Adversarial Machine Learning in Sequential Decision Making

KDD AdvML Workshop 2019

Jerry Zhu
University of Wisconsin-Madison

Awesome collaborators

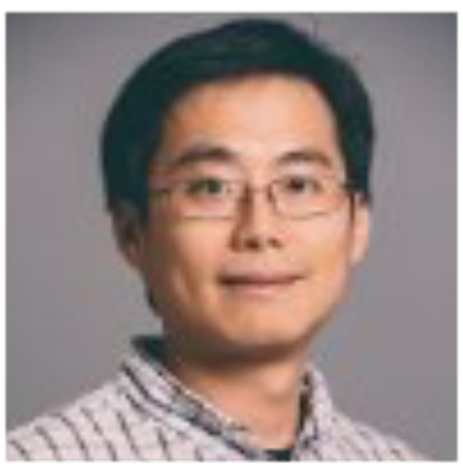
Scott Alfeld

Yiding Chen

Kwang-Sung Jun

Laurent Lessard

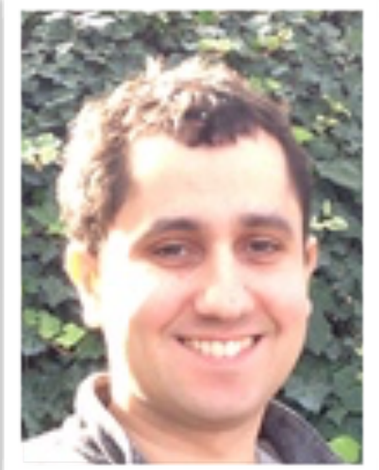
Owen Levin



Lihong Li

Yuzhe Ma

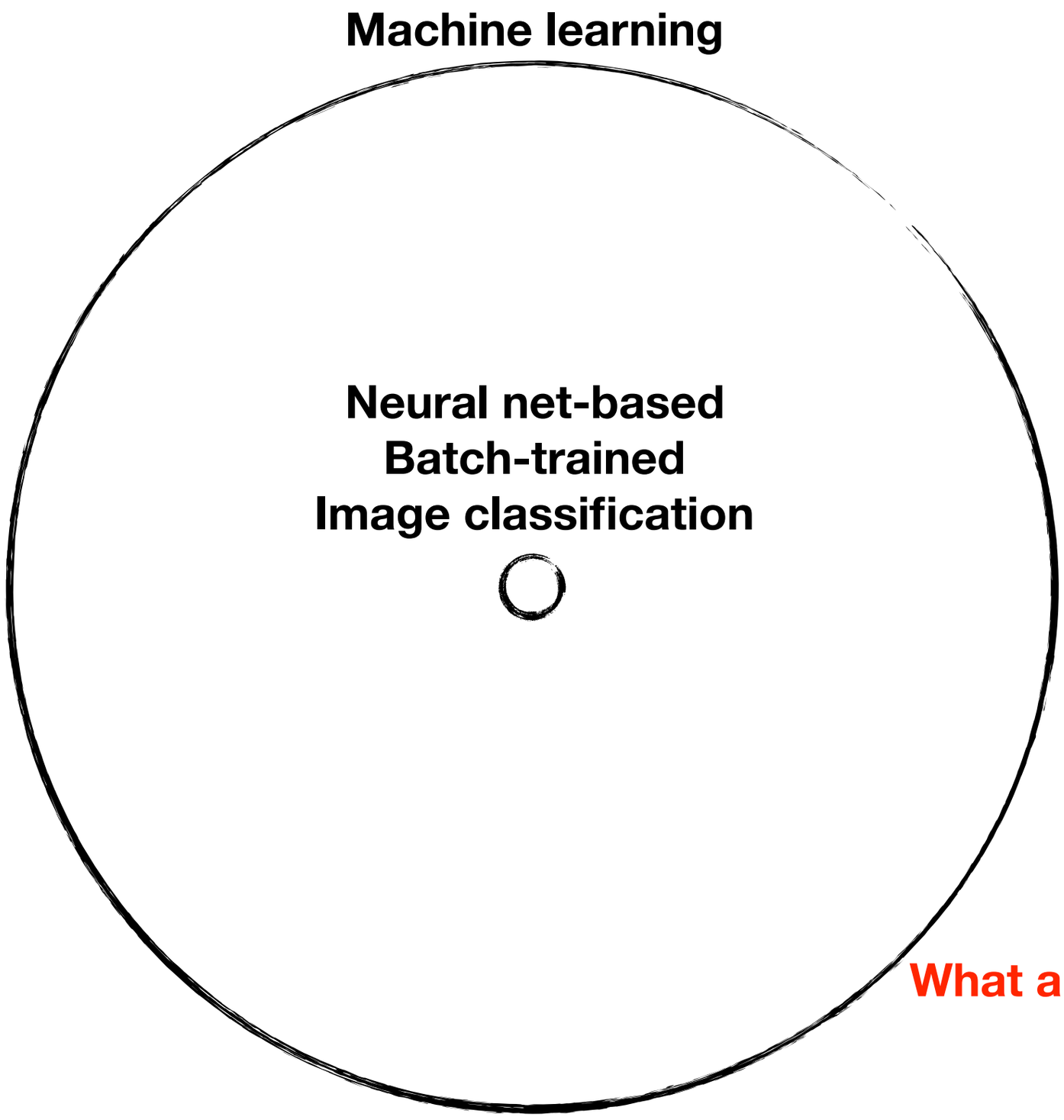
Ayon Sen



Ara Vartanian

Xuezhou Zhang

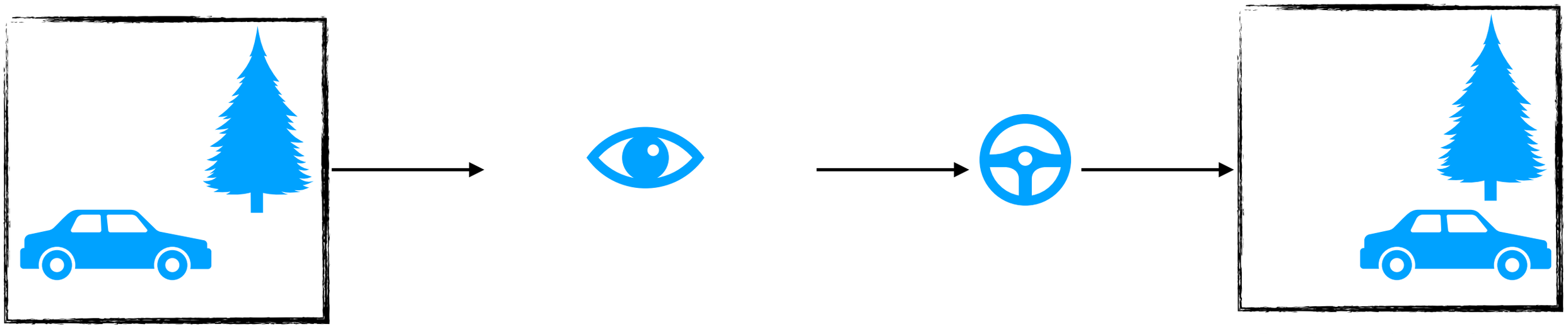
Comprehensive Adversarial Machine Learning



Control

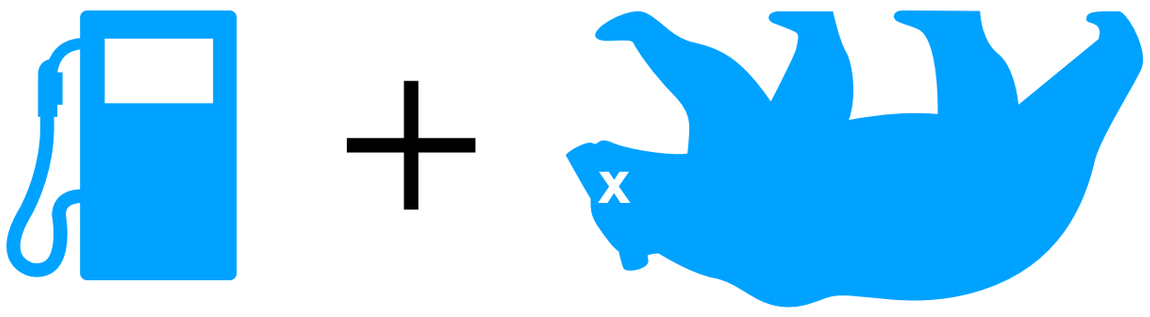
Control

(Alaska style)



state s **observation** o **action** a **state** $s' = f(s, a)$

cost $g(s, a) =$



1 **1,000,000**

$$\min_a g(s, a)$$

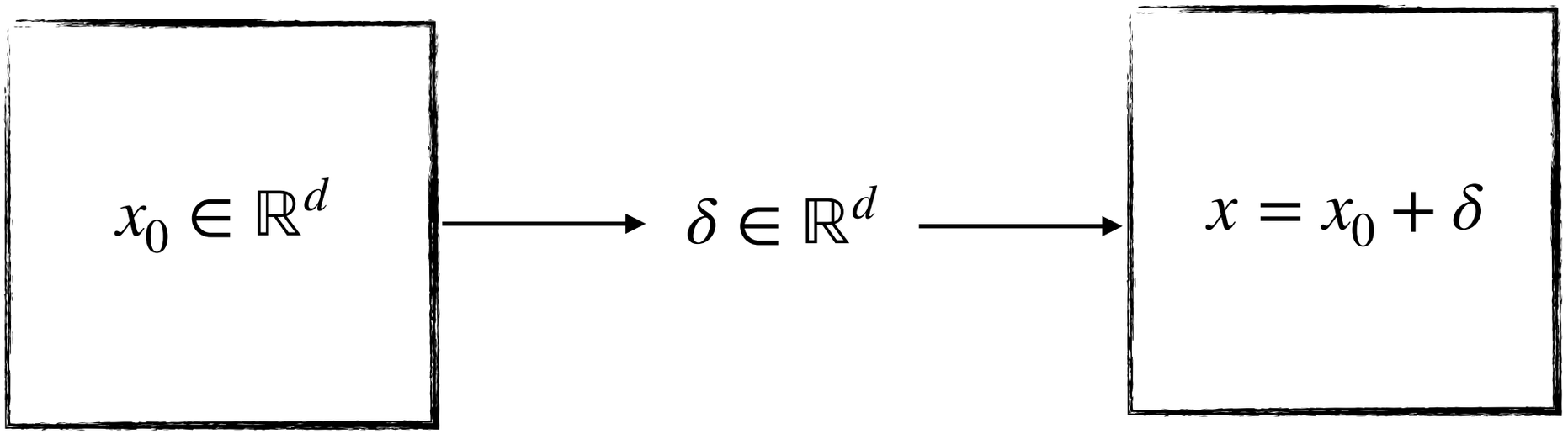
Control

- Plant
- State $s_0, \ldots, s_T \quad (x)$
- Action set $A \quad (U)$
- Action $a_t \in A_t \quad (u_t)$
- State transition $s_{t+1} = f(s_t, a_t)$
- Running cost $g_0(s_0, a_0) \ldots g_{T-1}(s_{T-1}, a_{T-1})$
- Terminal cost $g_T(s_T)$
- Policy $a_t = \phi(s_t)$

$$\min_{\phi} g_T(s_T) + \sum_{t=0}^{T-1} g_t(s_t, a_t)$$

s.t. s_0, f **given,** $s_{t+1} = f(s_t, a_t)$

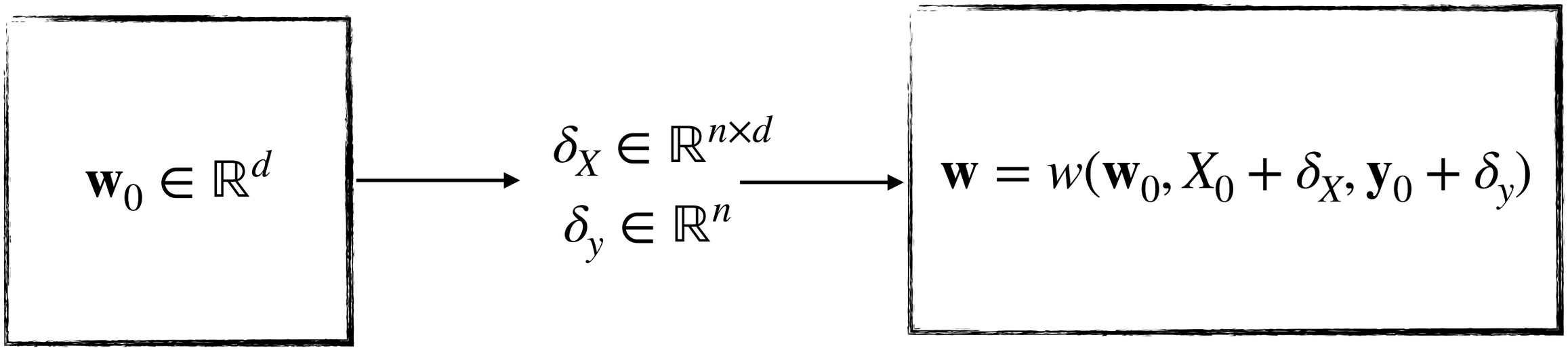
Test-Time Attack = Control



$$\textbf{cost}\ g(s,a) = \|\delta\|_p + \infty \cdot [y_w(x) \neq y^\dagger]$$

$$\min_a g(s,a)$$

Training Poisoning = Control



cost $g(s, a) = \|\delta\|_p + \infty \cdot [w \neq w^\dagger]$

$$\min_a g(s, a)$$

Adversarial Attack = Control

$$\min_{\phi} g_T(s_T) + \sum_{t=0}^{T-1} g_t(s_t, a_t)$$

s.t. s_0, f **given,** $s_{t+1} = f(s_t, a_t)$

- Good conceptual framework
- Doesn't mean easy to compute
- "One-step control is not control" — Steve Wright
  - Sequential models (today)
- Control or reinforcement learning?

Control or RL?

Whitebox attacks	Blackbox attacks
s_0, f **given**	s_0, f **unknown**
Control (planning)	**RL (system identification)**

This talk: 3 case studies

Case study 1

Optimal Sequential Attack = Optimal Control

- Example victim: online gradient descent with squared loss

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \frac{\eta}{2}\nabla(\mathbf{x}_t^\top \mathbf{w}_t - y_t)^2 = \mathbf{w}_t - \eta(\mathbf{x}_t^\top \mathbf{w}_t - y_t)\mathbf{x}_t$$

- Example attacker:
  - Modifies $\|\mathbf{x}_t\| \le 1, |y_t| \le 1$
  - Minimizes T such that $\mathbf{w}_T = \mathbf{w}^\dagger$

Optimal Sequential Attack = Optimal Control

$$\min_{T,(\mathbf{x},y)_{0:T-1}} \sum_{t=0}^{T-1} 1 + \infty \cdot [\mathbf{w}_T \neq \mathbf{w}^\dagger]$$

$$\textbf{s.t. } \mathbf{w}_{t+1} = \mathbf{w}_t - \eta(\mathbf{x}_t^\top \mathbf{w}_t - y_t)\mathbf{x}_t \quad \textbf{Nonlinear dynamics}$$

$$\|\mathbf{x}_t\| \leq 1, \, |y_t| \leq 1$$

$$\mathbf{w}_0 \textbf{ given}$$

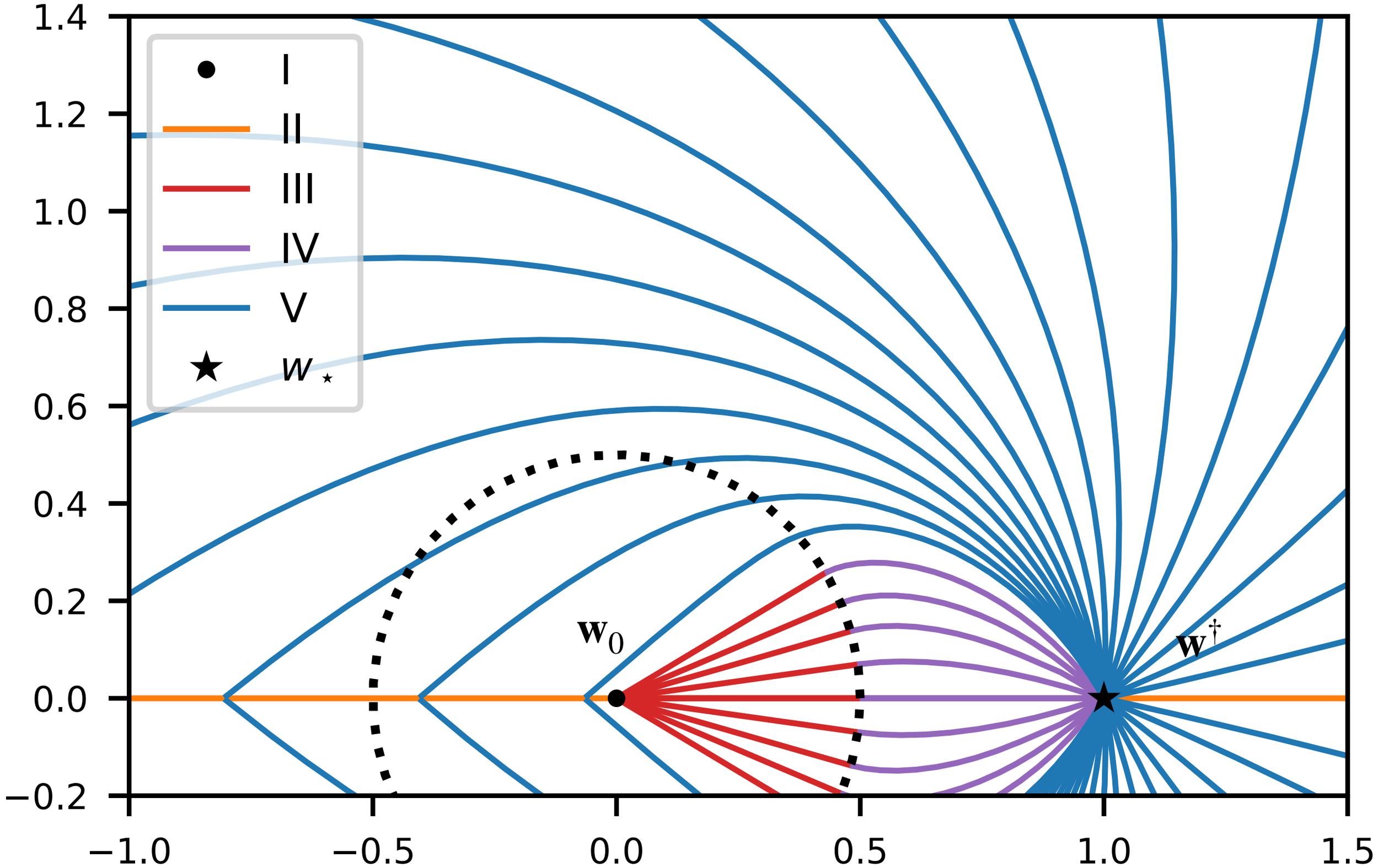
The continuous dynamics

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta(\mathbf{x}_t^\top \mathbf{w}_t - y_t)\mathbf{x}_t$$

$$\frac{\mathbf{w}_{t+1} - \mathbf{w}_t}{\eta} = -(\mathbf{x}_t^\top \mathbf{w}_t - y_t)\mathbf{x}_t$$

$$\dot{\mathbf{w}}(t) = -(\mathbf{x}(t)^\top \mathbf{w}(t) - y(t))\mathbf{x}(t)$$

Pontryagin maximum principle: necessary condition for optimality


I
II
III
IV
V
$w_\star$
$\mathbf{w}_0$
$\mathbf{w}^\dagger$
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−1.0
−0.5
0.0
0.5
1.0
1.5

Optimal vs greedy control

$$\min_{T,(\mathbf{x},y)_{0:T-1}} \sum_{t=0}^{T-1} 1 + \infty \cdot [\mathbf{w}_T \neq \mathbf{w}^\dagger]$$

s.t. $\mathbf{w}_{t+1} = \mathbf{w}_t - \eta(\mathbf{x}_t^\top \mathbf{w}_t - y_t)\mathbf{x}_t$

$\|\mathbf{x}_t\| \leq 1, |y_t| \leq 1$

$\mathbf{w}_0$ **given**

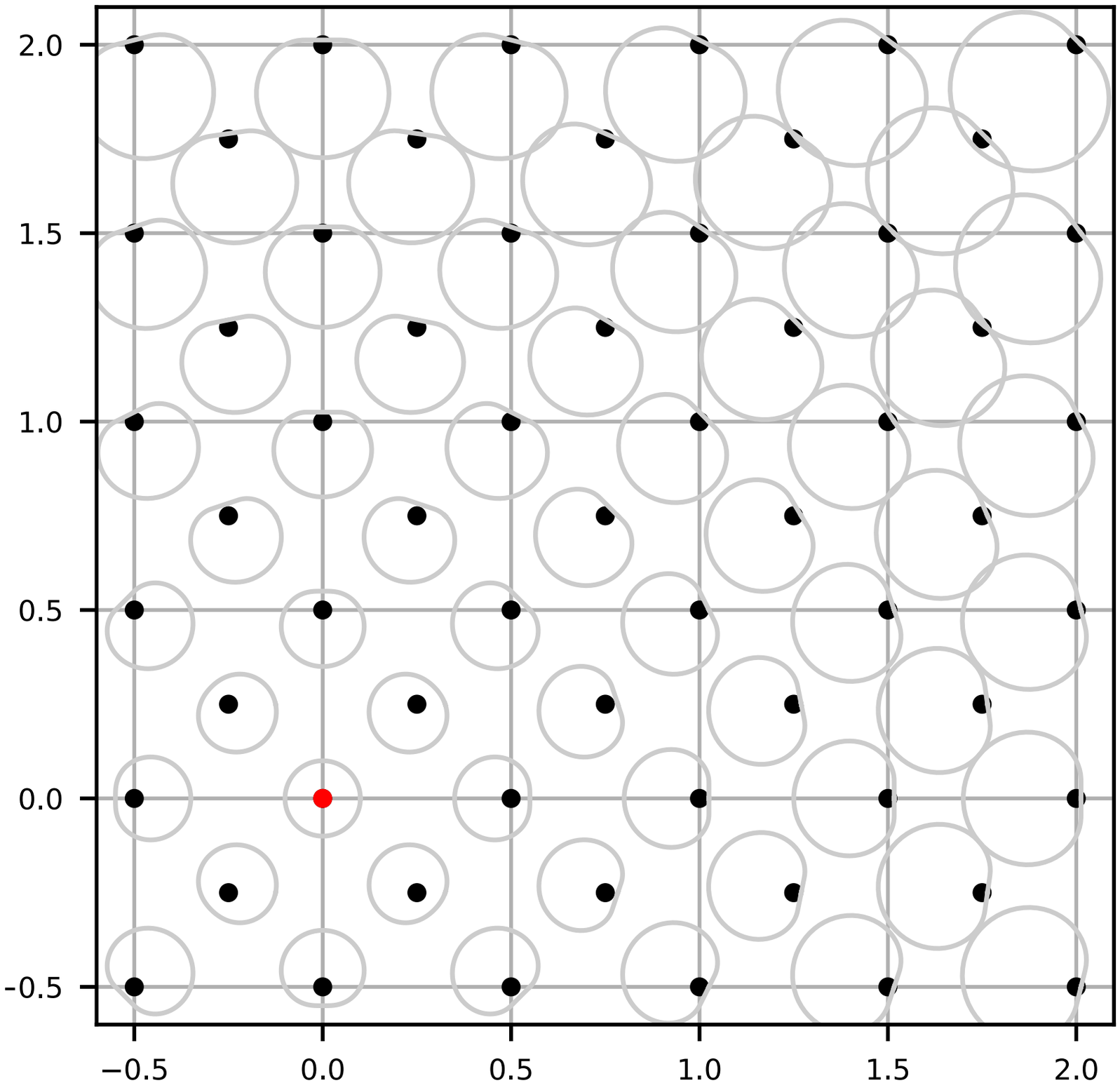
$$\min_{\mathbf{x}_t, y_t} \|\mathbf{w}_{t+1} - \mathbf{w}^\dagger\|$$

s.t. $\mathbf{w}_{t+1} = \mathbf{w}_t - \eta(\mathbf{x}_t^\top \mathbf{w}_t - y_t)\mathbf{x}_t$

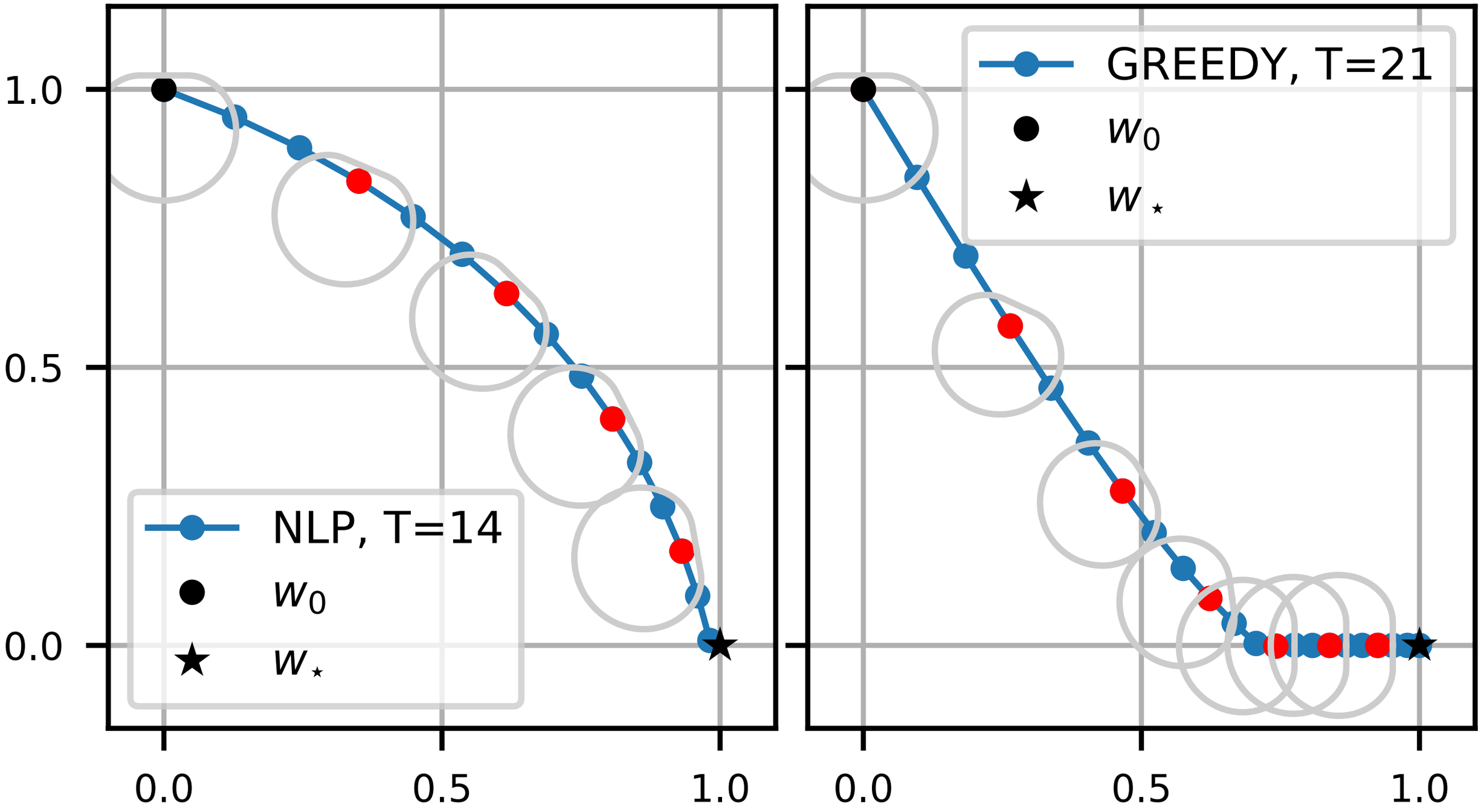
$\|\mathbf{x}_t\| \leq 1, |y_t| \leq 1$

$\mathbf{w}_t$ **given**

Reachable sets $\mathbf{w}_{t+1} = \mathbf{w}_t - \eta(\mathbf{x}_t^\top \mathbf{w}_t - y_t)\mathbf{x}_t$ **s.t.** $\|\mathbf{x}_t\| \le 1, |y_t| \le 1$



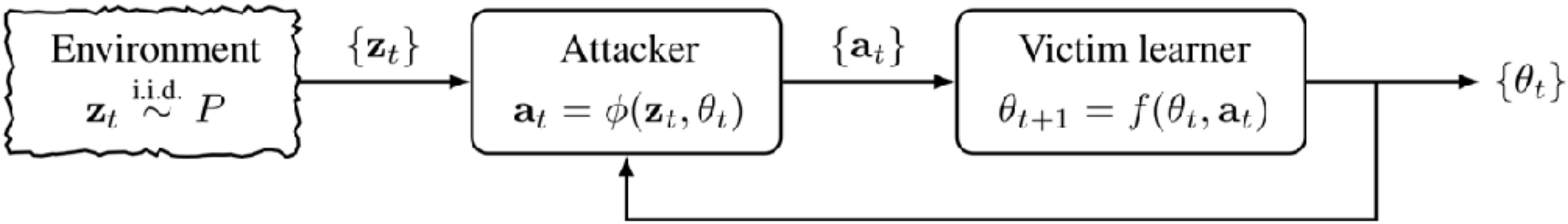
Optimal vs greedy control



Optimal vs greedy control

$\mathbf{w}_0$	$\mathbf{w}^\dagger$	OPTIMAL	GREEDY
$(0, 1)$	$(2, 0)$	148	233
$(0, 2)$	$(4, 0)$	221	721
$(0, 4)$	$(8, 0)$	292	2667
$(0, 8)$	$(16, 0)$	346	10581

Online data poisoning



Attacker does not know future data $z_{t+1}, z_{t+2}, \cdots$

Online data poisoning

- State $s_t = [\theta_t, z_t]$
- Action $a_t \in \mathcal{Z}$
- Transition $s_{t+1} = [f(\theta_t, a_t), z_{t+1} \sim P]$ **Stochastic disturbance**
- Running cost $g(s_t, a_t) = g_1(z_t, a_t) + \lambda g_2(\theta_t)$
  - example: $g_1(z_t, a_t) = \|z_t - a_t\|^2, \quad g_2(\theta_t) = \|\theta_t - \theta^\dagger\|^2$ **Tracking**
- Discount factor γ
- Policy $a_t = \phi(s_t)$

Online data poisoning

$$\min_{\phi} \mathbb{E}_P \sum_{t=0}^{\infty} \gamma^t g(s_t, \phi(s_t))$$

s.t. θ_0, f **given**

But P unknown…

$$\mathbb{E}_P$$

$$s_0 = [\theta_0, z_0 \sim P]$$

$$s_{t+1} = [f(\theta_t, a_t), z_{t+1} \sim P]$$

Solution: Model Predictive Control

- At each time step t
  - Estimate $\hat{P}_t$ from $z_0, \ldots, z_t$
  - Plan ahead from $s_t = [\theta_t, z_t]$: $\hat{\phi}_t = \arg\min_{\phi} \mathbb{E}_{\hat{P}_t} \sum_{\tau=t}^{\infty} \gamma^{\tau-t} g(s_\tau, \phi(s_\tau))$
  - But only execute one step

$$a_t = \hat{\phi}_t(s_t)$$

Subproblem: planning ahead

$$\hat{\phi}_t = \arg\min_{\phi} \mathbb{E}_{\hat{P}_t} \sum_{\tau=t}^{\infty} \gamma^{\tau-t} g(s_\tau, \phi(s_\tau))$$

Nonlinear programming (NLP):

$\mathbb{E}_{\hat{P}_t}$ **Monte Carlo** $z_{t:t+h-1} \sim \hat{P}_t$

Truncation $\tau = t \ldots t+h-1$

Action sequence instead of policy

$$\min_{a_{t:t+h-1}} \sum_{\tau=t}^{t+h-1} \gamma^{\tau-t} g(s_\tau, a_\tau)$$

Or reinforcement learning (DDPG)

It's OK to estimate $\hat{P}_t$

With probability at least $1 - \delta$

$$\sup_{s \in S} V^{\phi^*_{\hat{P}_t}}(s) - V^{\phi^*_P}(s) \leq \frac{2\gamma C_{\max}}{(1-\gamma)^2} \sqrt{\frac{1}{2t} \ln \frac{2^{|\mathcal{Z}|+1}}{\delta}} = O(t^{-1/2})$$

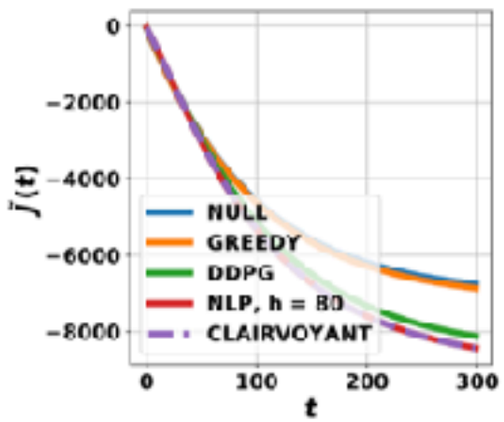
where

$$V^{\phi}(s) := \mathbb{E}_P \sum_{t=0}^{\infty} \gamma^t g(s_t, \phi(s_t)) \Big|_{s_0 = s}$$

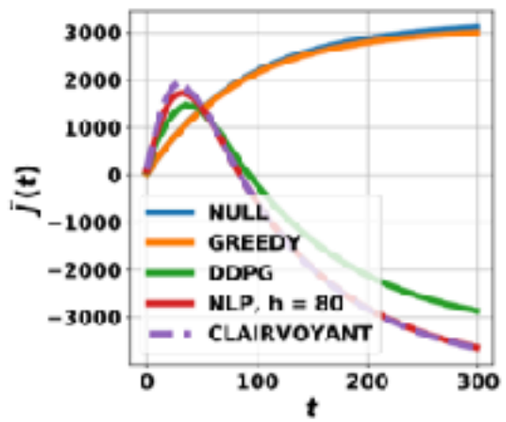
Online data poisoning experiments

(random $\theta^\dagger$)

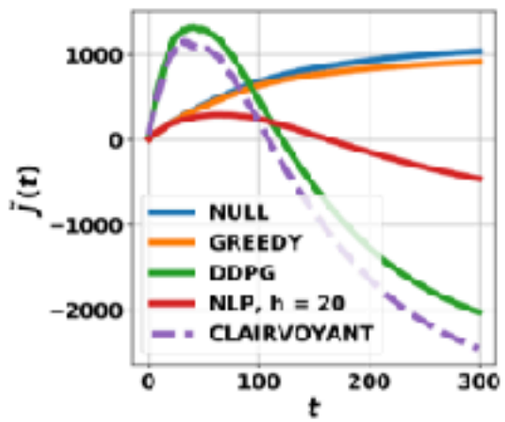
Online logistic regression victims



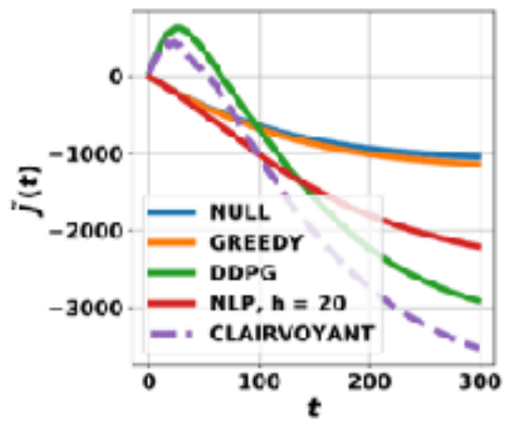
(a) Banknote



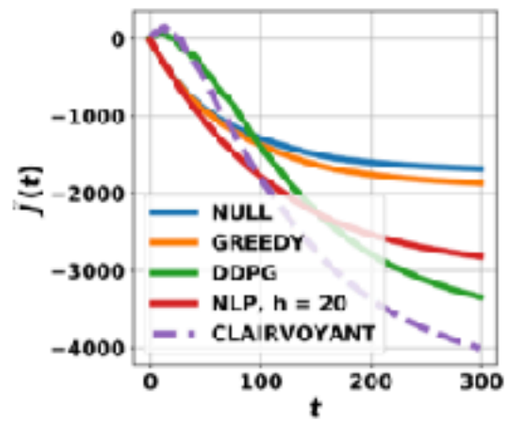
(b) Breast



(c) CTG

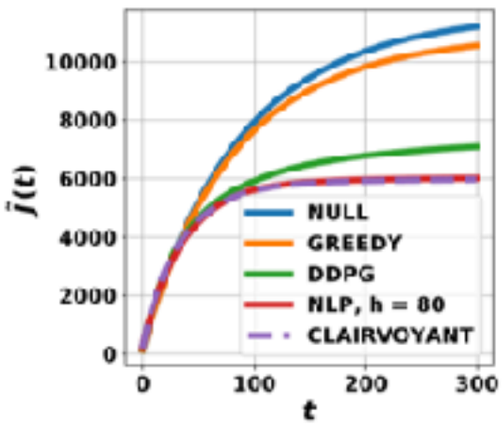


(d) Sonar

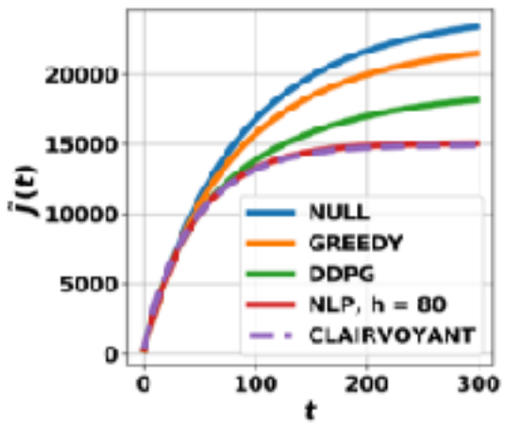


(e) MNIST 1 vs. 7

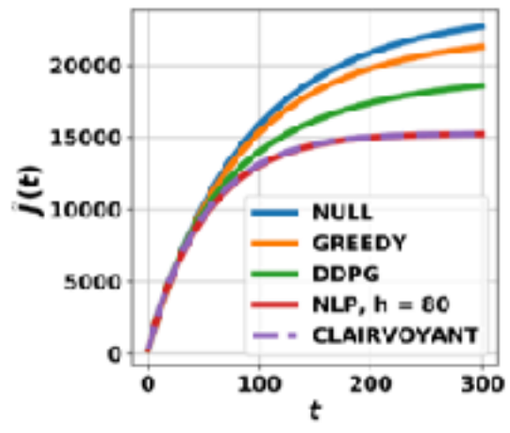
Online k-means victims



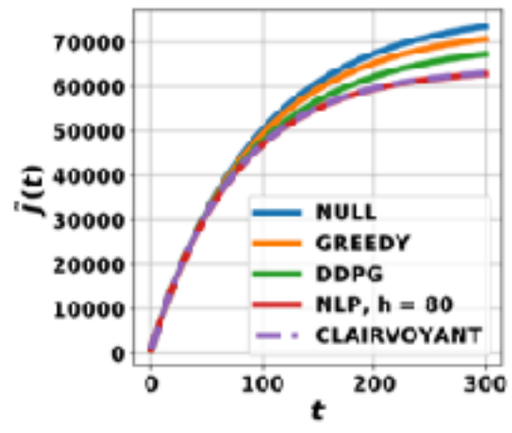
(f) Knowledge



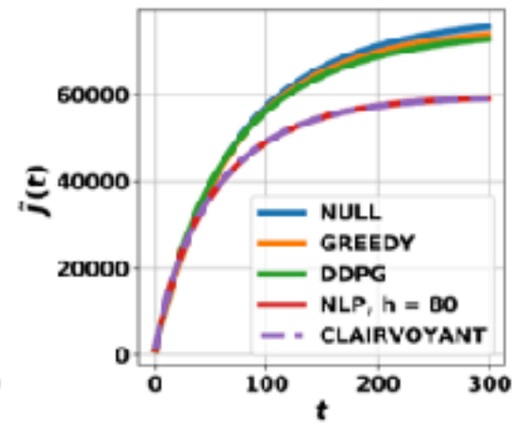
(g) Breast



(h) Seeds



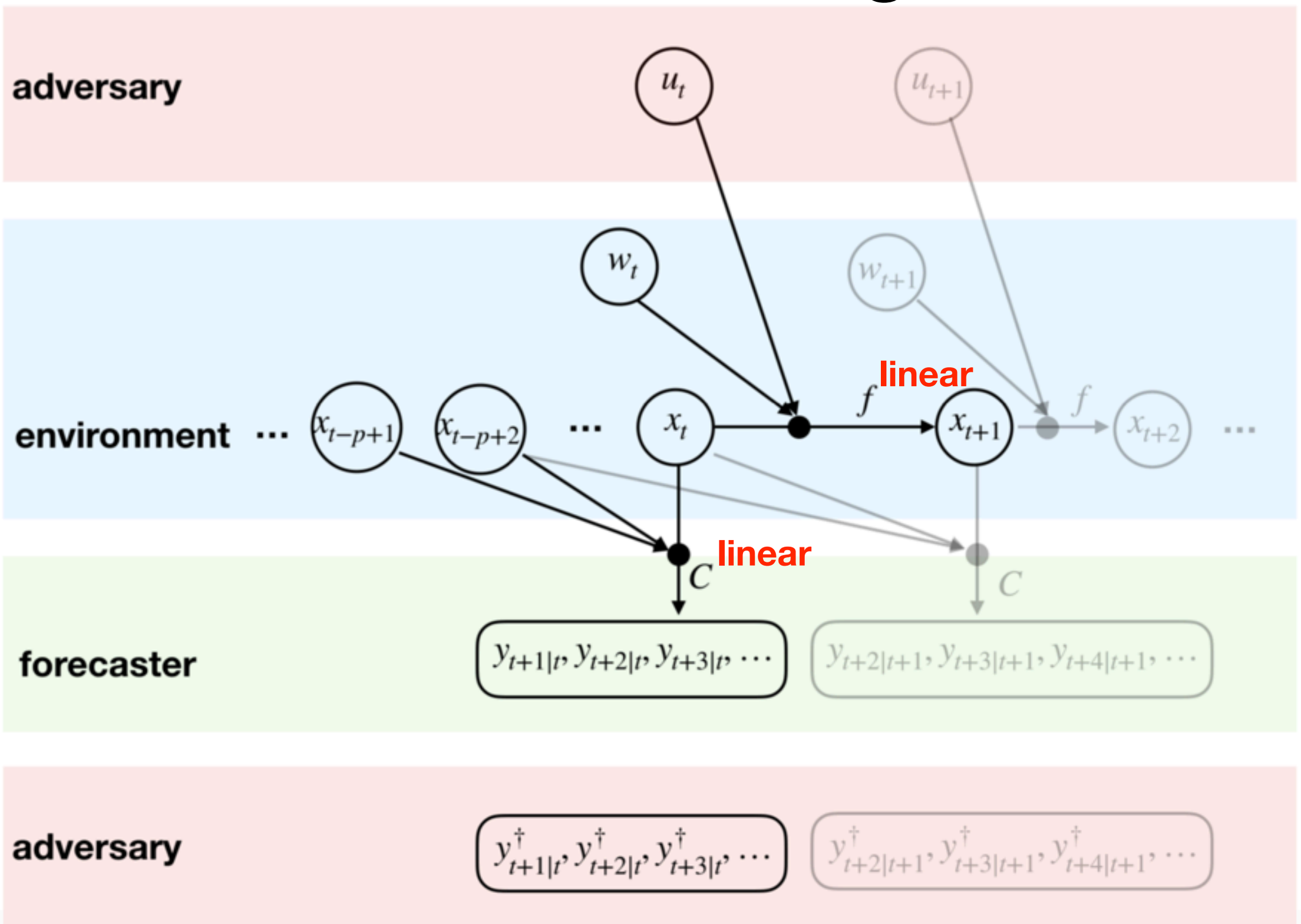
(i) Posture



(j) MNIST 1 vs. 7

Case study 2

When we know the optimal attack: Linear Quadratic Regulator



Linear dynamics

- Linear environment

$$x_{t+1} = f(x_t, \ldots, x_{t-q+1}, w_t)$$

Stochastic disturbance

- Linear forecaster AR(p) model

$$y_{t+1|t} = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i y_{t+1-i|t}$$

- Quadratic costs

$$(y_{t'|t} - y^{\dagger}_{t'|t})^2 + \lambda u_t^2$$

Optimal attack as LQR

- Dynamics can be written in matrix form

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B(u_t + w_t)$$

$$\mathbf{y}_{t'|t} = C^{t'-t}\mathbf{x}_t$$

- So can the cost

$$\mathbb{E}_{w_{0:(T-1)}}\left[\sum_{t=1}^{T-1} \|C\mathbf{x}_t - \mathbf{y}^{\dagger}_{t+1|t}\|^2_{Q_{t+1|t}} + \sum_{t=0}^{T-1} \|u_t\|^2_R \,\middle|\, \mathbf{x}_0\right]$$

Optimal solution: Riccati equations

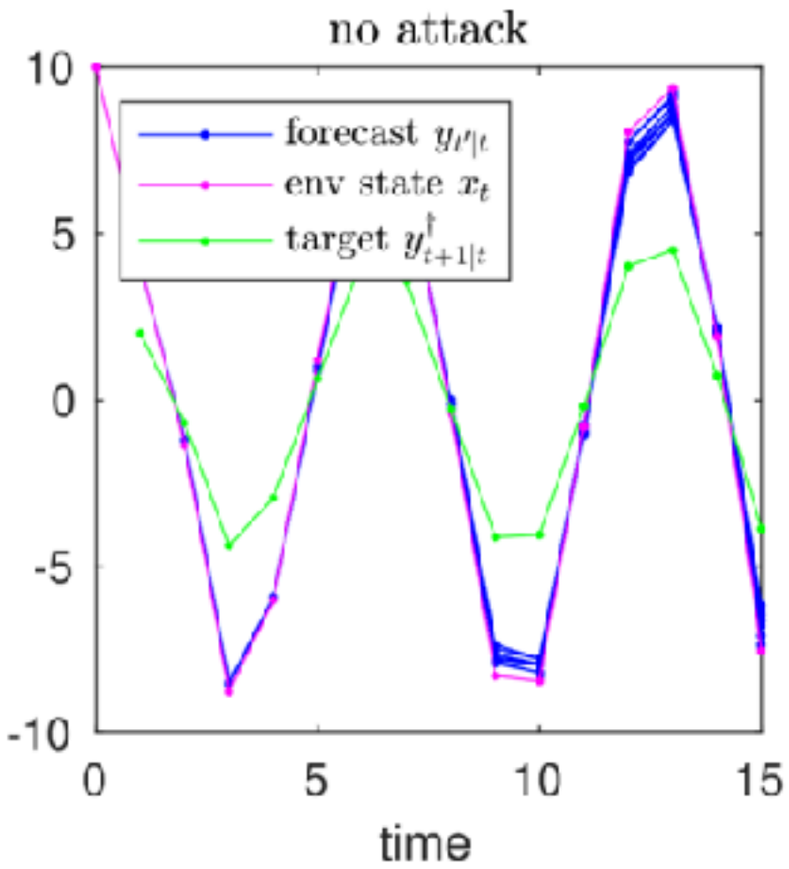
$$\phi_t(\mathbf{z}) = -\frac{B^\top \mathbf{q}_{t+1} + 2B^\top P_{t+1} A\mathbf{z}}{2(\lambda + B^\top P_{t+1} B)}, \quad t = 0 \ldots T-1.$$

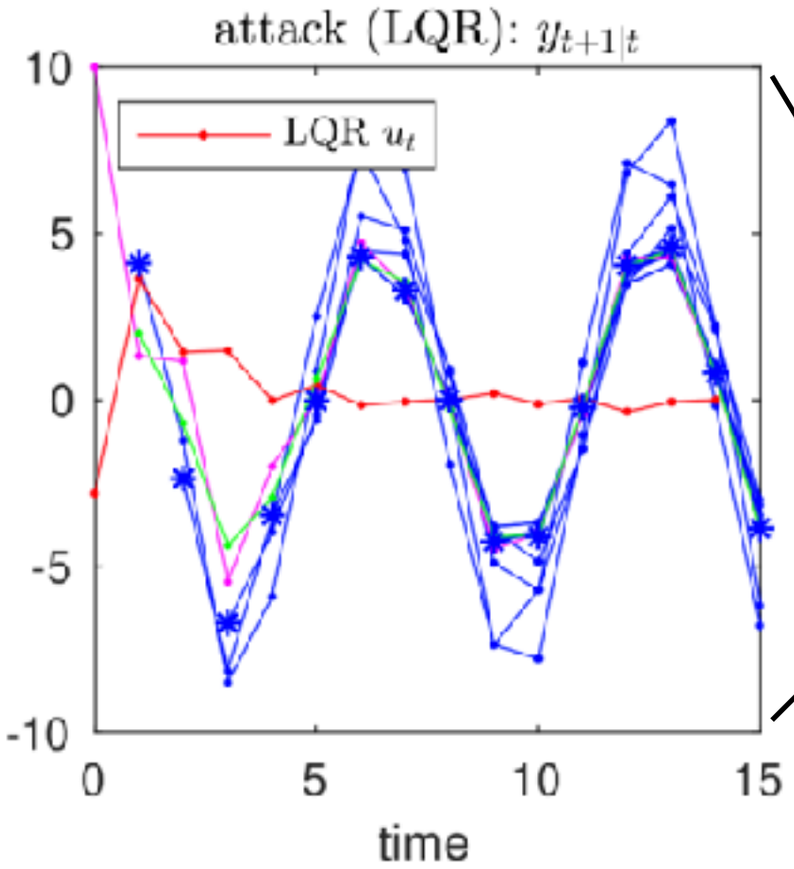
$$P_T = 0, \quad \mathbf{q}_T = 0$$

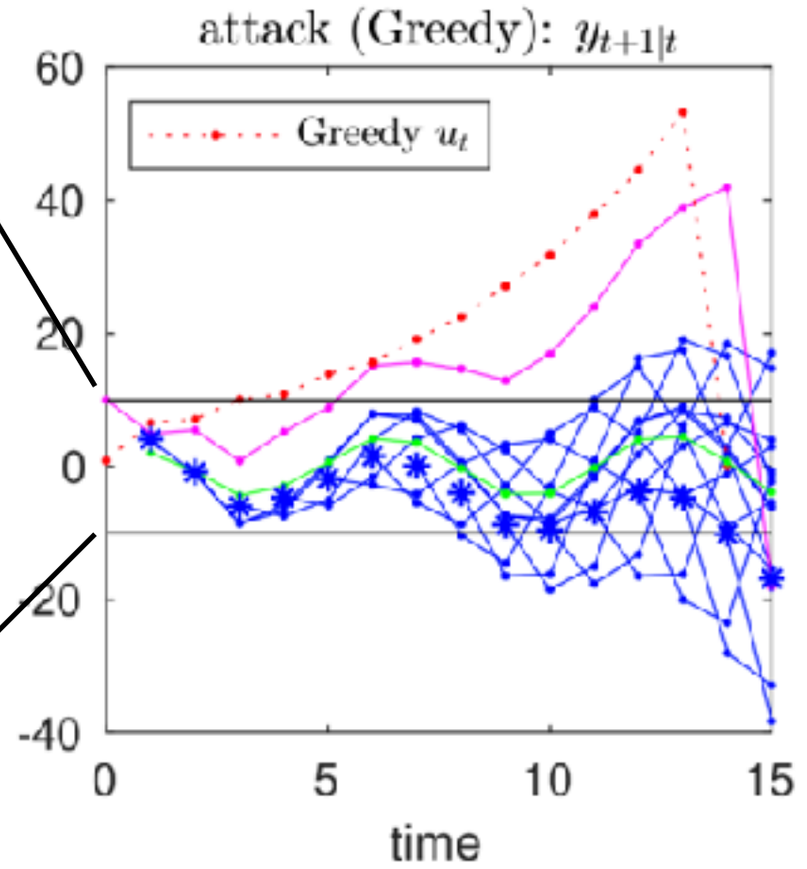
$$P_t = CA^\top Q_{t+1|t} C + A^\top (I + \frac{1}{\lambda} P_{t+1} BB^\top)^{-1} P_{t+1} A$$

$$\mathbf{q}_t = -2C^\top Q_{t+1|t} \mathbf{y}^\dagger_{t+1|t} + A^\top \mathbf{q}_{t+1} - \frac{1}{\lambda + B^\top P_{t+1} B} A^\top P_{t+1}^\top BB^\top \mathbf{q}_{t+1}$$

Optimal vs. greedy attacks







Case study 3

Attacking Multi-Armed Bandits

Input: Bob's bandit algorithm, target arm K
for $t = 1, 2, \ldots$ **do**
    Bob chooses arm I_t to pull.
    World generates pre-attack reward r_t^0.
    Alice observes I_t and r_t^0, and then decides the attack α_t.
    Alice gives $r_t = r_t^0 - \alpha_t$ to Bob.
end for

Goal: make Bob frequently pull suboptimal (not necessarily the worst) arm K.

Attacking Multi-Armed Bandits

- State

$$s_t = \begin{bmatrix} N_1(t) \\ \vdots \\ N_K(t) \\ \hat{\mu}_1(t) = \frac{1}{N_1(t)} \sum_{\tau:I_\tau=1} r_\tau \\ \vdots \\ \hat{\mu}_K(t) \end{bmatrix}$$

Number of pulls on arm 1

Empirical average of arm 1

- Action

$$a_t \in \mathbb{R}$$

Attacking Multi-Armed Bandits

- (stylized) cost

$$\sum_{t=1}^{\infty} |a_t| + \lambda \cdot [I_{t+1} \neq K]$$

- Bob's arm choice depends on his bandit algorithm:
  - epsilon-greedy

$$I_t = \begin{cases} \text{draw uniform}[K], & \text{w.p. } \epsilon_t \quad \text{(exploration)} \\ \arg\max_i \hat{\mu}_i(t-1), & \text{otherwise (exploitation)} \end{cases}$$

  - UCB

$$I_t = \begin{cases} t, & \text{if } t \leq K \\ \arg\max_i \left\{ \hat{\mu}_i(t-1) + 3\sigma \sqrt{\frac{\log t}{N_i(t-1)}} \right\}, & \text{otherwise.} \end{cases}$$

Heuristic solution

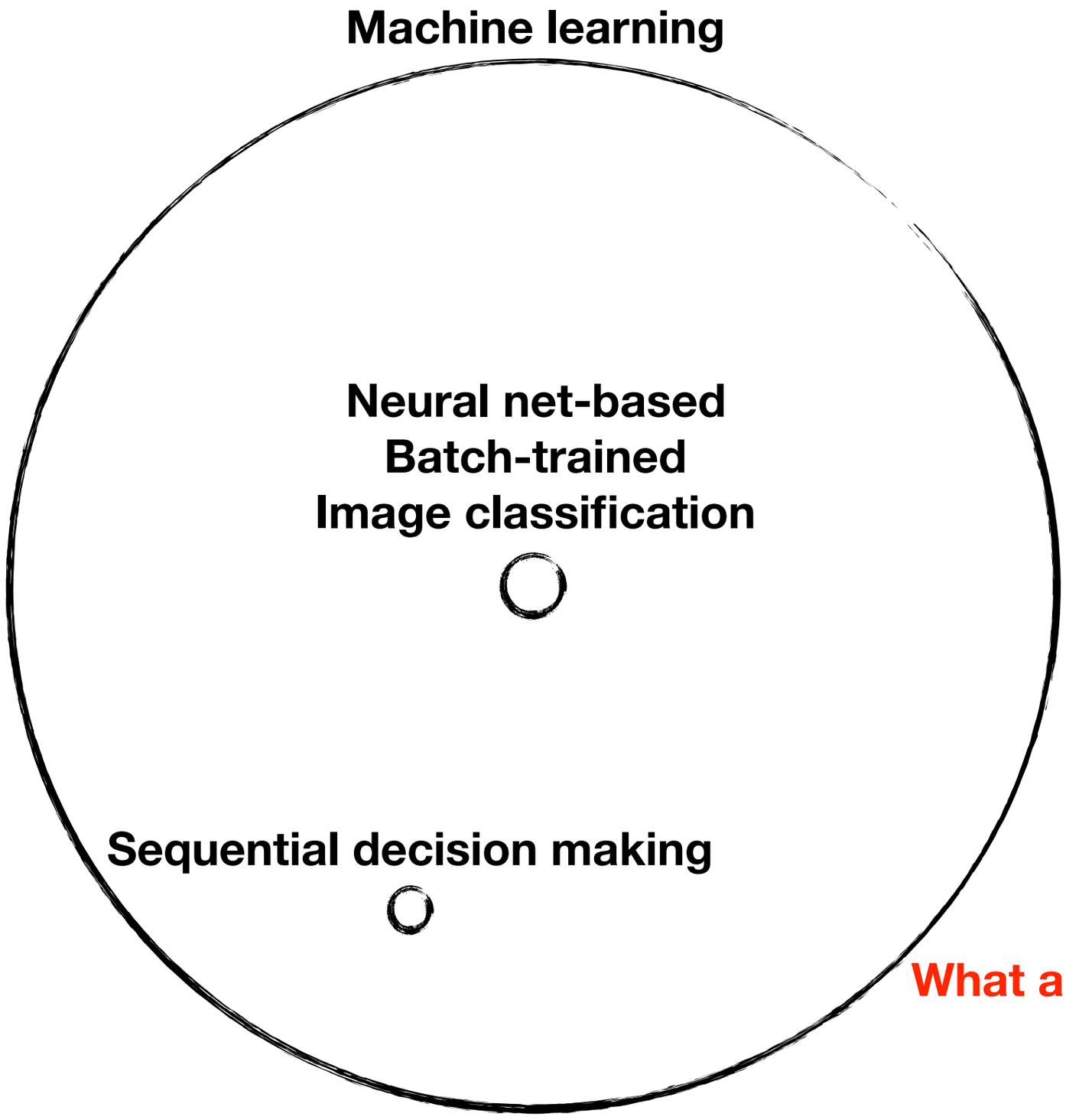
- Whenever Bob pulls a non-target (not K) arm, make it look worse than arm K.

- Theorem:

  - Alice can force Bob to pull arm K for *T-o(T)* times

  - Alice's control expenditure $\sum_{t=1}^{T} |a_t| = O(\log T)$

Beyond Attacks on Sequential Models



What are other threats?

References

- Y Chen, X Zhu. **Optimal Adversarial Attack on Autoregressive Models**. arXiv:1902.00202, 2019

- K Jun, L Li, Y Ma, X Zhu. **Adversarial attacks on stochastic bandits**. NeurIPS, 2018

- L Lessard, X Zhang, X Zhu. **An optimal control approach to sequential machine teaching**. AISTATS, 2019

- X Zhang, X Zhu, L Lessard. **Online Data Poisoning Attacks**. arXiv:1903.01666, 2019

- X Zhu. **An optimal control view of adversarial machine learning**. arXiv:1811.04422, 2018