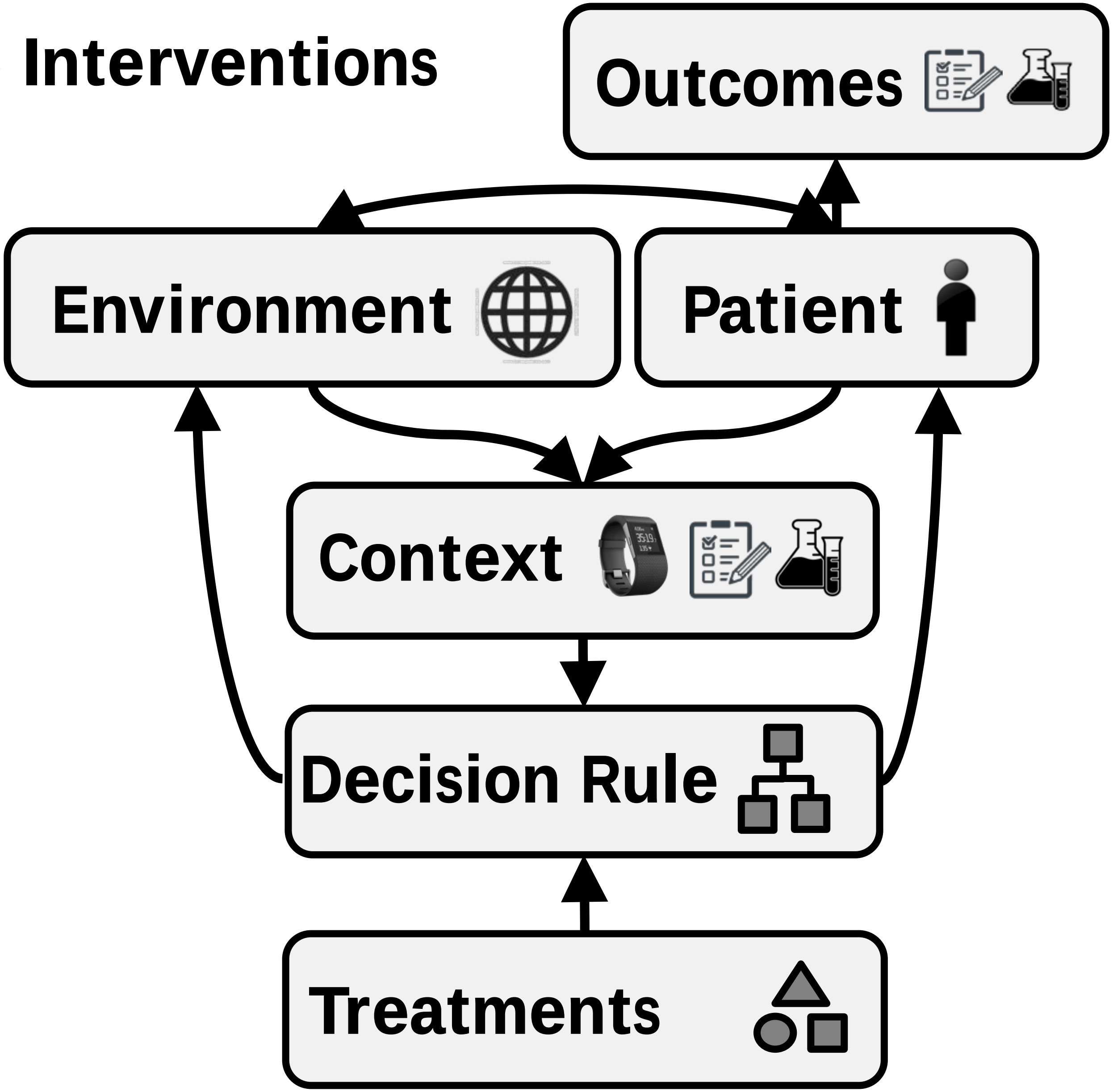
Machine Learning for Behavioral Interventions

Benjamin M. Marlin
College of Information and Computer Sciences
University of Massachusetts Amherst
marlin@cs.umass.edu

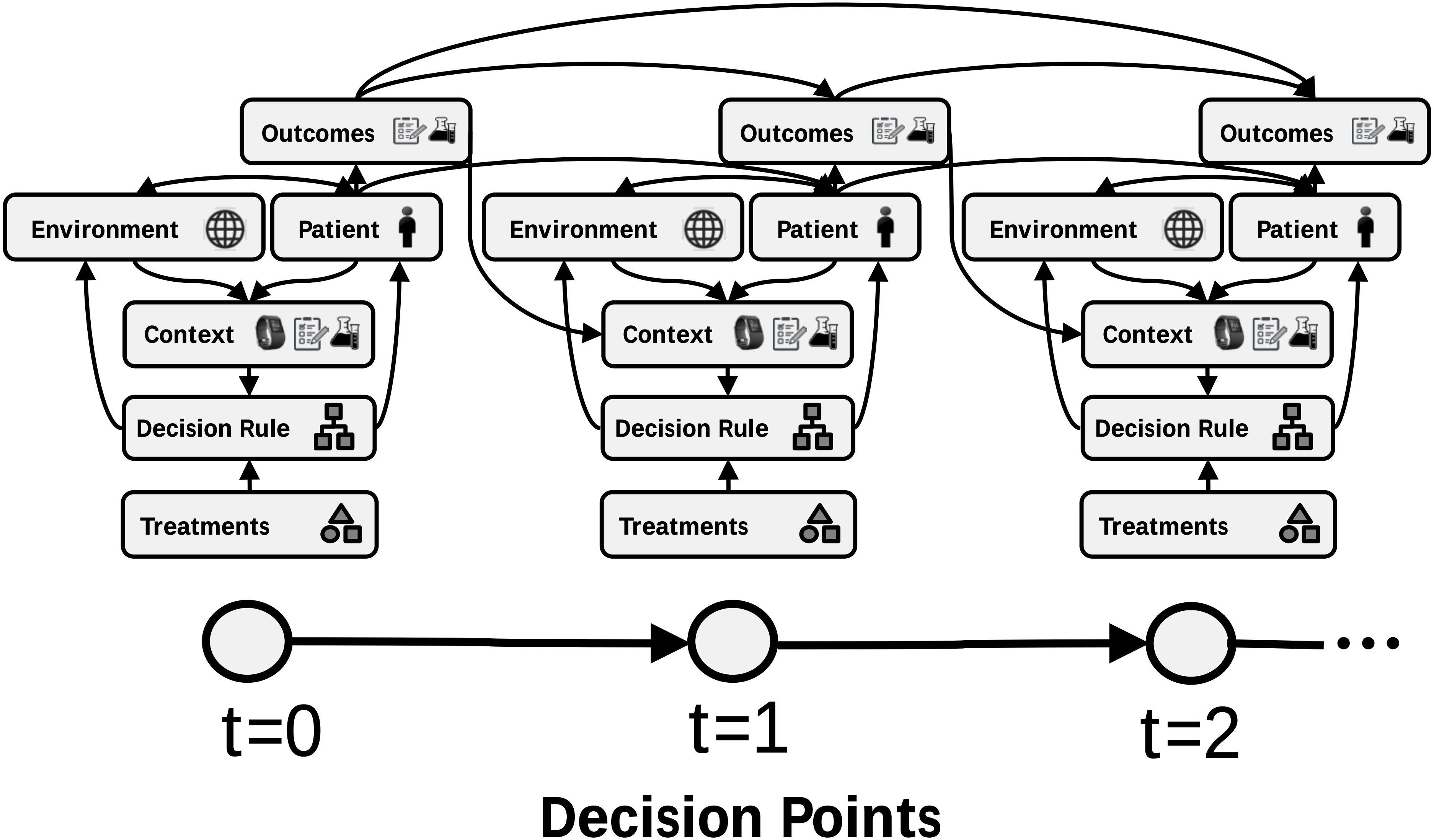
Mobile Health Training Institute
Aug 9, 2017

Introduction

Adaptive Interventions



Adaptive Interventions



Adaptive Intervention Design Questions:

1. **Treatments:** What should the set of treatment options or intervention components be?

2. **Context:** How to assess or measure the context/tailoring variables? What context variables are relevant to making treatment decisions?

3. **Decision Rule:** What should the decision rule be?

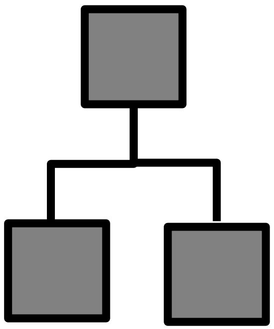
4. **Decision Points:** How to decide how many decision points to have and when they should be?

5. **Outcome Measures:** How to assess or measure the (proximal and distal) outcome variables?

What Can Machine Learning Help With?

1. Context and Outcomes: Inferring context variables and outcomes from sensor data.

2. Decision Rules: Inducing decision rules from outcome and context data.



3. And more...

Context & Outcomes

Inference for Context and Outcomes

Sensor Data

Contexts & Outcomes

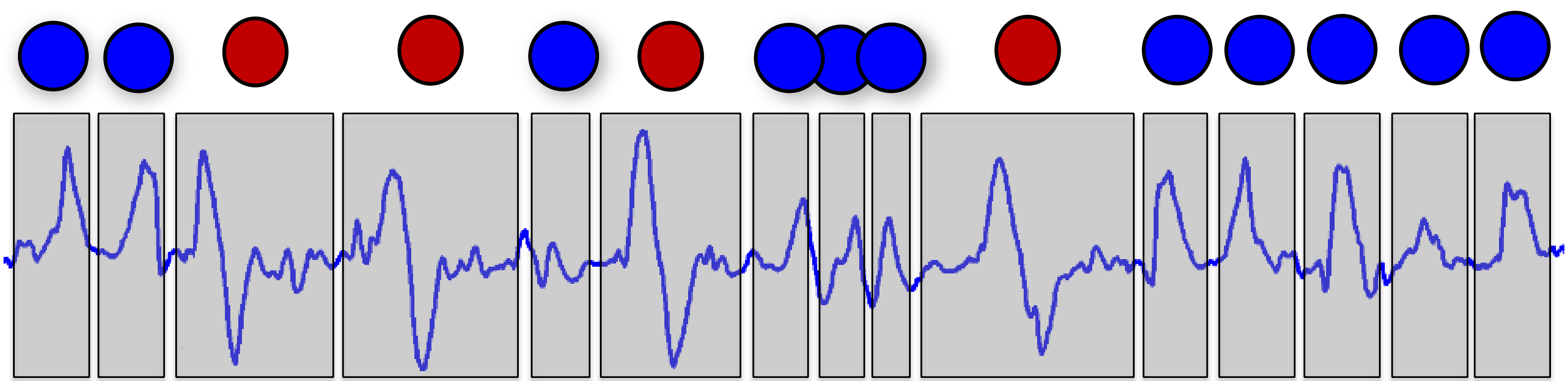
The Detection Problem: Suppose we have a dynamical system in which an event of interest either occurs or does not occur at each time instant t. Given a partial description $\mathbf{x}_t \in \mathbb{R}^D$ of the state of the system at time t, infer whether the event occurred at time t or not.



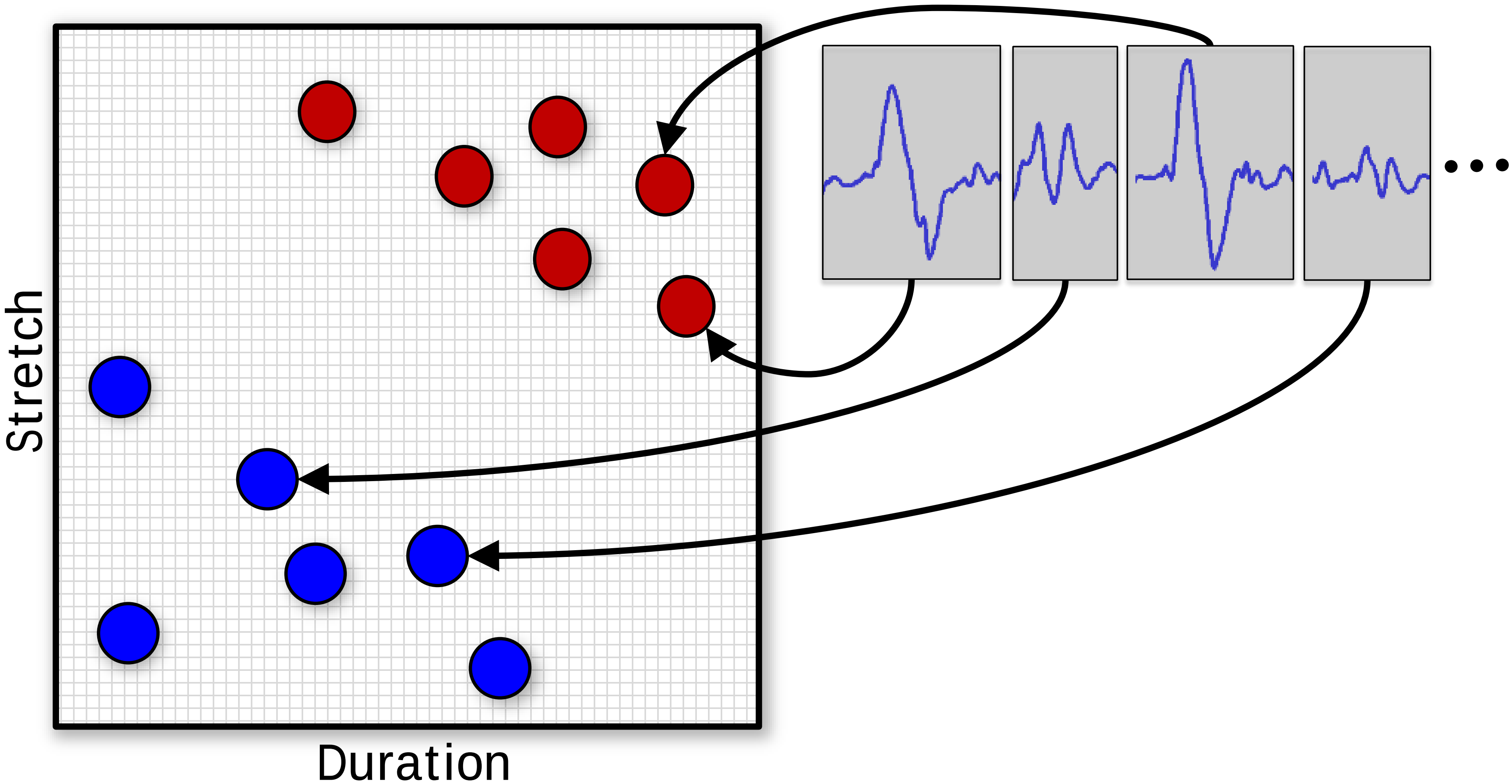




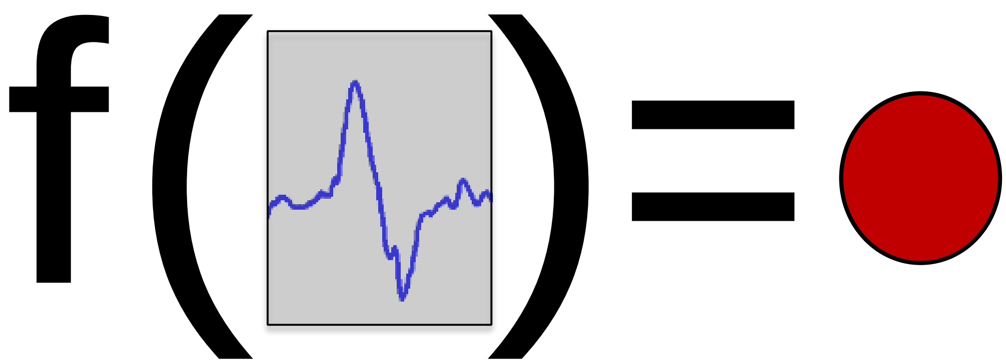

The Detection Problem: Suppose we have a dynamical system in which an event of interest either occurs or does not occur at each time instant t. Given a partial description of the state of the system at time t, infer whether the event occurred at time t or not.



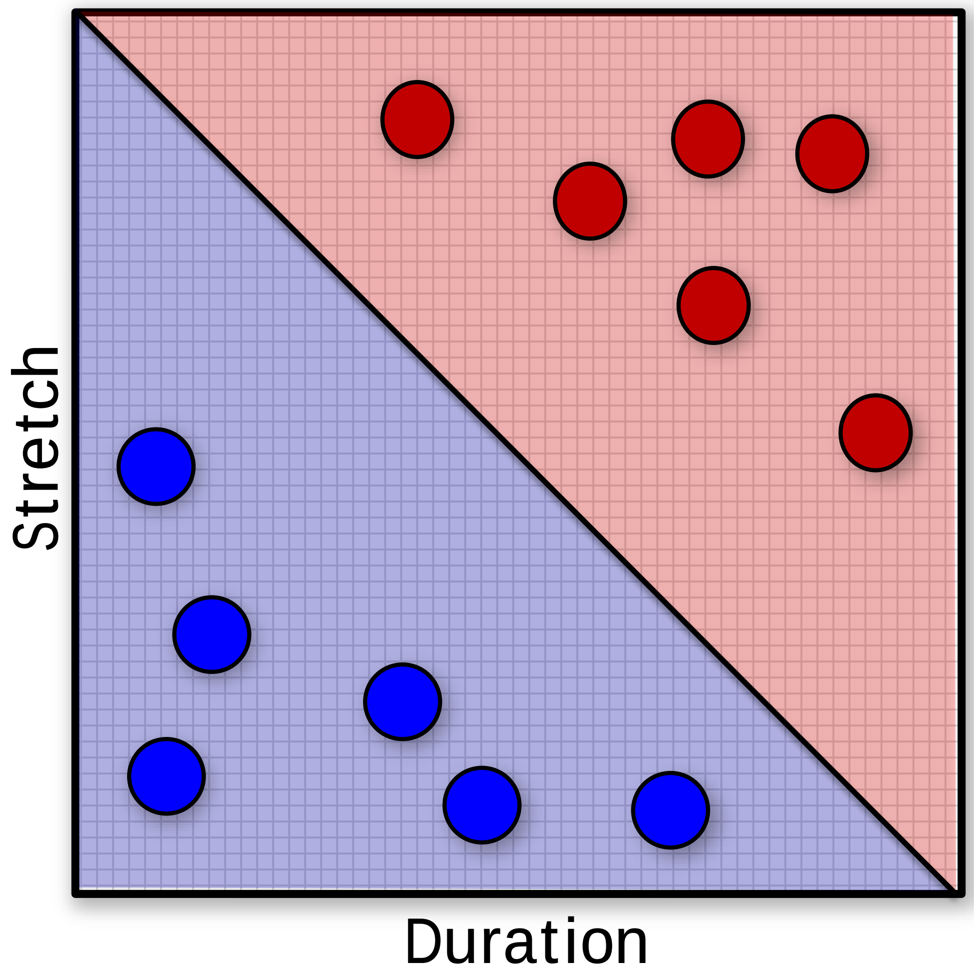
Feature Space



A Detection Function: Given a feature vector $\mathbf{x}_t \in \mathbb{R}^D$ that partially describes the state of a dynamical system at time t, a detection function f: $\mathbb{R}^D \rightarrow \{0,1\}$. 0 indicates the event of interest did not occur, and 1 indicates that the event of interest did occur. These are the **event labels.**



Detection Function in Feature Space

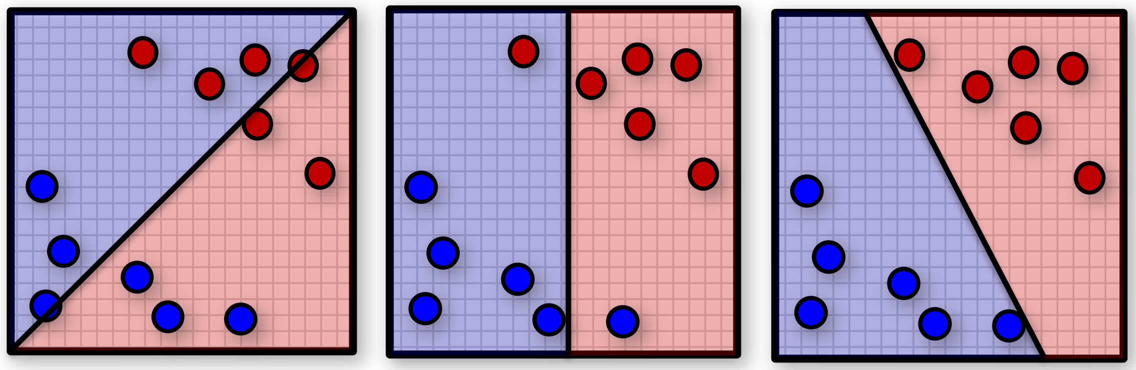


Detection Model: A detection model is a set of detection functions F where for each f ∈ F,
f: $\mathbb{R}^D \rightarrow \{0,1\}$.

$$f_w(\mathbf{x}) = \begin{cases} 1 & \ldots \text{ if } \displaystyle\sum_{d=1}^{D} w_d x_d > w_0 \\ 0 & \ldots \text{ otherwise} \end{cases}$$

Detection Model in Feature Space

$F=\{$



$\ldots\}$

Detector Learning Problem: Given a data set $D=\{(\mathbf{x}_i, y_i) | 1<i<N\}$ consisting of feature vectors $\mathbf{x}_i \in \mathbb{R}^D$ and event labels $y_i \in \{0,1\}$, select a function f: $\mathbb{R}^D \rightarrow \{0,1\}$ from F that maps feature vectors $\mathbf{x} \in \mathbb{R}^D$ to their event labels as accurately as possible.

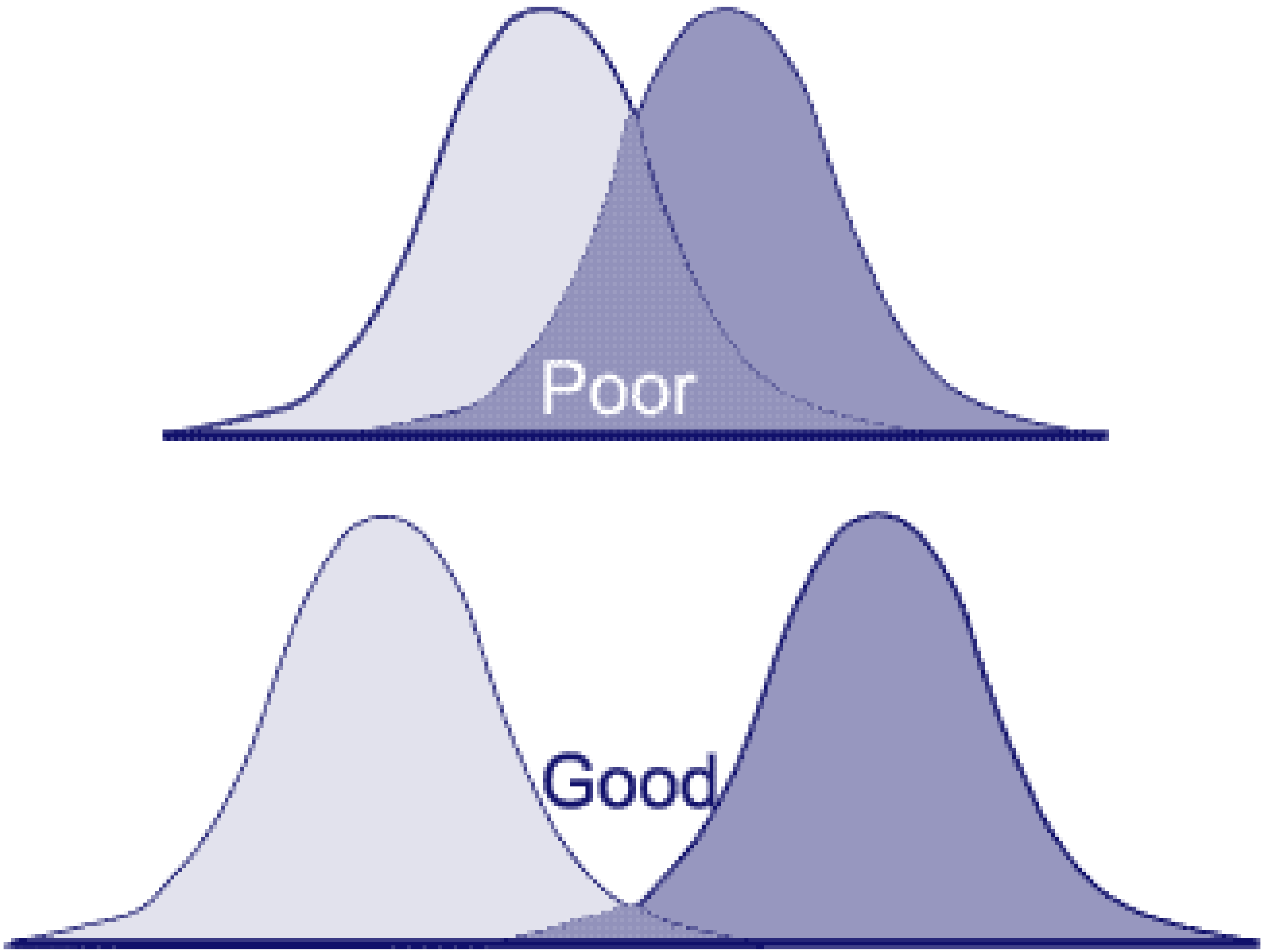
$$w_* = \arg\min_w \sum_{n=1}^{N} [f_w(\mathbf{x}_n) \neq y_n]$$

Detector Learning Problem: Given a data set $D=\{(\mathbf{x}_i, y_i) \mid 1<i<N\}$ consisting of feature vectors $\mathbf{x}_i \in \mathbb{R}^D$ and event labels $y_i \in \{0,1\}$, select a function f: $\mathbb{R}^D \rightarrow \{0,1\}$ from F that maps feature vectors $\mathbf{x} \in \mathbb{R}^D$ to their event labels as accurately as possible.

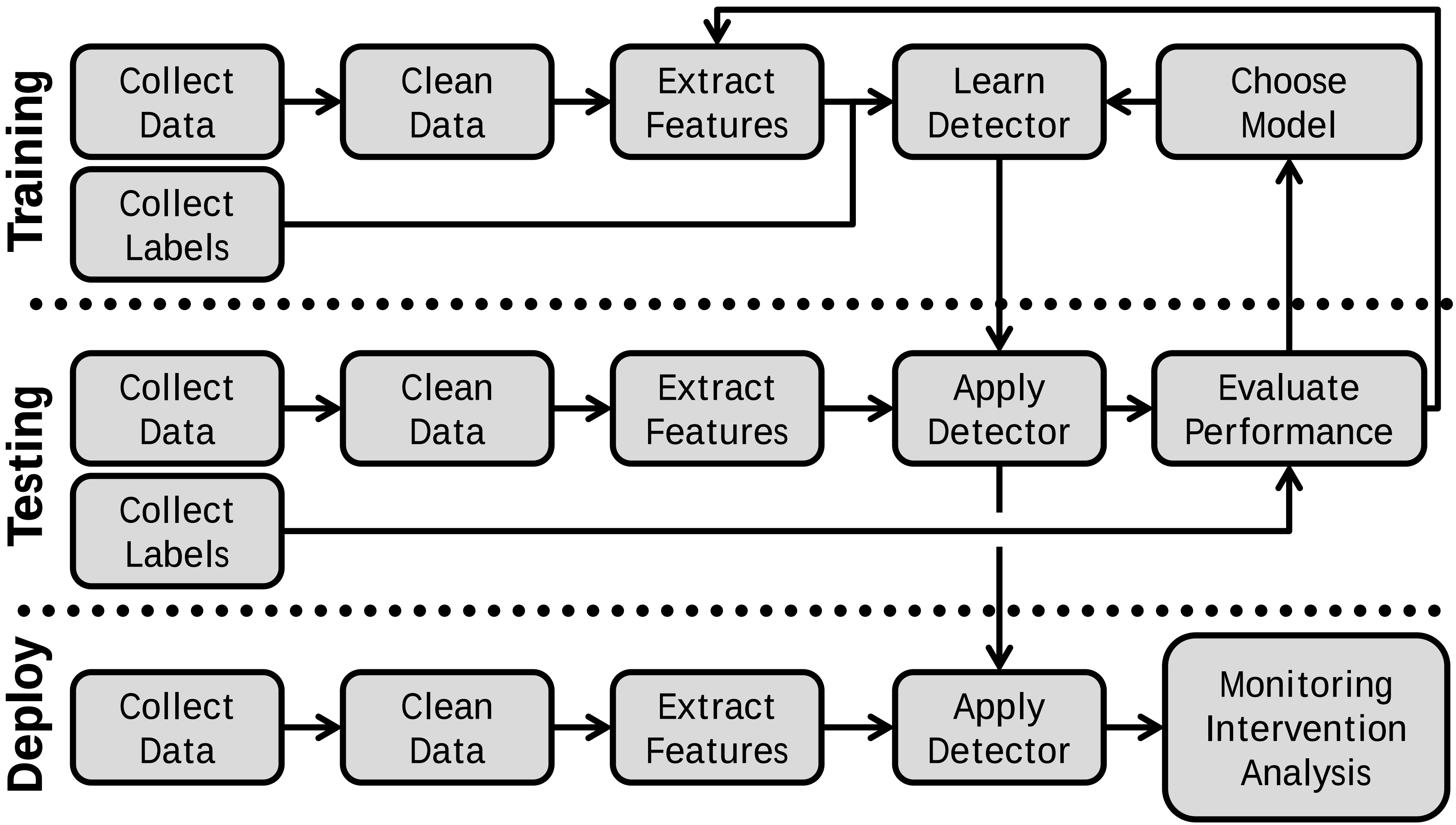
$$w_* = \arg\min_w \sum_{n=1}^{N} \mathrm{loss}(f_w(\mathbf{x}_n), y_n)$$

Detector Learning Demo

Class Overlap



Full Detector Learning Process



Data-Driven Model Selection

mHealth Experiment Designs

- **Within-Subject Design:** Use train-test with temporal splits. Estimates generalization error for future data from the same subject.
- **Between-Subject Design:** Randomly partition subjects into train and test sets. Estimates ability to generalize to new subjects.
- **Transfer Design:** Use train-test with temporal splits for an individual subject, but start from previously learned between-subject model.

Detector Learning Challenges:

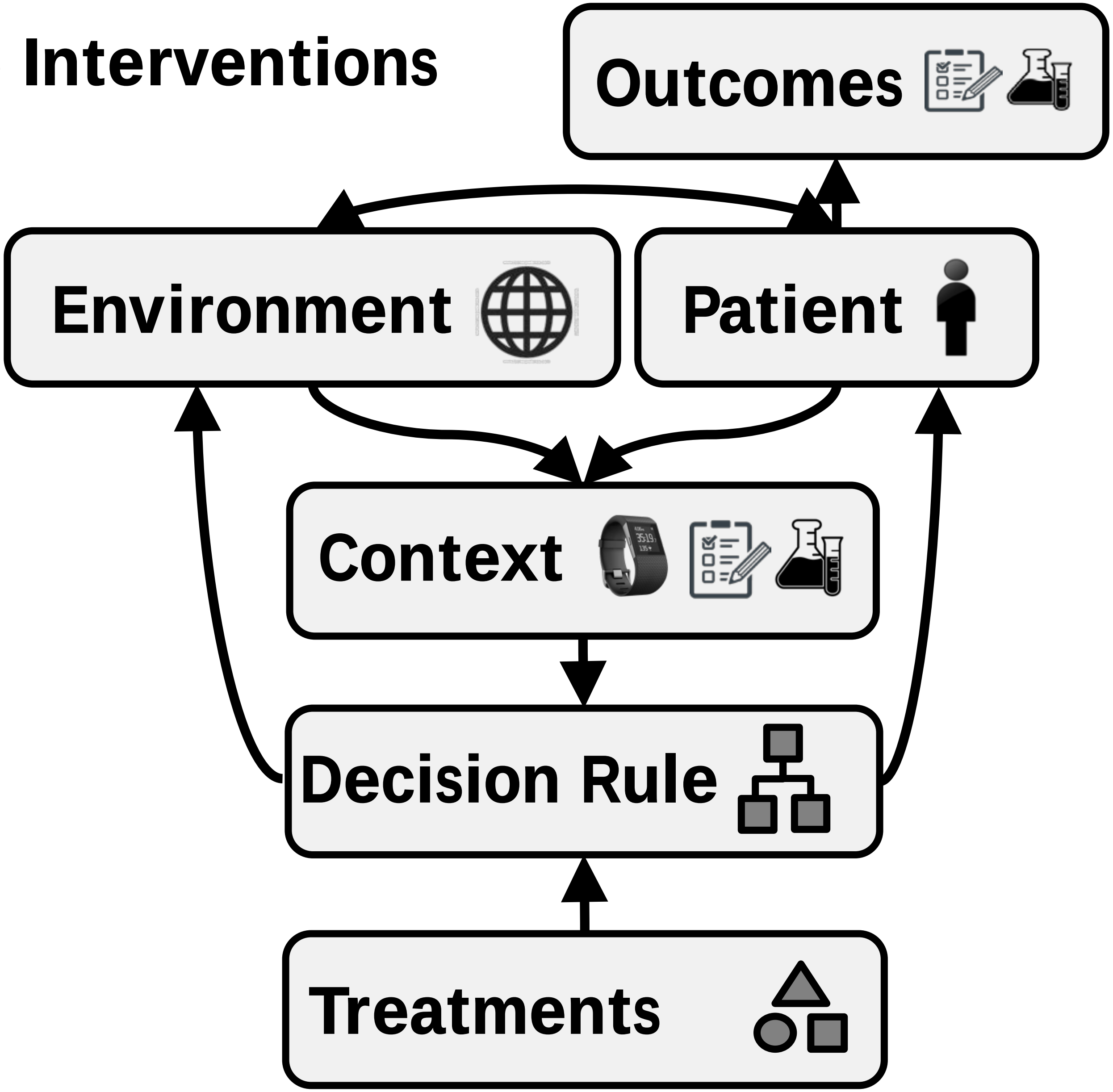
1. **Obtaining Labels:** Collecting accurate labels can be difficult and costly. Tradeoff between ecological validity and accuracy.

2. **Lab-to-Field Generalizability:** Models built on data collected in the lab often fail to generalize well in the field.

3. **Between Subject Variability:** Big cross-sectional data often not sufficient. Need big longitudinal data. The N=me problem.

Decision Rules

Adaptive Interventions



A Decision Rule: Given a context vector $\mathbf{c}_t \in \mathbb{R}^D$ that partially describes the context at time t, a decision rule f: $\mathbb{R}^D \rightarrow \{0,1,\ldots,K\}$. 0 indicates no treatment, and 1 to K indicate a choice of one of K treatments.

Question: How is decision rule learning different than detector learning?

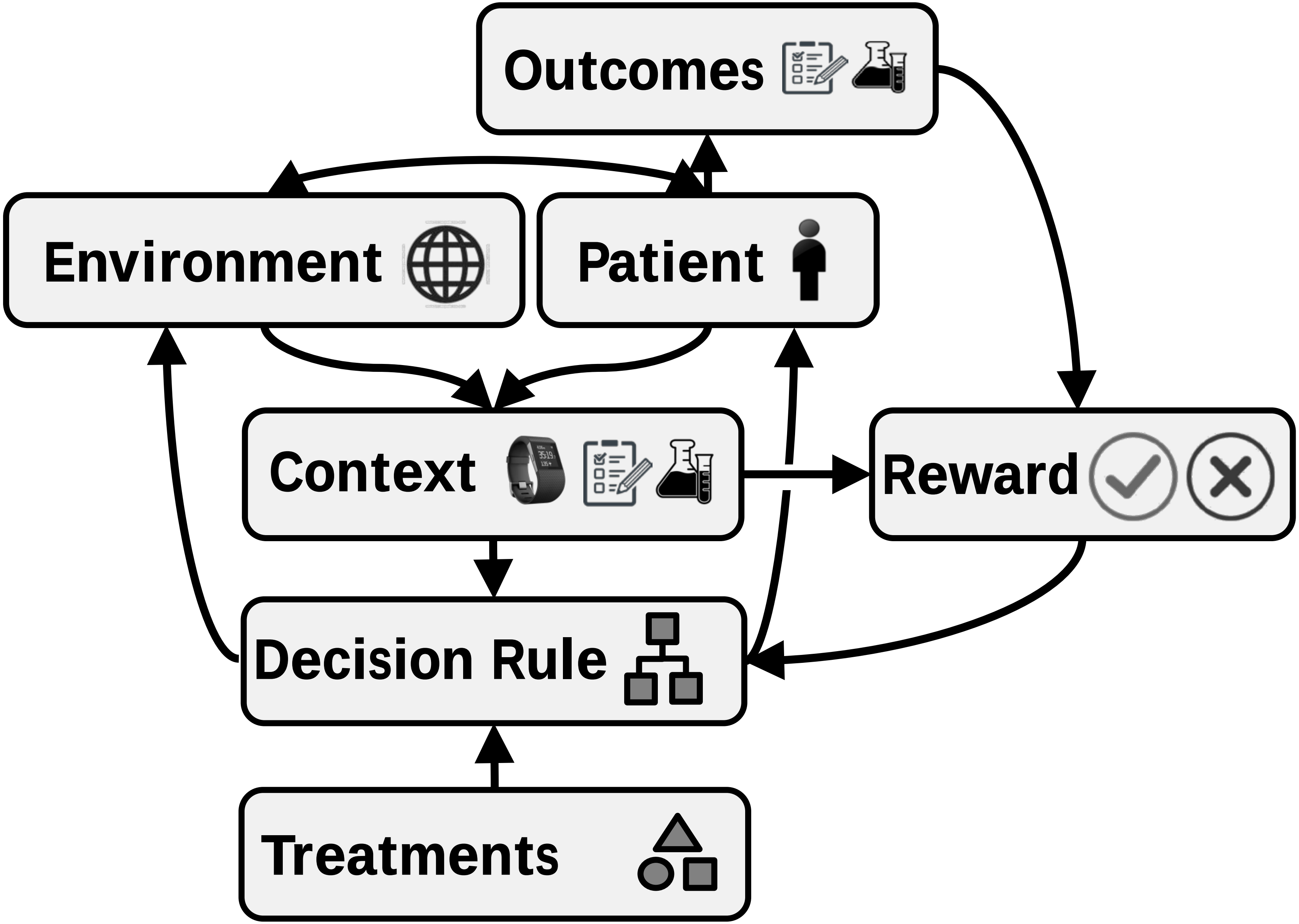
Answer: There is nothing to tell us what the *right* treatment option is at each time point!

Predicting Proximal Outcomes:

- One solution is to learn a model that maps from context and intervention/treatment component selections to proximal outcomes.
- You then select the intervention component that leads to the desired proximal outcome at each decision point.
- Can not be learned from observational data. Need to apply SMART/MRT designs.
- **What are the issues with this approach?**

Reinforcement Learning:

- Define a *reward function* that assigns higher scores to “better” states of the system.
- Chose the decision rule that maximizes the long-term sum of rewards.
- Requires iteratively exploring the space of decision rules to improve the long-term sum of rewards.
- Can be applied in within, between and transfer modes as with detector learning.


Outcomes
Environment
Patient
Context
Reward
Decision Rule
Treatments

Closely Related Ideas:

- **Optimal Control:** Finds a control law (decision function) for a given system such that a certain optimality condition is achieved.

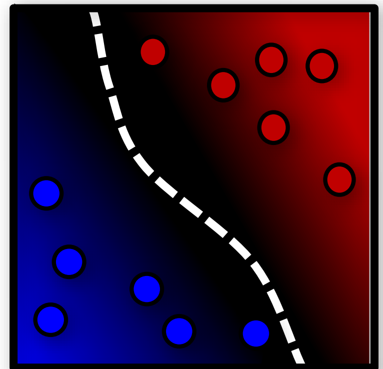
And More...

Machine Learning

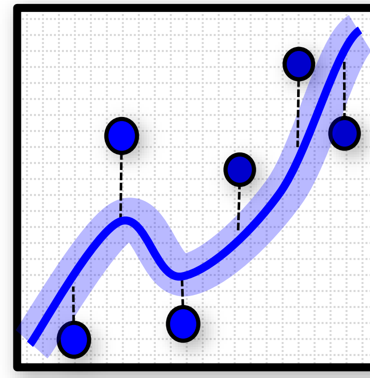
Supervised

Learning to detect and predict.

Classification

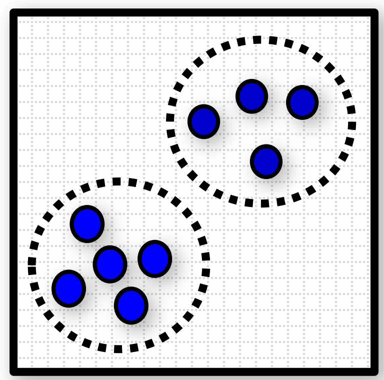


Regression

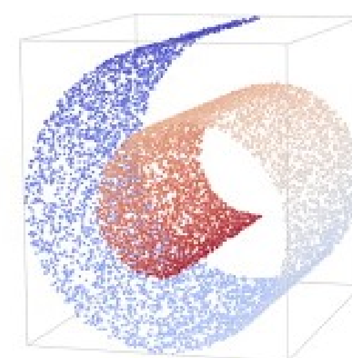


Unsupervised

Learning to organize and represent.



Clustering



Dimensionality Reduction

Conclusions

- Machine learning provides a powerful data analysis toolkit that can be applied solve many problems.
- As an ML user, you have to make sure you design valid ML experiments that correctly assess the type of generalization performance you care about.
- Keep in mind that mHealth presents a number of challenges for machine learning. Stay tuned as we work on developing generalized solutions.

Thank You!

Collaborators and Sponsors

Roy Adams
UMass CS PhD candidate

Malai Natarajan
UMass CS PhD candidate

Abhinav Parate
UMass CS PhD graduate

Nazir Saleheen
U of Memphis PhD Candidate
Computer Science

Prof. Deepak Ganesan
UMass Amherst
Computer Science

Prof. Santosh Kumar
U of Memphis
Computer Science

Prof. Donna Spruijt-Metz
USC

MD2K Center of Excellence for Mobile Sensor Data-to-Knowledge

NIH National Institutes of Health

NSF

ARL