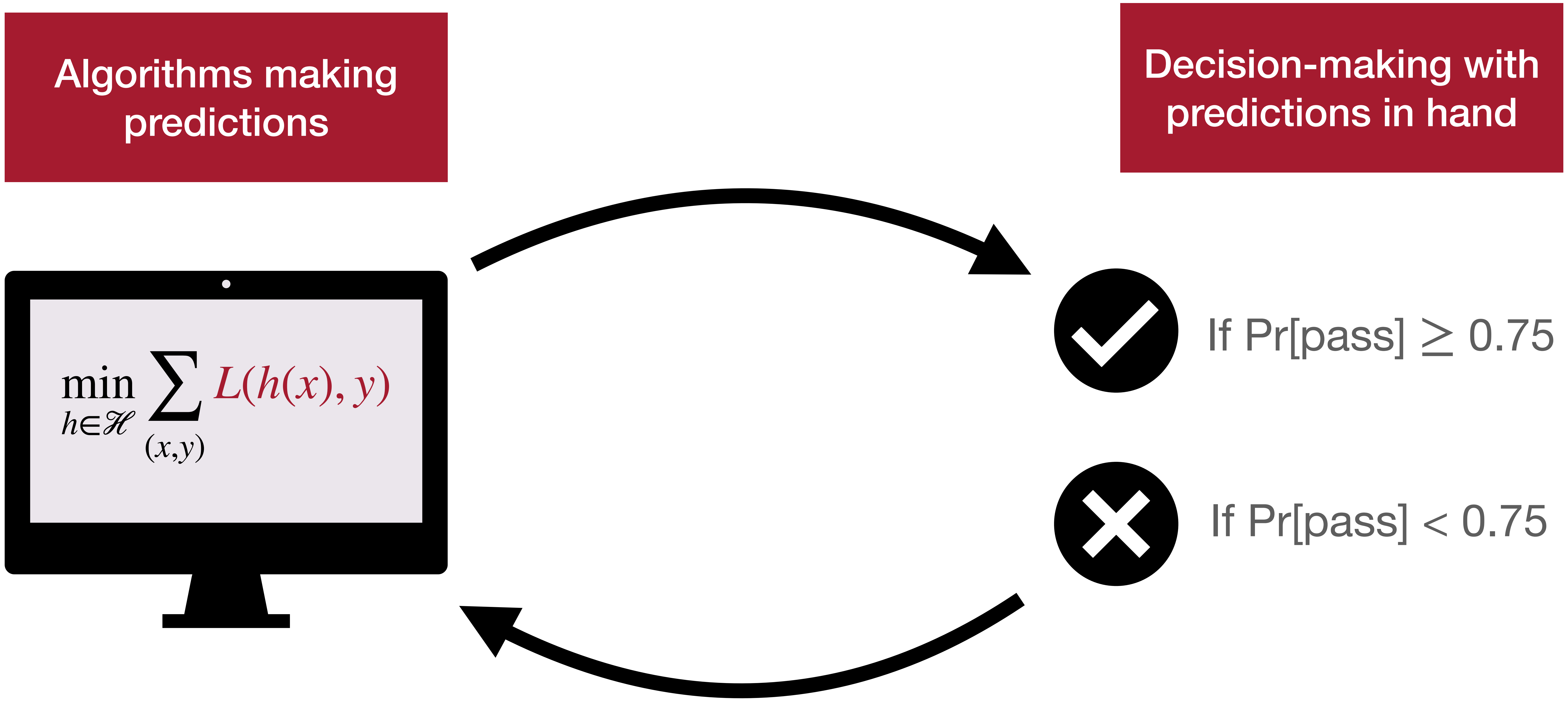
# Why predict decisions?

**Goal:** design algorithms (loss functions) that incorporate decision problem to make "smarter" errors

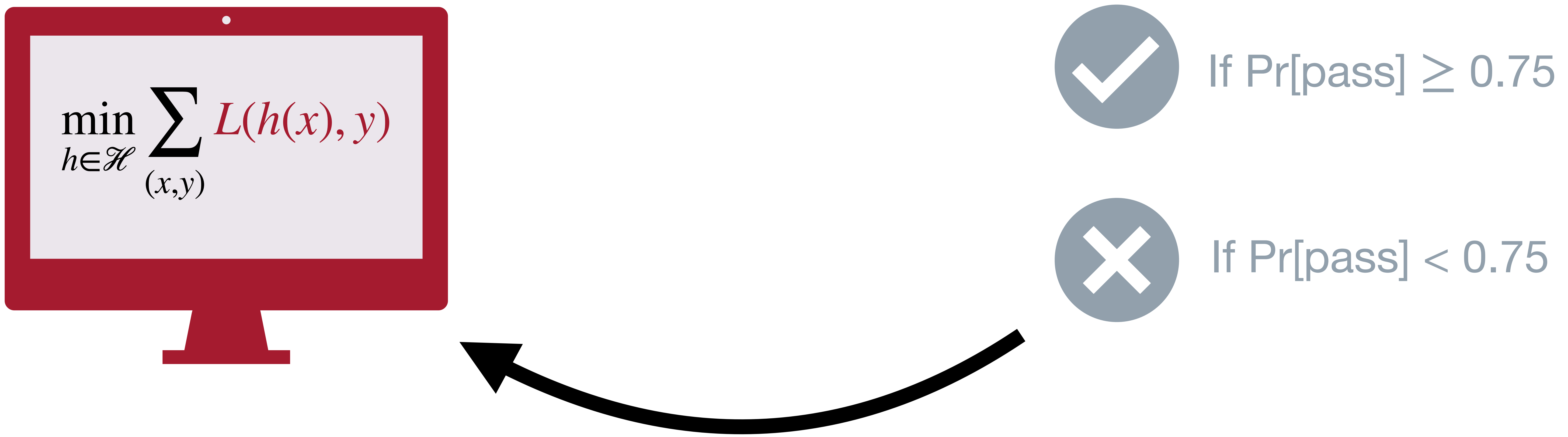


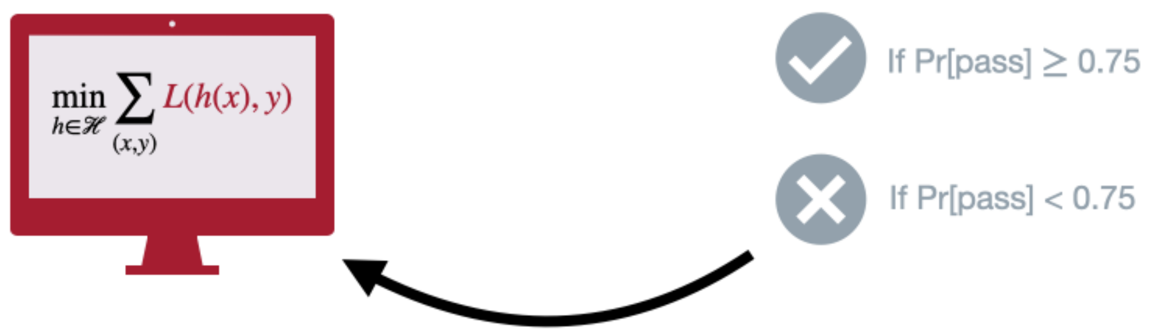

Challenges:
- Lots of decision problems!
- *How to construct* good loss functions?

# Outline


Algorithms making predictions
Decision-making with predictions in hand
$\min_{h \in \mathcal{H}} \sum_{(x,y)} L(h(x), y)$
If Pr[pass] ≥ 0.75
If Pr[pass] < 0.75

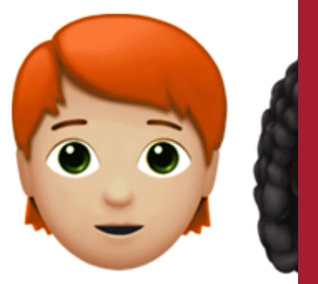
# Algorithm design: incorporating decision structure






# What is a decision task: common structure?

Ranking:



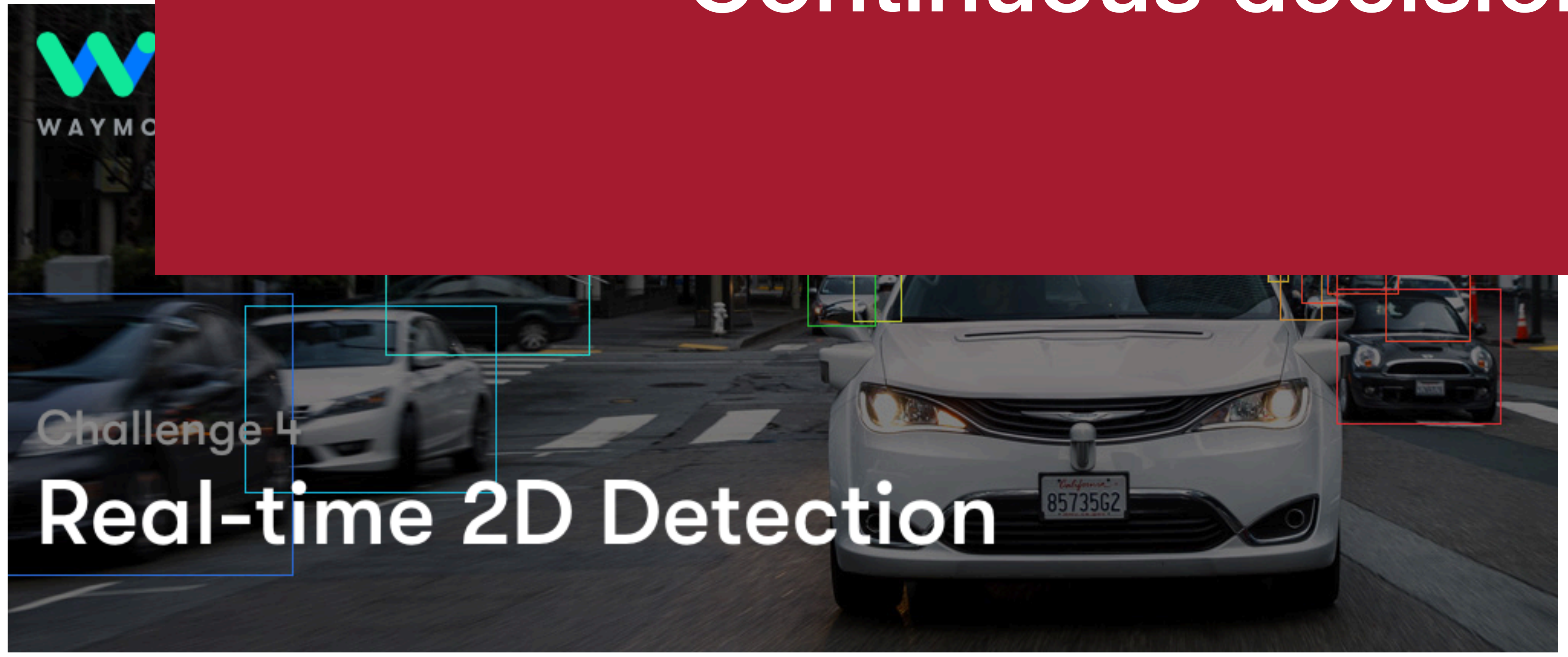
Classification:

arch-for-the-best-
i-cbda4d6b425d/

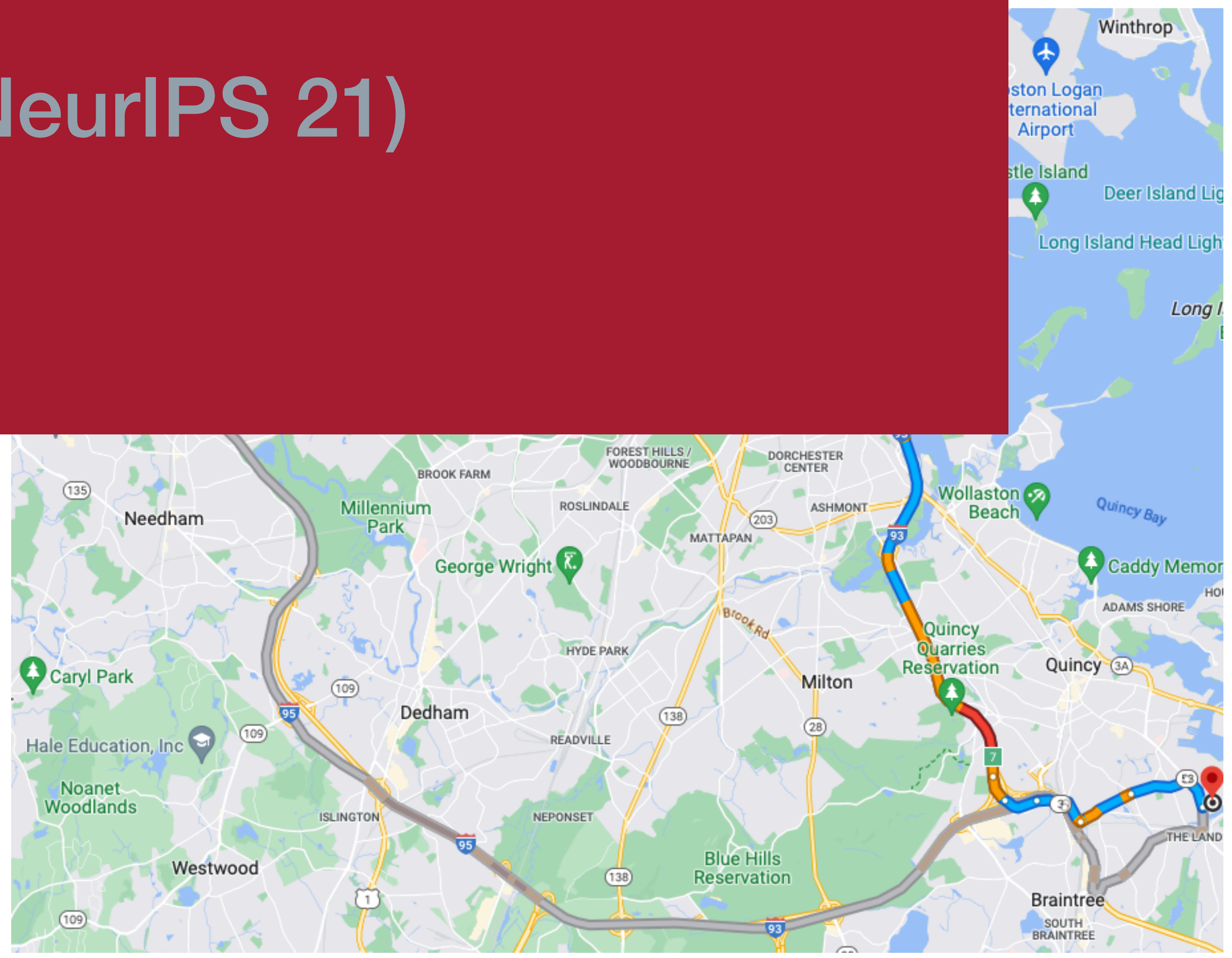
**Today: discrete decisions** (FFW NeurIPS 19→JMLR, FFGT ICML 22, FFN COLT 22)

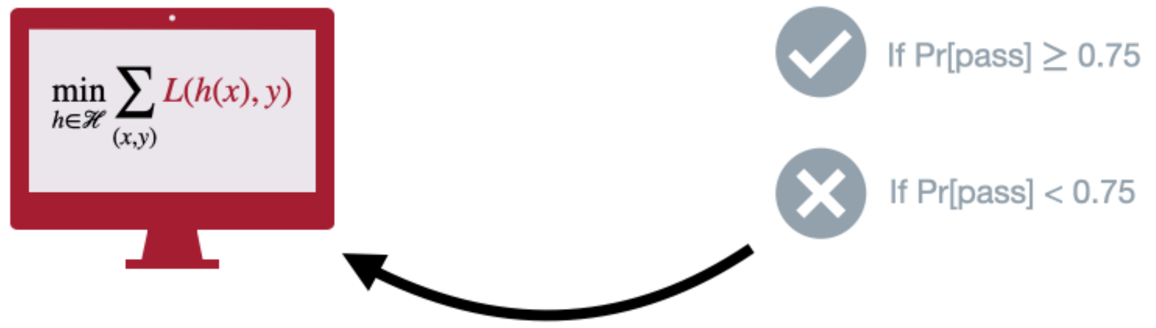
Continuous decisions (FF NeurIPS 18, FFW NeurIPS 21)

Segmenta




https://waymo.com/open/challenges/2021/real-time-2d-prediction/#



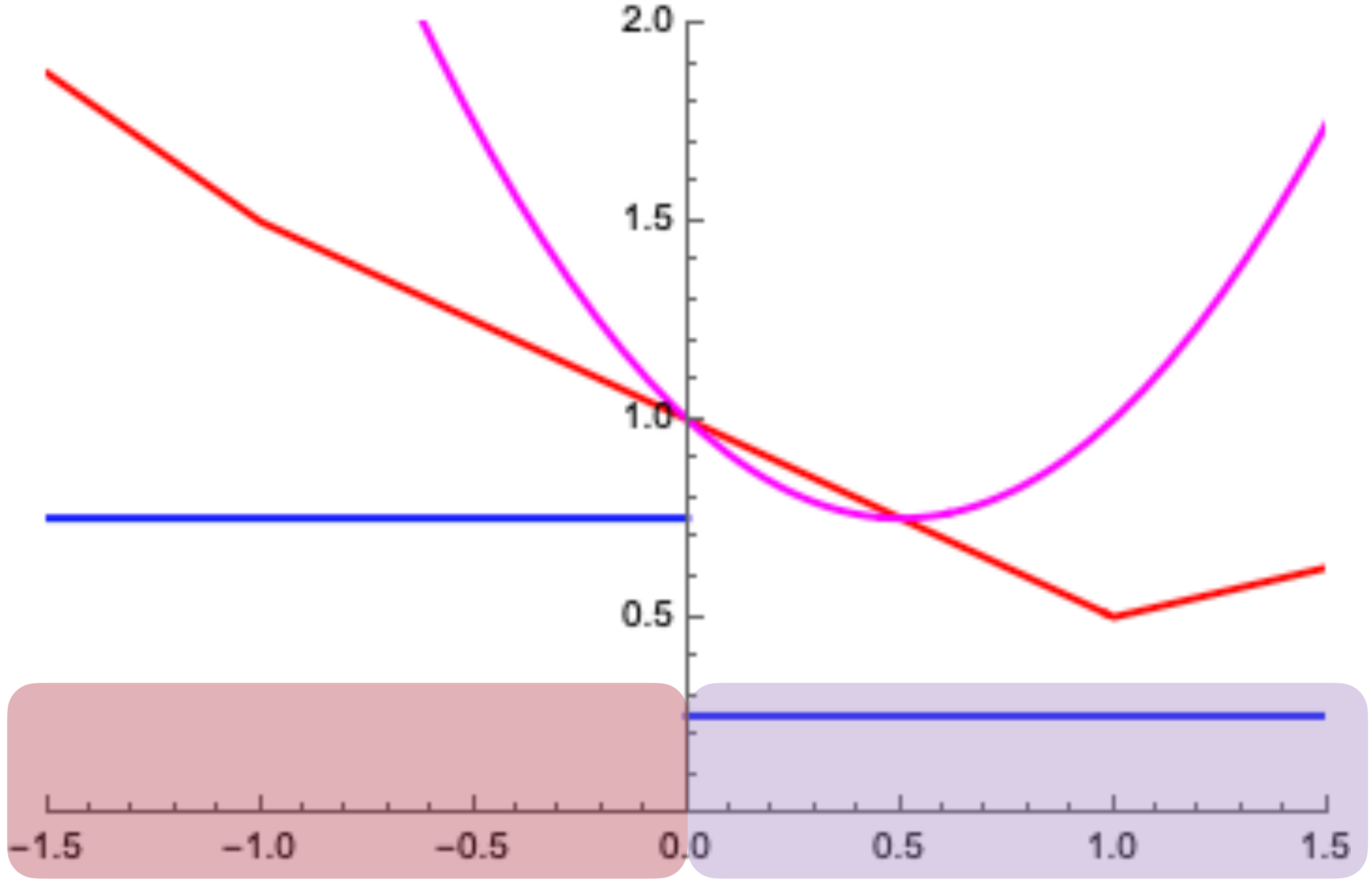



# Common structure: decision loss matrix

$$\ell(r, y) = \ell_{r,y}$$

| | y=1 | y=-1 |
|---|---|---|
| **r=1** | 0 | 1 |
| **r=-1** | 1 | 0 |

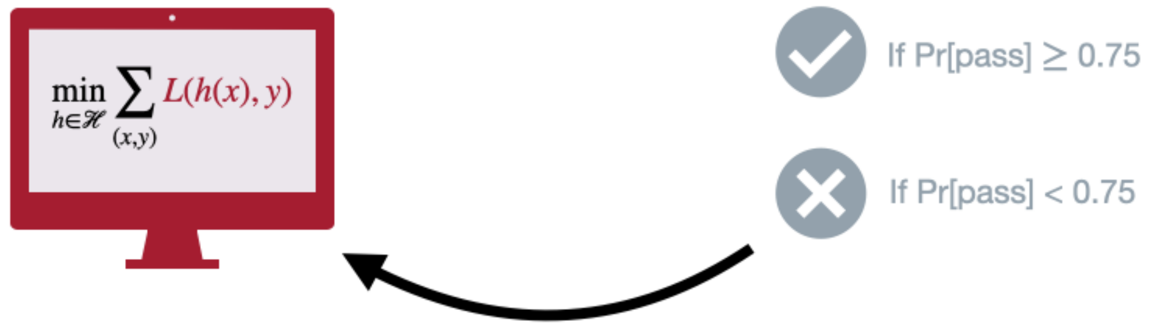
$\mathbb{E}_{Y \sim p} L(u, Y)$




$dec : \mathbb{R} \to \mathscr{R}$

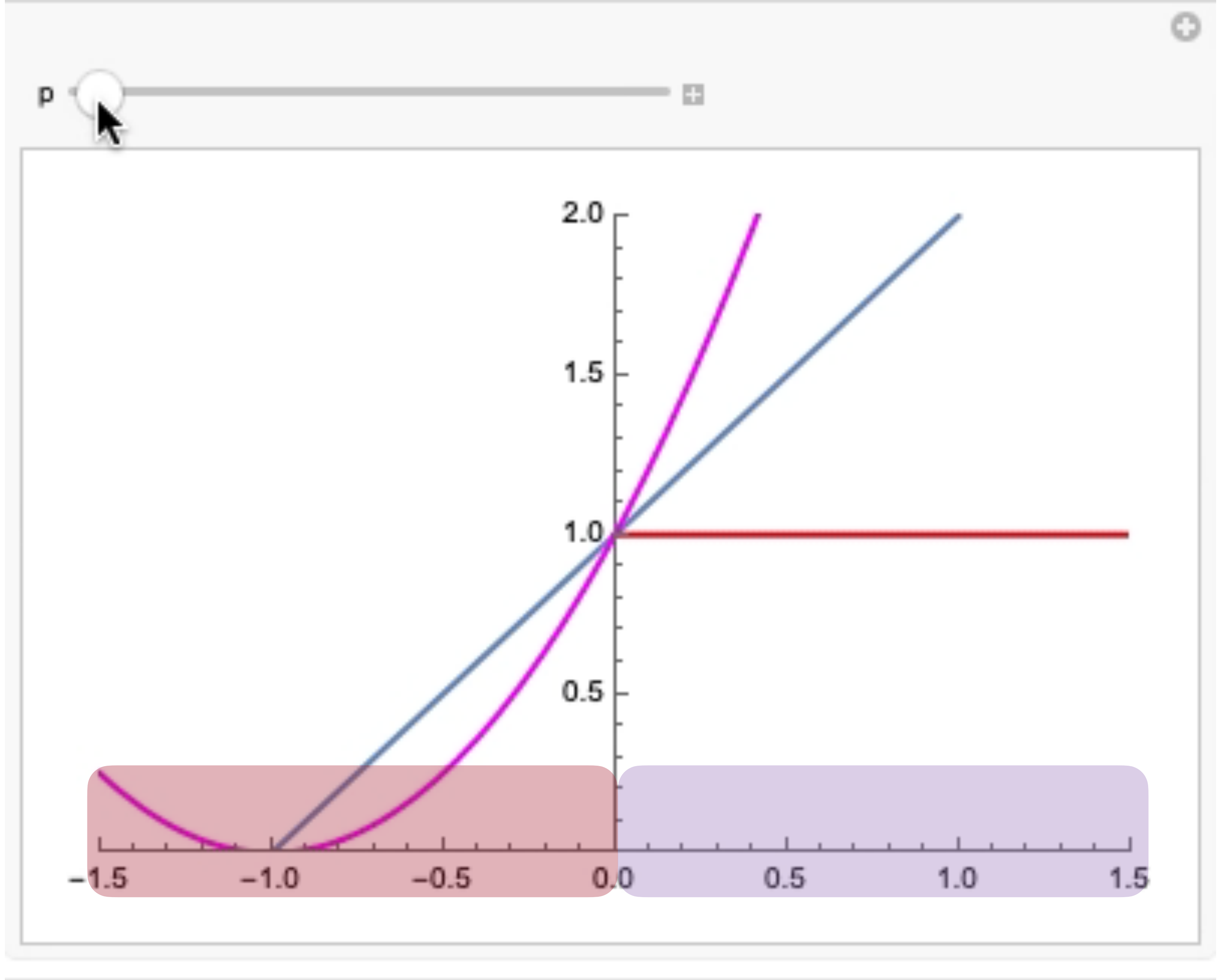
$u \in \mathbb{R}$

Decision loss $\ell$ easy (relatively) to analyze, but intractable to optimize.




# Good losses: consistent and convex

$\mathbb{E}_{Y \sim p} loss(u, Y)$

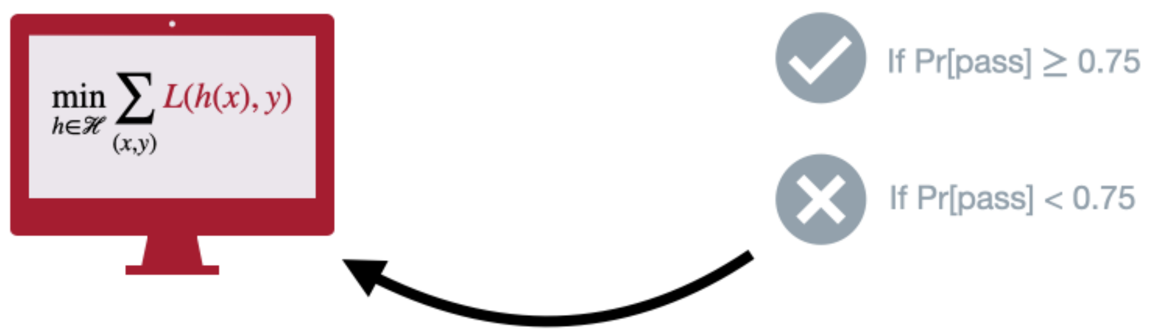



$u \in \mathbb{R}$

A surrogate loss $L$ and decision $dec$ pair $(L, dec)$ is consistent with respect to a decision loss $\ell$ if minimizing the expected surrogate loss $L$ then applying $dec$ yields the same decision as minimizing expected $\ell$ directly

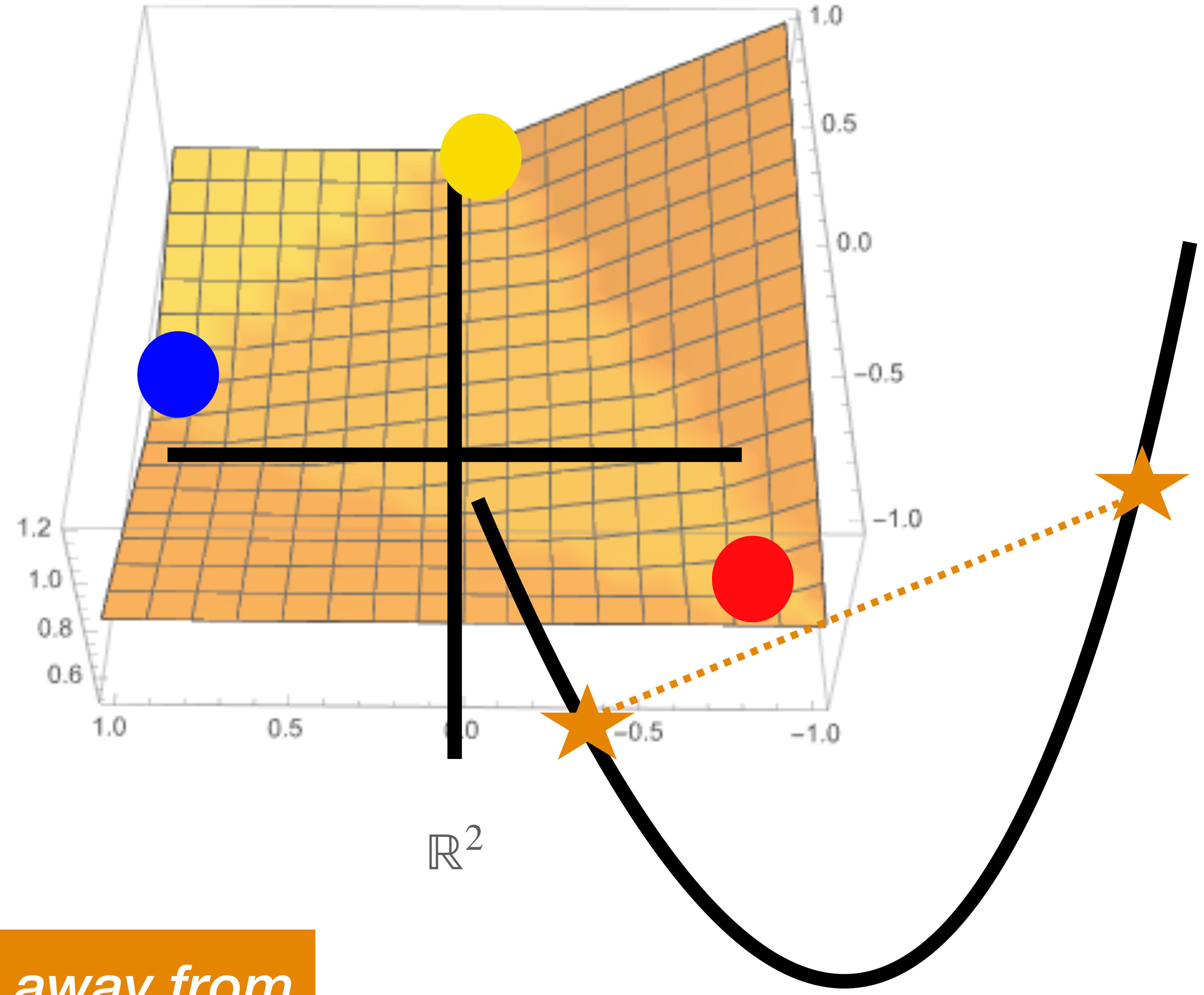
Challenge: for a given decision task, design *one* surrogate loss and decision pair that works *for all* data distributions

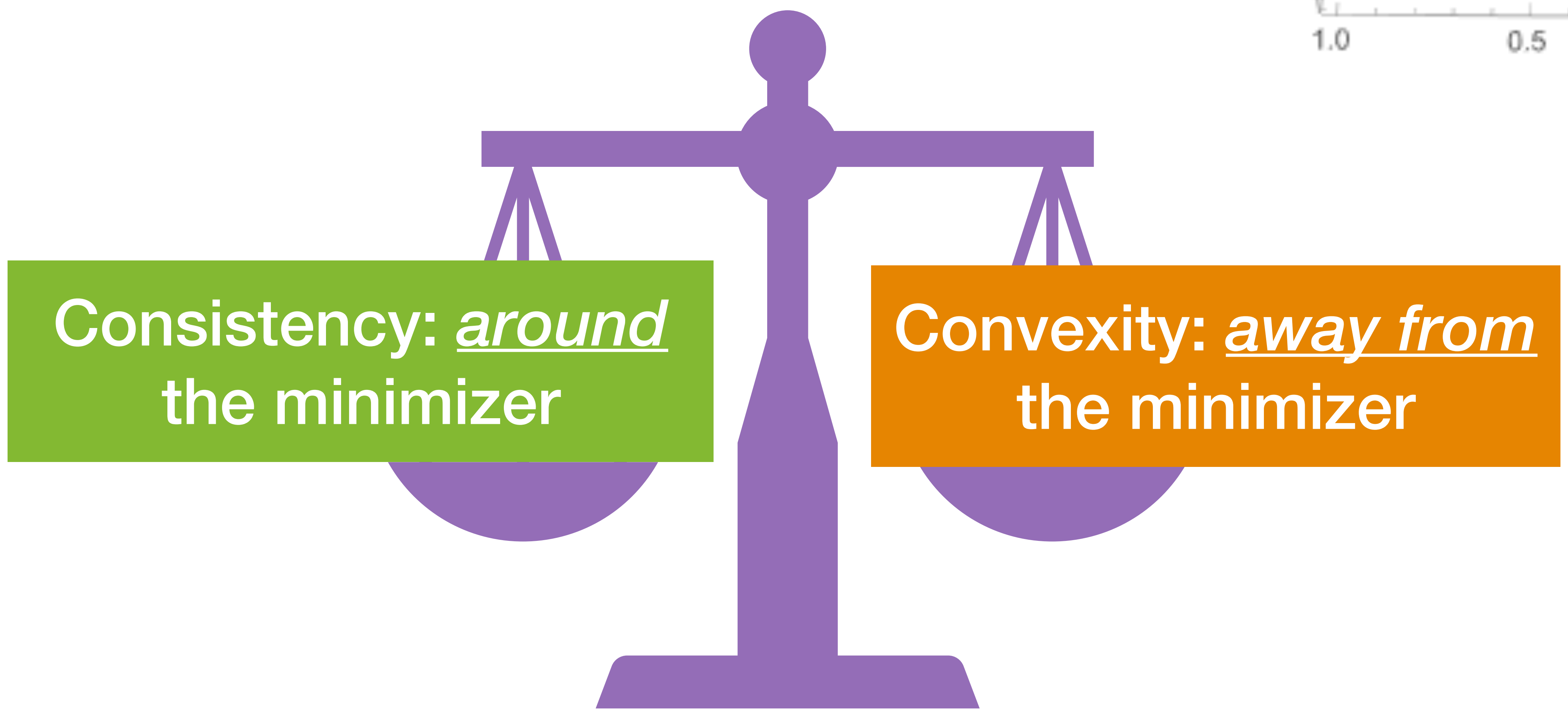
Example: $L$ logistic loss, hinge loss, squared loss
$dec$ = sign
$\ell$ is 0-1 loss



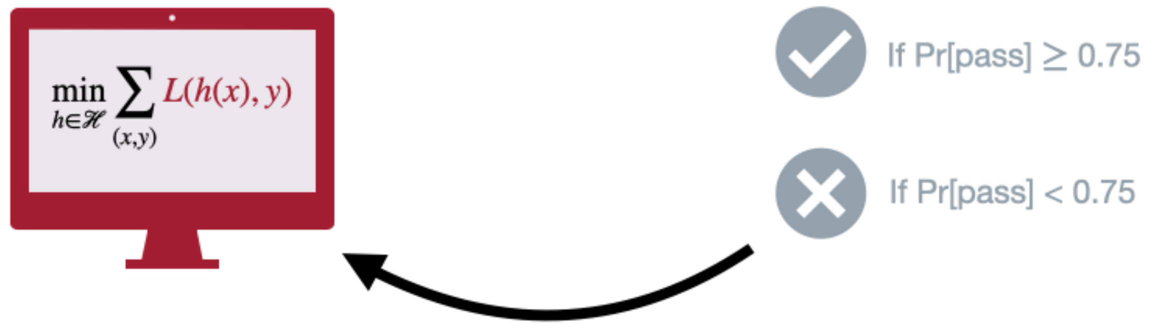

# Good losses: consistent and convex

Convex:
If decisions are ***discrete***, but $\mathbb{R}^d$ is ***infinite***, what do we do in the infinite space in between?



# Our contributions

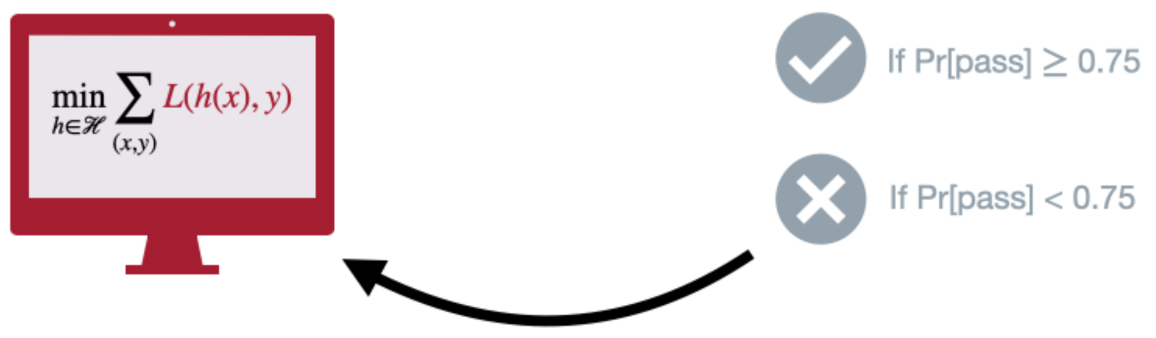



**Our proposal**: a framework to analyze the consistency of piecewise linear and convex (PLC) surrogates for discrete decision losses

Introduce the definition of *embeddings*

Show embedding $\implies$ consistency

A *much simpler* tool for analyzing consistency




# Hinge loss embeds (twice) 0-1 loss

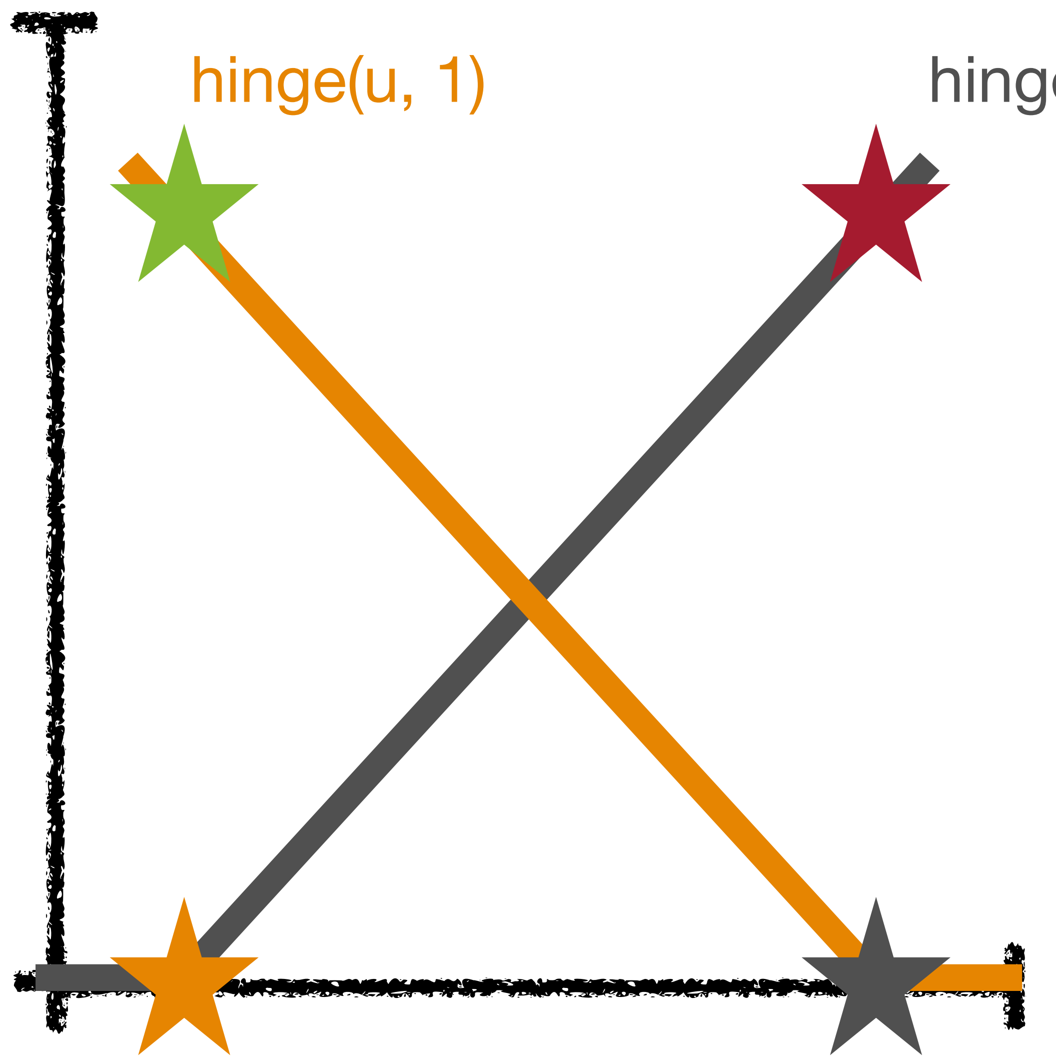
**Surrogate loss** $L : \mathbb{R}^d \times \mathscr{Y} \to \mathbb{R}_+$ embeds a **Decision loss** $\ell : \mathscr{R} \times \mathscr{Y} \to \mathbb{R}_+$ if there exists an $embedding : \mathscr{R} \to \mathbb{R}^d$…

**1. Loss values match**

$L(u, y)$

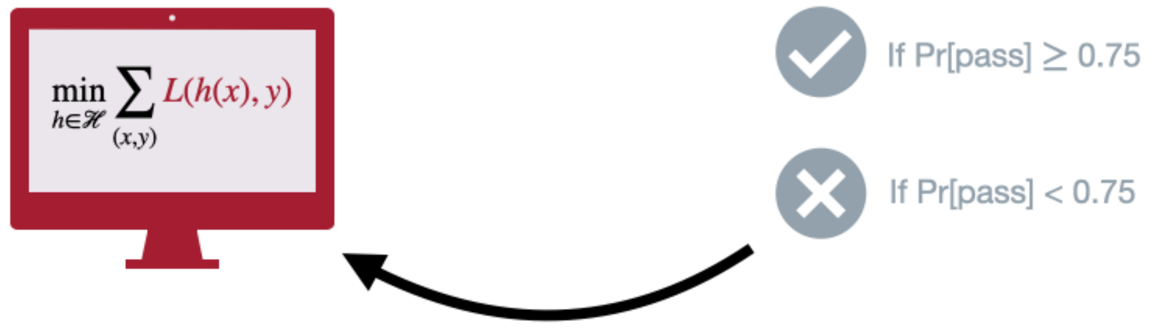
hinge(u, 1)

hinge(u, -1)



$embedding$(no) = -1 $embedding$(yes) = 1

| | Y = 1 | Y = -1 |
|---|---|---|
| **Yes** | 0 | 1 |
| **No** | 1 | 0 |




# Hinge loss embeds (twice) 0-1 loss

**Surrogate loss** $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_+$ embeds a **Decision loss** $\ell : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}_+$ if there exists an $embedding : \mathcal{R} \to \mathbb{R}^d \ldots$

**2. Optimal reports match on embeddings**

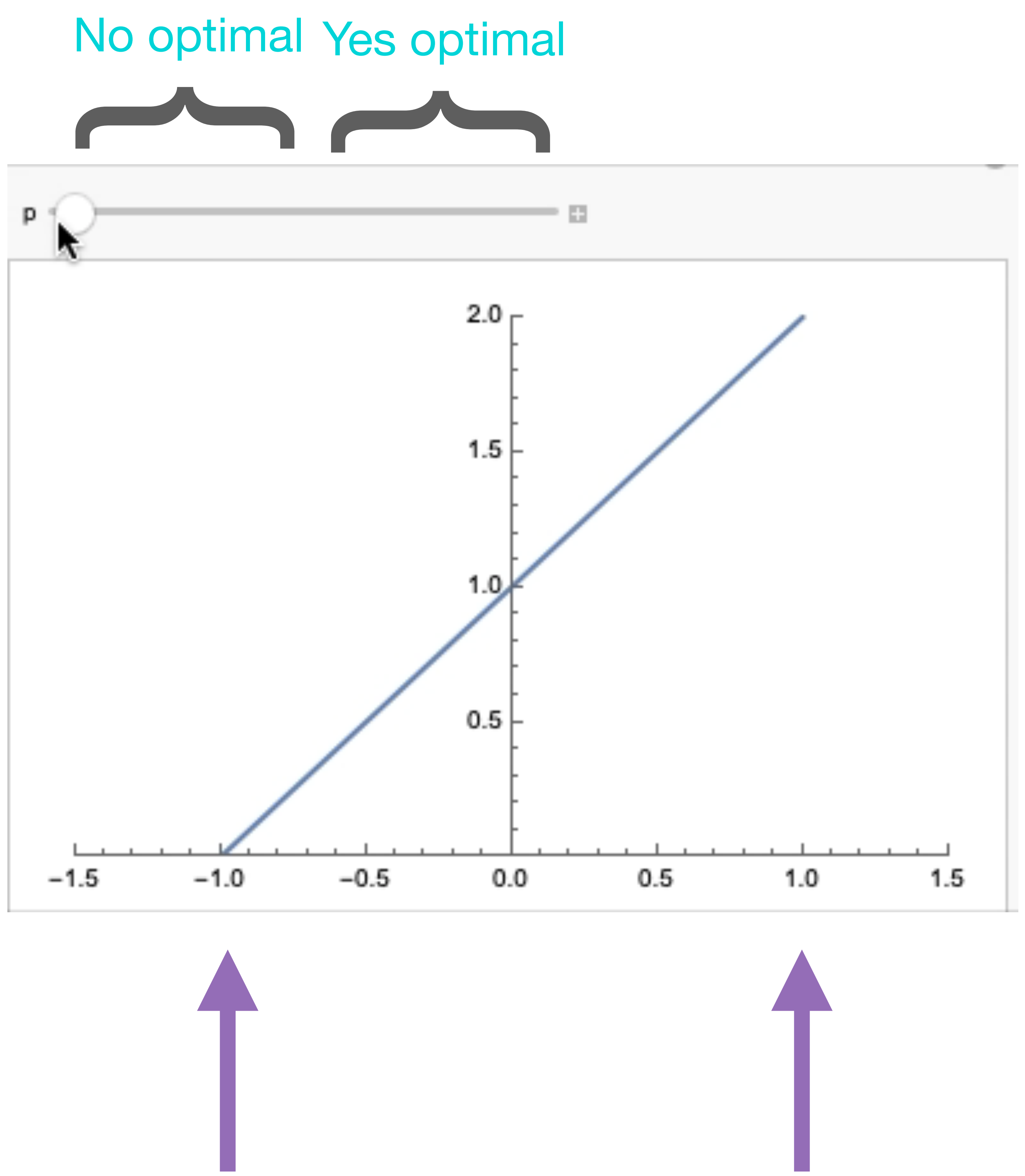
| | **Y = 1** | **Y = -1** |
|---|---|---|
| **Yes** | 0 | 1 |
| **No** | 1 | 0 |

Let $p$ be $\Pr[Y = 1]$. Then $1 - p = \Pr[Y = -1]$

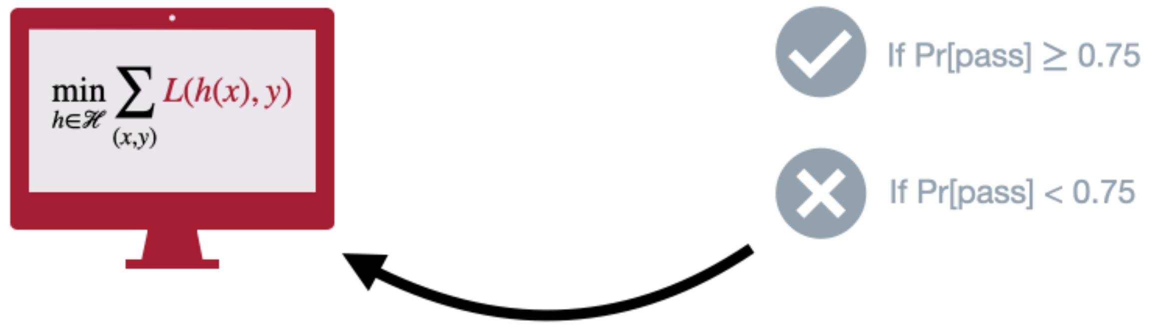
$$\mathbb{E}_{Y \sim p} \ell(Yes, Y) = \sum_y p_y \ell(Yes, y) = 0 \times p + (1 - p) \times 1 = 1 - p$$

$$\mathbb{E}_{Y \sim p} \ell(No, Y) = \sum_y p_y \ell(No, y) = 1 \times p + 0 \times (1 - p) = p$$

Yes optimal iff $p \geq \frac{1}{2}$




embedding(No) embedding(Yes)




# PLC embeddings

**Theorem (FFW19):** Every (PLC) surrogate embeds a decision loss

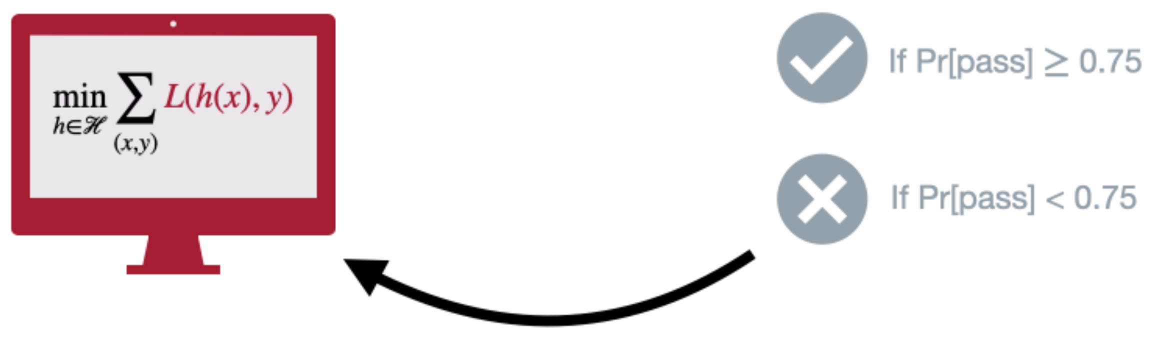
Consistency

Piecewise linear and convex (PLC) surrogate

Decision loss

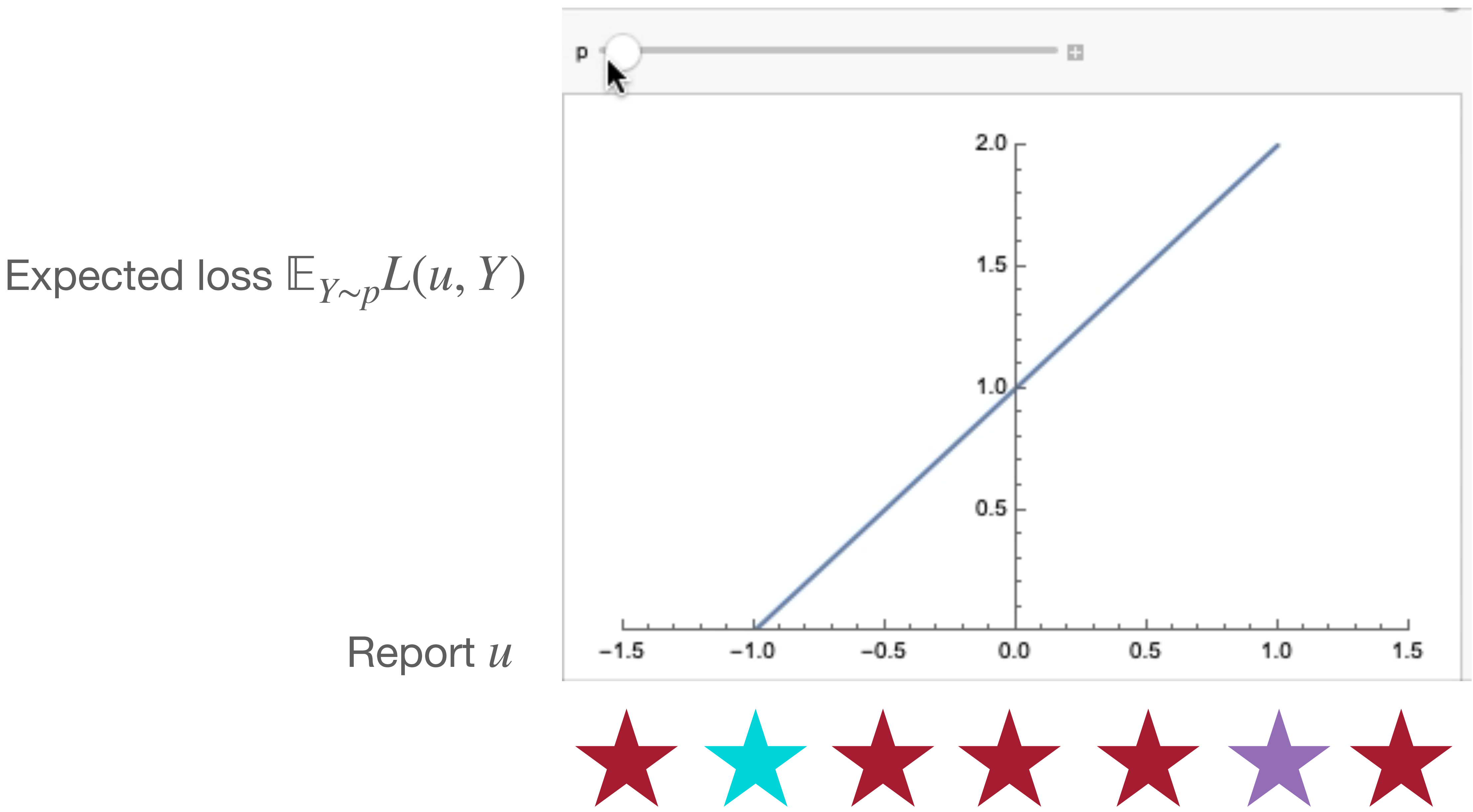
**Theorem (FFW19):** Every decision loss can be embedded by a PLC loss

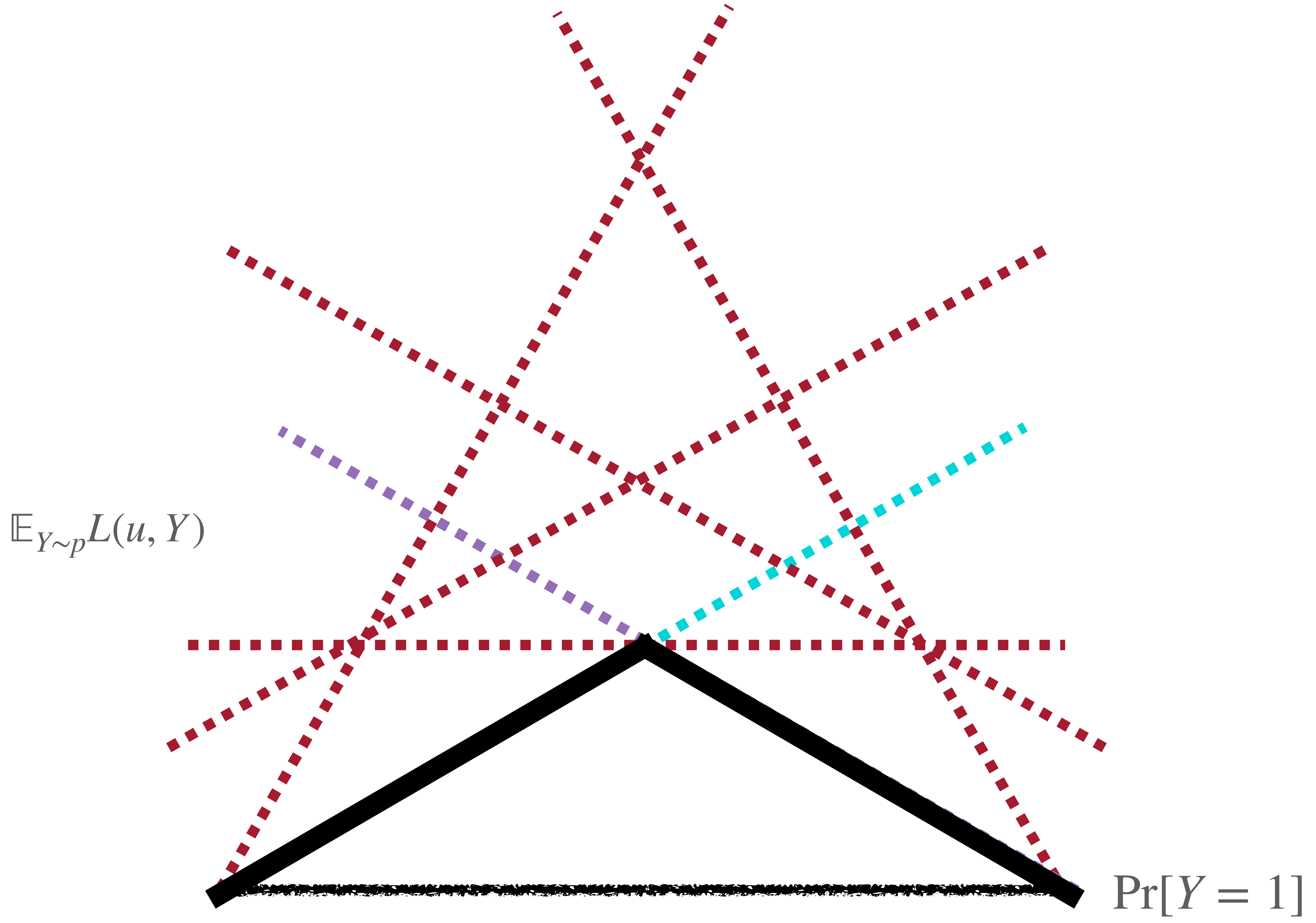
# Theorem (FFW19): Every (PLC) surrogate embeds a decision loss



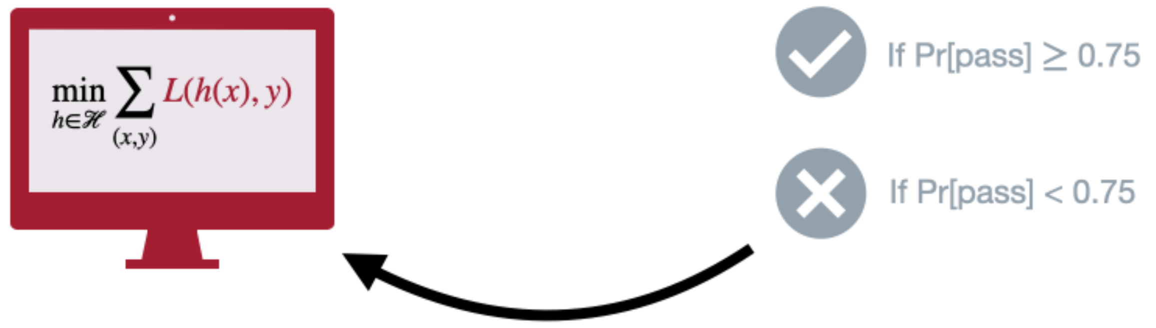

**Intuition: PLC surrogates have PLC Bayes risks $\underline{L} : p \mapsto \inf_u \mathbb{E}_{Y \sim p} L(u, Y)$**

$$\mathbb{E}_{Y \sim p} L(u, Y) = \sum_y p_y L(u, y) = \sum_y p_y c_y = \langle p, c \rangle$$

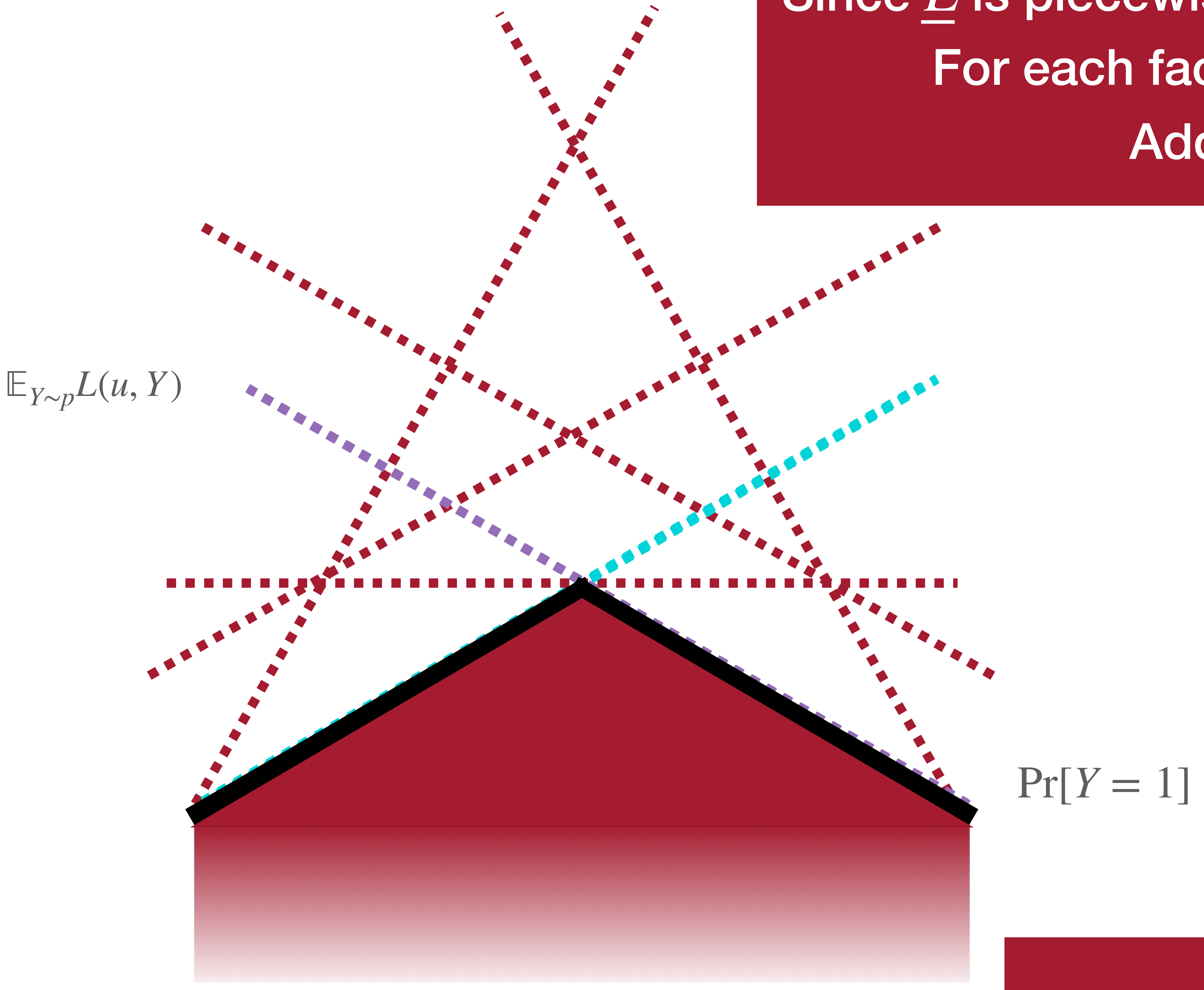


# <u>Theorem</u> (FFW19): Every (PLC) surrogate embeds a decision loss




Since $\underline{L}$ is piecewise linear and concave, its hypograph $hypo(\underline{L})$ has finitely many facets.
For each facet $F$, pick one report $u$ such that $\langle u, p \rangle$ supports $hypo(\underline{L})$ on $F$.
Add the row $\{L(u, y) \mid y \in \mathcal{Y}\}$ to the decision loss matrix.

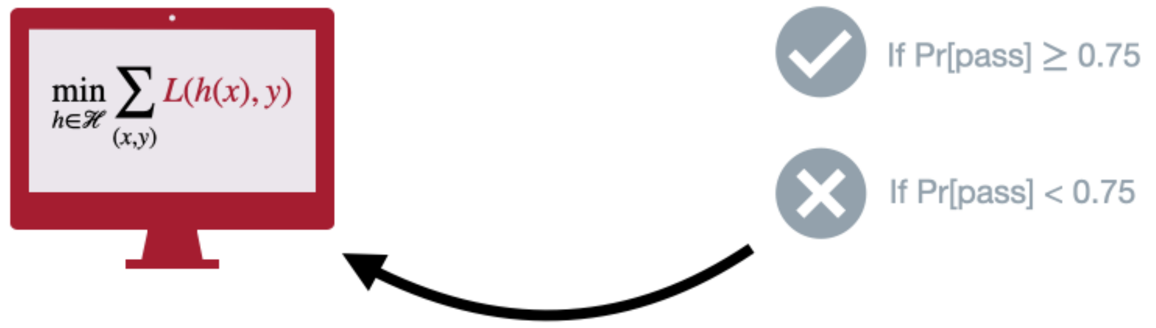



True outcome $y \in \mathcal{Y}$

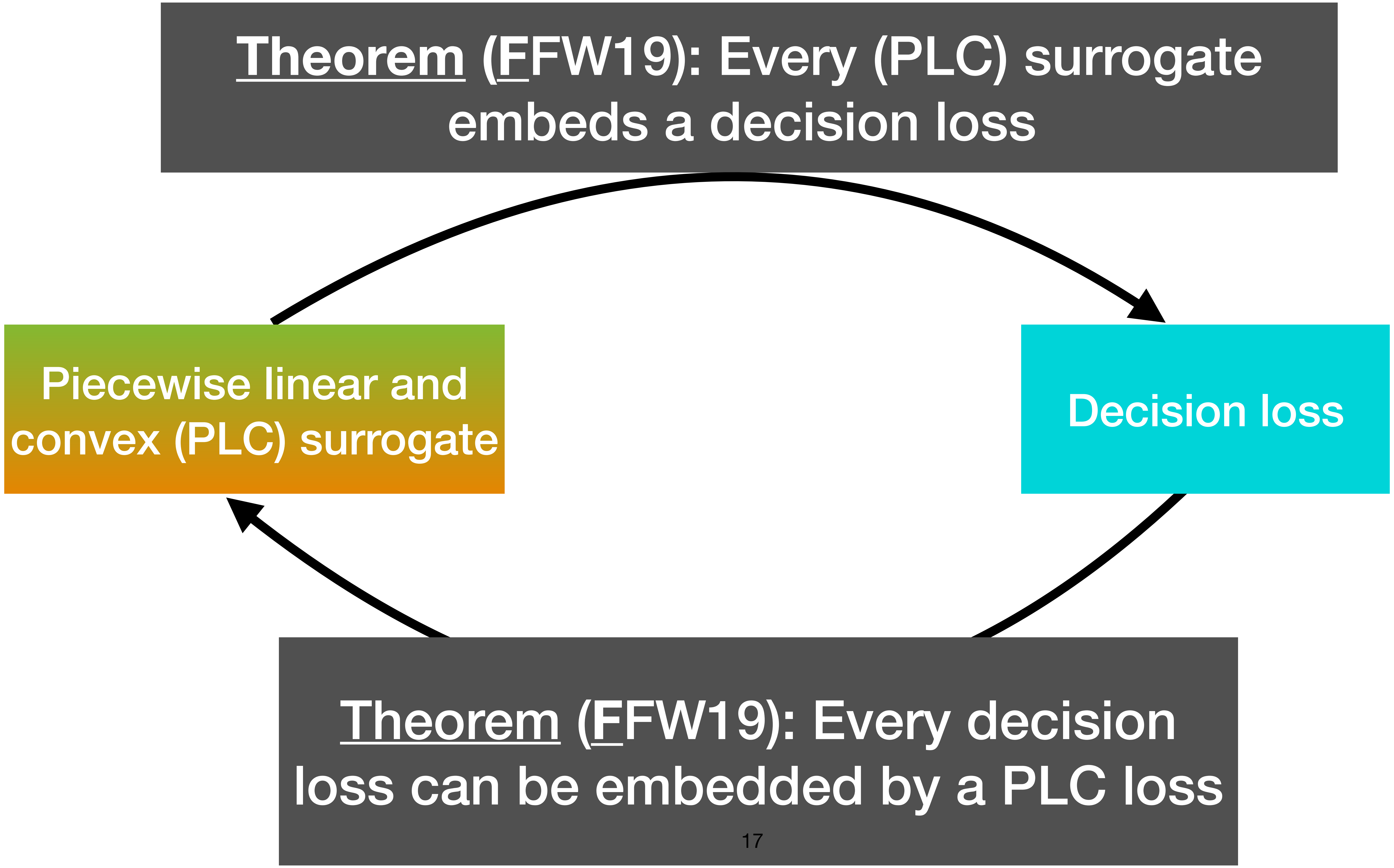
Discrete report $r \in \mathcal{R}$

| | -1 | 1 |
|---|---|---|
| -1 | 0 | 2 |
| 1 | 2 | 0 |

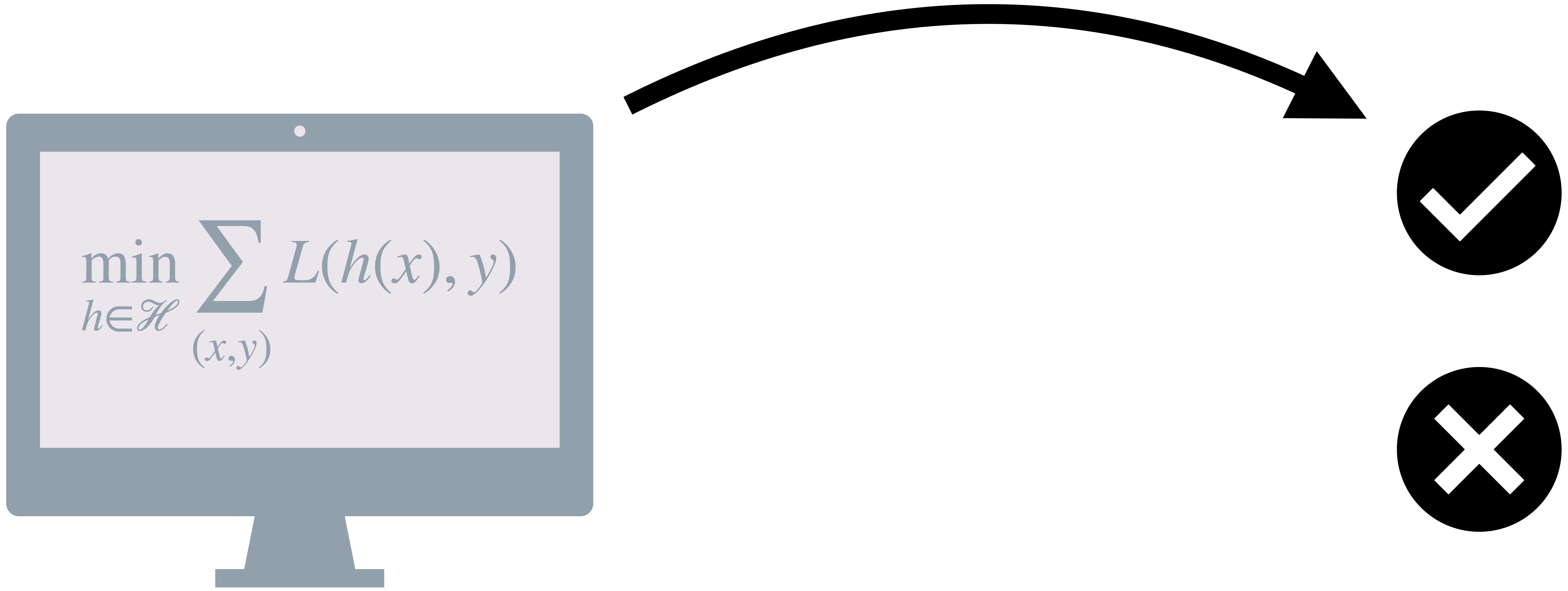
Is it an embedding?
Match loss values: by construction ✓
Match optimality: Bayes risks match, which means optimality matches ✓




# PLC embeddings

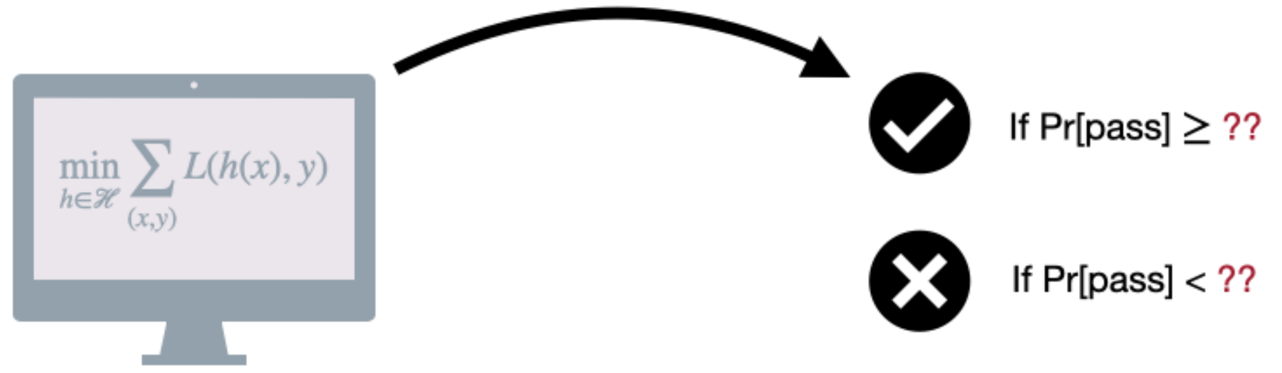


# Analyzing fixed *embeddings*



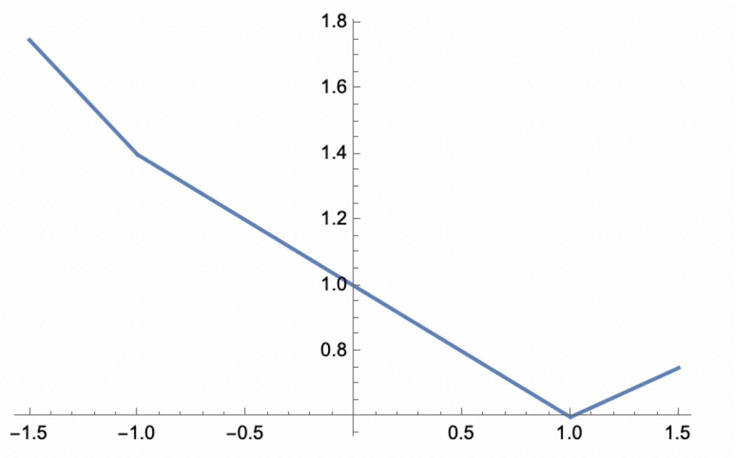

If Pr[pass] $\geq$ ??

If Pr[pass] < ??



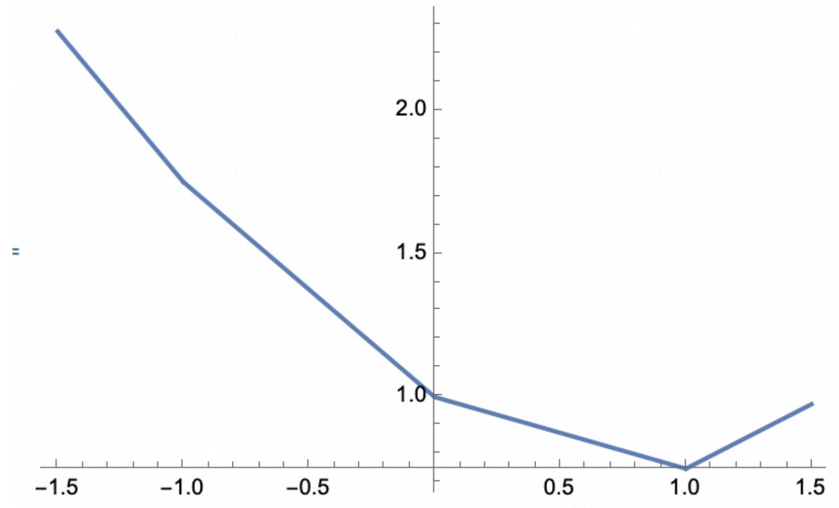

# Analyzing inconsistency of proposed embeddings

| Proposed surrogate | Embedded loss | Desired decision loss |
|---|---|---|



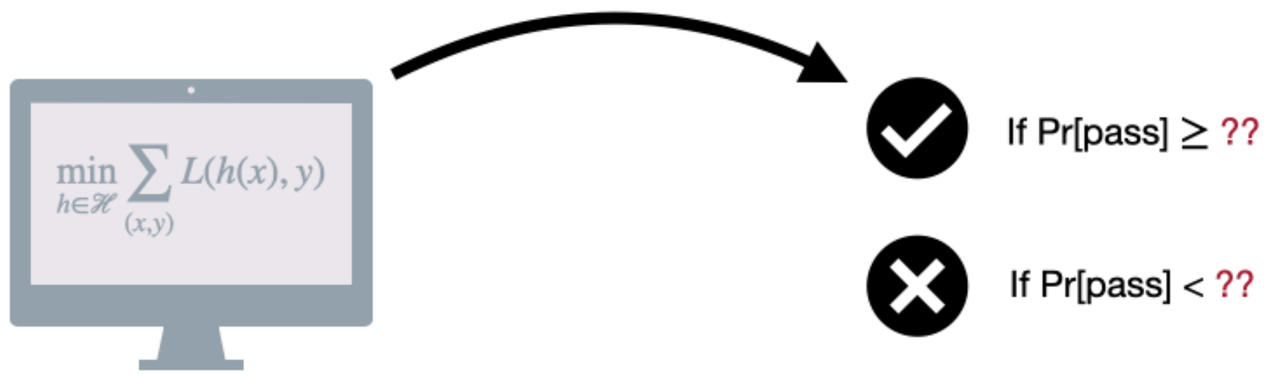

| | Y = 1 | Y = -1 |
|---|---|---|
| **Yes** | 0 | 2 |
| **No** | 2 | 0 |

| | Y = 1 | Y = -1 |
|---|---|---|
| **Yes** | 0 | 1 |
| **No** | 1 | 0 |

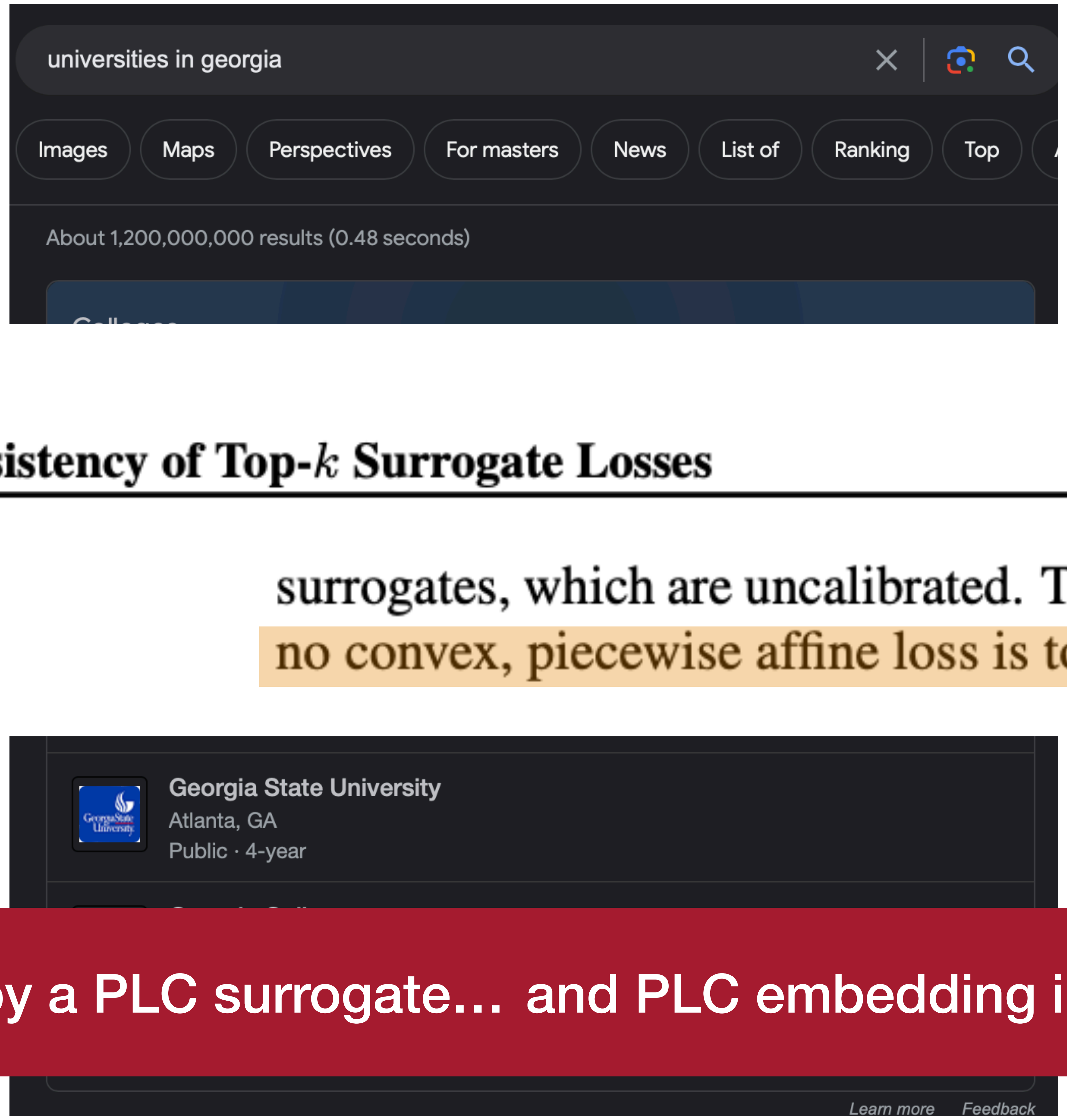



| | Y = 1 | Y = -1 |
|---|---|---|
| **Yes** | 0 | 1 |
| **Maybe** | 1/3 | 1/3 |
| **No** | 1 | 0 |

| | Y = 1 | Y = -1 |
|---|---|---|
| **Yes** | 0 | 1 |
| **No** | 1 | 0 |




# Analyzing (in)consistency for common decision tasks


$$\ell(r, y) = \ell_{r,y}$$

y = 3

1

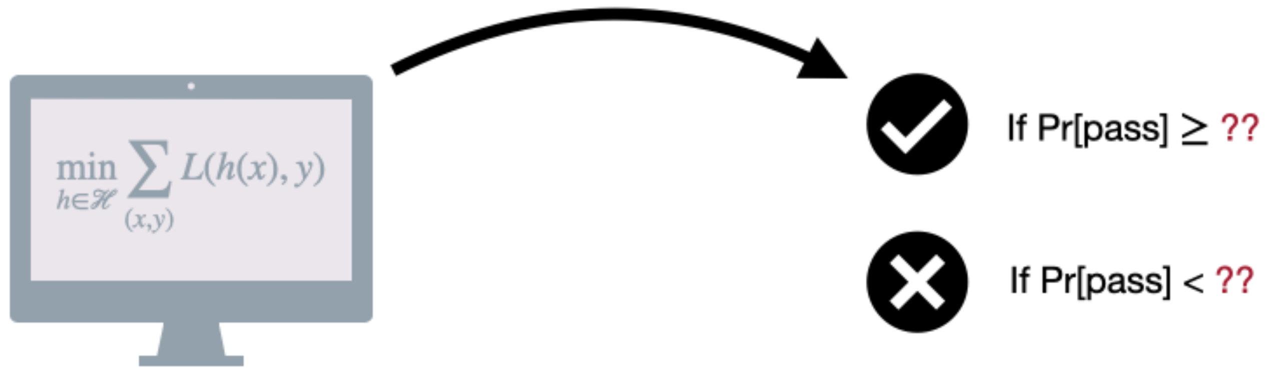
0

YK ICML 20

**On the Consistency of Top-$k$ Surrogate Losses**

surrogates, which are uncalibrated. Thus, we conjecture that no convex, piecewise affine loss is top-$k$ calibrated.

But we know top-*k* is embedded by a PLC surrogate... and PLC embedding implies consistent (calibrated)



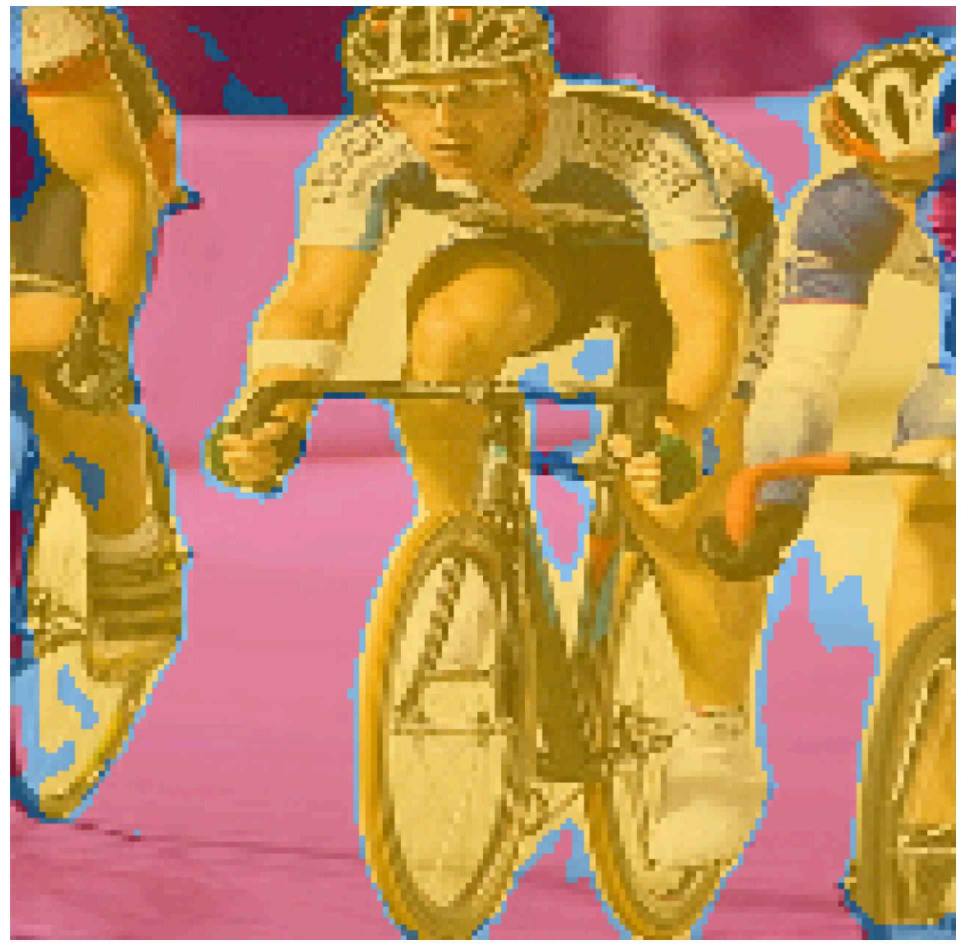

# Analyzing (in)consistency for common decision tasks

PLC surrogates for top-$k$ prediction (FFGT ICML 22)

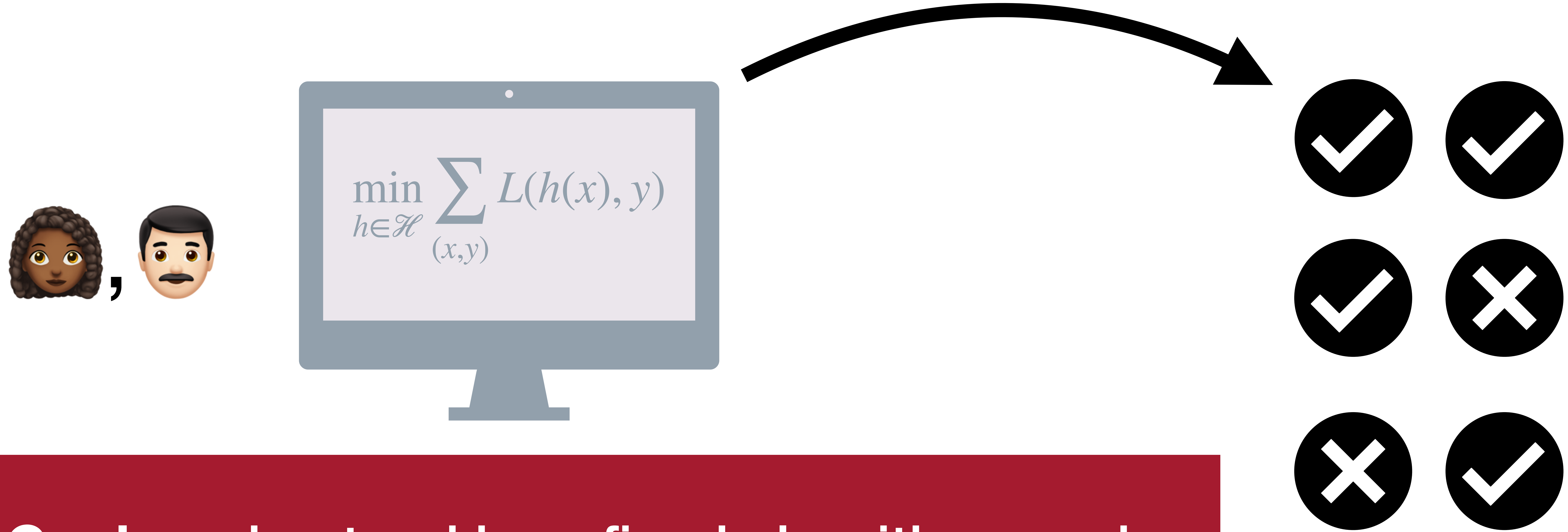
SVM generalizations for structured prediction (NBR, ICML 20)

Weston-Watkins hinge embeds the ordered partition (WS NeurIPS 20)

Lovász hinge for structured prediction (FFN COLT 22)



# Analyzing fixed algorithms: beyond pointwise predictions




If P(👩🏾, 👨🏻) …?

If P(👩🏾, 👨🏻) …?

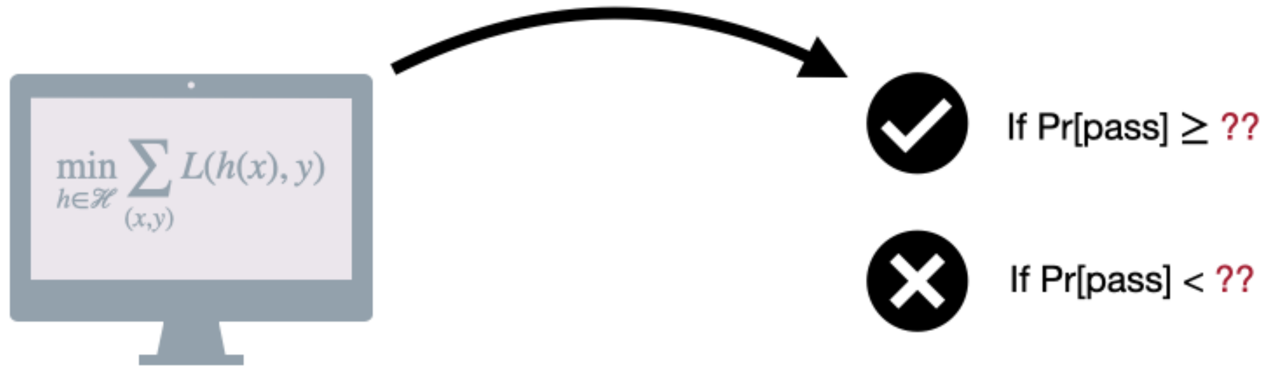
If P(👩🏾, 👨🏻) …?

**Goal: understand how fixed algorithms make decisions in various settings**

Challenges:
- Need to codify inherently abstract concepts
- Limitations on expressing utility

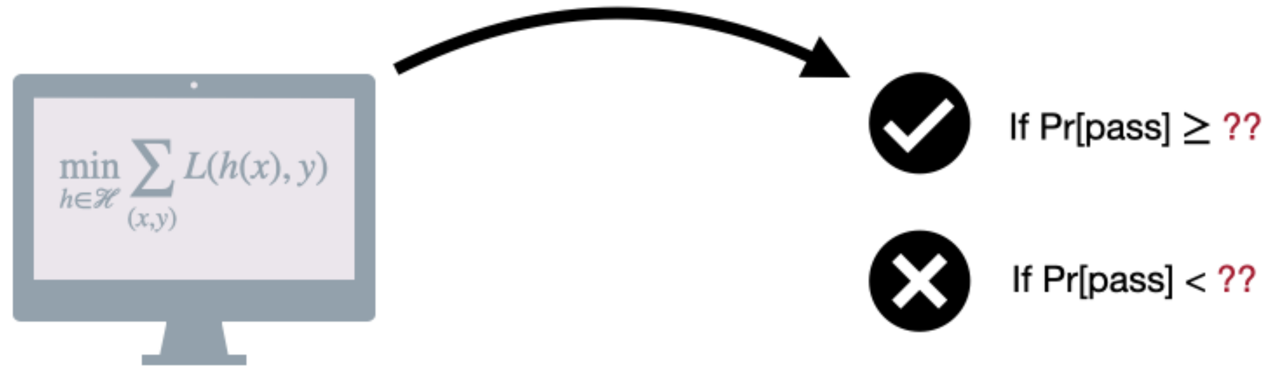
F23 arXiv

# Sometimes algorithm is fixed




$$\min_{prediction} \quad (1-\lambda) loss(prediction, outcome) + \lambda\ unfairness(prediction, outcome)$$

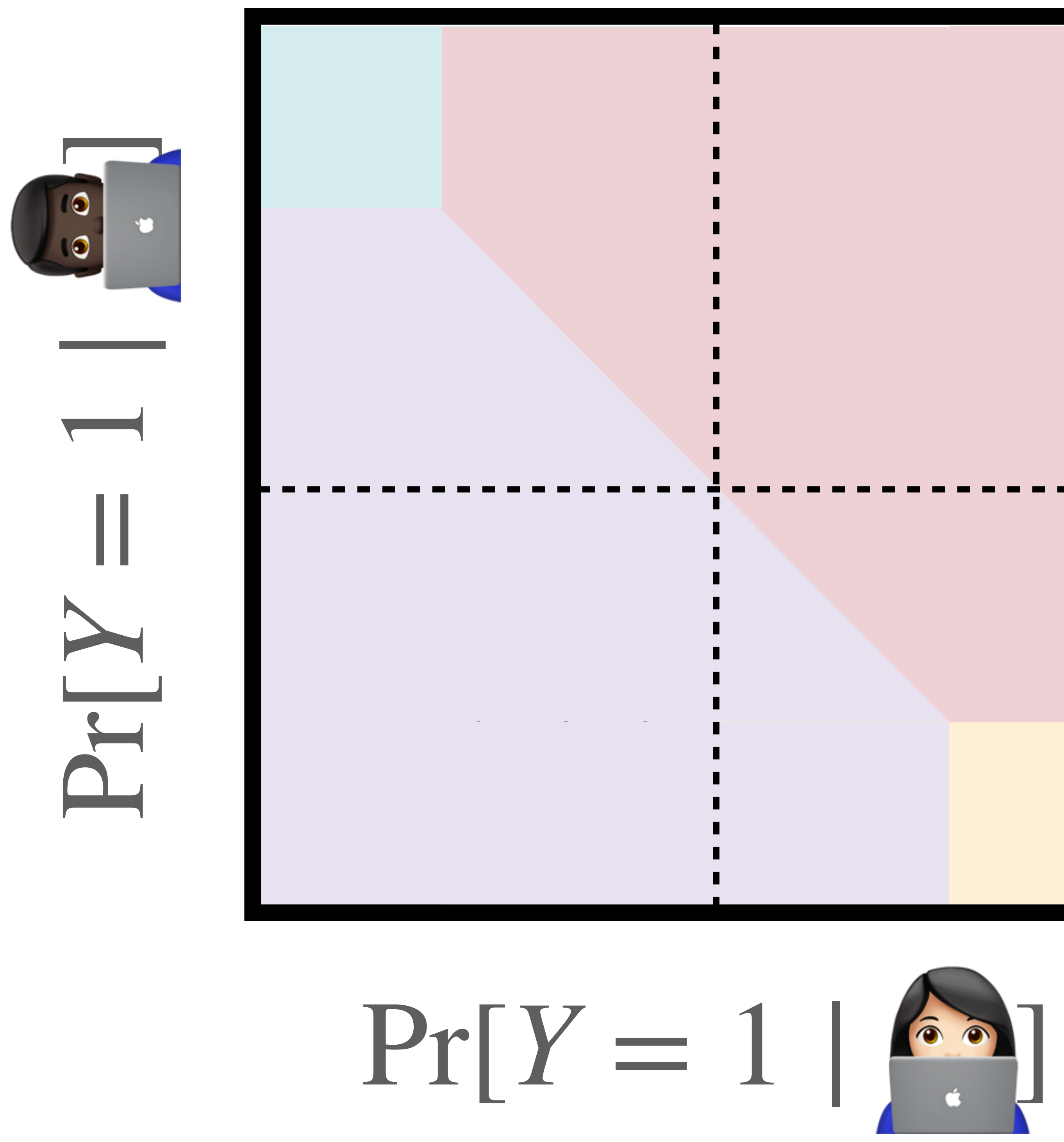
| | "True probability" | **Unconstrained decision** | Constrained (DP) decision | Constrained (FPR) decision |
|---|---|---|---|---|
|  | 0.52 | ✅ | ✅ | ❌ |
|  | 0.49 | ✅ | ✅ | ❌ |

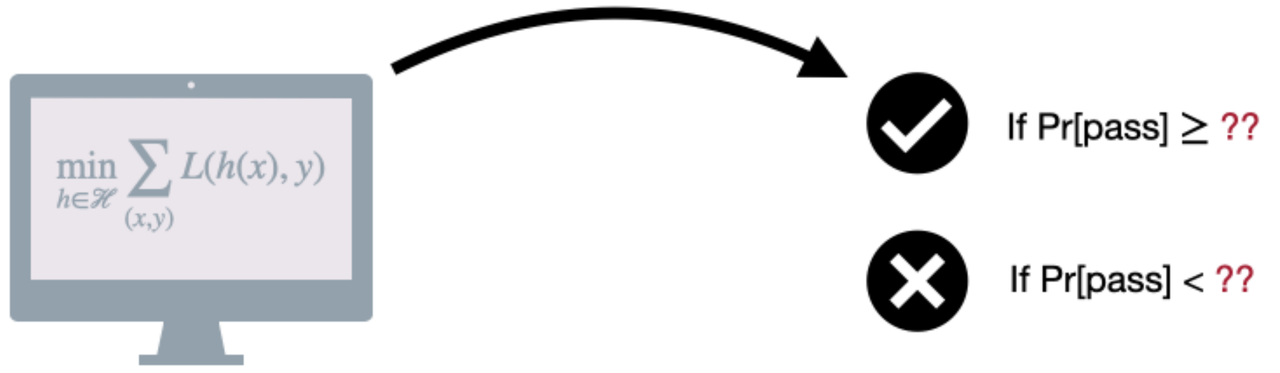



# How do fairness constraints change decisions?

**(Theorem F23): Decision-making is the same for every distribution iff the unfairness metric is "basically the same" as the loss $L$**

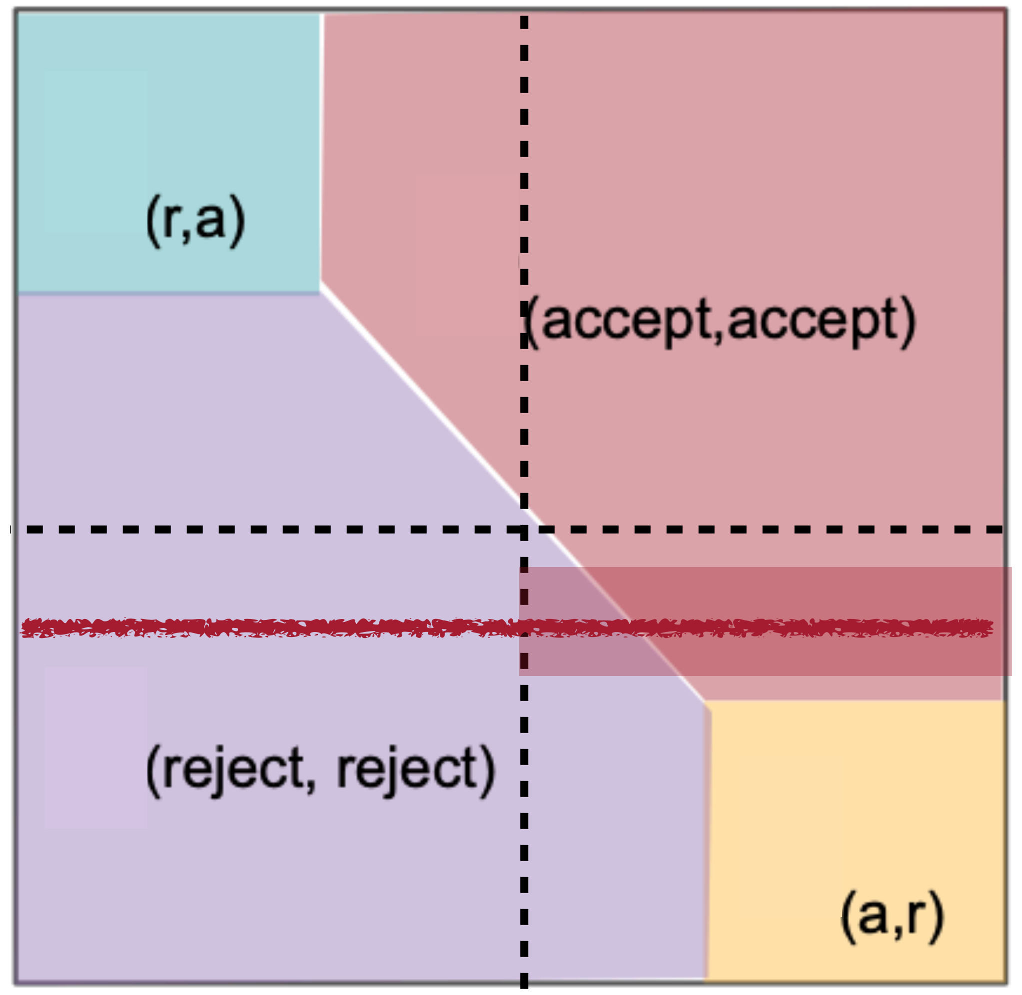
Demographic Parity



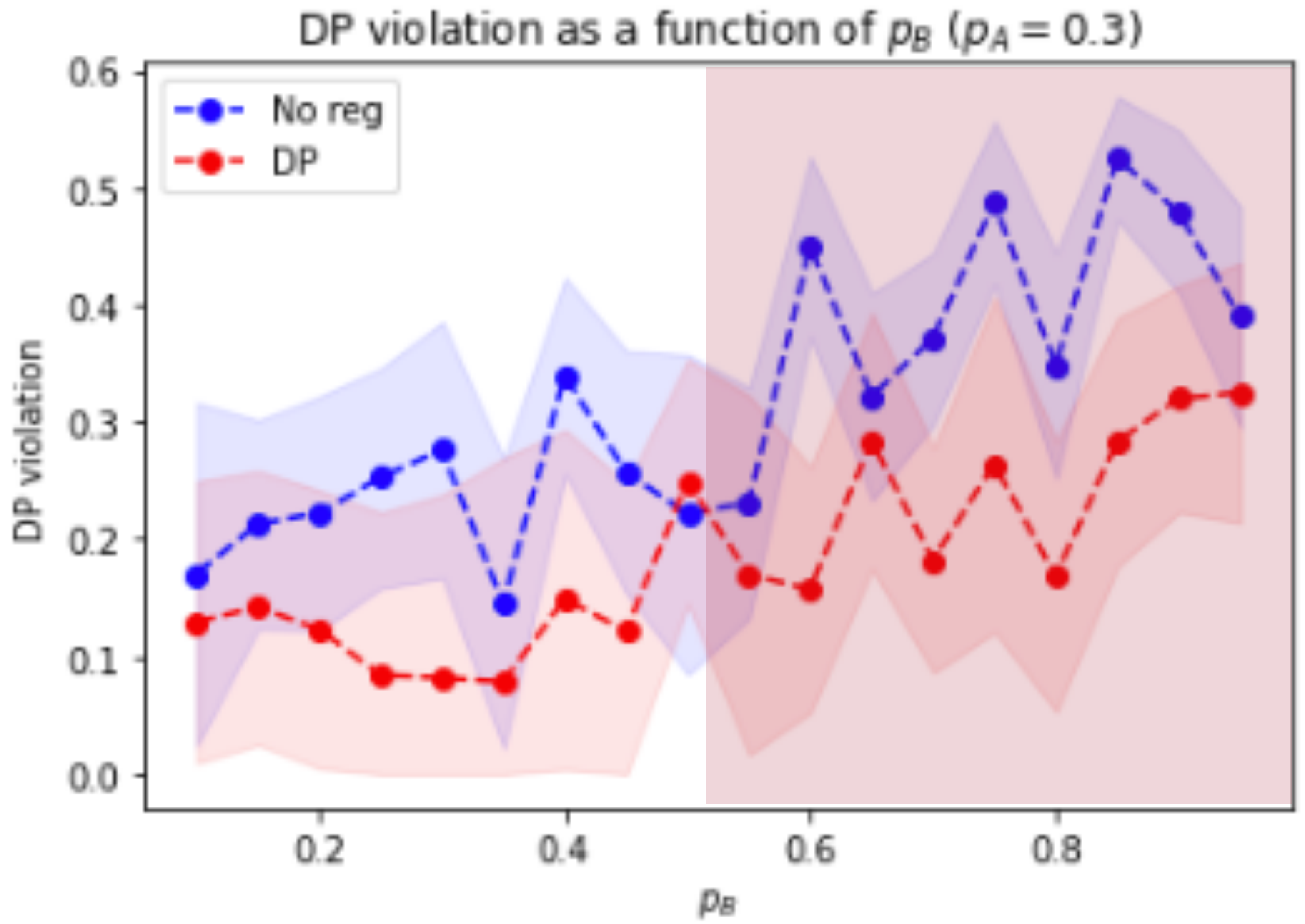


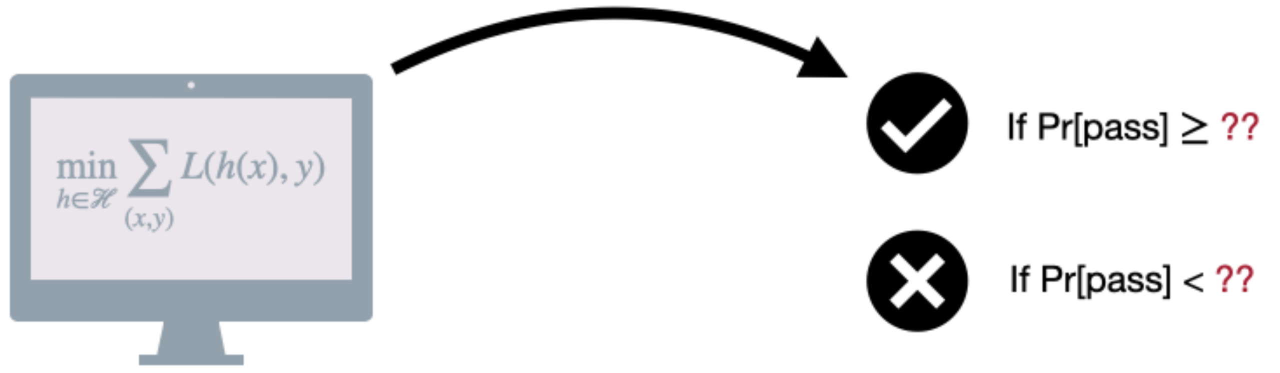

# How do fairness constraints change decisions?

(**Theorem F23**): Decision-making is the same for every distribution iff the unfairness metric is “basically the same” as the loss $L$

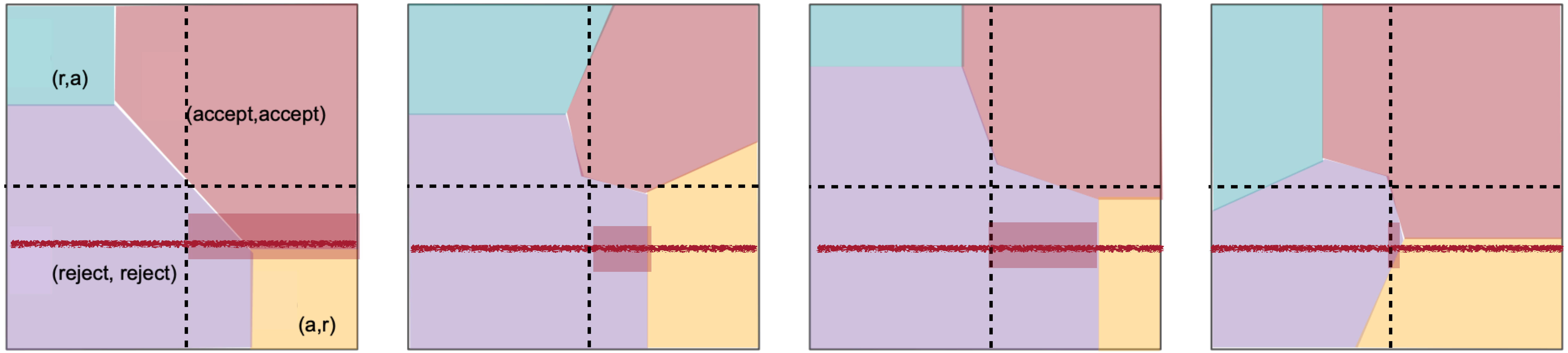



Demographic Parity

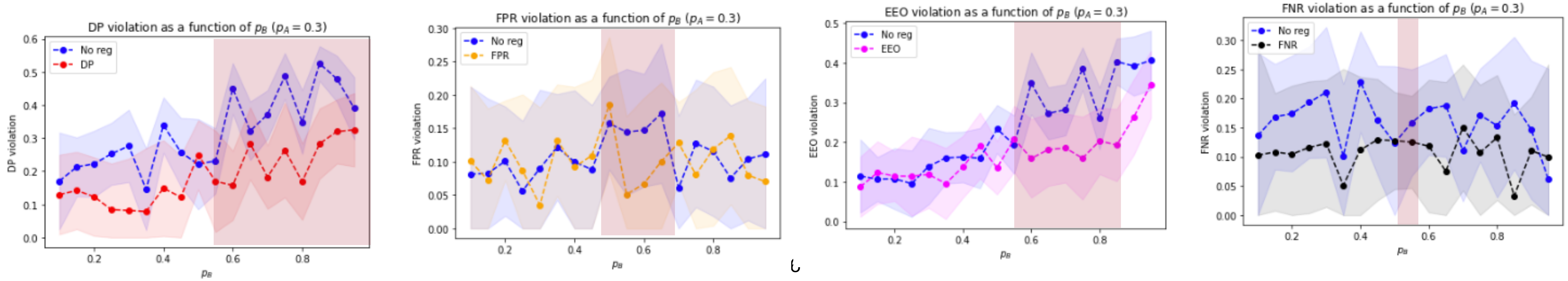





# Comparing unfairness metrics




Demographic Parity | False Positive Rates | Equalized Odds | False Negative Rates



# Beyond today's talk: research

**Machine Learning/AI**

**Algorithmic Game Theory**

Bridging Fairness in Machine Learning and Mechanism Design
FMMPRST21 FAccT

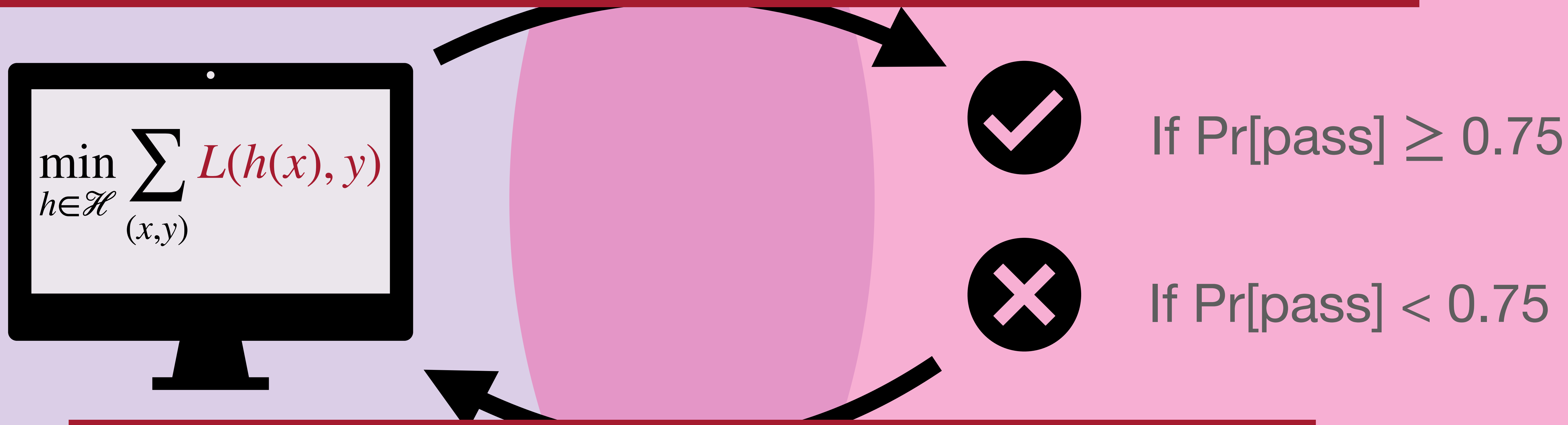
Impacts of fairness constraints in information sharing
SFMNRJ23 AAAI

Voting algorithms with anchoring bias
CF in submission

Robustness of predict-then-optimize algorithms
JFWSVTT23 GameSec

Resource allocation with inequality-averse communities
SFA in submission

Holistically analyzing decisions made by fixed algorithms




Designing objective and decision functions

Convex losses for continuous decisions
FF18 NeurIPS

Computational challenges around loss efficiency
FFW20 COLT, FFW21 NeurIPS

Designing decision functions for structured prediction
FFN22 COLT

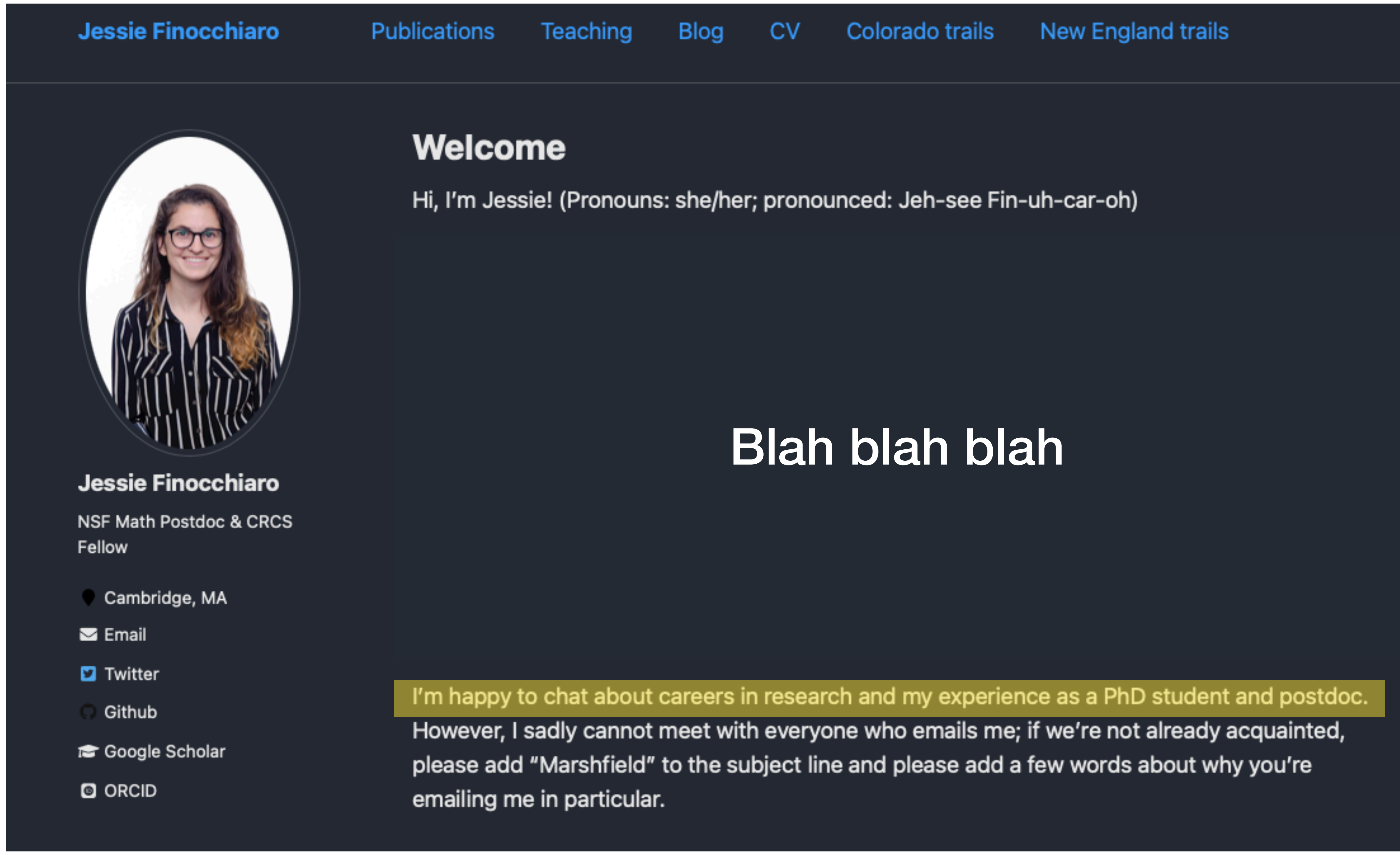
Learning to cooperate in competitive games
FM20 IEEE ToG

# Beyond research: outreach and mentorship

MD4SG
Mechanism Design for Social Good

Community engagement lead
Working group on fairness and discrimination co-lead

Chair, vice-chair,
Neural network
Piloting PhD applicant feedback program




PhD App mentorship and general Q+A!

QUEER in AI

PhD App mentorship
AAAI 2023 invited talk

PhD App mentorship

LXAI

PhD App mentorship

Optimization design is a *value choice*, often made difficult by *computational costs*.

My work *designs objectives* that aligns with stated values and *evaluates the consequences* of objective choice on algorithmic decision-making.

# Future work

Understand consequences of objective function choice

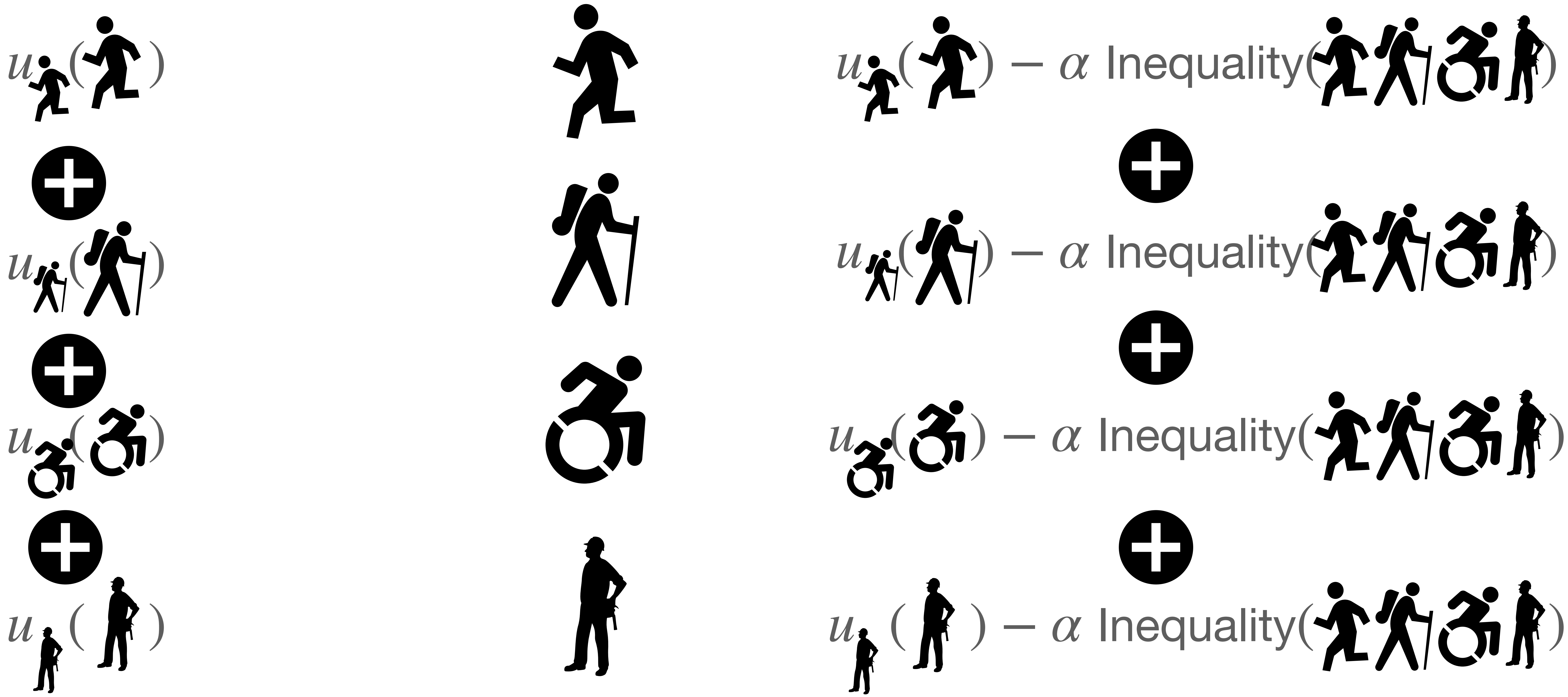

Under… s and limitations of using “smart” loss functions

Understand how to incorporate value choices into algorithm design

## Understand consequences of objective function choice

**Design algorithms to maximize…**

**But what if utilities are actually…?**



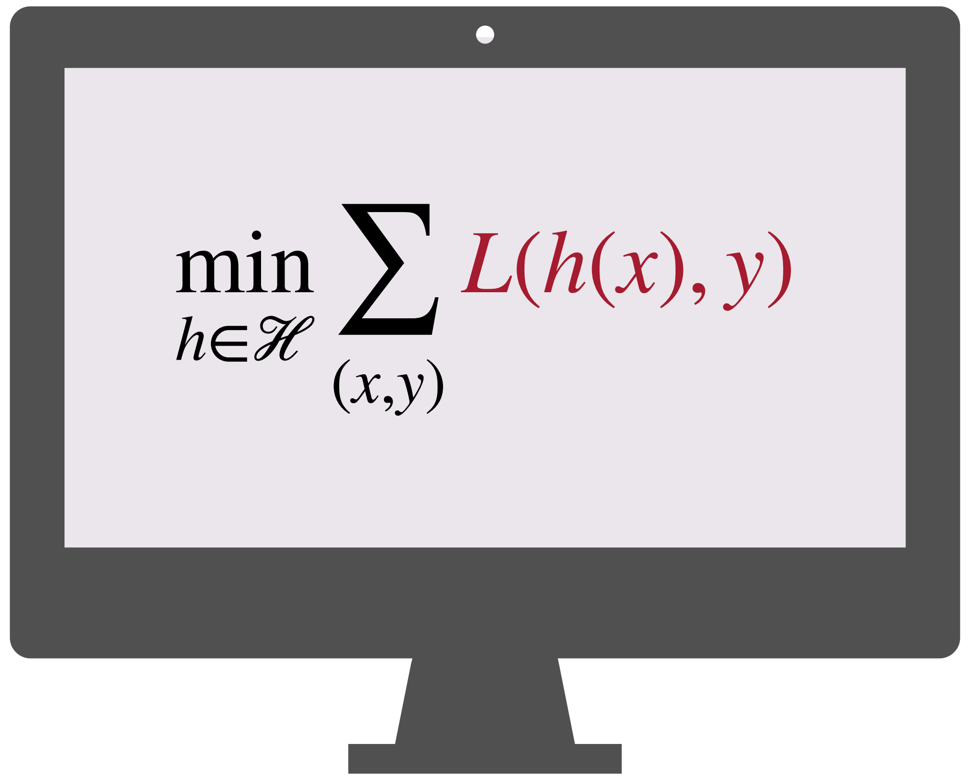

SFA in submission

# Future work: Understand how to incorporate value choices into algorithm design

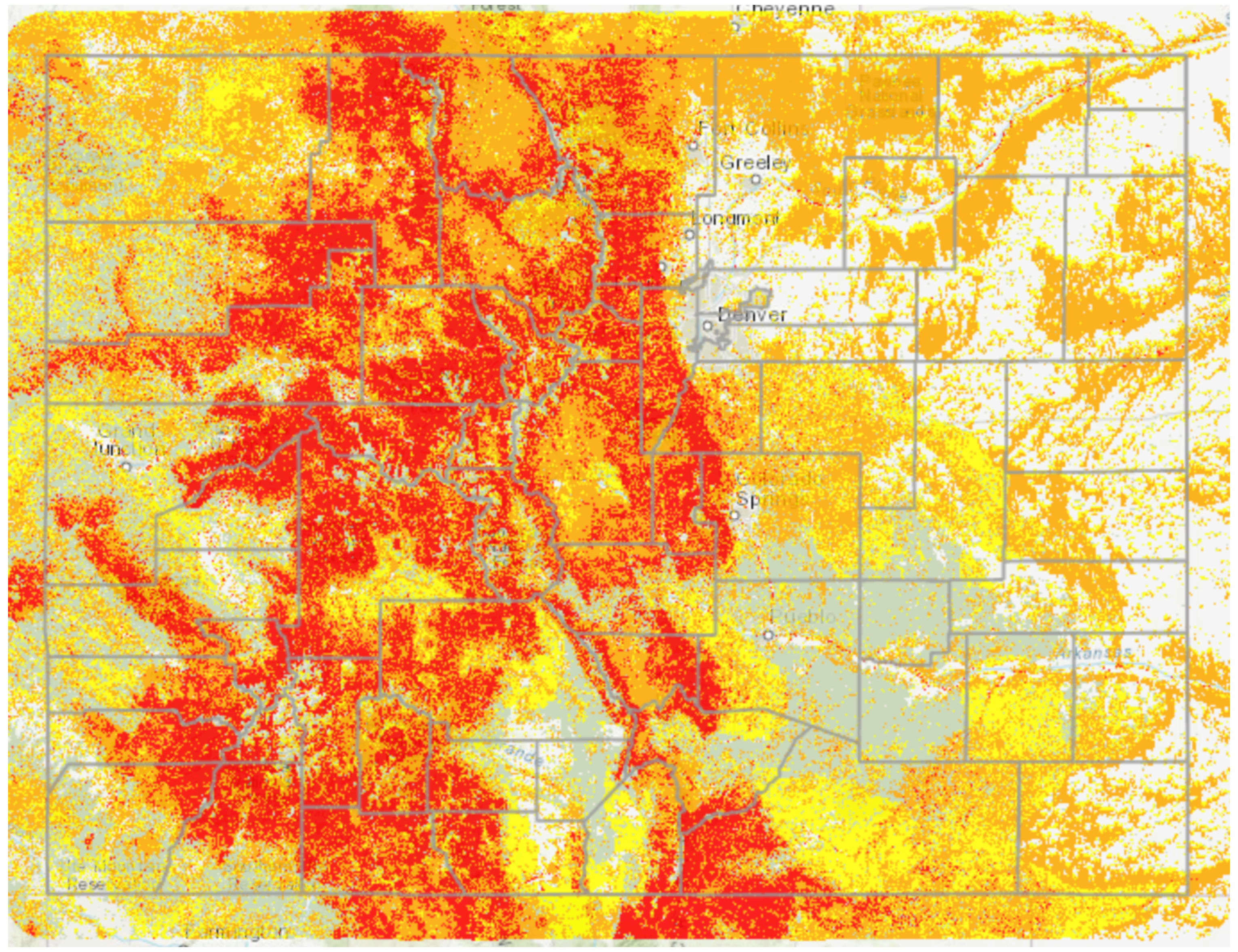
Table 1,
Private Forest Land Protection Criteria, 2020

| Criteria | Priority |
|---|---|
| Water Quality/Quantity | 1 |
| Wildlife Habitat | 2 |
| Growth/Sprawl Control | 3 |
| Large Continuous Forest | 4 |
| Wetland/Riparian Areas | 5 |
| Unique Ecological Areas | 6 |
| Wildfire Control Issues | 7 |
| Private Property Rights | 8 |
| Forest Timber Products | 9 |
| Lifestyle Protection for Landowner | 10 |

https://csfs.colostate.edu/wp-content/uploads/2020/11/FINAL2020_FLP_AON-.pdf



$$\min_{h \in \mathcal{H}} \sum_{(x,y)} L(h(x), y)$$




https://co-pub.coloradoforestatlas.org/#/

# Future work: Understand advantages and limitationss of using “smart” loss functions

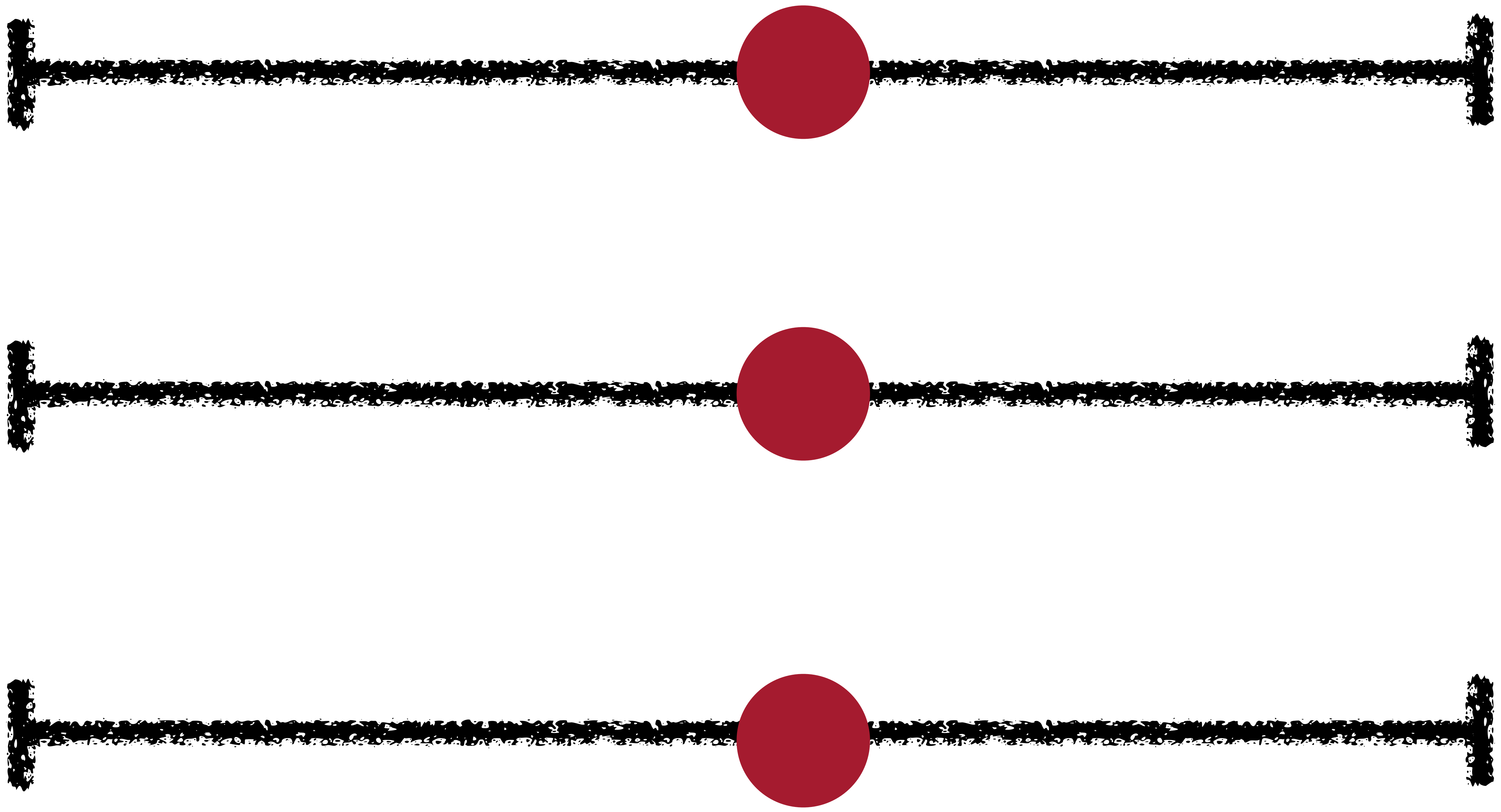
Clever loss functions help a lot

Don’t need clever loss functions!

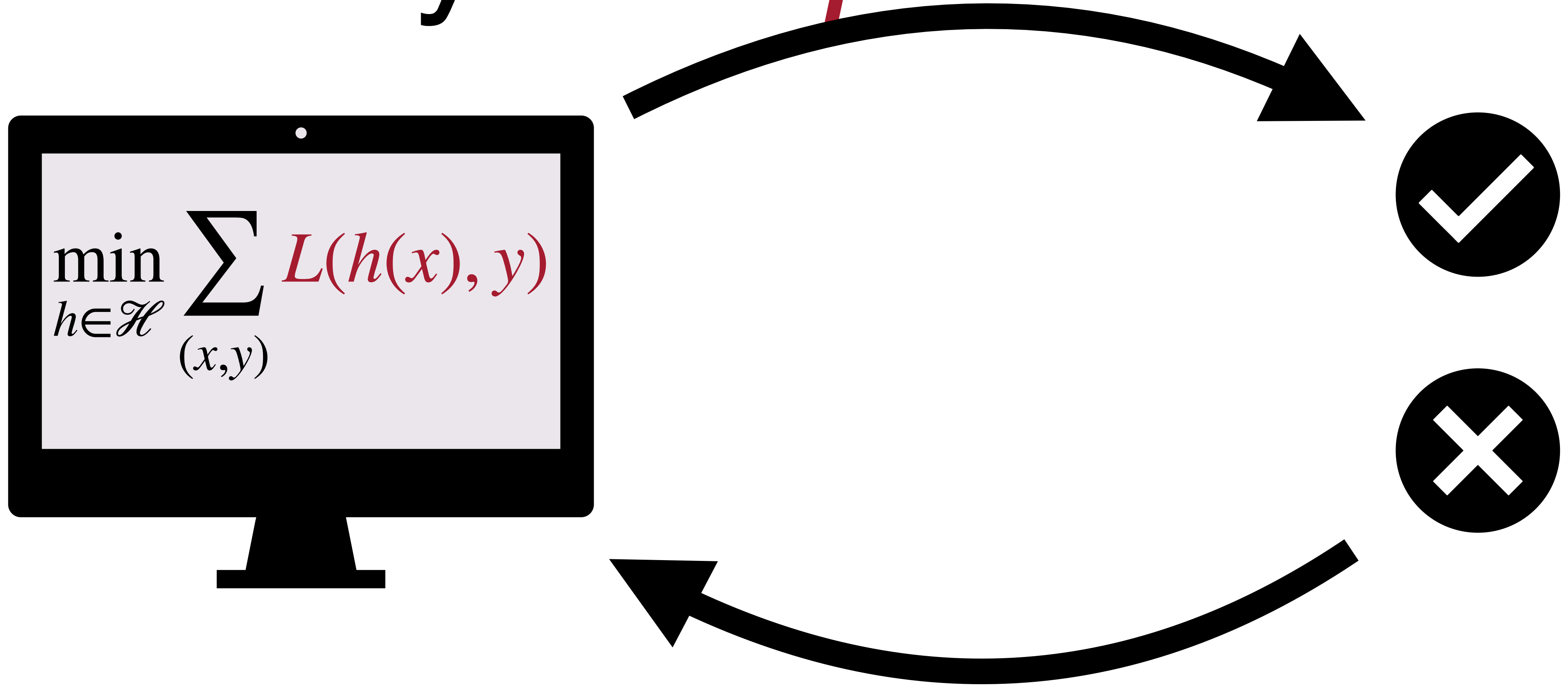
Model size

Training data size

Complexity of decision

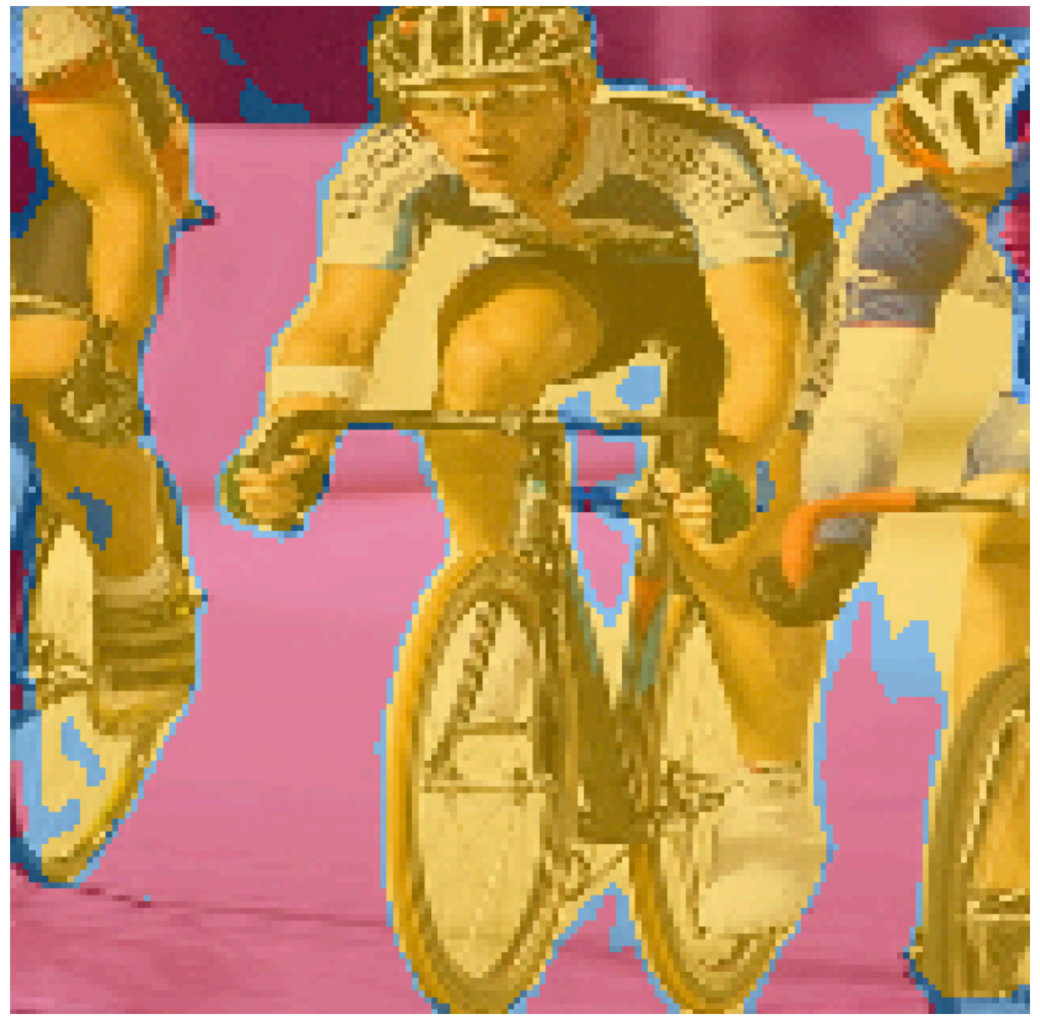


# Optimization design is a *value choice*, often made difficult by *computational costs*.




If Pr[pass] $\geq 0.75$

If Pr[pass] $< 0.75$



Handbook of Computational Social Choice
BCELP 2016

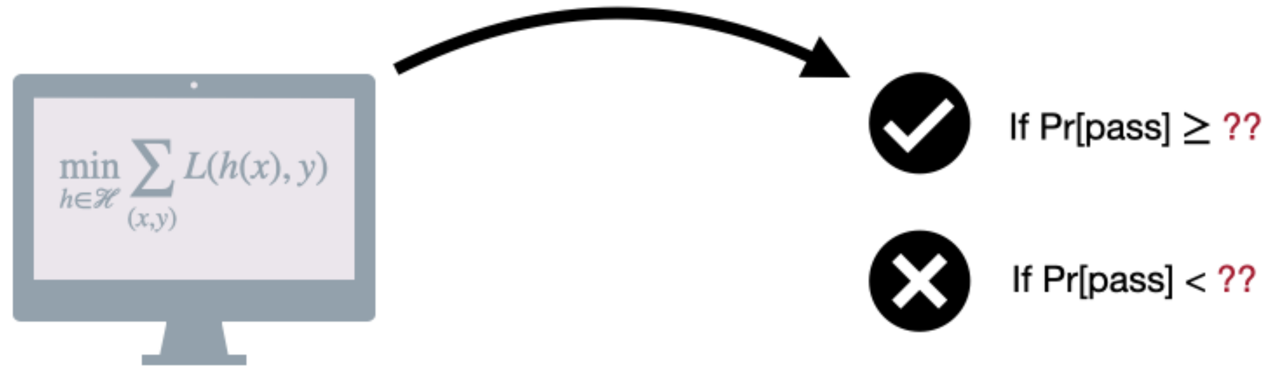
Wealthfront

https://www.wealthfront.com/

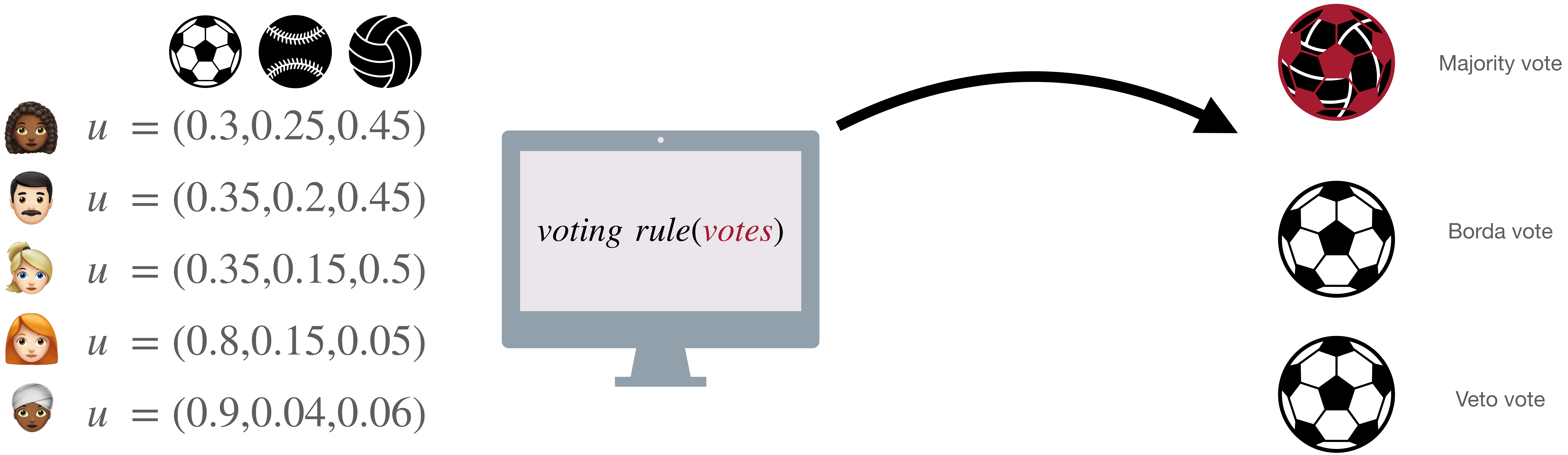
Collaborations with you!

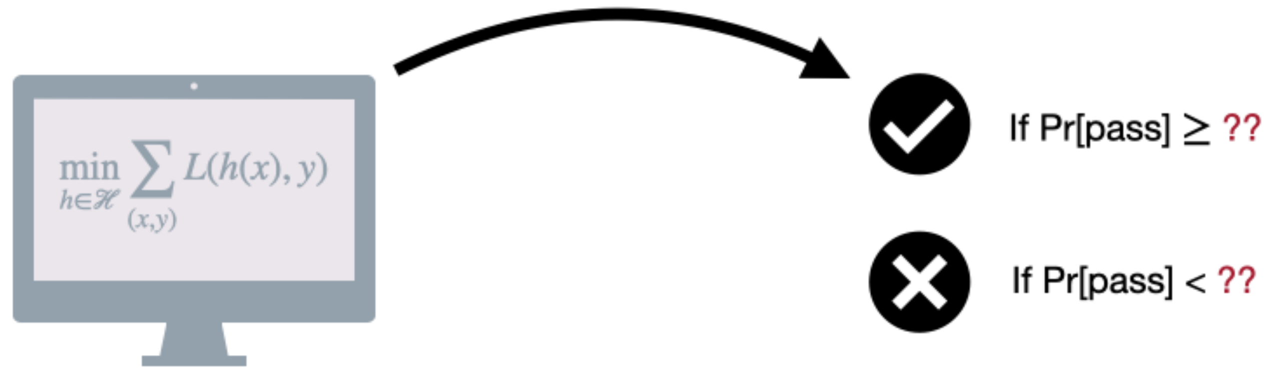
**Thank you**
**www.jessiefin.com**

# Appendix



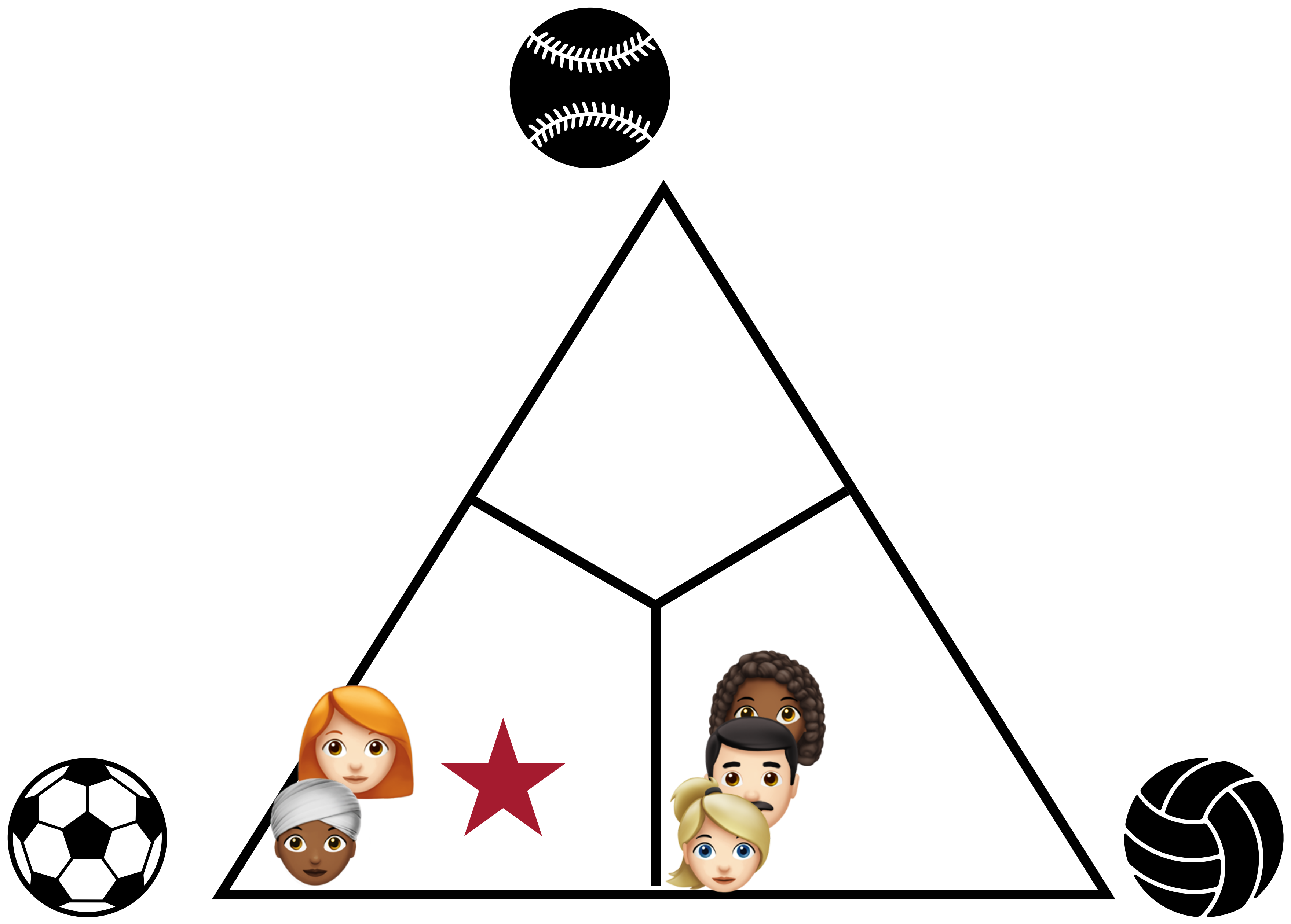

# Analyzing fixed algorithms: beyond prediction

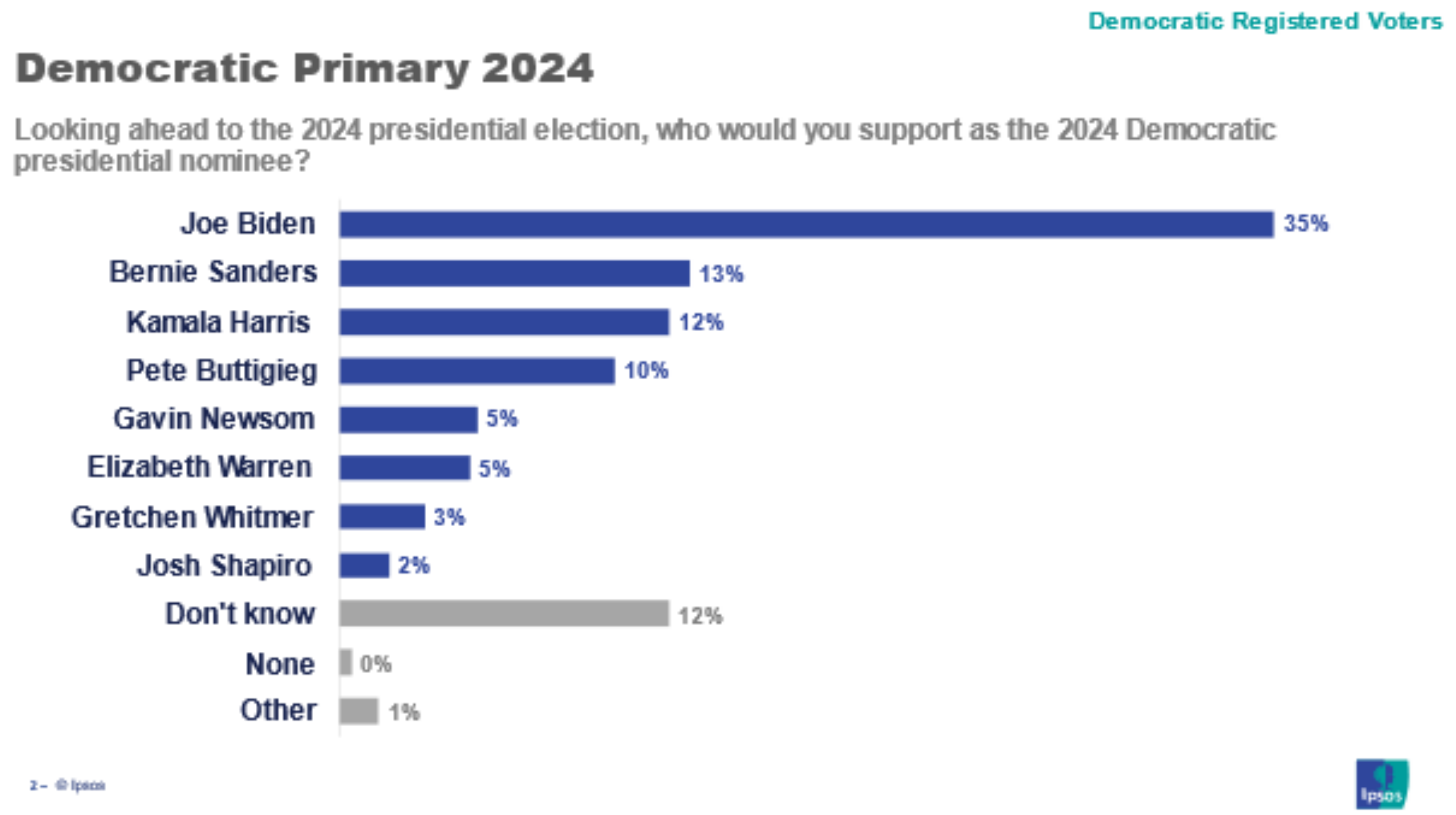




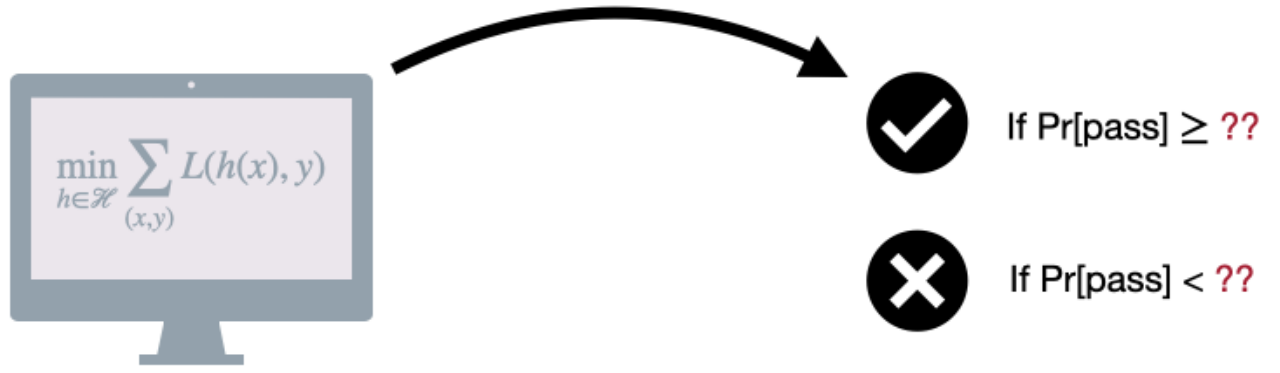

# Analyzing fixed algorithms with anchored play

How do algorithmic decisions change when inputs (peoples opinions) shift according to anchored preferences?



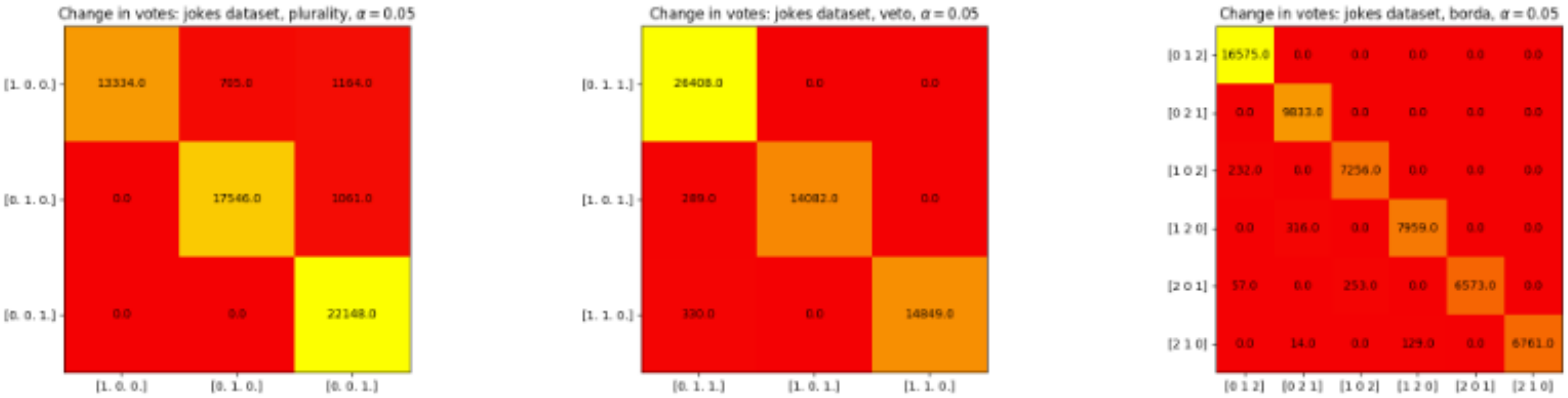



https://www.ipsos.com/en-us/trump-leads-republican-primary-field-biden-leads-democrats

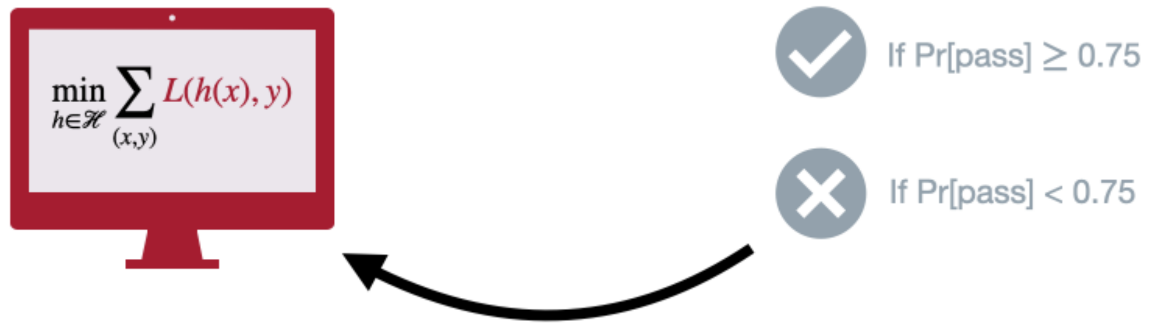



# Analyzing fixed algorithms with social play

(Proposition CF23): Individual votes align more closely with the anchoring point




(Theorem CF23): Borda is more robust to external information than plurality




# Why do we need to construct a decision function

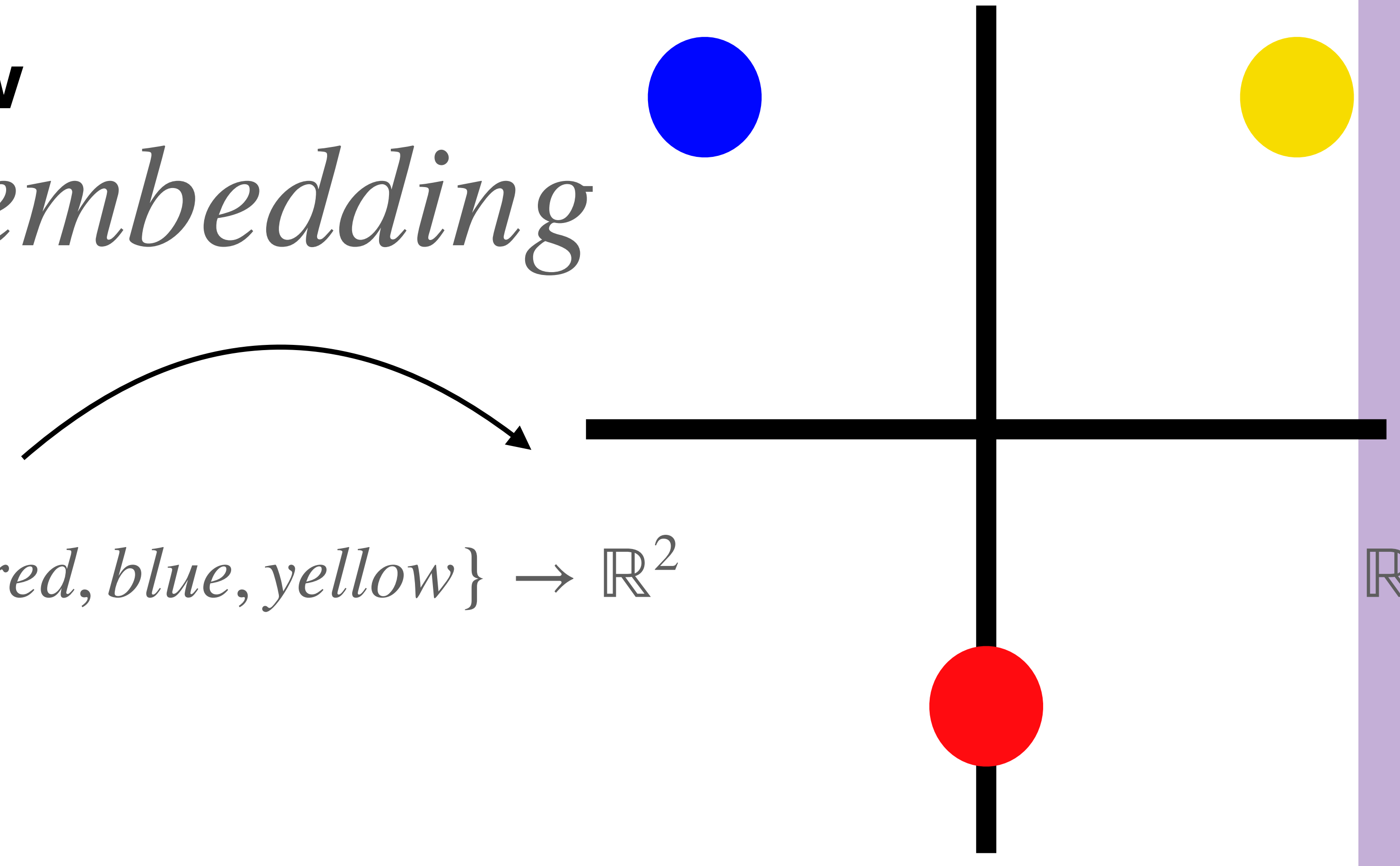
**Theorem** (FFW19): If a PLC surrogates $L$ embeds $\ell$, there exists a decision function $dec$ such that $(L, dec)$ is consistent with respect to $\ell$

Need $dec$ defined over all of $\mathbb{R}^d$

| | Red | Blue | Yellow |
|---|---|---|---|
| **Red** | 0 | 1 | 1 |
| **Blue** | 1 | 0 | 1 |
| **Yellow** | 1 | 1 | 0 |

*embedding*

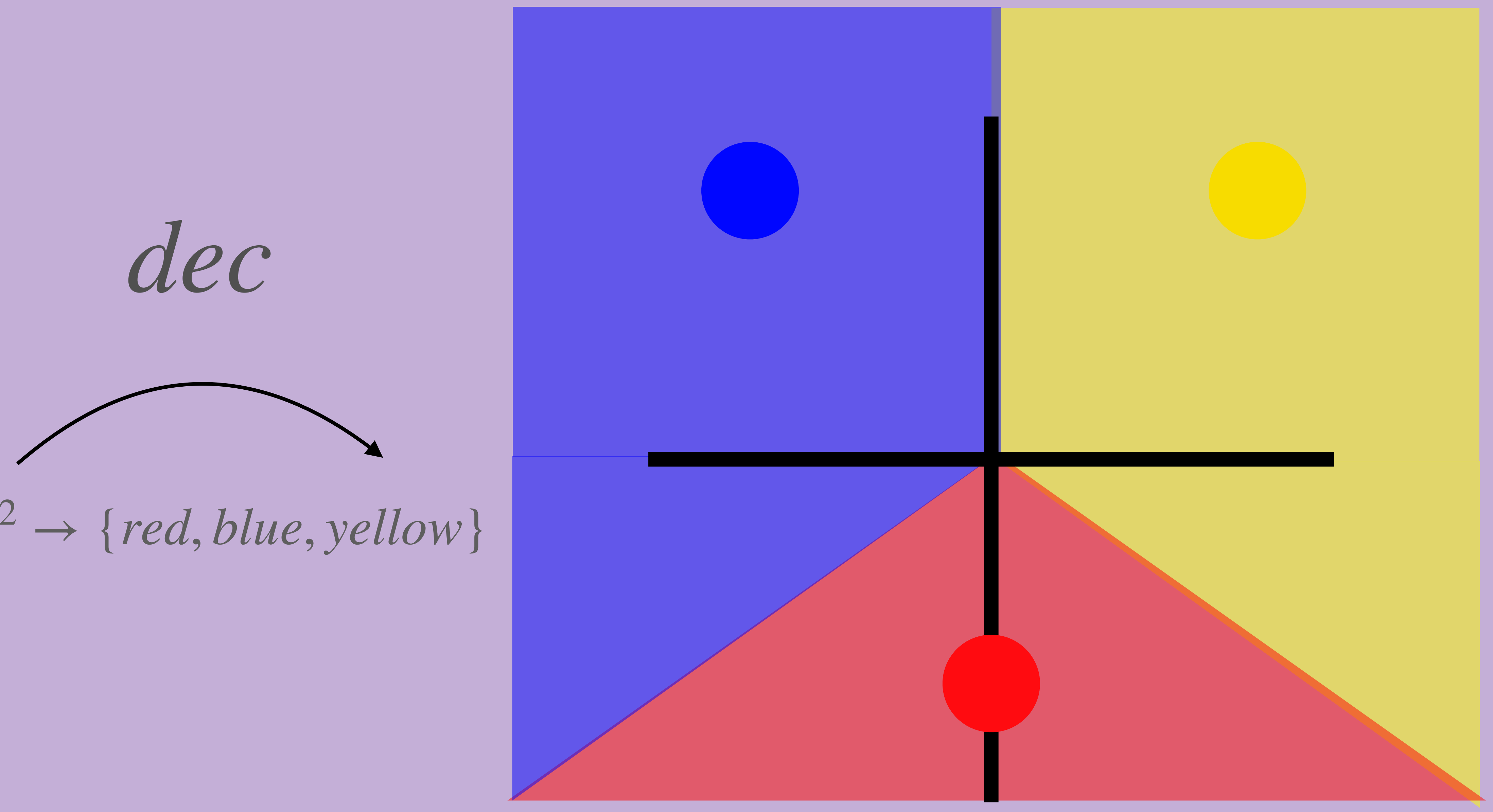
$\{red, blue, yellow\} \to \mathbb{R}^2$

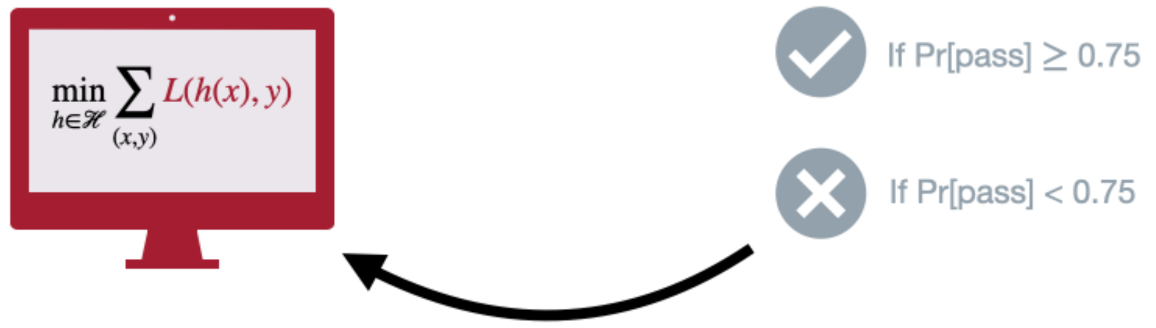


$\mathbb{R}^2$

*dec*

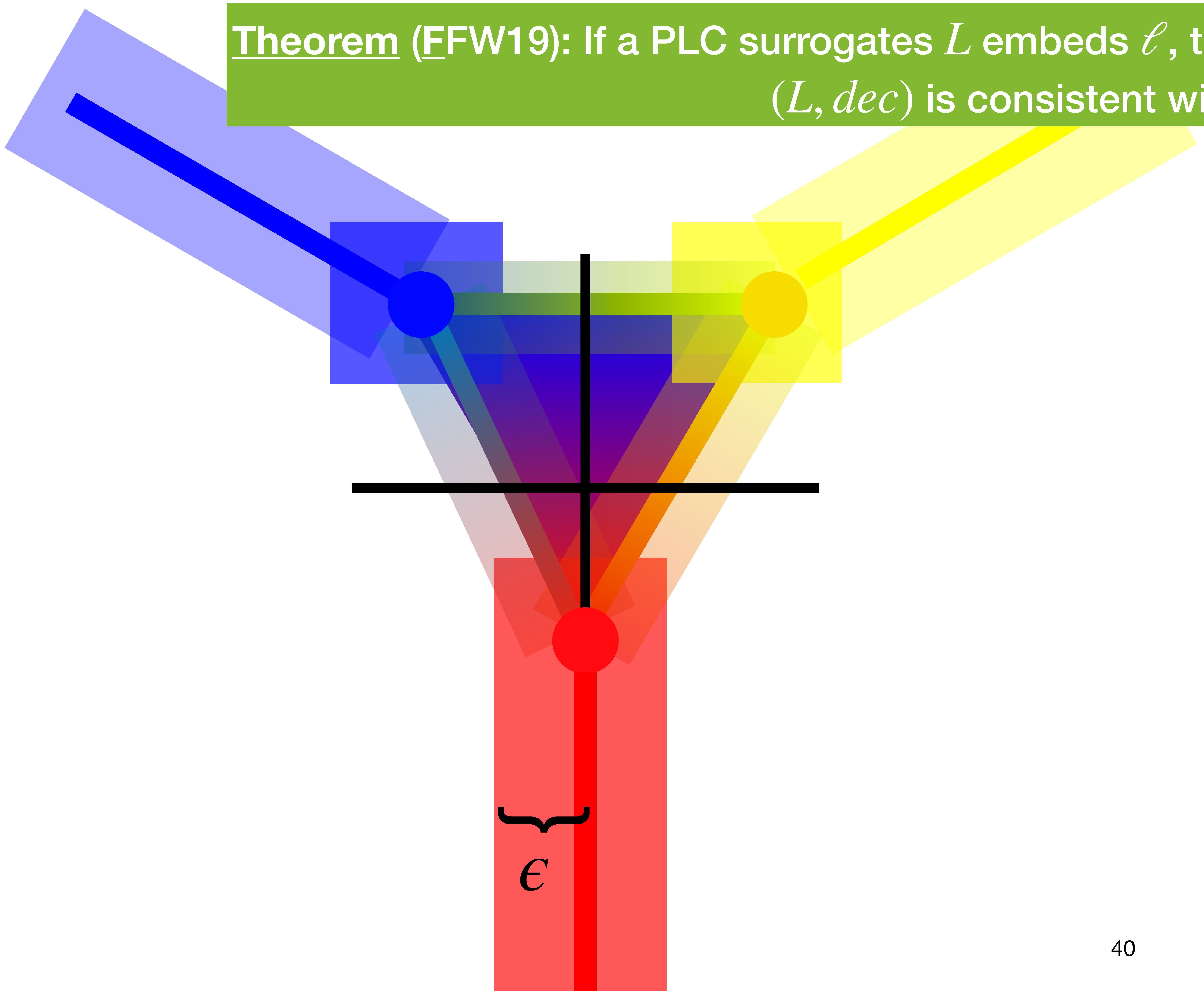
$\mathbb{R}^2 \to \{red, blue, yellow\}$






# Constructing a consistent decision function

**Theorem** (FFW19): If a PLC surrogates $L$ embeds $\ell$, there exists a decision function $dec$ such that $(L, dec)$ is consistent with respect to $\ell$




Consistency focused on *approaching* the optimum
Embedding focuses on the *exact minimizer*

# Dimensional efficiency

$L : \mathbb{R}^{d} \to \mathbb{R}$

Roughly: complexity of gradient computation linear in d

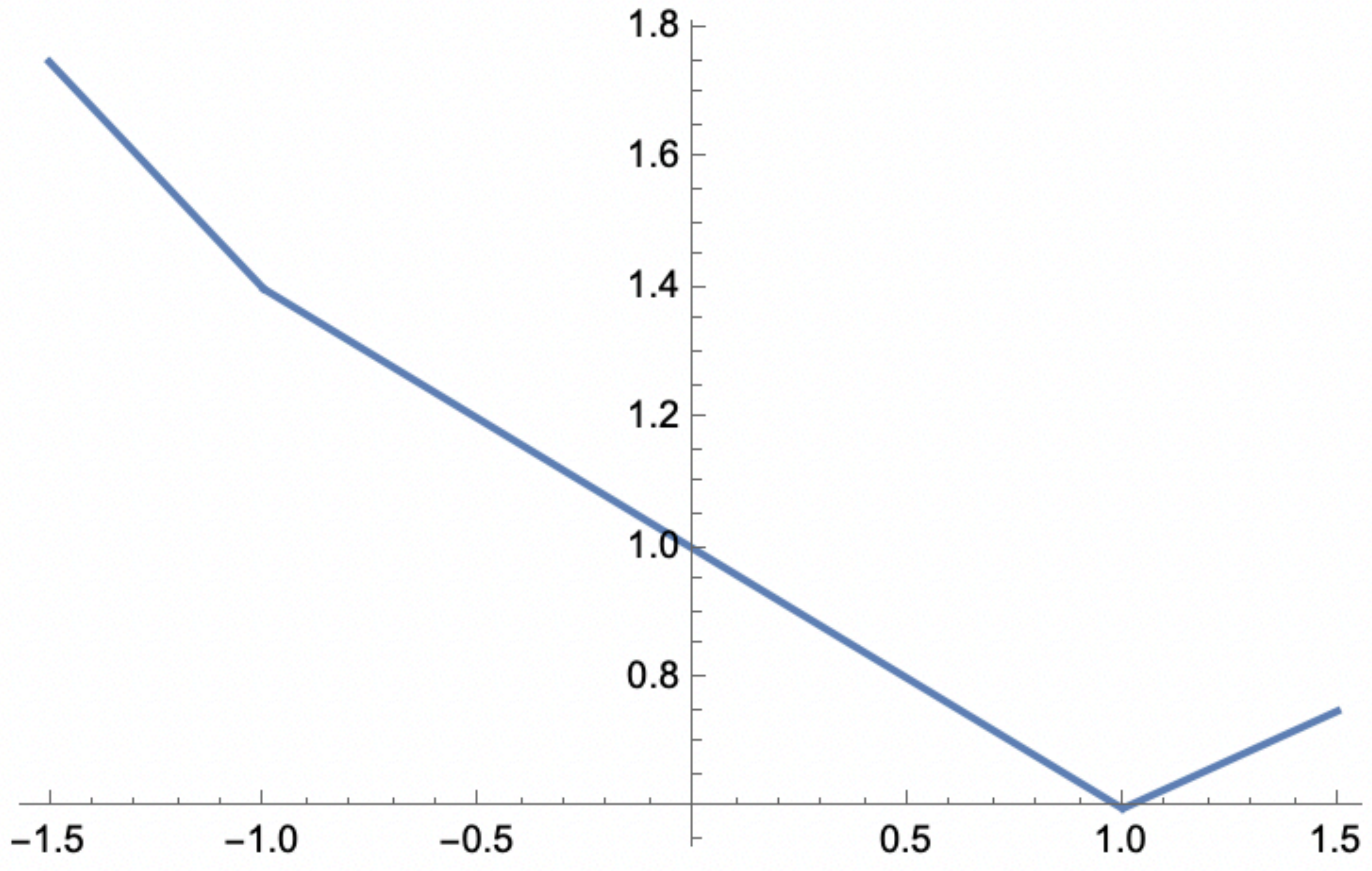
$d$ smaller —> better

Most “naive” losses are score-based: d = number of alternatives.

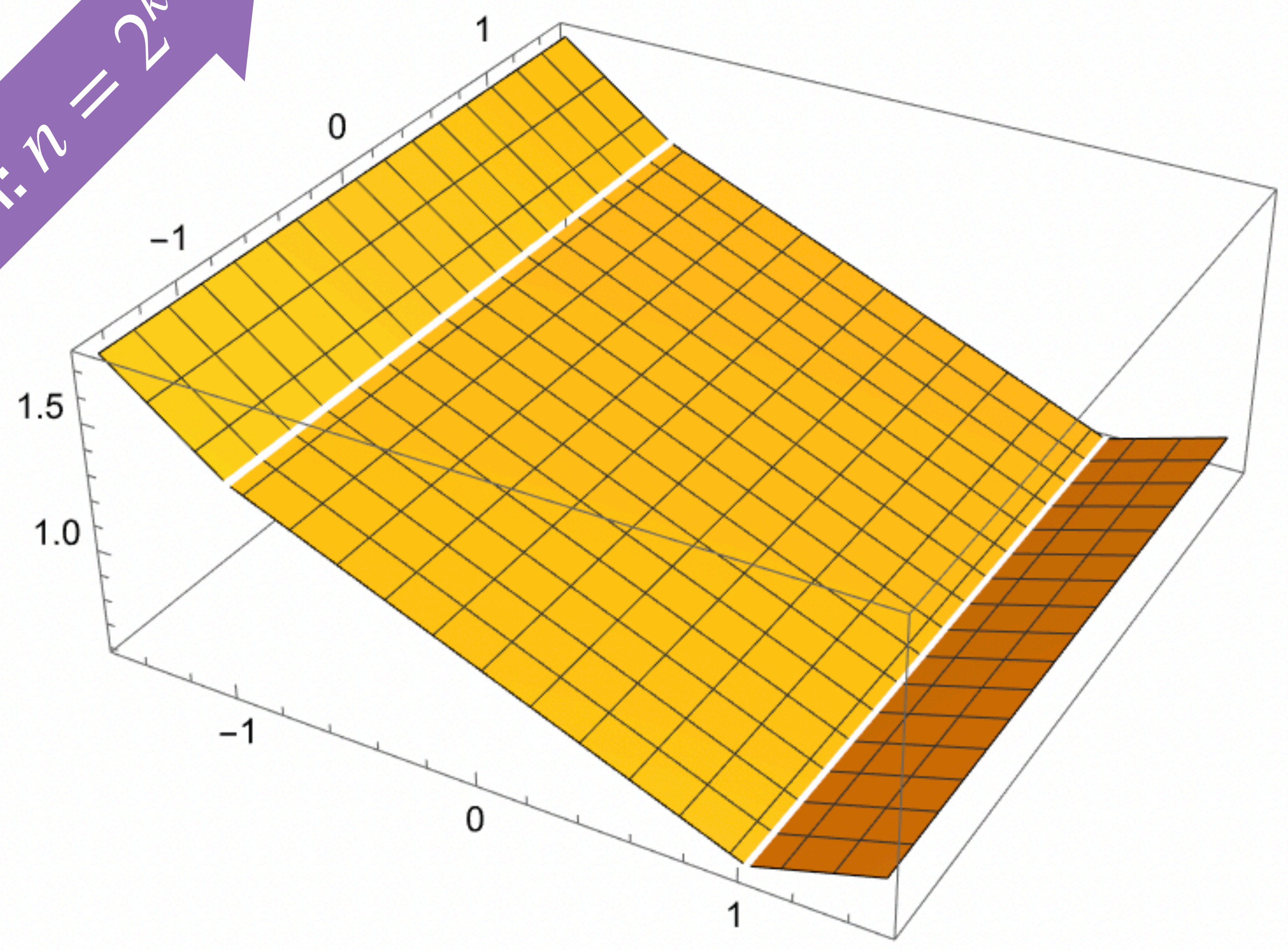
$d$ dimensions needed for consistent surrogate:

$$?? \leq d \leq n - 1$$

Structured prediction: $n = 2^k$




$d = 1$




$d = 2$

# Analyzing consistency via embeddings in image segmentation

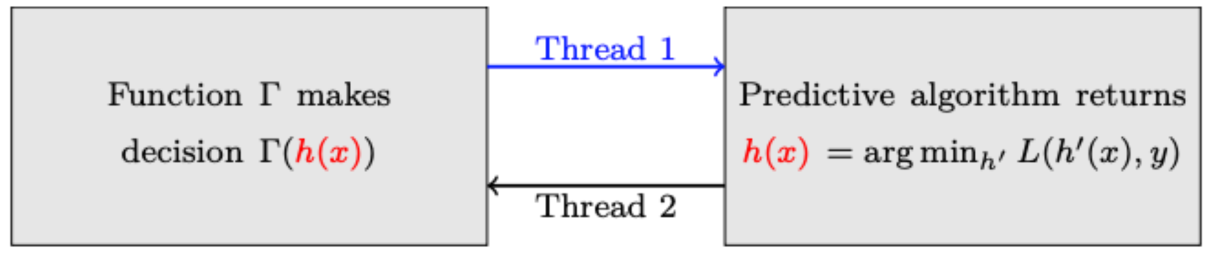
$$\ell(r, y) = \frac{|\{i : r_i = y_i\}|}{|\{i : y_i = 1\} \cup \{i : r_i \neq y_i\}|} = \frac{\text{num. correct pixels}}{\text{num. foreground or incorrect}}$$

$k$ pixels: $L : \mathbb{R}^{k} \times 2^k \to \mathbb{R}$ inconsistent

$L : \mathbb{R}^{2^k} \times 2^k \to \mathbb{R}$ consistent

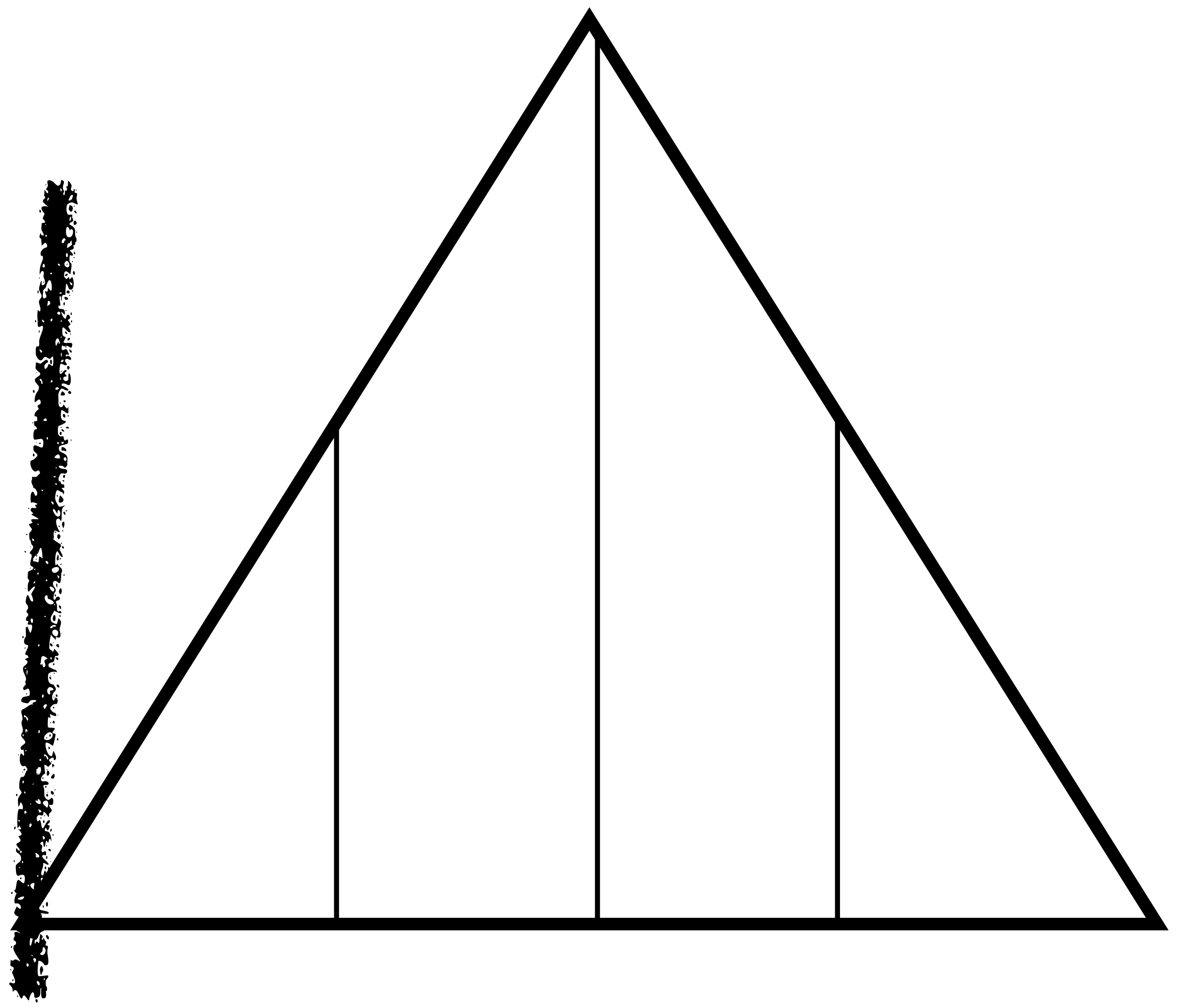
Note: didn't construct consistent surrogate because of dimension

Future work: trade off consistency for efficiency?



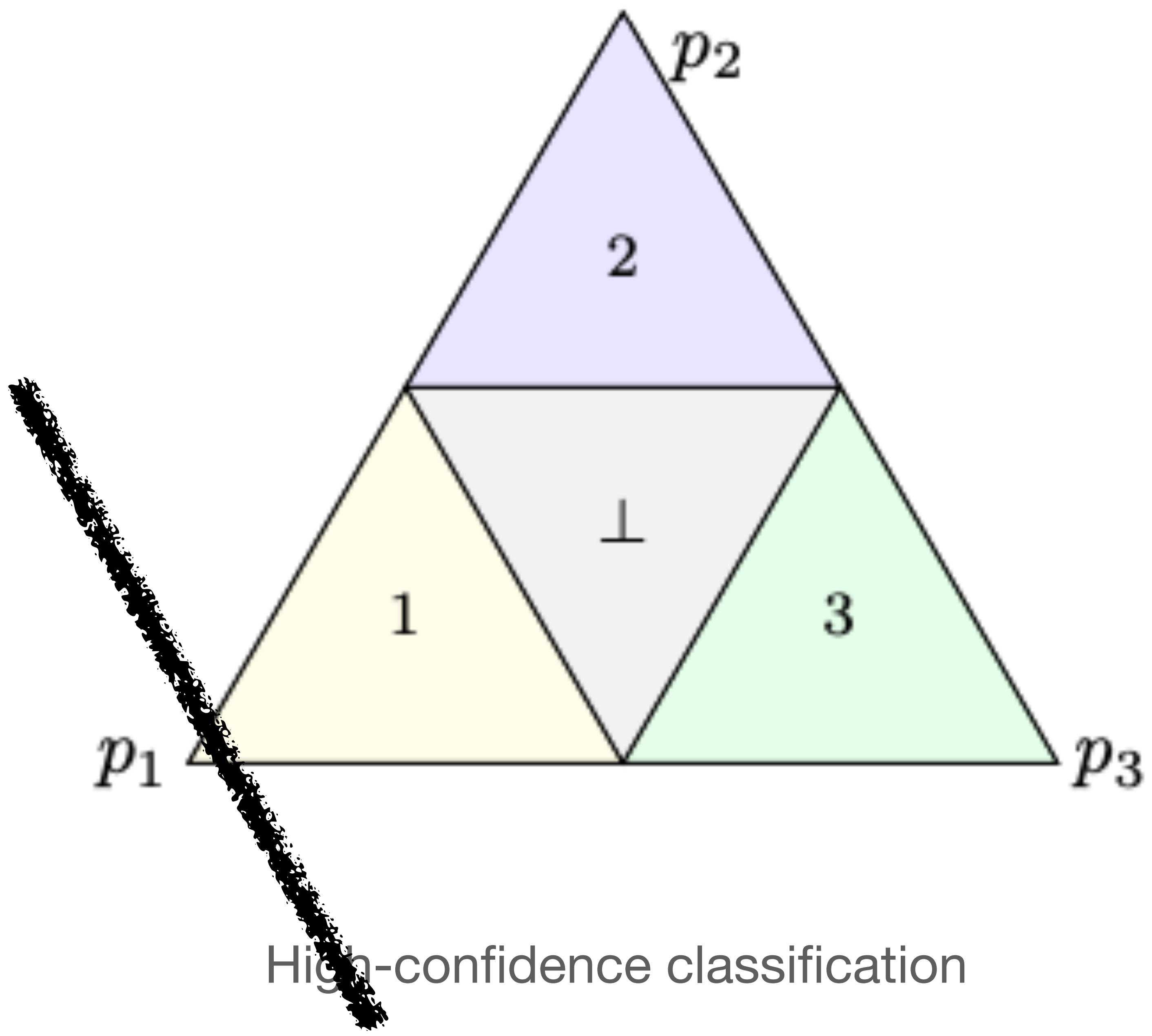

# Lower bounds on prediction dimension from the property

Convex flats depend on *global* features of property rather than *local* to improve lower bounds



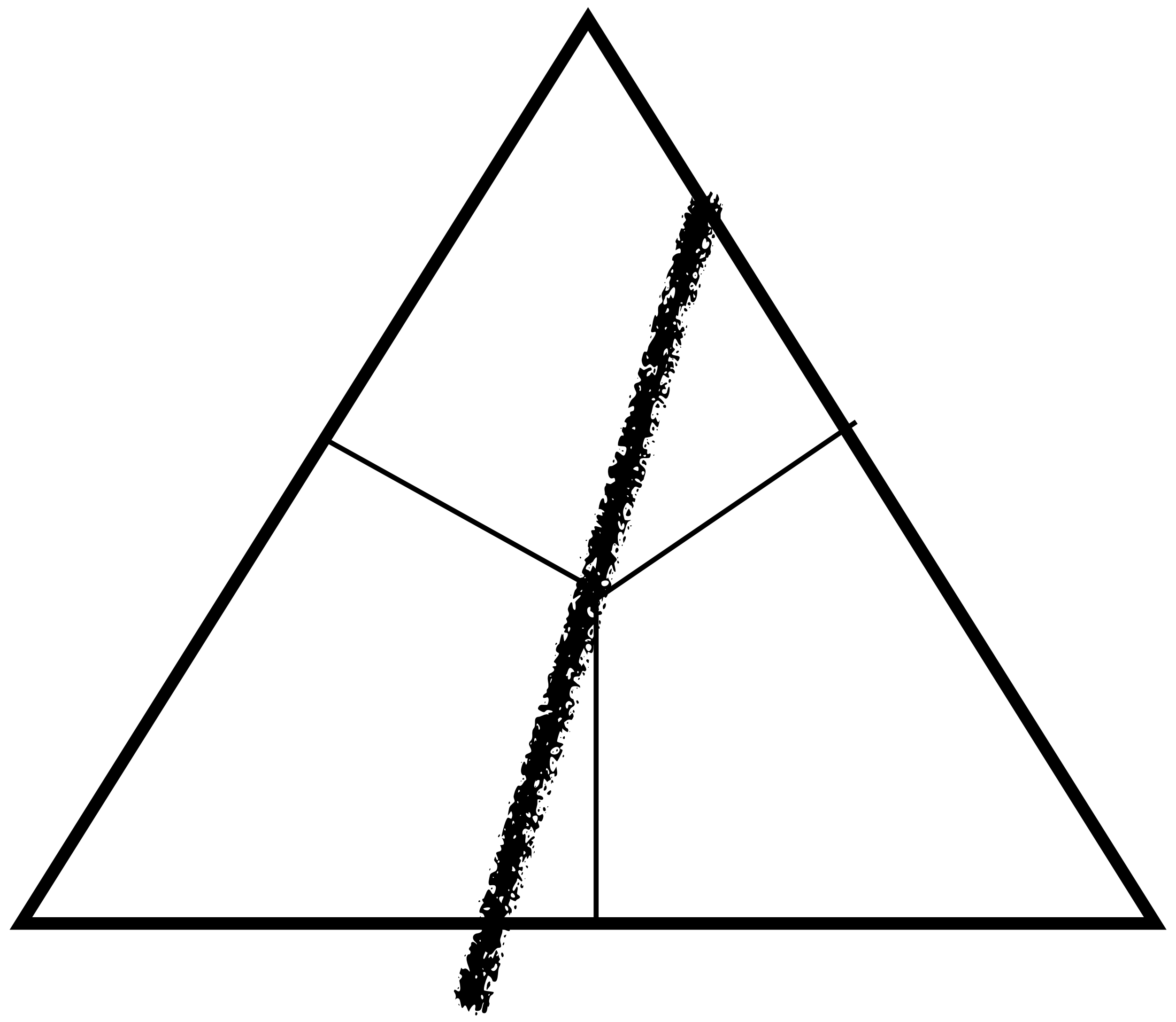
Truncated mean

$1 \leq d \leq 1$




High-confidence classification

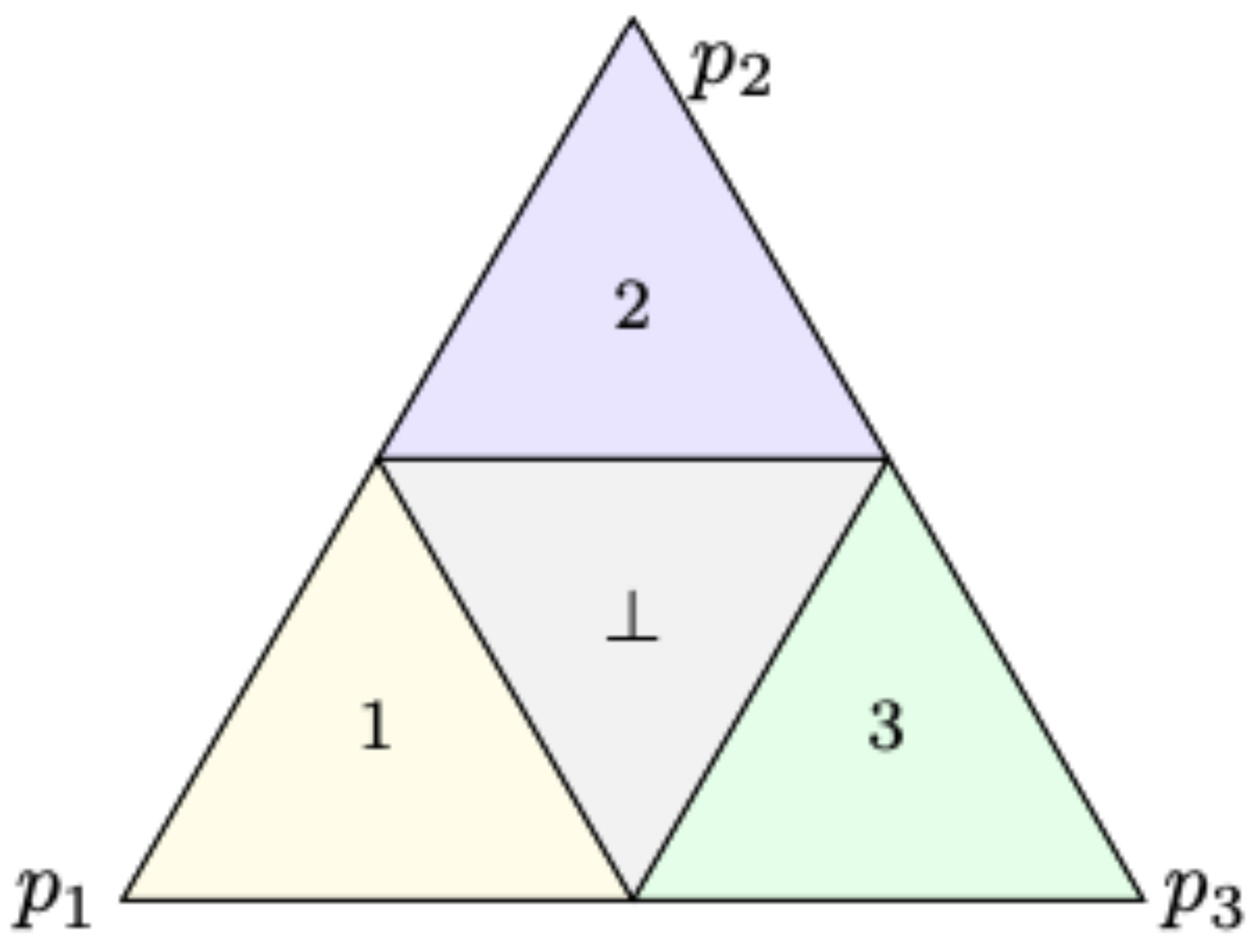
$2 \leq d \leq \log_2(n)$



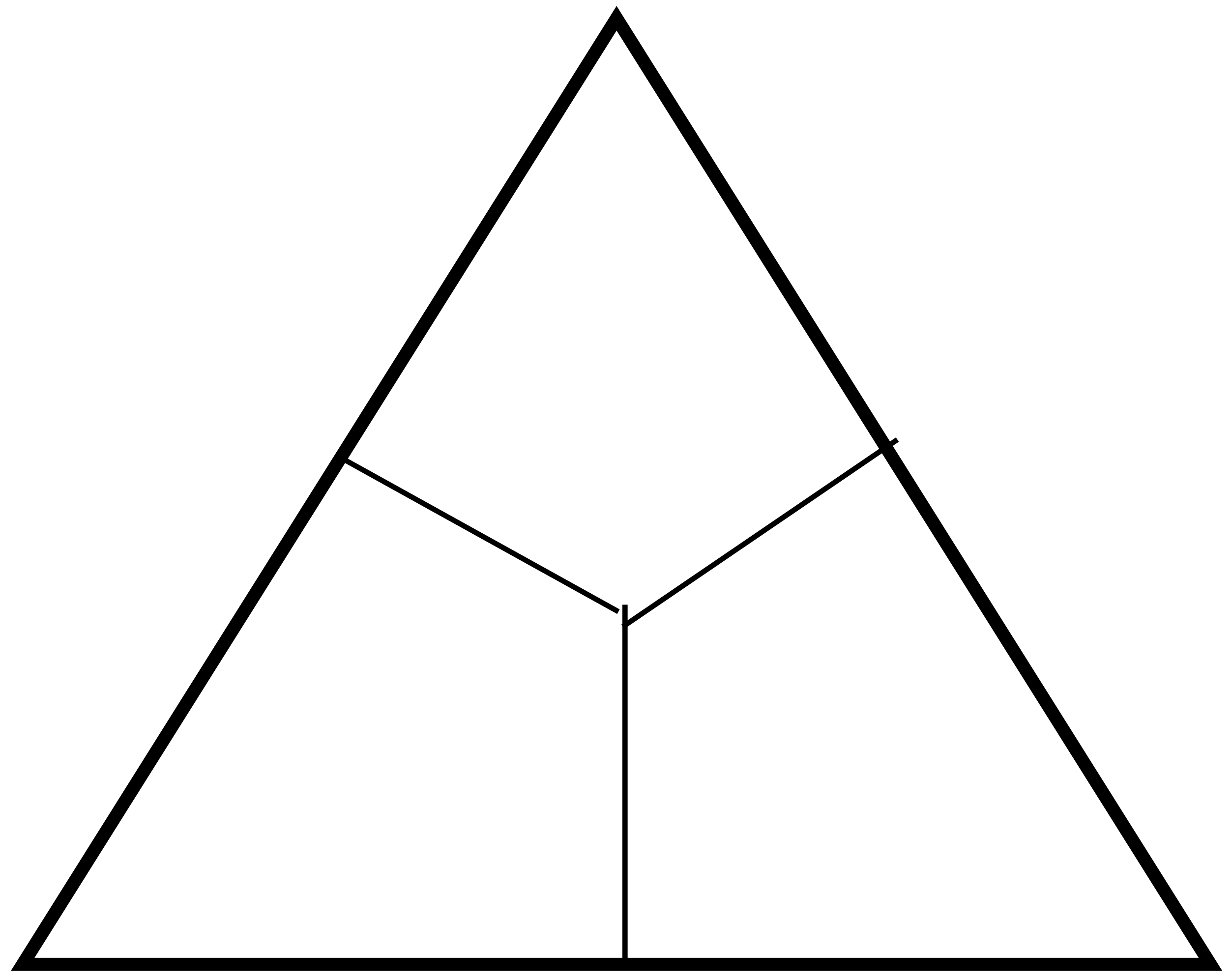
Classification

$n - 1 \leq d \leq n - 1$

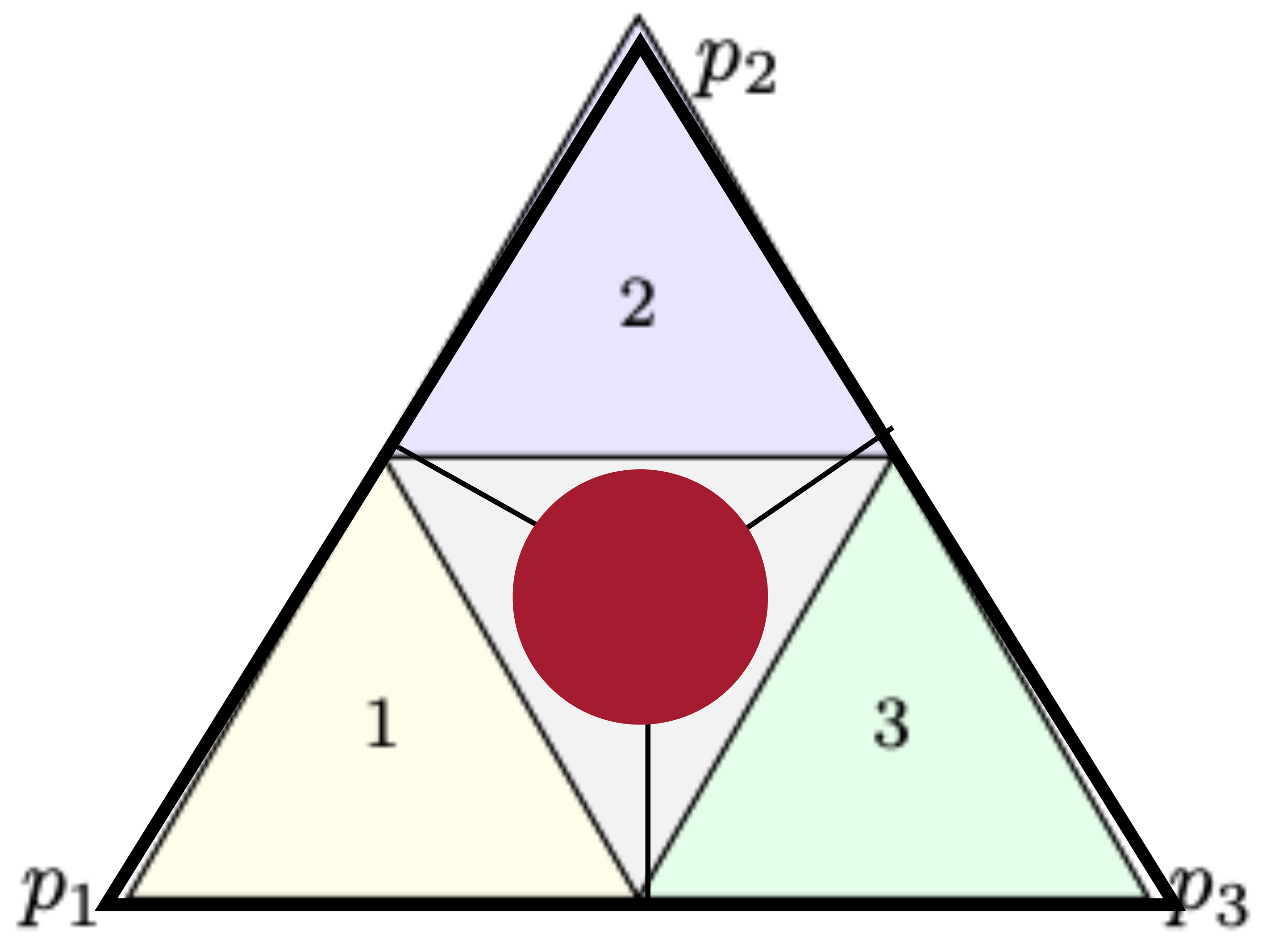
# Future work: trading off consistency and efficiency




$d \leq \log_2(n)$



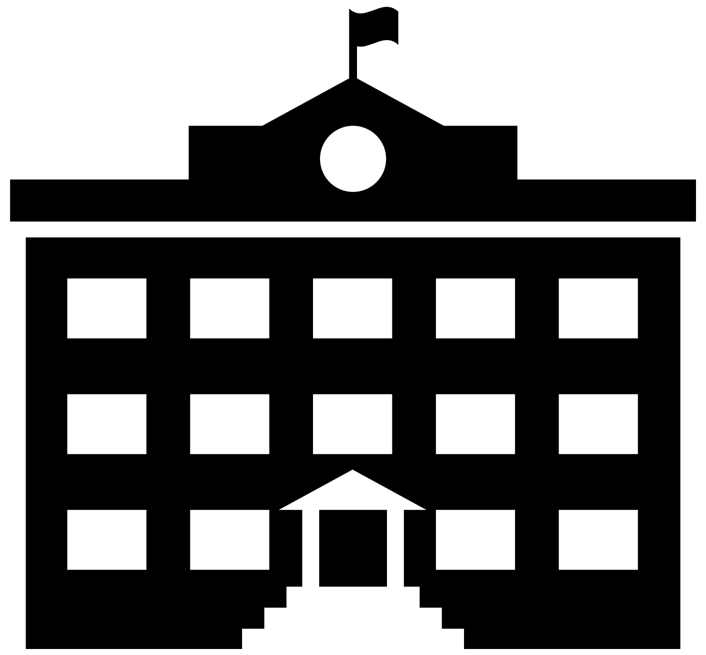
$n - 1 \leq d$

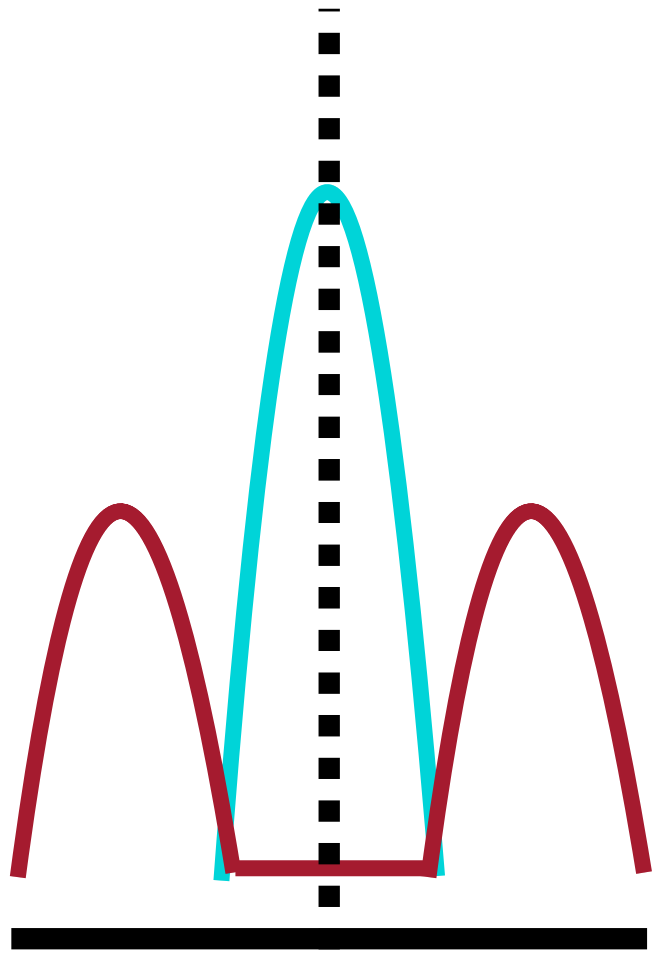


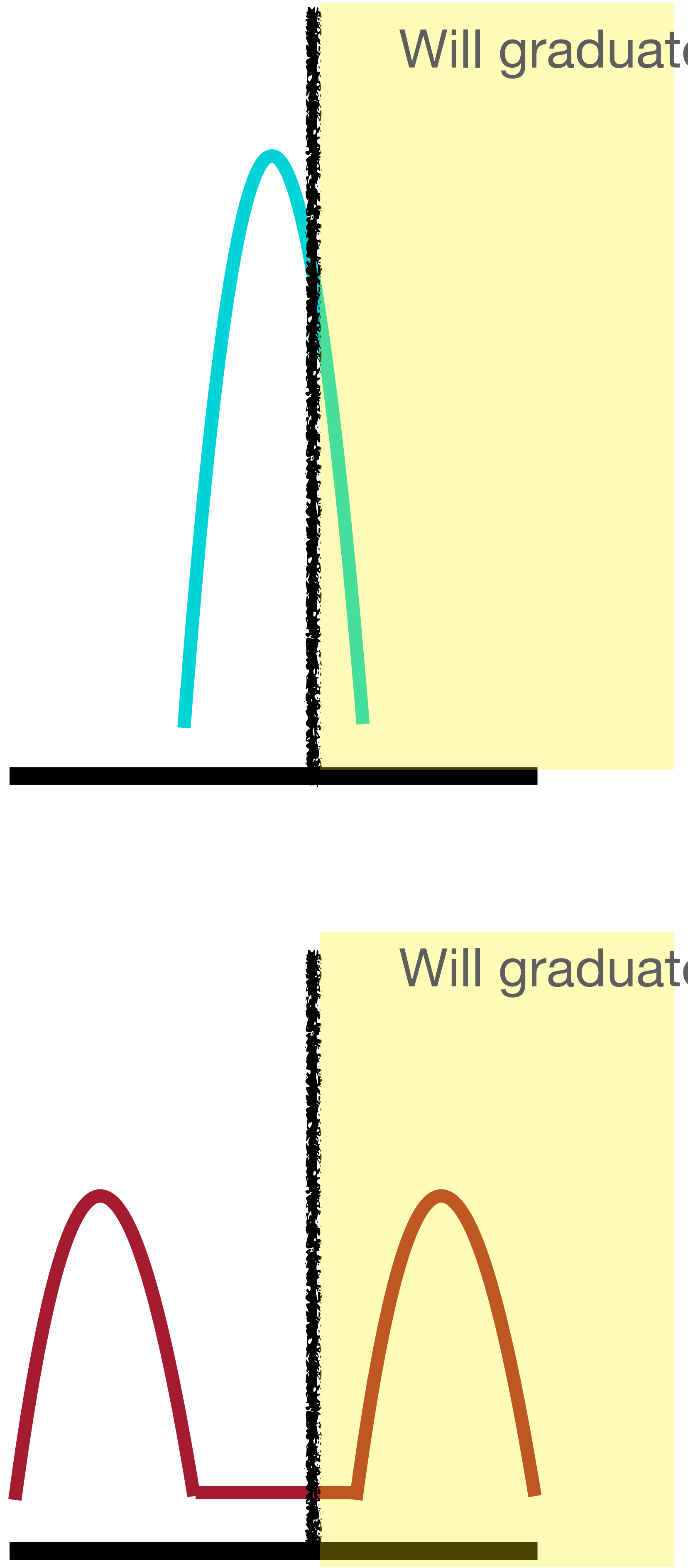

$d = \log_2(n)$ and *usually* makes right decision, but not always

# Future work: When to predict more granular information?

Access to property value, can (noisily) predict more granular information. How to trade off noise in prediction vs








Predict, even if noisy

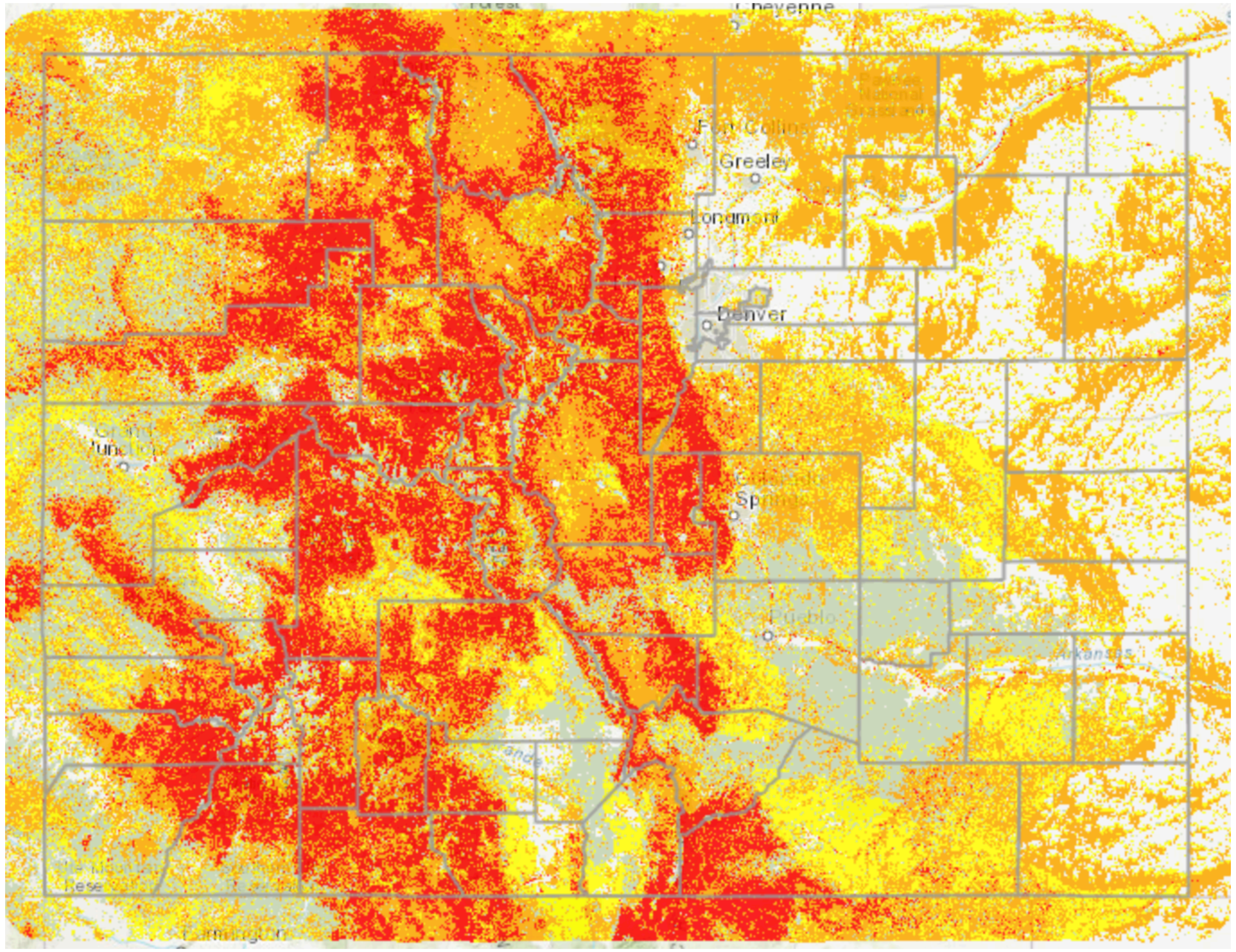
# Future work: Wildfire risk prediction

**Knowing how predictions are used to prescribe burns, how do we design predictive algorithms for fire intensity?**

Table 1,
Private Forest Land Protection Criteria, 2020

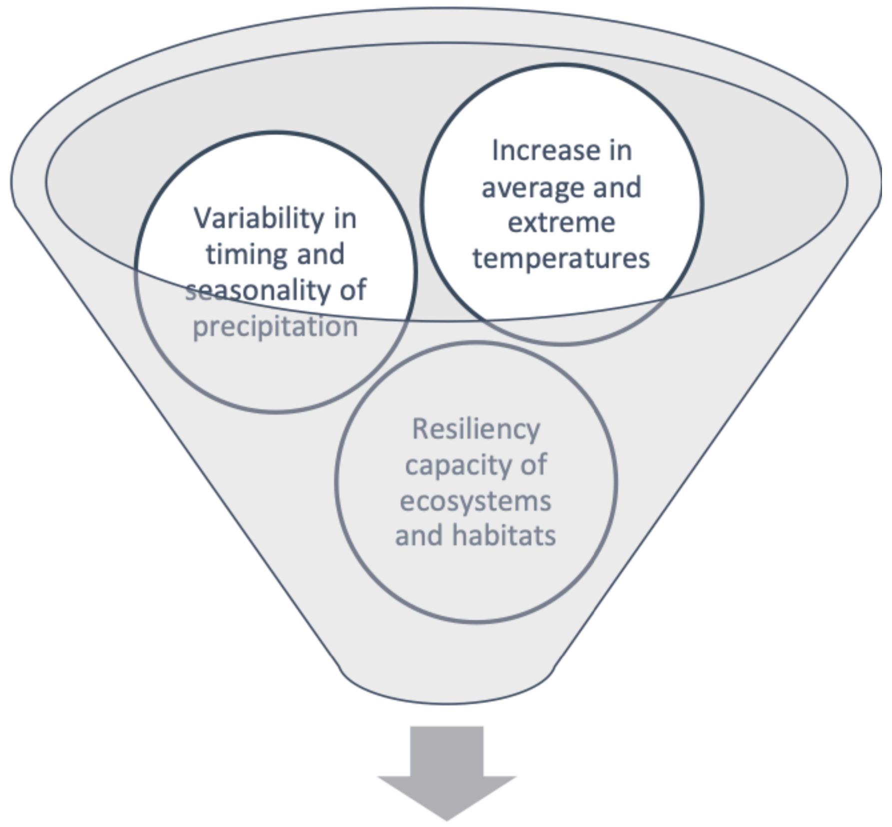
| Criteria | Priority |
|---|---|
| Water Quality/Quantity | 1 |
| Wildlife Habitat | 2 |
| Growth/Sprawl Control | 3 |
| Large Continuous Forest | 4 |
| Wetland/Riparian Areas | 5 |
| Unique Ecological Areas | 6 |
| Wildfire Control Issues | 7 |
| Private Property Rights | 8 |
| Forest Timber Products | 9 |
| Lifestyle Protection for Landowner | 10 |

https://csfs.colostate.edu/wp-content/uploads/2020/11/




https://co-pub.coloradoforestatlas.org/#/

# Decisions —> Algorithms: Wildfire risk prediction




Climate Change Risk Matrix

Severity of Impacts

| Likelihood | Negligible | Minor | Moderate | Major | Severe |
|---|---|---|---|---|---|
| Very Likely | Med. Low | Medium | Med. High | High | High |
| Likely | Low | Med. Low | Medium | Med. High | High |
| Possible | Low | Med. Low | Medium | Med. High | Med. High |
| Unlikely | Low | Med. Low | Med. Low | Medium | Med. High |
| Very Unlikely | Low | Low | Med. Low | Medium | Med. High |

Table 1,
Private Forest Land Protection Criteria, 2020

| Criteria | Priority |
|---|---|
| Water Quality/Quantity | 1 |
| Wildlife Habitat | 2 |
| Growth/Sprawl Control | 3 |
| Large Continuous Forest | 4 |
| Wetland/Riparian Areas | 5 |
| Unique Ecological Areas | 6 |
| Wildfire Control Issues | 7 |
| Private Property Rights | 8 |
| Forest Timber Products | 9 |
| Lifestyle Protection for Landowner | 10 |

https://csfs.colostate.edu/wp-content/uploads/2020/11/FINAL2020_FLP_AON-.pdf

https://cdphe.colorado.gov/clean-water-gis-maps

https://co-pub.coloradoforestatlas.org/#/