Lecture 10: Advanced Topics in Control

Professor Katie Driggs-Campbell

March 4, 2021

ECE484: Principles of Safe Autonomy

Administrivia

- Introducing half-time breaks
- Safety training information posted on discord
- MP1 due this week
  - Demo due today
  - Report due Friday
- Milestone report due Friday 3/19 by 5pm
  - Rubric now online!

Today's Plan

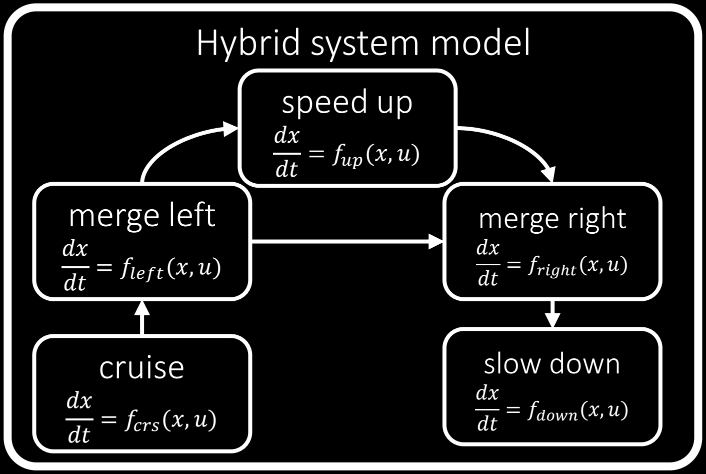
- Quick discussion of future topics in advanced control theory
- Introduction to "optimal control"
  - Linear Quadratic Regulation (LQR)
  - Model Predictive Control (MPC)
- End-to-end learning

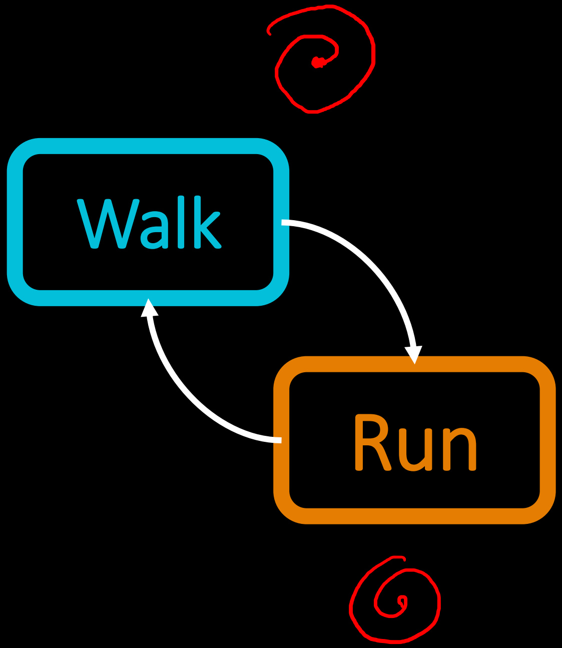
Today's Plan

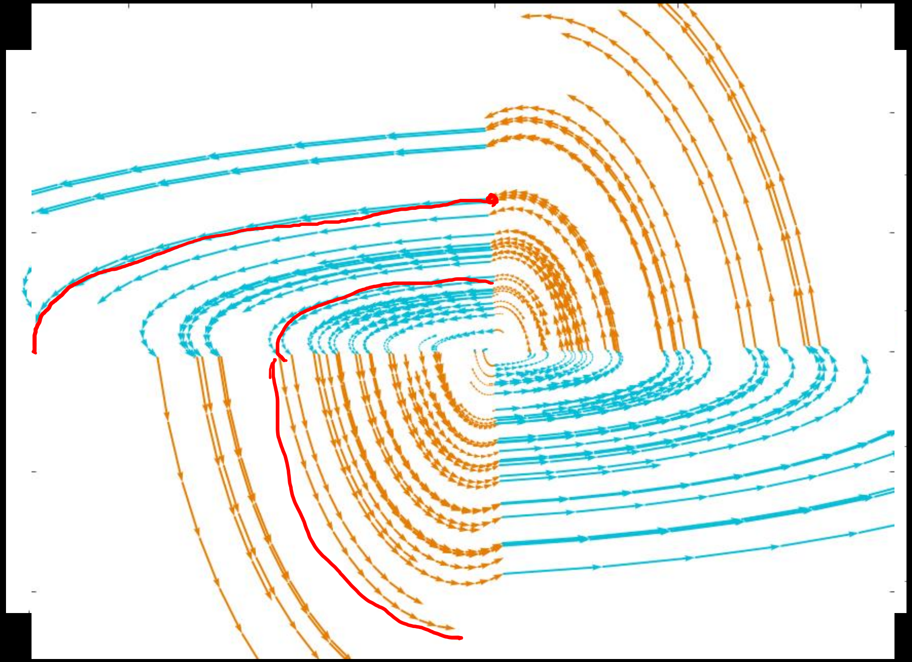
- Quick discussion of future topics in advanced control theory
- Introduction to optimal control
  - Linear Quadratic Regulation (LQR)
  - Model Predictive Control (MPC)
- End-to-end learning

Extensions from Control Theory

1. Hybrid Control
   - Given discrete modes of continuous behavior, can we guarantee stability?



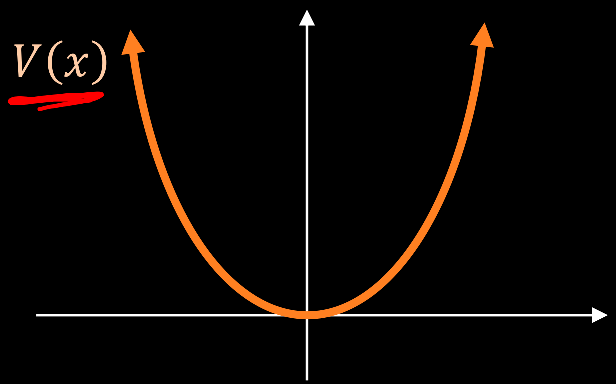




UIUC has one of the best
control programs in the country!
Consider some grad courses in this area!

Extensions from Control Theory

1. Hybrid Control
   - Given discrete modes of continuous behavior, can we guarantee stability?
2. Lyapunov Stability
   - The system is said to be Lyapunov stable about an equilibrium if
     $\forall \varepsilon > 0 \; \exists \delta_\varepsilon > 0$ such that $|x_0| \leq \delta_\varepsilon \Rightarrow \forall t \geq 0, |\xi(x_0, t)| \leq \varepsilon$



Today’s Plan

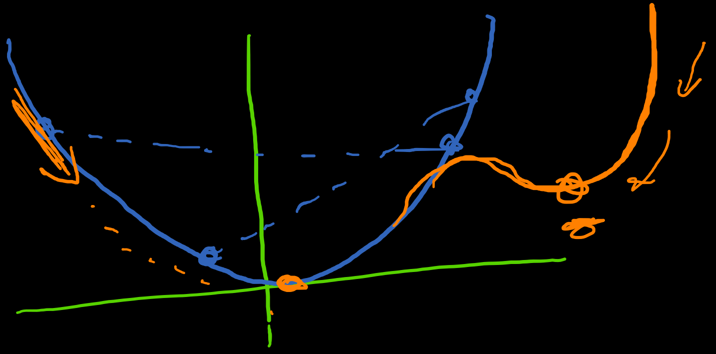
- Quick discussion of future topics in advanced control theory
- Introduction to optimal control
  - Linear Quadratic Regulation (LQR)
  - Model Predictive Control (MPC)
- End-to-end learning

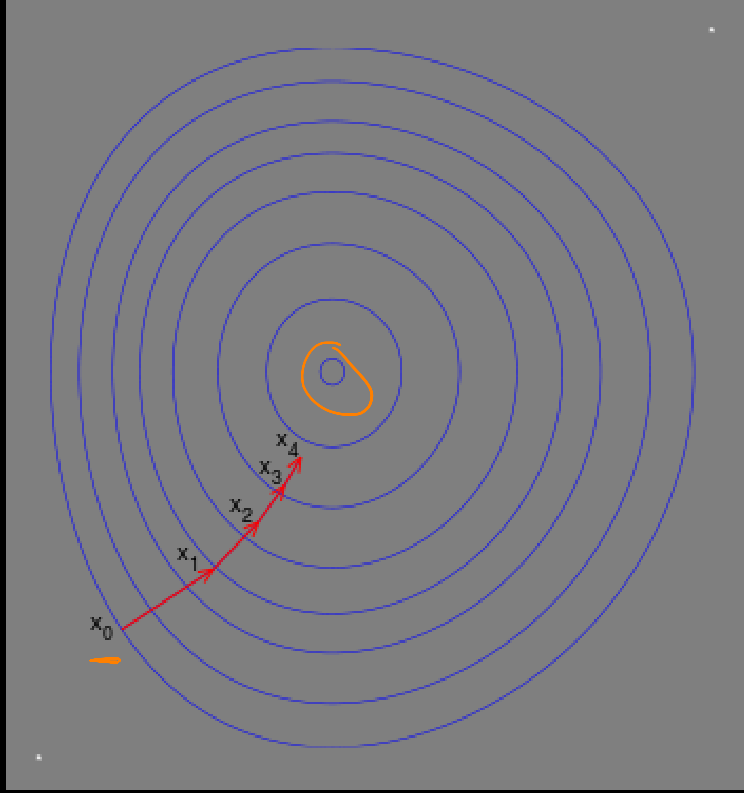
Convex Optimization

minimize$_x$ $f(x)$

subject to
- $g_i(x) \le 0, \ \forall i$
- $h_j(x) = 0, \ \forall j$

if f is convex:
→ every local min is a global min





Linear Quadratic Regulation (LQR)

$x_{t+1} = Ax_t + Bu_t$ // $x = [x_1 \cdots x_T]^T$, $u = [u_1 \cdots u_T]^T$

goal: find input to min cost:

→ $J(x,u) = x_T^T Q_F x_T + \sum_t x_t^T Q x_t + u_t^T R u_t$

where $Q, R, Q_F \in \mathbb{R}^{n \times n}$ → solve w/ Least Squares!

$x_2 = Ax_1 + Bu_1$ // $x_3 = Ax_2 + Bu_2 = A(Ax_1 + Bu_1) + Bu_2$

→ full traj: $x = Gu + Hx_1$

$J = x^T \begin{bmatrix} Q & & 0 \\ & Q & \\ 0 & \ddots & Q_F \end{bmatrix} x + u^T \begin{bmatrix} R & & \\ & \ddots & \\ & & R \end{bmatrix} u$

$\vdots$

$u = (Q^{1/2} G)^{\dagger} Q^{1/2} H x$

Is Optimal Enough?

Deploying a PID Controller

Video Credit: Jonathan Hui

Is Optimal Enough?

Deploying a PID Controller

Model Predictive Control

Video Credit: Jonathan Hui

Model Predictive Control

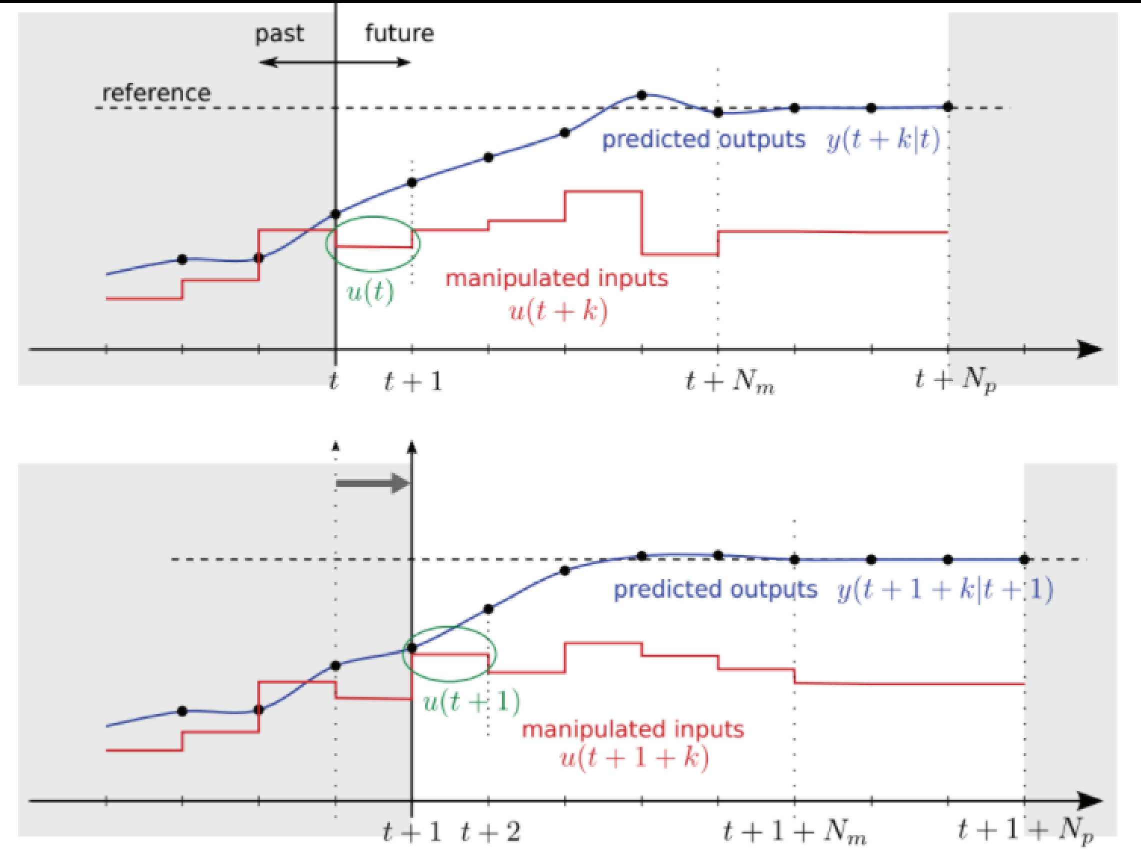
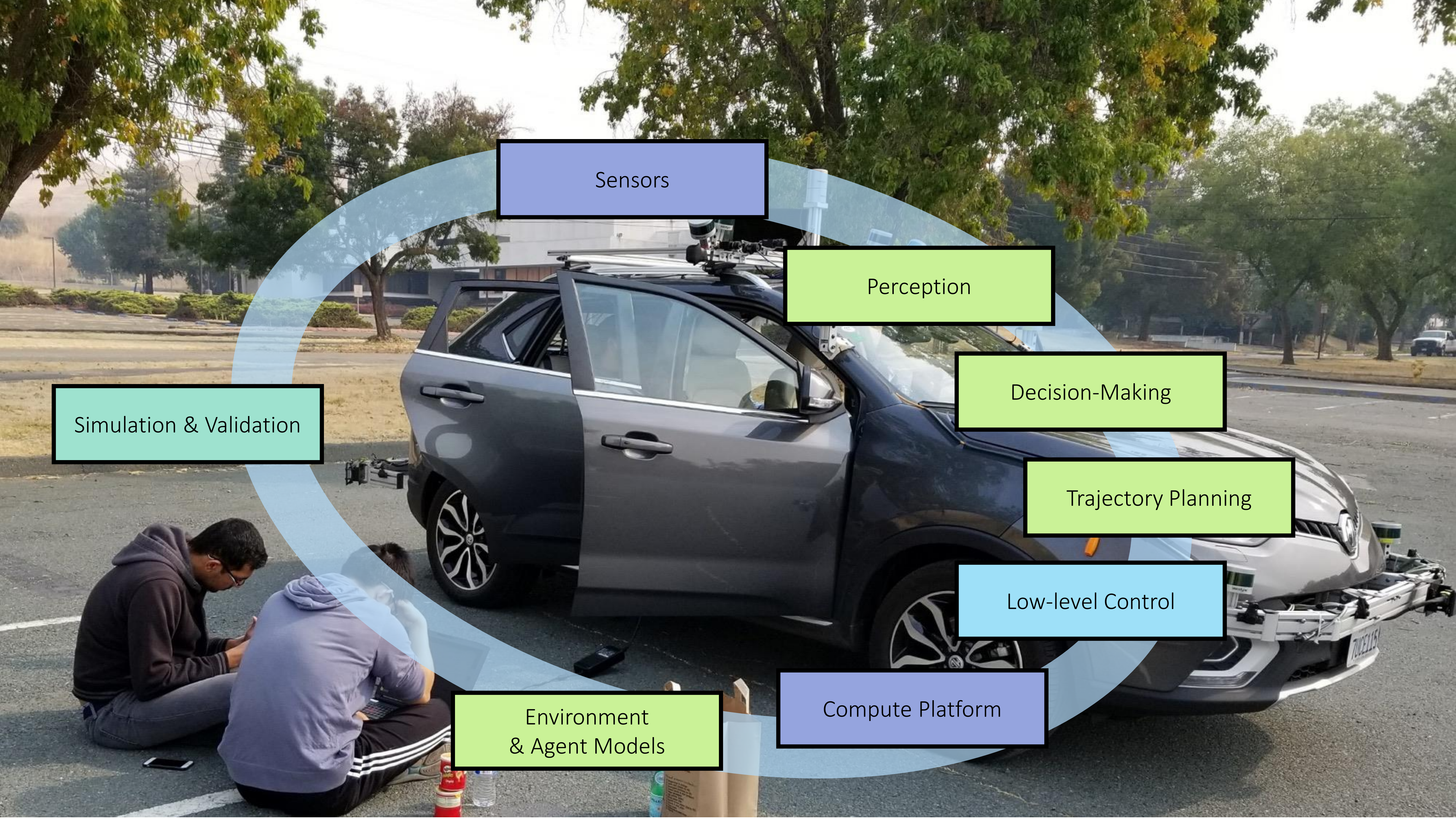


Image Credit: F. Borrelli

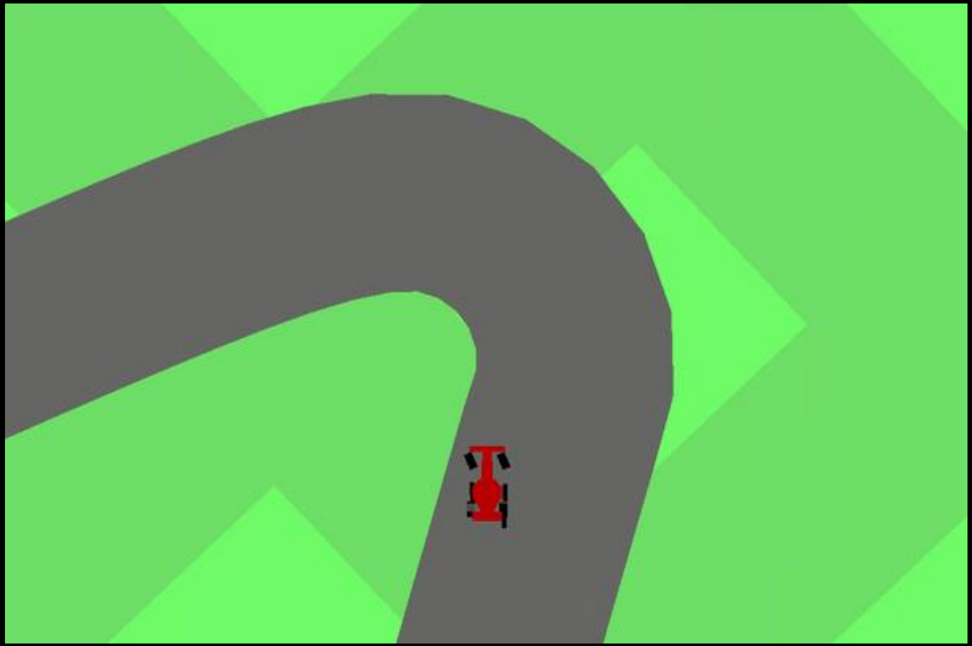

Sensors
Perception
Decision-Making
Trajectory Planning
Low-level Control
Compute Platform
Environment
& Agent Models
Simulation & Validation

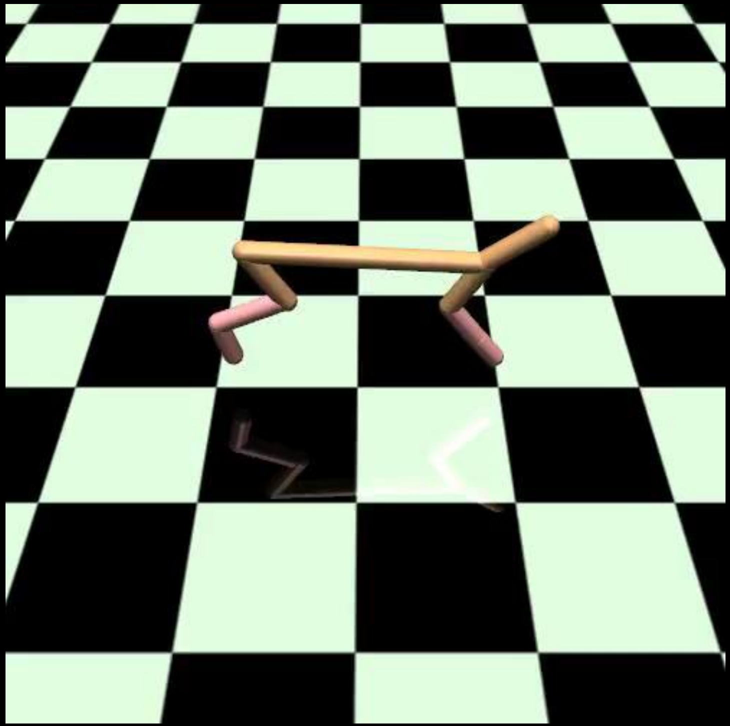
Today's Plan

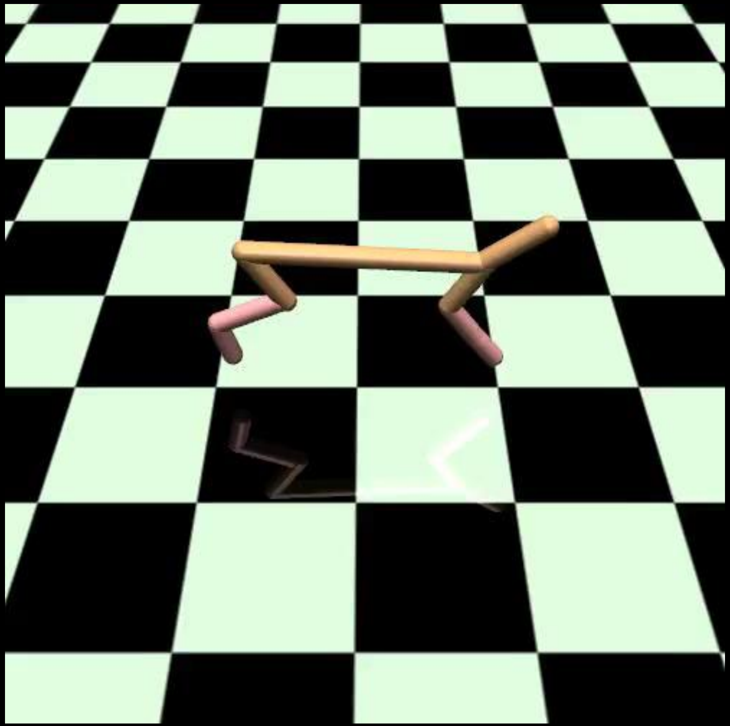
- Quick discussion of future topics in advanced control theory
- Introduction to optimal control
  - Linear Quadratic Regulation (LQR)
  - Model Predictive Control (MPC)
- End-to-end learning

RL Approaches: Hand Specifying Rewards

OpenAI Gym Racecar Environment



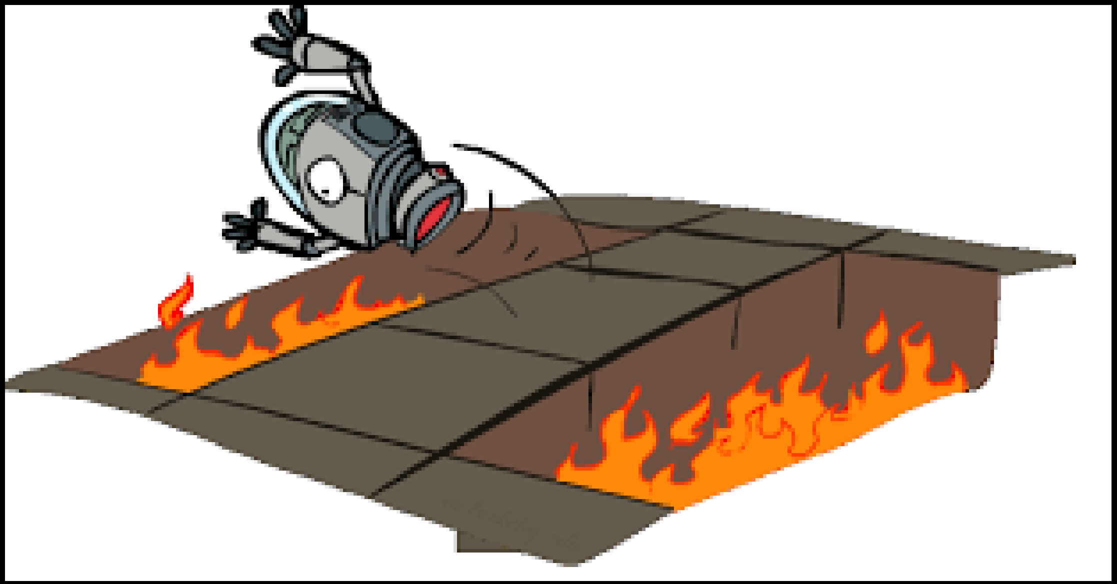




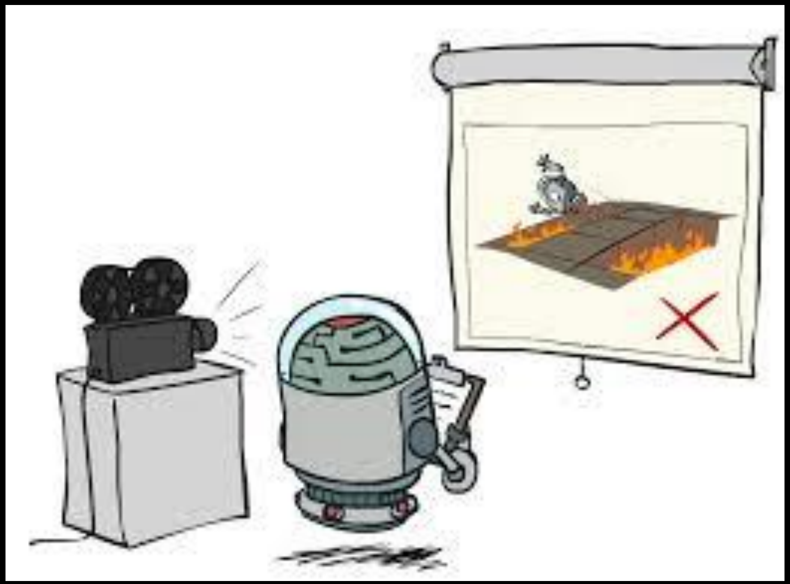
Video Credit: A. Irpan and https://notanymike.github.io/Solving-CarRacing/

Experience vs. Demonstrations

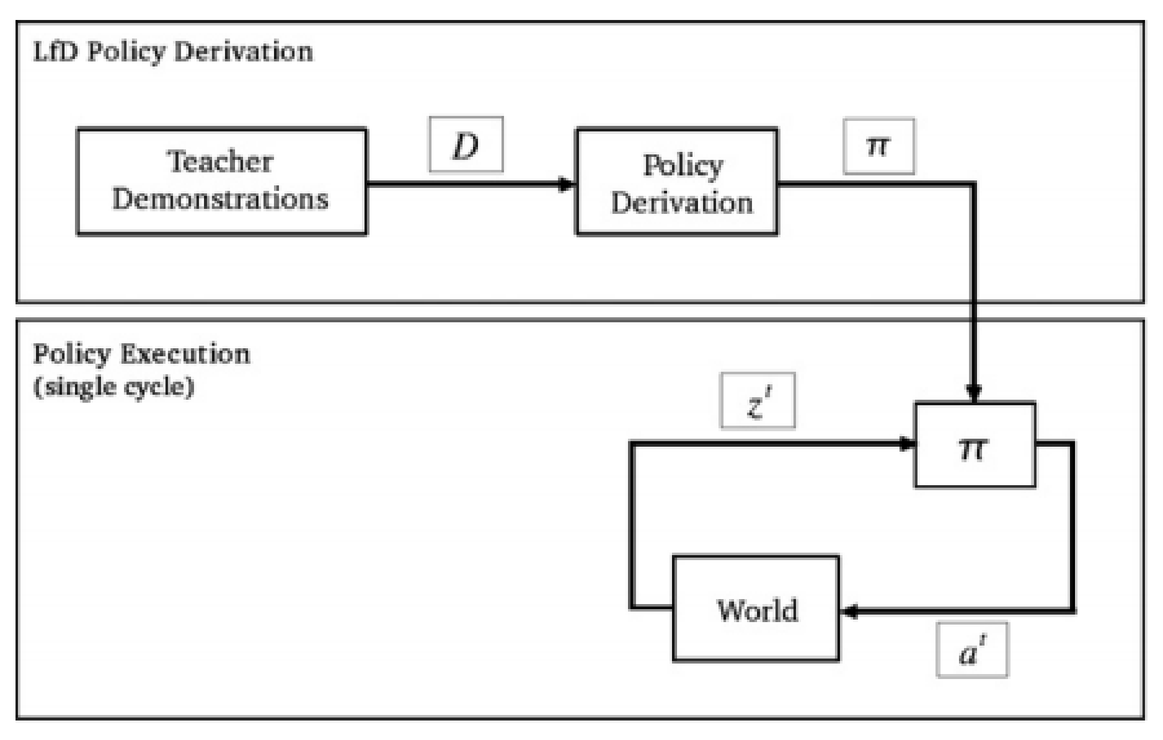
Reinforcement Learning



Demonstrations (sort of)



LfD: Framework and Design Choices

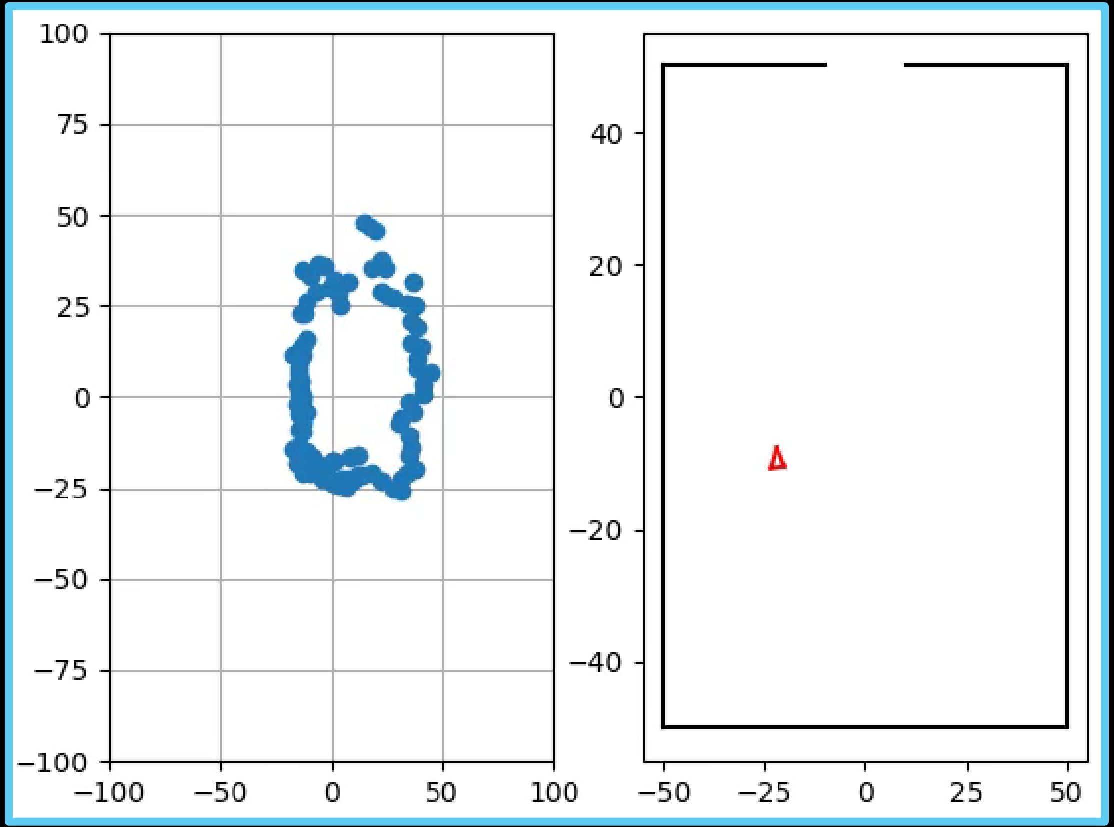


$d_j = \{(z_j^i, a_j^i)\} \in D$
$z_j^i \in Z, a_j^i \in A$
$i = 0, \ldots, k_j$

- Demonstration approach
  - Choice of demonstrator (expert)
  - Demonstration technique (offline, online, iterative)
- Problem space continuity
- Dataset gathering (and limitations)
  - Correspondence (recording, embodiment)
  - Demonstration (teleoperation, shadowing)
- Policy derivation

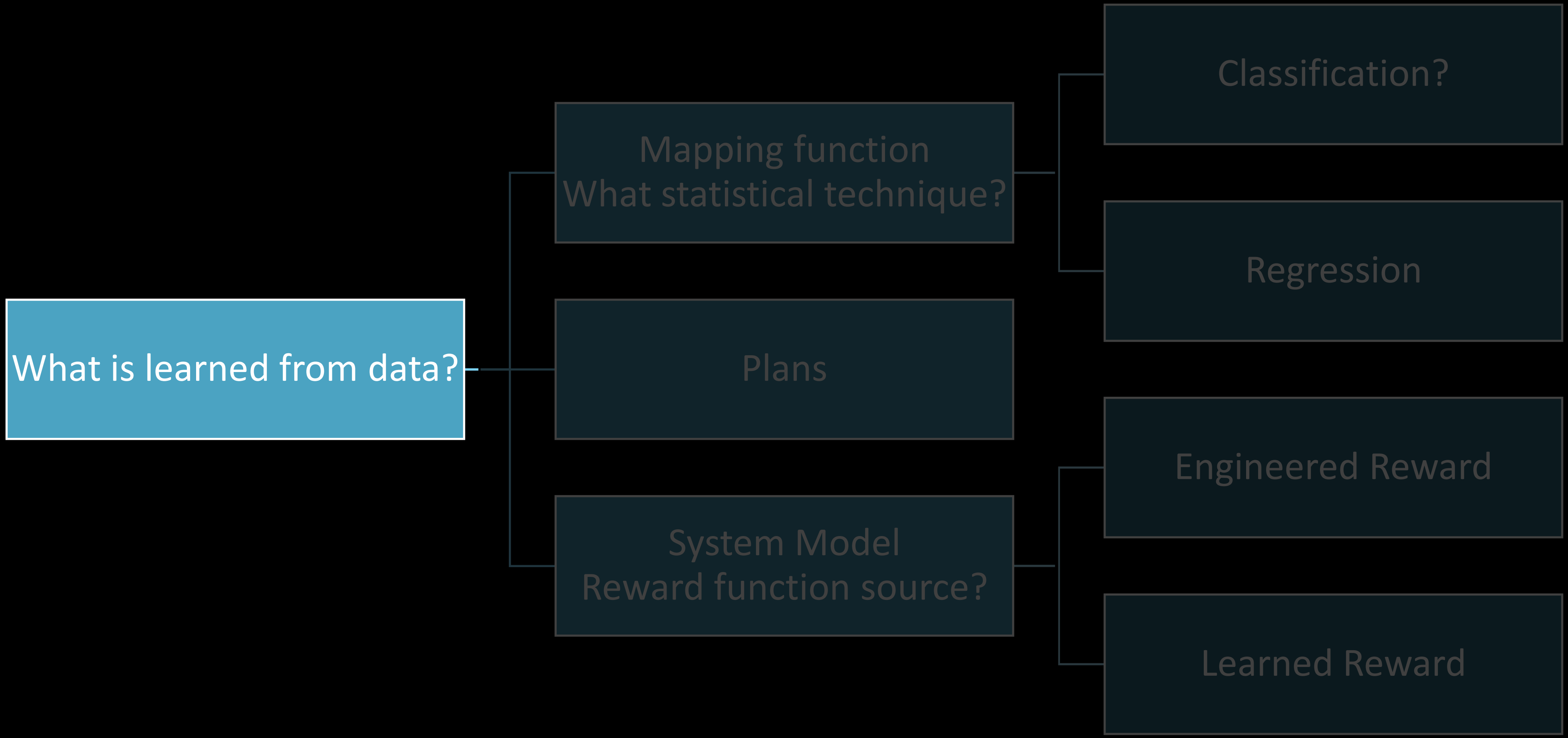
B. Argall, et al., "A survey of robot learning from demonstration," Robotics and Autonomous Systems, 2009.

LfD: Framework and Design Choices

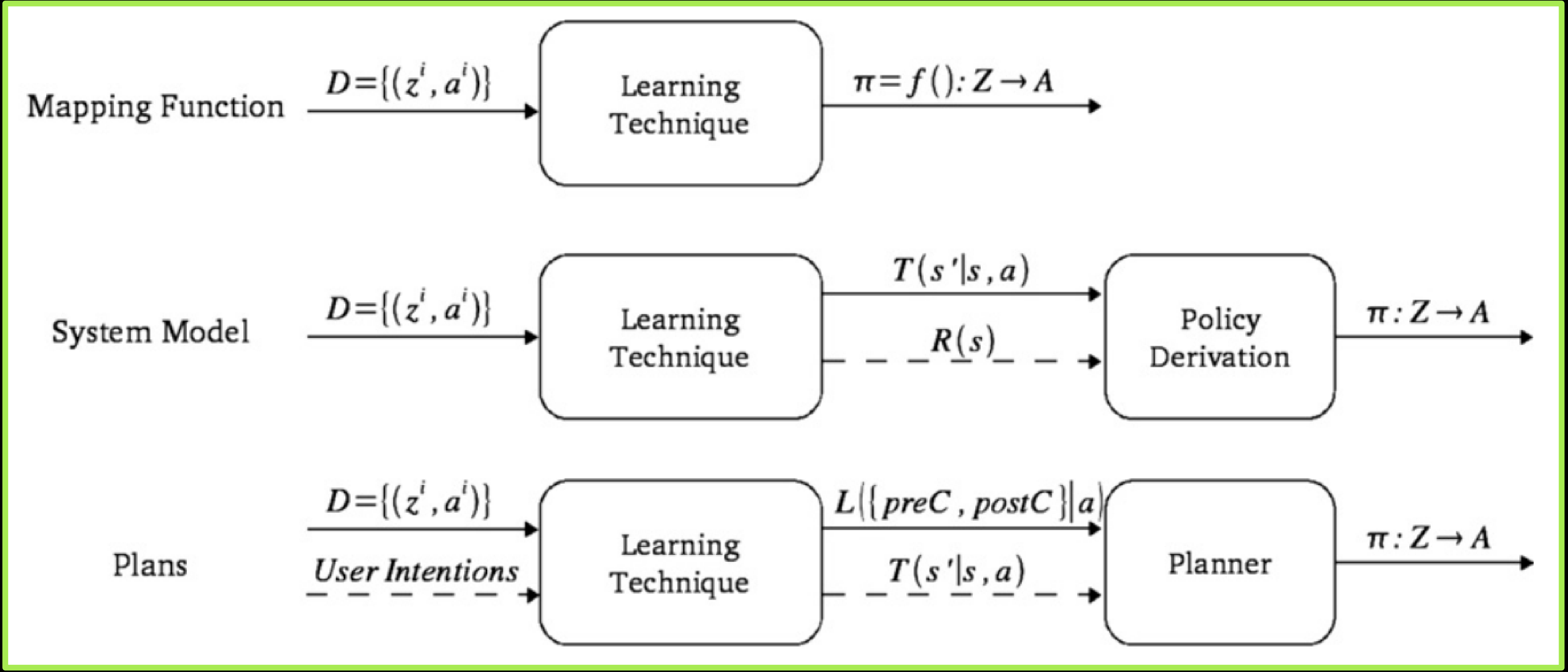


- Demonstration approach
  - Choice of demonstrator (expert)
  - Demonstration technique (offline, online, iterative)
- Problem space continuity
- Dataset gathering (and limitations)
  - Correspondence (recording, embodiment)
  - Demonstration (teleoperation, shadowing)
- Policy derivation

LfD: Deriving a Policy

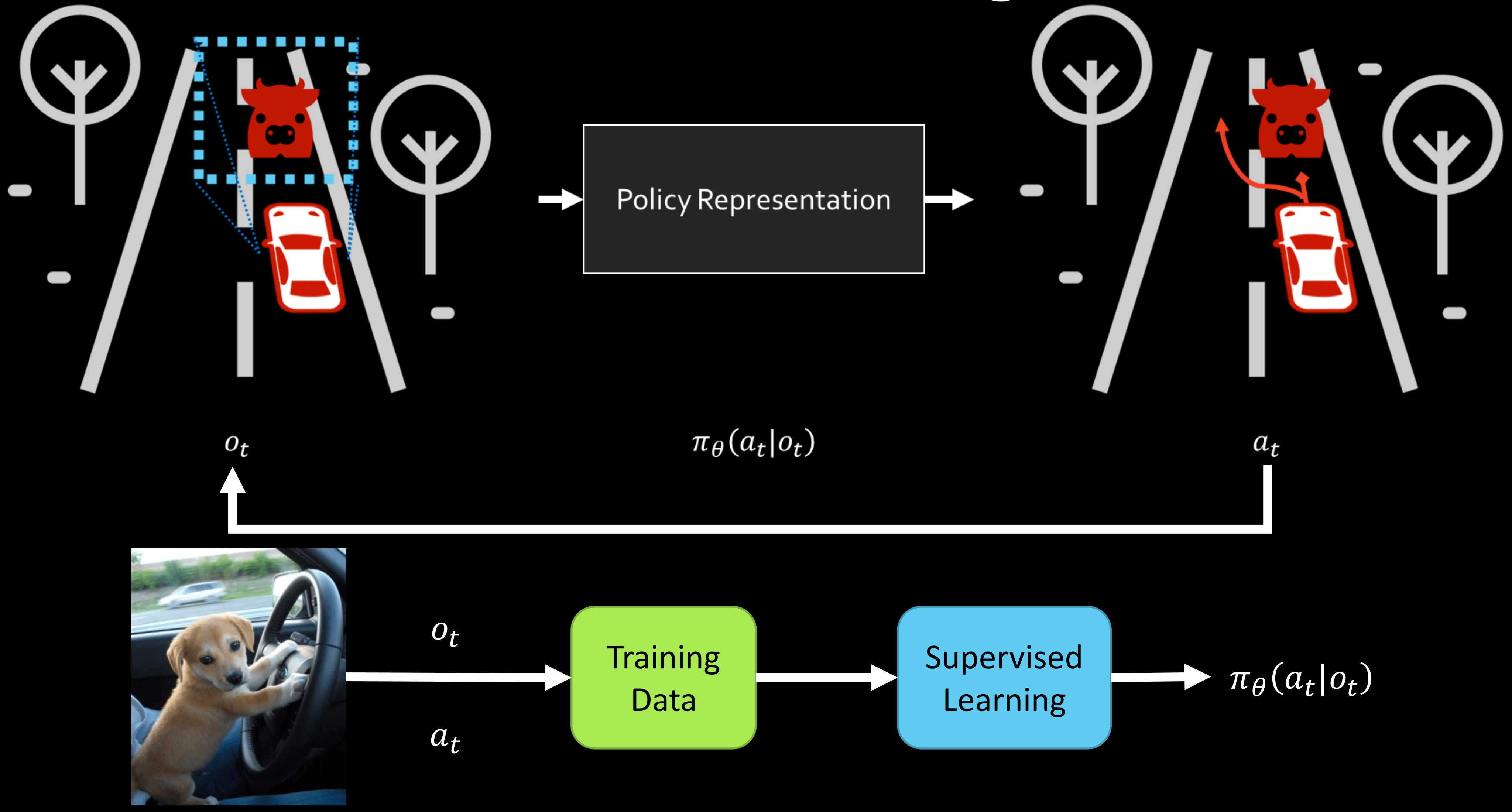


LfD: Deriving a Policy



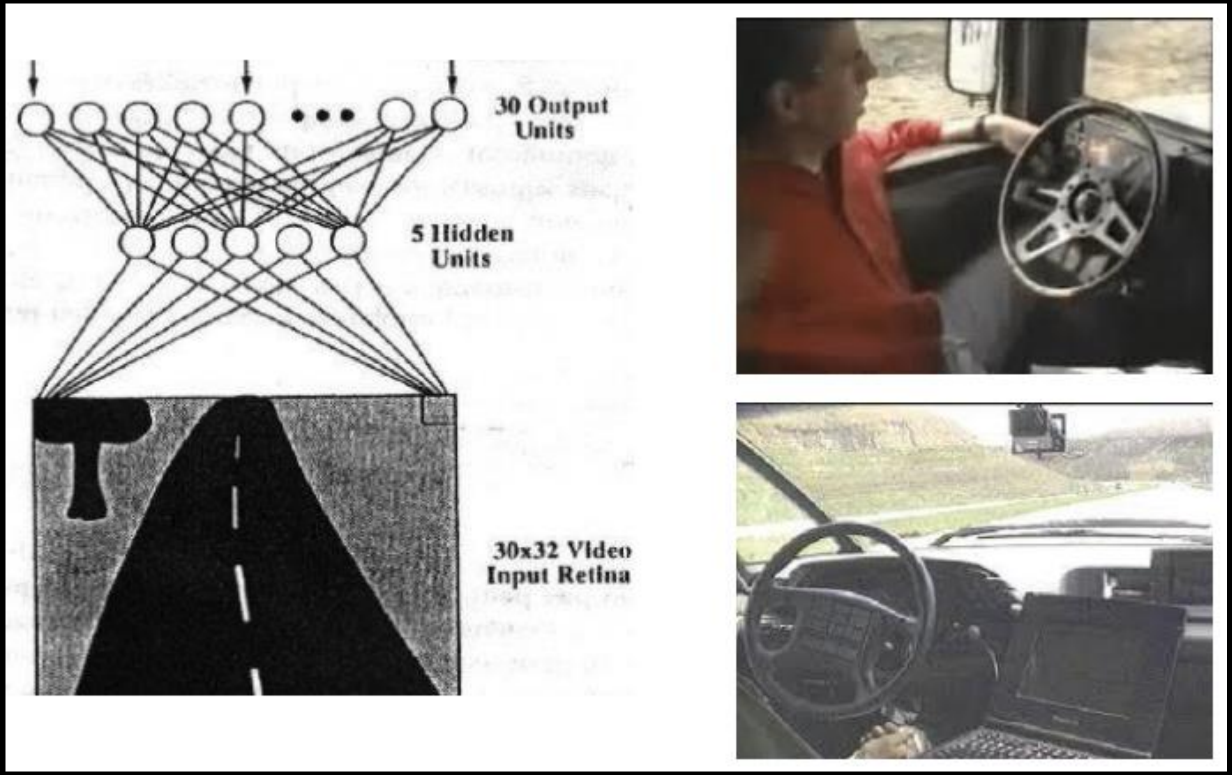
B. Argall, et al., "A survey of robot learning from demonstration," Robotics and Autonomous Systems, 2009.

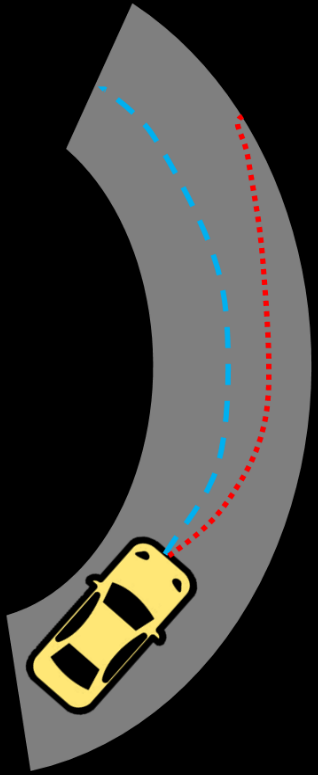
Behavior Cloning



Behavior Cloning

ALVINN: Autonomous Land Vehicle In a Neural Network (1989)

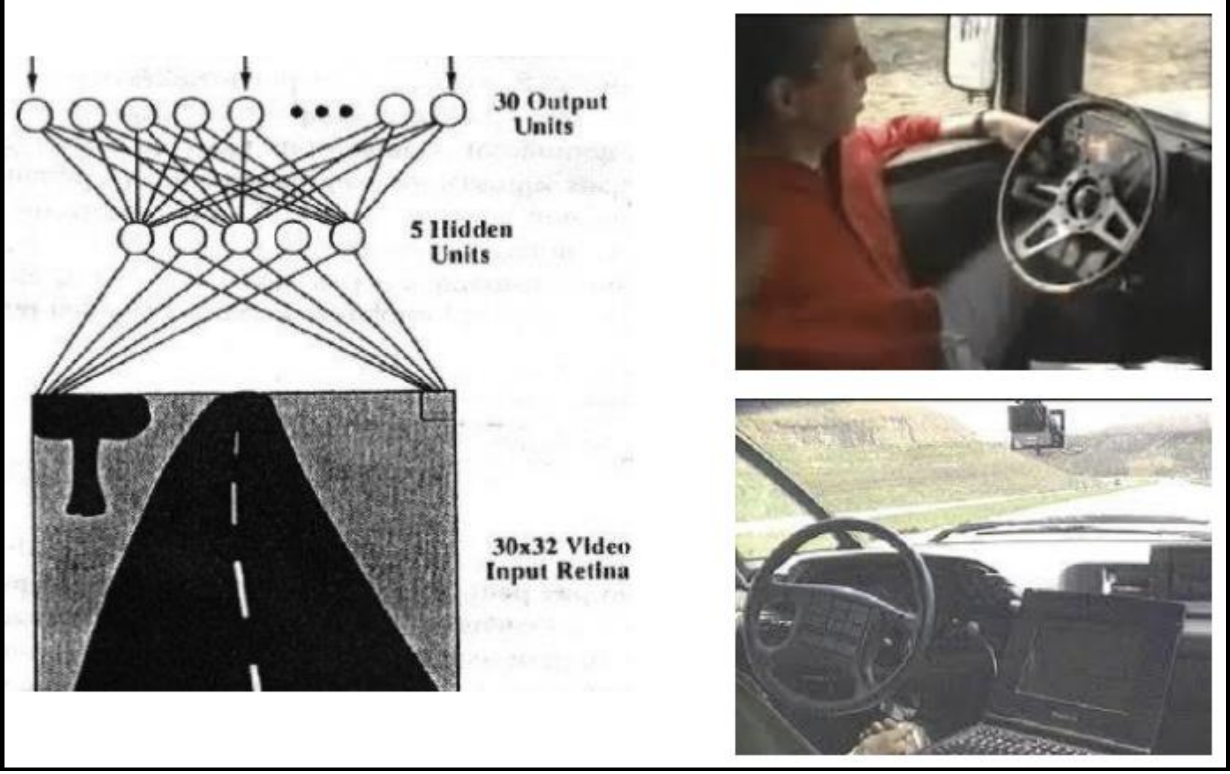




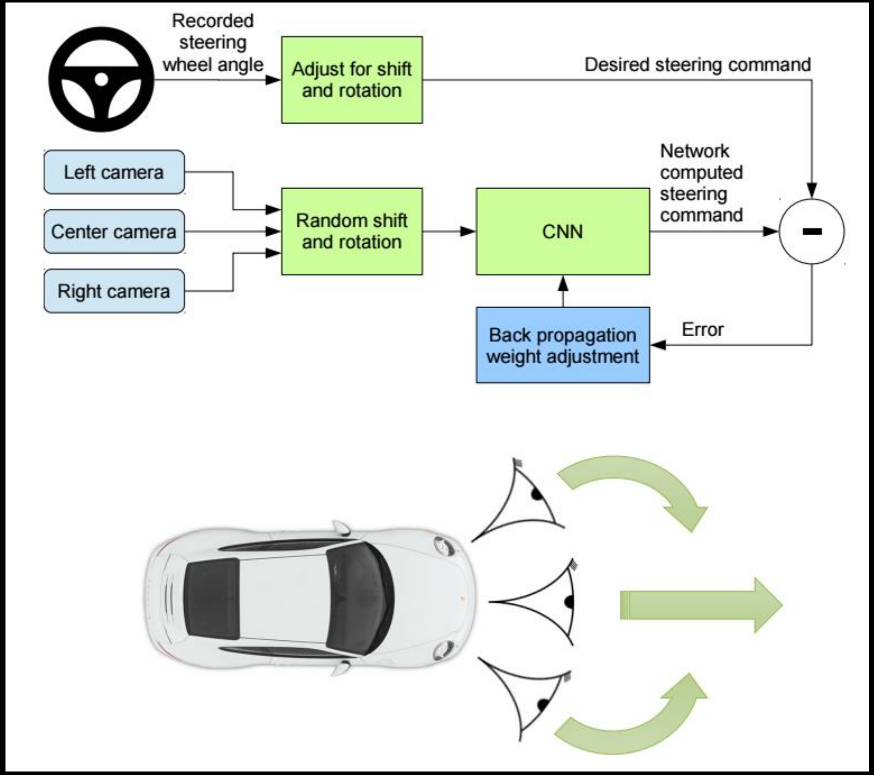
Pomerleau, Dean A. "Alvinn: An autonomous land vehicle in a neural network." *Advances in neural information processing systems*. 1989.

Behavior Cloning

ALVINN: Autonomous Land Vehicle In a Neural Network (1989)



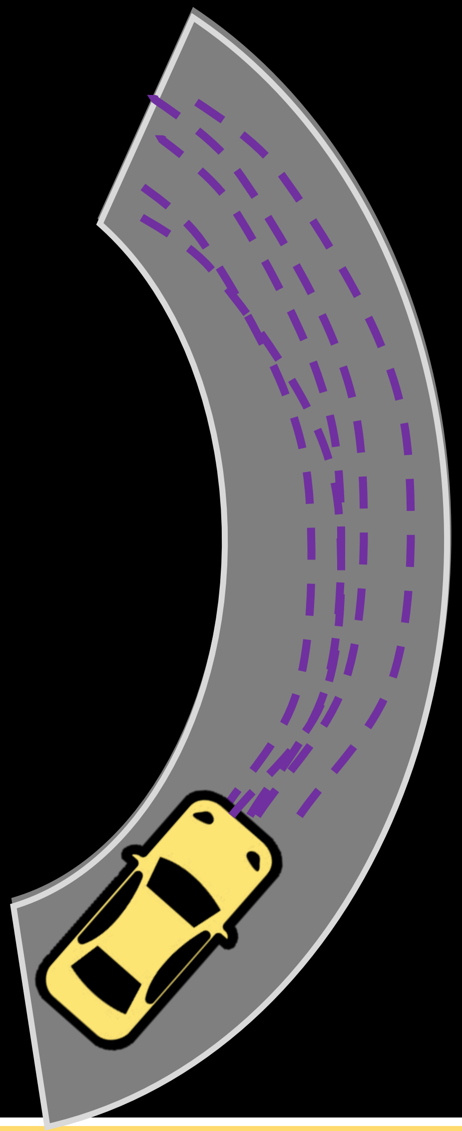
End-to-End Deep Learning for Self-Driving Cars (2016)



Pomerleau, Dean A. "Alvinn: An autonomous land vehicle in a neural network." *Advances in neural information processing systems*. 1989.
M. Bojarski, et al. "End to end learning for self-driving cars." arXiv:1604.07316 (2016).

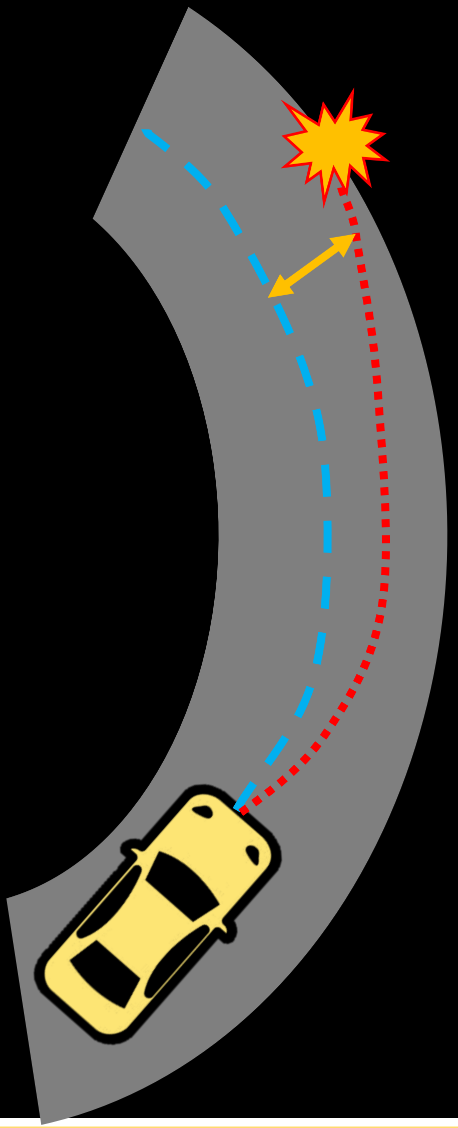
(Deep) Imitation Learning

- Given sample trajectories from an expert, try to learn the underlying policy



(Deep) Imitation Learning

- Given sample trajectories from an expert, try to learn the underlying policy
- Tends to suffer from distribution shift, compounding errors, model mismatch



NN Policy Baseline

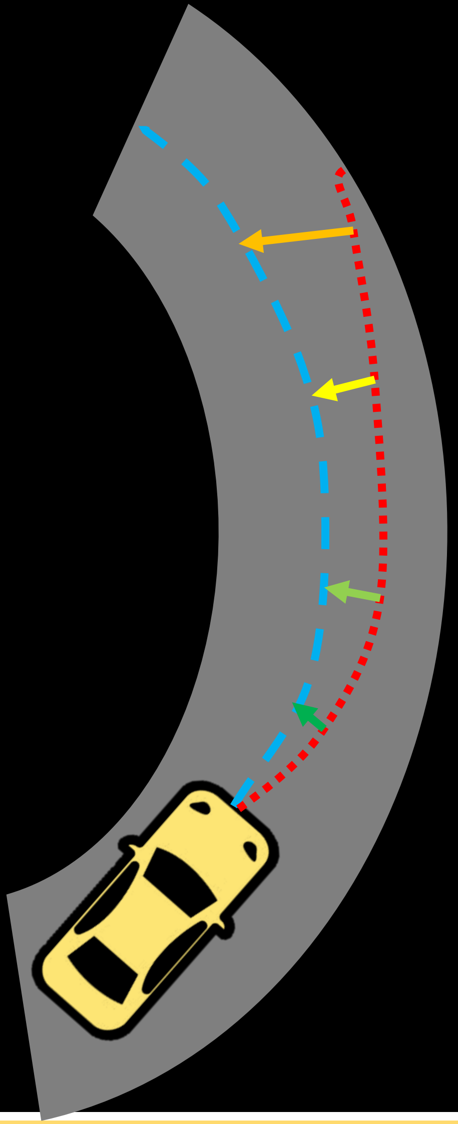
HG-DAgger:

Interactive Imitation Learning with Human Experts

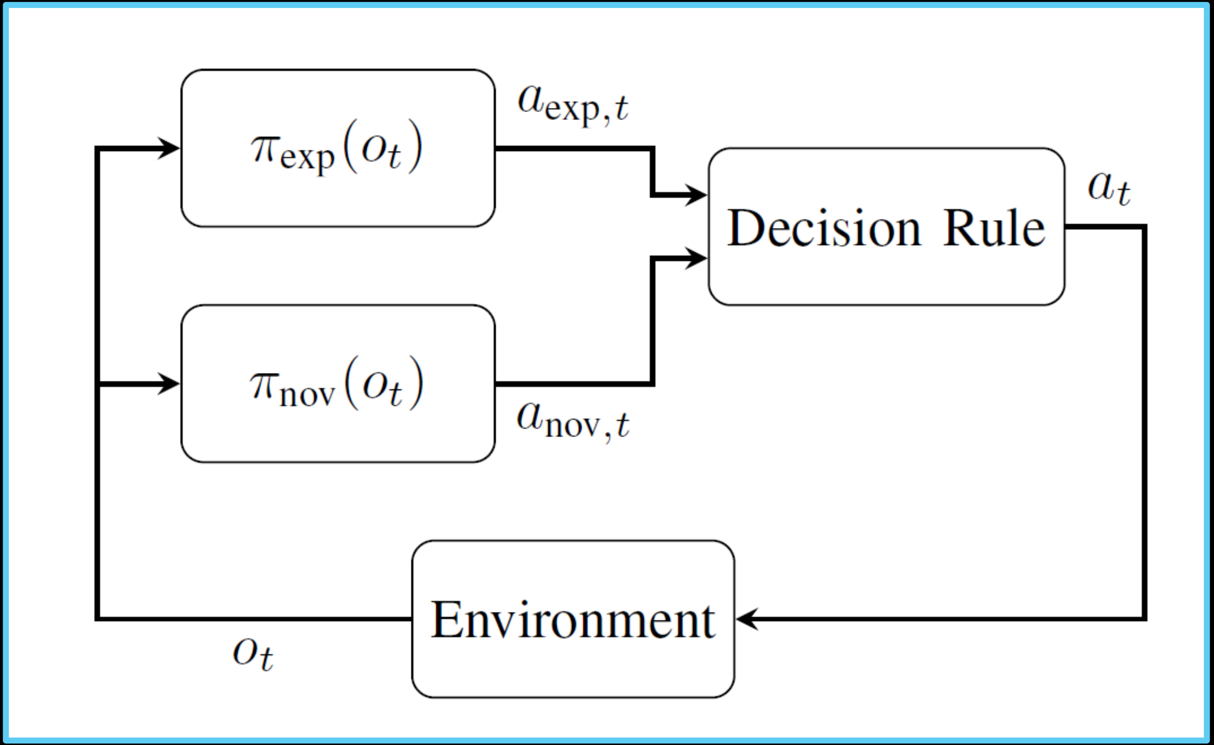
M. Kelly, C. Sidrane, K. Driggs-Campbell, M. Kochenderfer

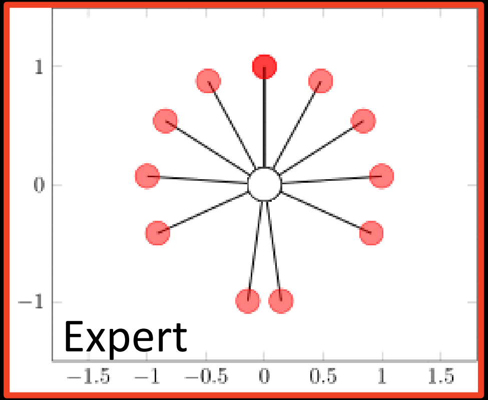
(Deep) Imitation Learning

- Given sample trajectories from an expert, try to learn the underlying policy
- Tends to suffer from distribution shift, compounding errors, model mismatch
- By improving how we collect the data, we can improve the resulting policy!

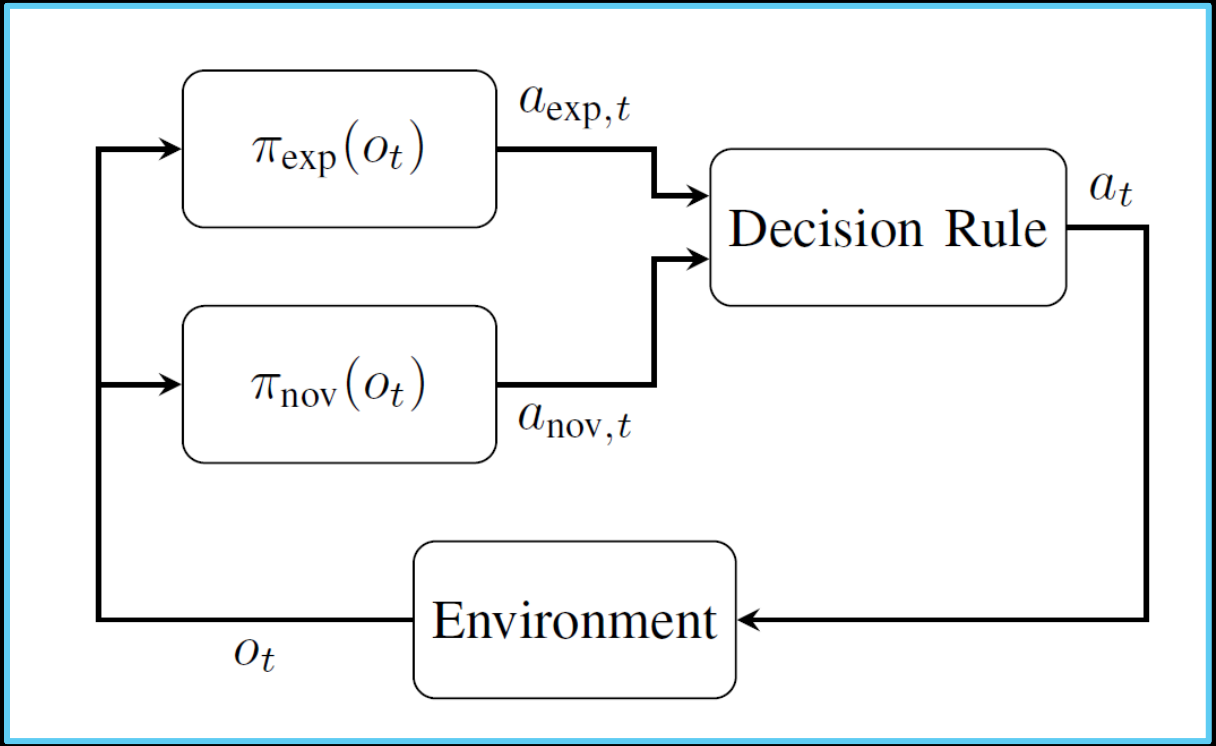


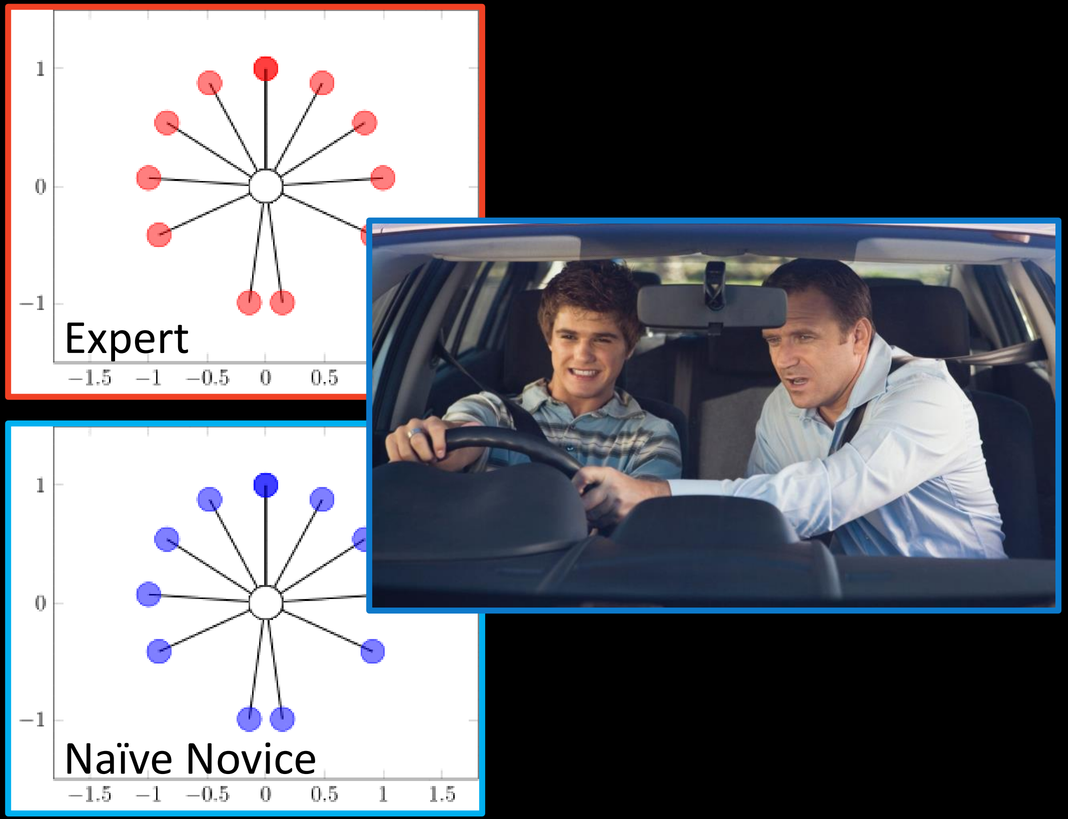
"Safe" Imitation Learning



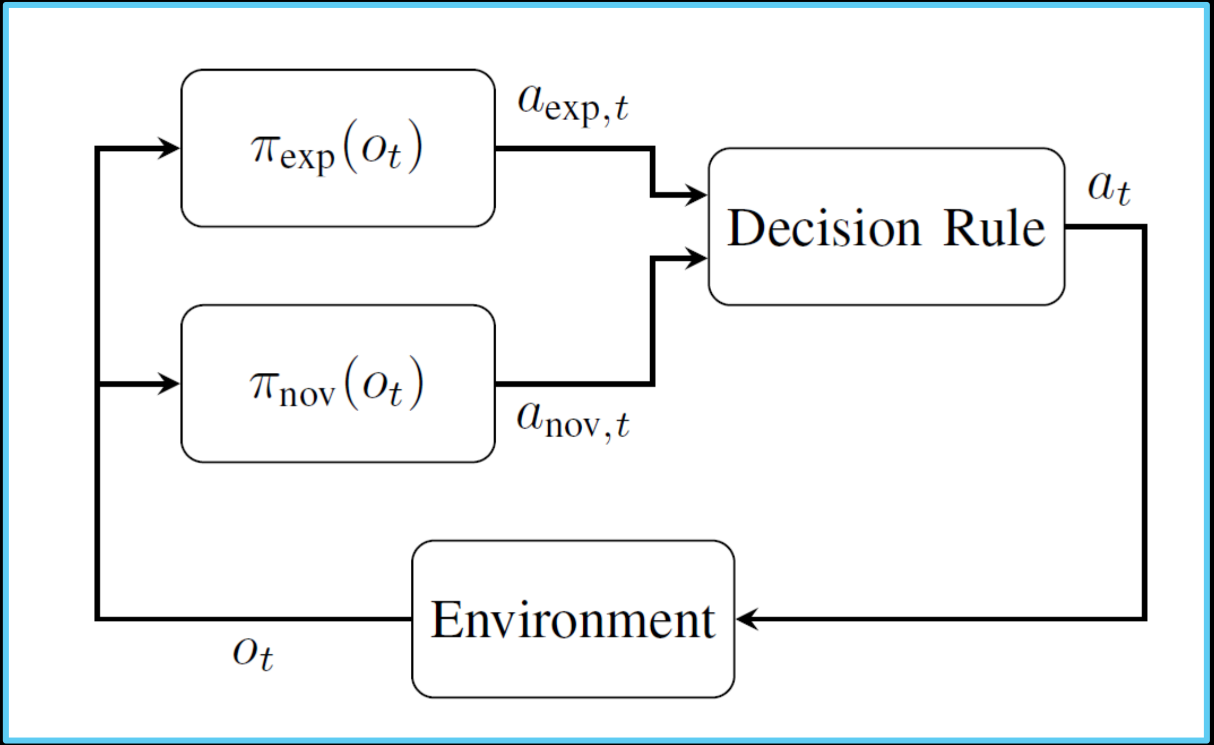


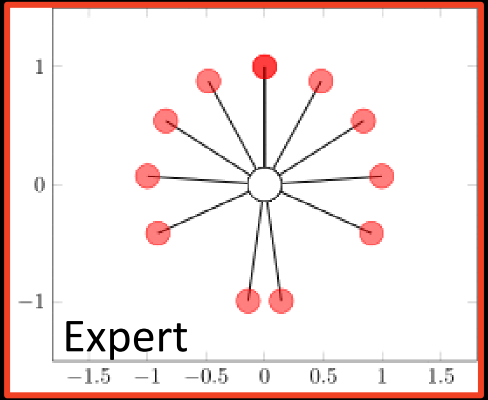
"Safe" Imitation Learning

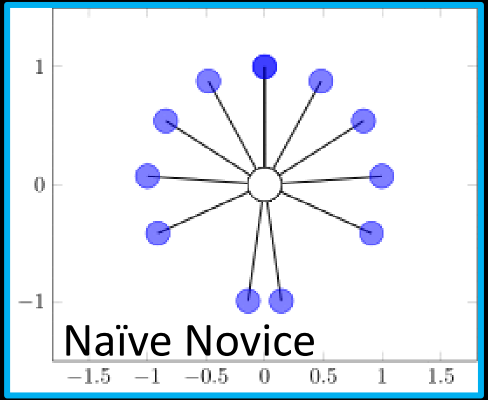




“Safe” Imitation Learning

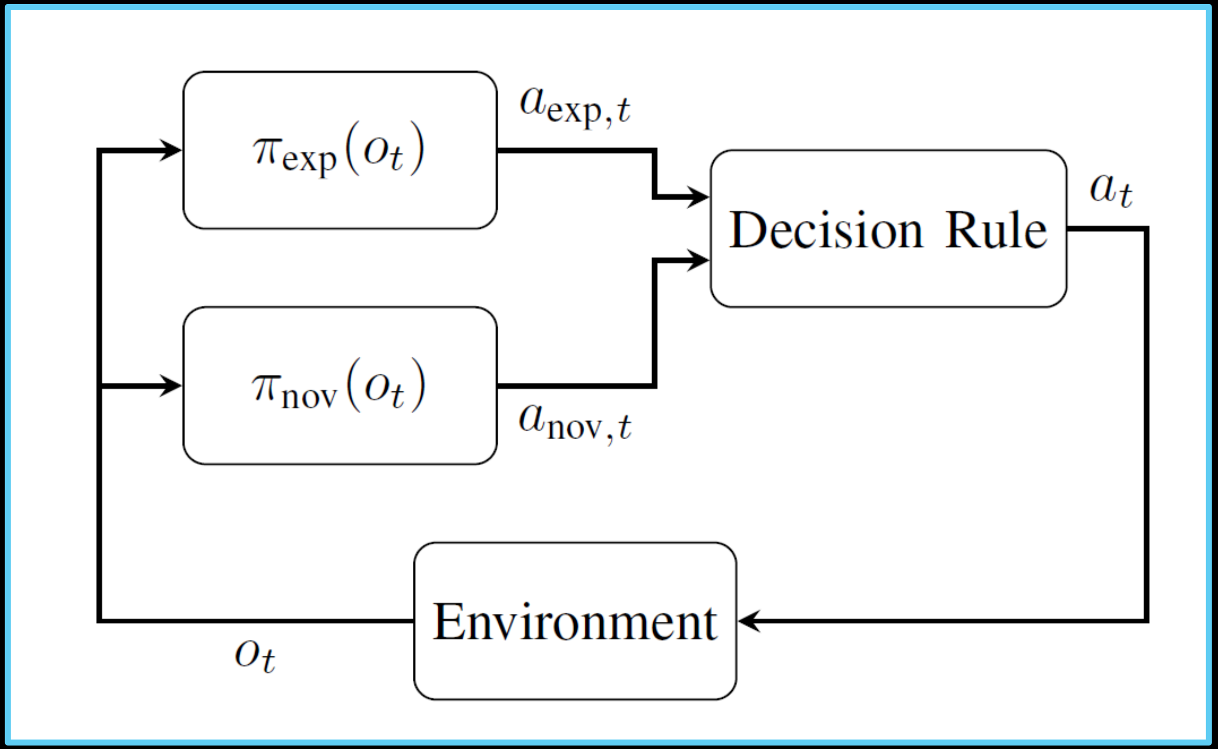


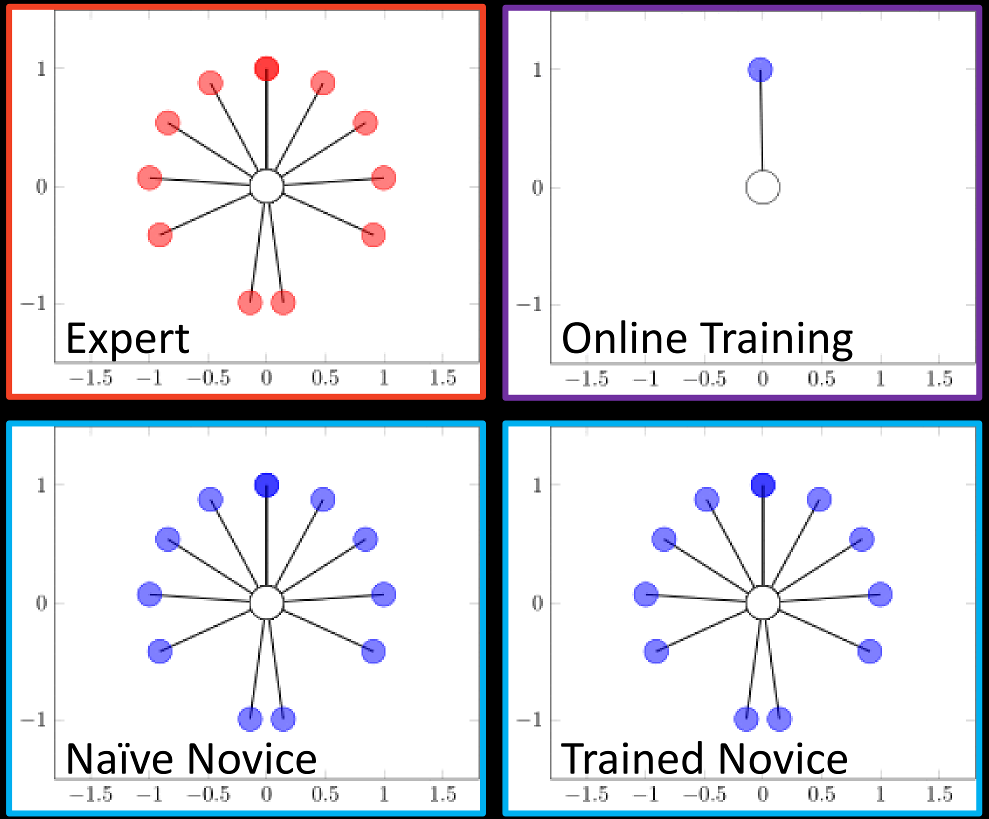




Visualization Credit: K. Menda

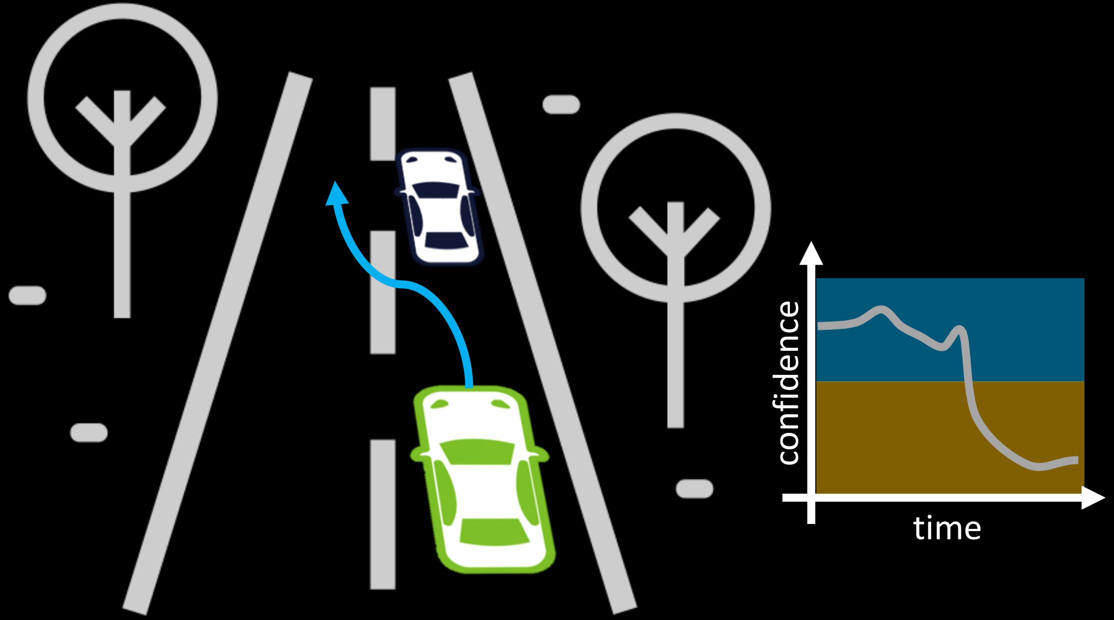
“Safe” Imitation Learning



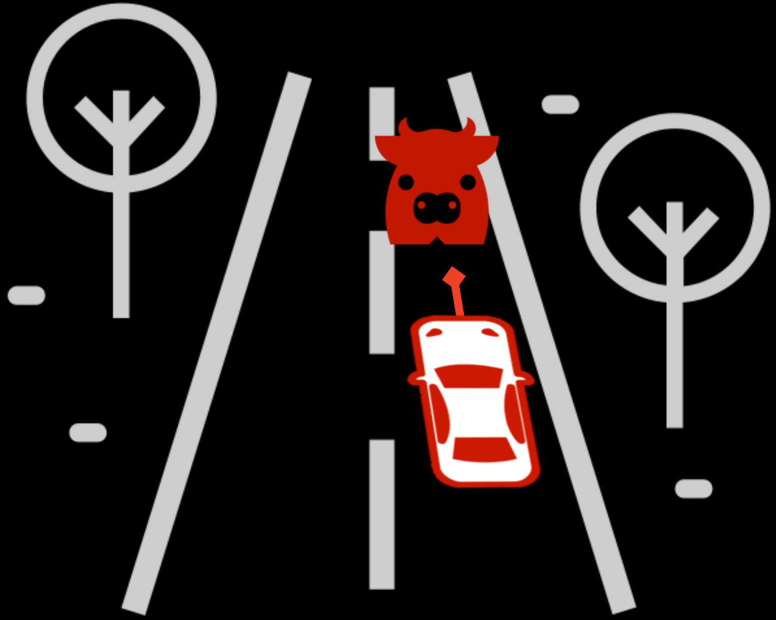


Self-Driving Demonstration

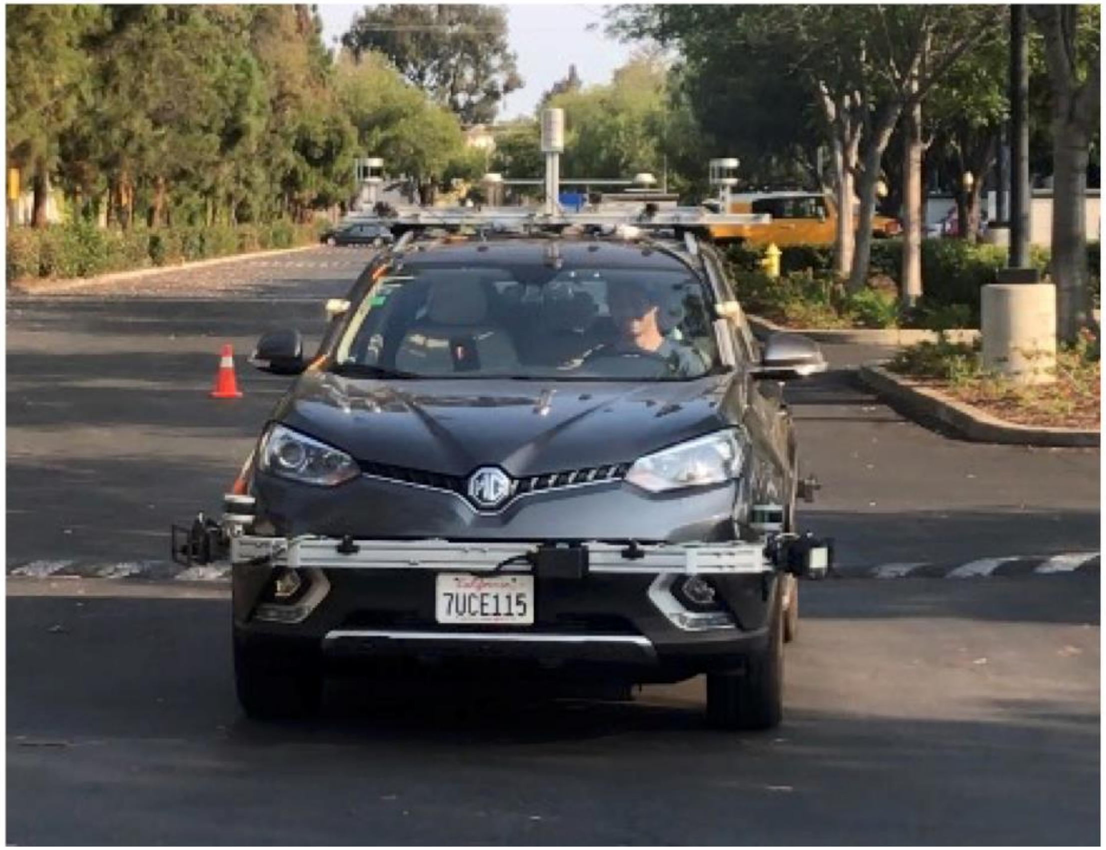
High Confidence in NN Policy



Unseen scenario → Resume control



Human-Gated Imitation Learning



M. Kelly, C. Sidrane, K. Driggs-Campbell, and M.J. Kochenderfer. HG-DAgger: Interactive Imitation Learning with Human Experts, ICRA 2019.

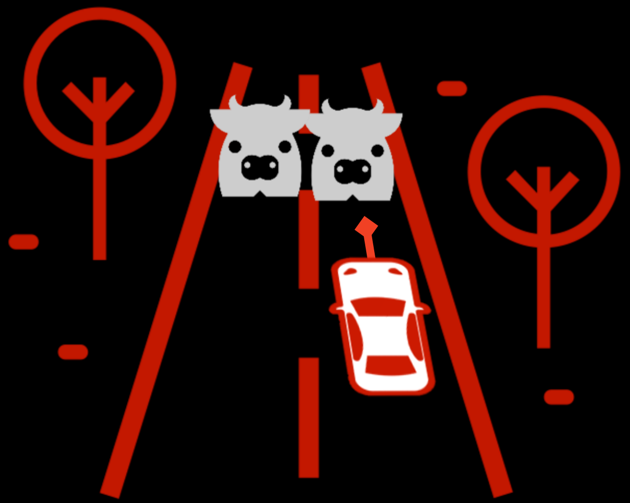
Demo Test Scenarios

Deep NN Policy:

Weaving Lane Changes

No Feasible Lane Change:

Emergency Stop Required



No Obstacles Detected:

Lane Keeping Required

HG-DAgger:

Interactive Imitation Learning with Human Experts

M. Kelly, C. Sidrane, K. Driggs-Campbell, M. Kochenderfer

HG-DAgger:
Interactive Imitation Learning with Human Experts

M. Kelly, C. Sidrane, K. Driggs-Campbell, M. Kochenderfer

HG-DAgger:
Interactive Imitation Learning with Human Experts

M. Kelly, C. Sidrane, K. Driggs-Campbell, M. Kochenderfer

Lane Keeping Required

HG-DAgger:
Interactive Imitation Learning with Human Experts

M. Kelly, C. Sidrane, K. Driggs-Campbell, M. Kochenderfer

No Feasible Lane Change

Summary

- Introduced a few advanced topics on model-based control
- Discussed learning and end-to-end (model-free) approaches
- Note that all of the methods discussed require some low-level controller (i.e., PID) and some high-level input (i.e., decision-making)
- Did not discuss the safety implications of different control methods! What do you think are the hazards and advantages of different approaches?
- *Next time:* Filtering and localization!

Extra Slides

Inverse Reinforcement Learning

- Given an optimal trajectory, we want to find the cost function:

$$\xi_D \rightarrow \mathcal{U}: \Xi \rightarrow \mathbb{R}_+ \; s.t. \, \mathcal{U}[\xi_D] \leq \mathcal{U}[\xi], \forall \xi$$

- Rewrite as: $\mathcal{U}[\xi_D] \leq \min_{\xi} \mathcal{U}[\xi]$ → Suffers from trivial solutions!
- Modify to find cost function that gives minimum cost by a margin:

$$\mathcal{U}[\xi_D] \leq \min_{\xi} \mathcal{U}[\xi] - l(\xi, \xi_D), \text{ where } l(\xi, \xi_D) = \begin{cases} 0 \text{ if } \xi = \xi_D \\ 1 \text{ otherwise} \end{cases}$$

- To make this hold true for the *maximum margin*:

$$\max_{\mathcal{U}} \min_{\xi} \mathcal{U}[\xi] - l(\xi, \xi_D) - \mathcal{U}[\xi_D]$$

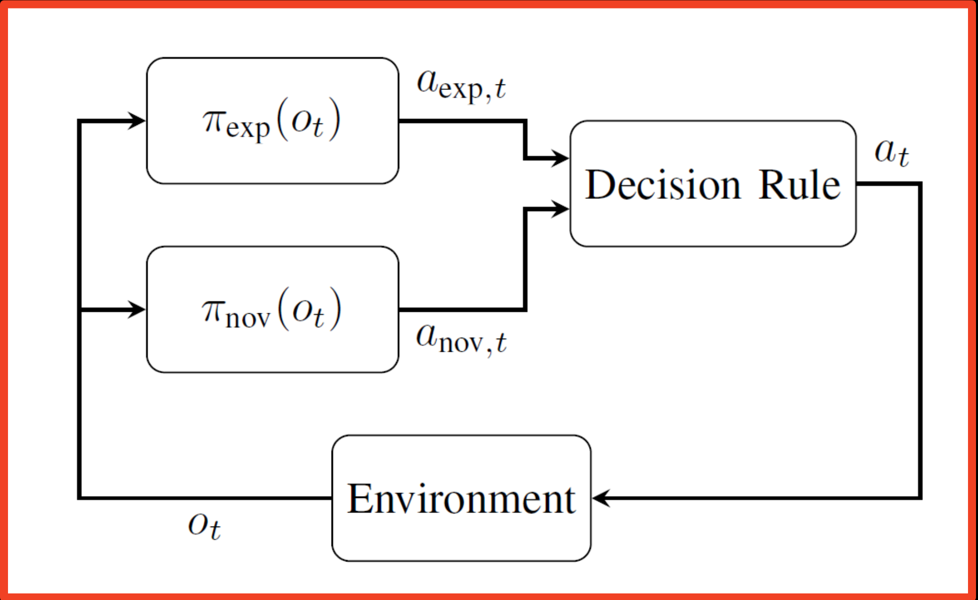
$$\min_{\mathcal{U}} \left[\mathcal{U}[\xi_D] - \min_{\xi} [\mathcal{U}[\xi] - l(\xi, \xi_D)] + \lambda R(\mathcal{U}) \right]$$

- To solve this problem, parameterize the function $\mathcal{U}$ → often a linear combination of features

P. Abbeel and A. Ng. "Apprenticeship learning via inverse reinforcement learning." International Conference on Machine learning. 2004.

DAgger:
Dataset Aggregation

1. Train π_{nov} from human data $\mathcal{D}$
2. Run π_{nov} to get dataset $\mathcal{D}_{\pi_{nov}}$
3. Obtain corrected labels
4. Aggregate: $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}_{\pi_{nov}}$
5. Repeat!

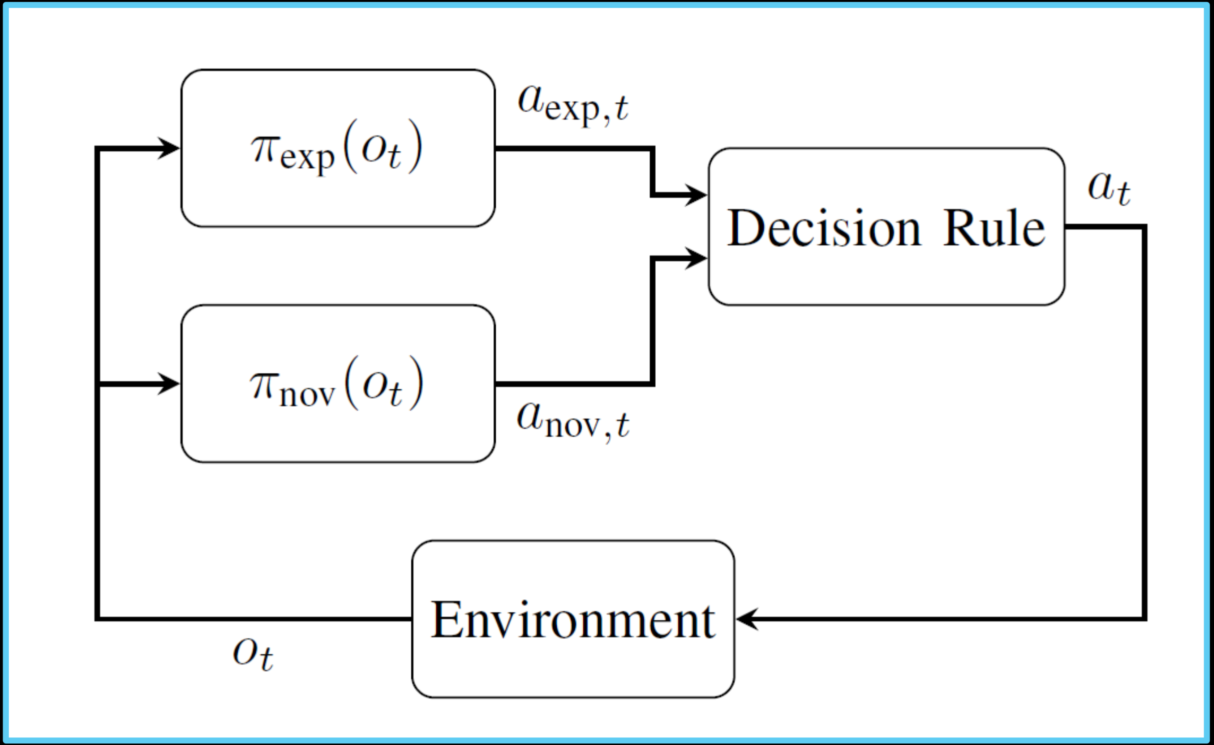


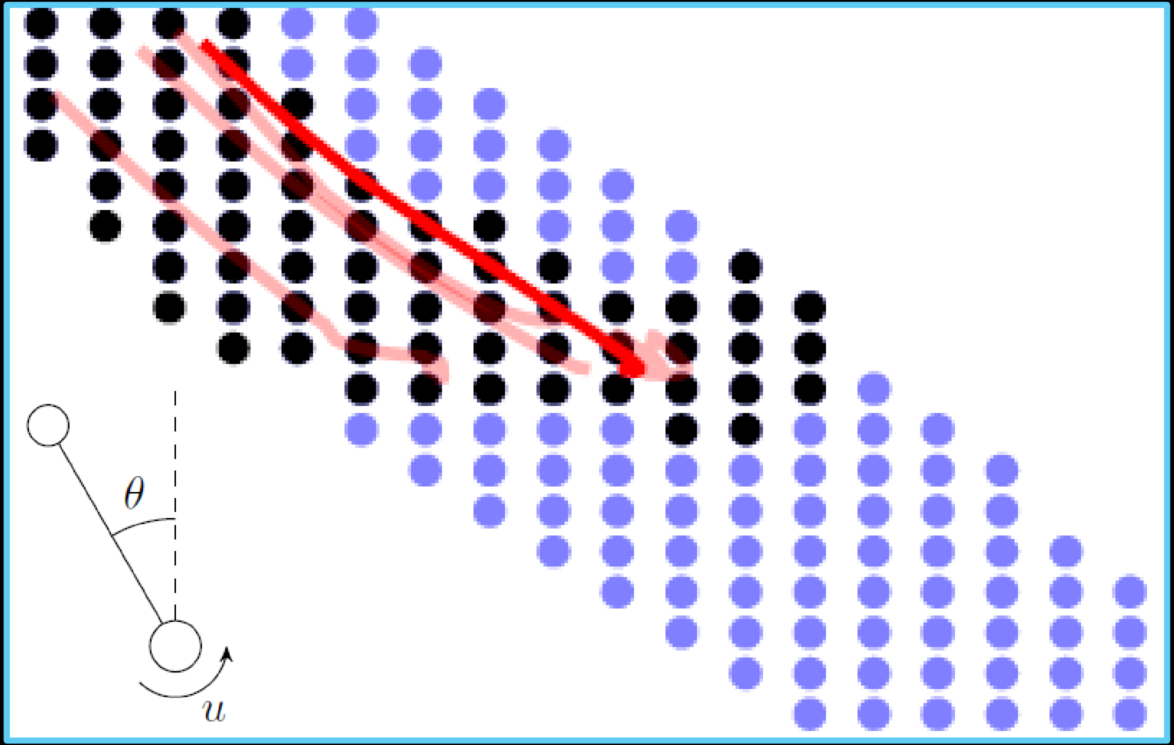
Algorithm 2 VanillaDAgger Decision Rule

procedure DR$(o_t, i, \beta_0, \lambda)$
  $a_{\text{nov},t} \leftarrow \pi_{\text{nov}}(o_t)$
  $a_{\text{exp},t} \leftarrow \pi_{\text{exp}}(o_t)$
  $\beta_i \leftarrow \lambda^i \beta_0$
  $z \sim \text{Uniform}(0, 1)$
  if $z \leq \beta_i$
    return $a_{\text{exp},t}$
  else
    return $a_{\text{nov},t}$

S. Ross, G. Gordon, and A. Bagnell. "A reduction of imitation learning and structured prediction to no-regret online learning." International Conference on Artificial Intelligence and Statistics. 2011.

Safe Imitation Learning





Visualization Credit: K. Menda

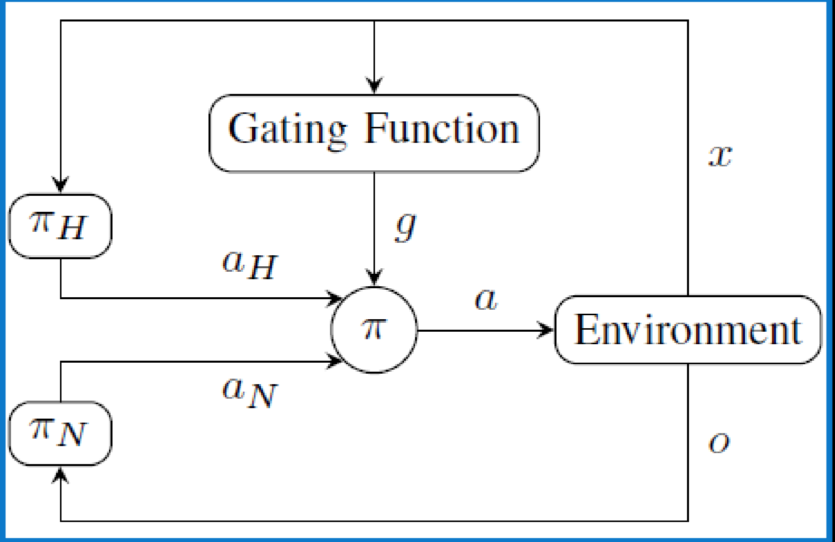
Methods for Determining the Decision Rule?

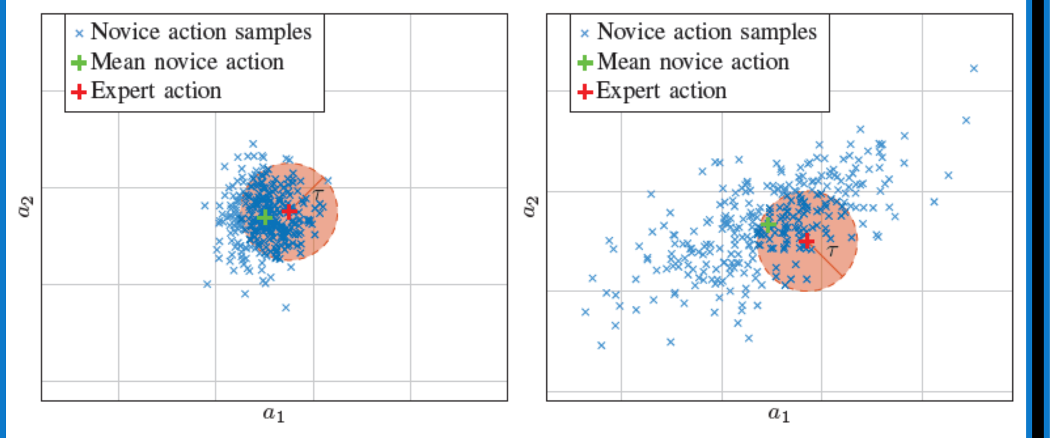
Algorithm 3 SafeDAgger* Decision Rule

1: **procedure** DR(o_t, τ)
2: $a_{\text{nov},t} \leftarrow \pi_{\text{nov}}(o_t)$
3: $a_{\exp,t} \leftarrow \pi_{\exp}(o_t)$
4: **if** $\|a_{\text{nov},t} - a_{\exp,t}\|^2 \leq \tau$
5: **return** $a_{\text{nov},t}$
6: **else**
7: **return** $a_{\exp,t}$

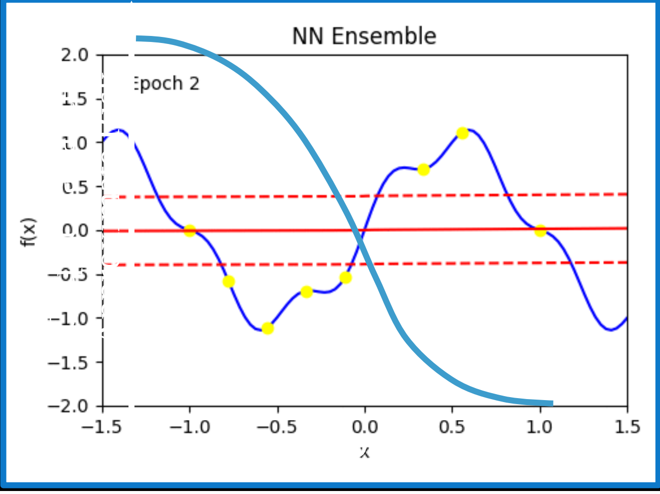
Algorithm 4 EnsembleDAgger Decision Rule

1: **procedure** DR(o_t, τ, χ)
2: $\bar{a}_{\text{nov},t}, \sigma^2_{a_{\text{nov},t}} \leftarrow \pi_{\text{nov}}(o_t)$
3: $a_{\exp,t} \leftarrow \pi_{\exp}(o_t)$
4: $\hat{\tau} \leftarrow \|\bar{a}_{\text{nov},t} - a_{\exp,t}\|^2$
5: $\hat{\chi} \leftarrow \sigma^2_{a_{\text{nov},t}}$
6: **if** $\hat{\tau} \leq \tau$ **and** $\hat{\chi} \leq \chi$
7: **return** $\bar{a}_{\text{nov},t}$
8: **else**
9: **return** $a_{\exp,t}$





(a) Well-represented state (b) Poorly-represented state



...hang, a... K. Cho. "Query-efficient imitation learning for end-to-end autonomous driving." arXiv:1605.06450, 2016.
Menda, ...