Introduction to Reinforcement Learning

AI in Industry

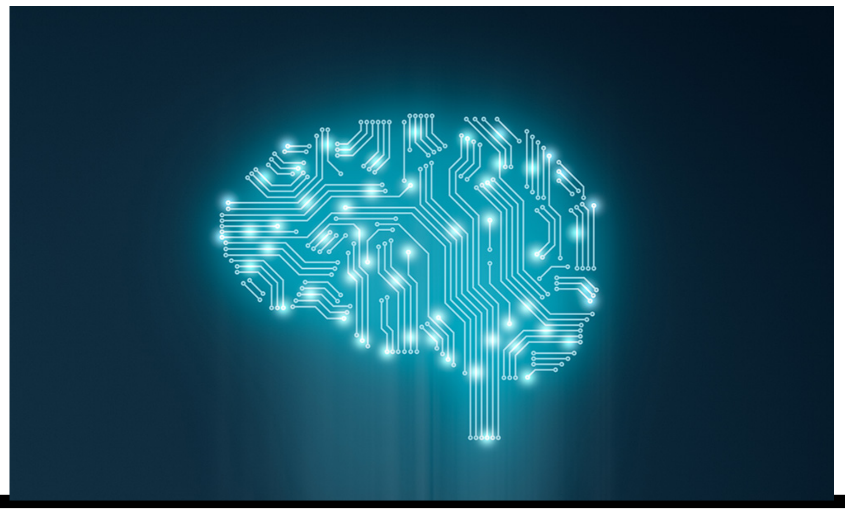
09.12.2024

Outline

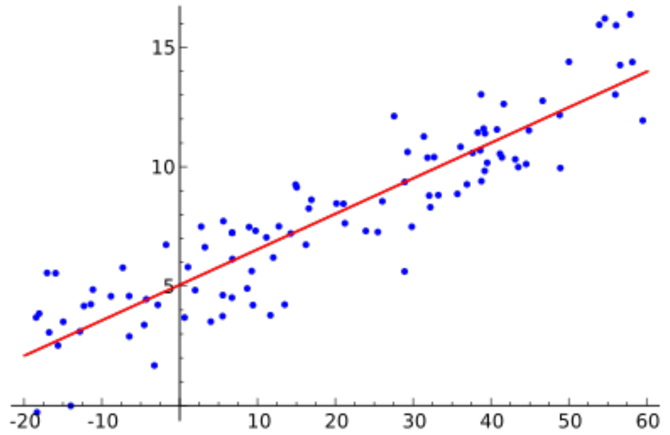
- Hour 1
  - AI Ethics feedback and assignment
  - Introduction to discrete-time dynamical system

- Hour 2: AI in Industry

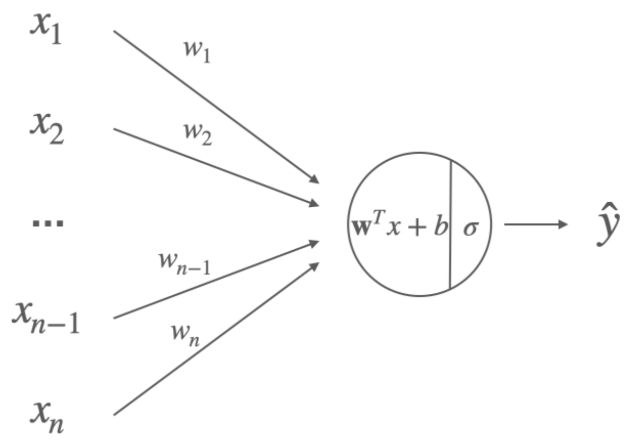
Introduction



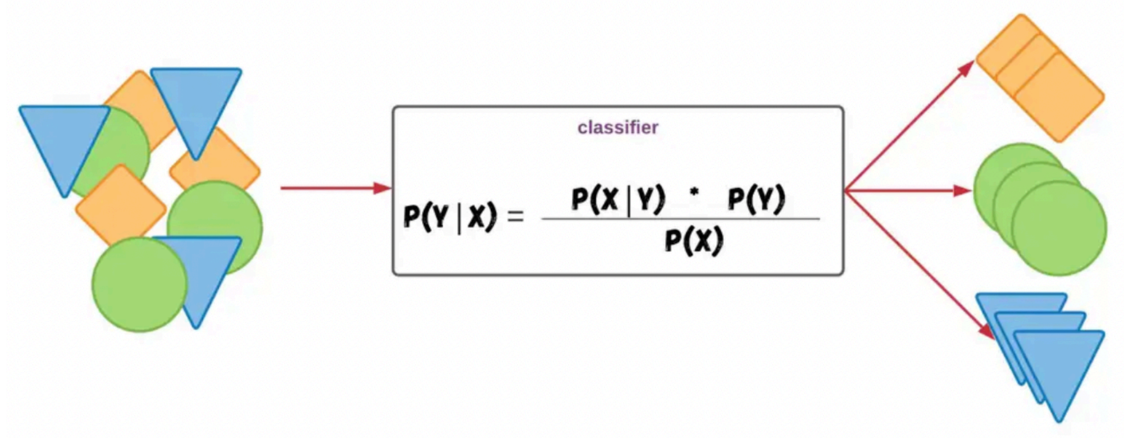
Linear regression



Logistic regression

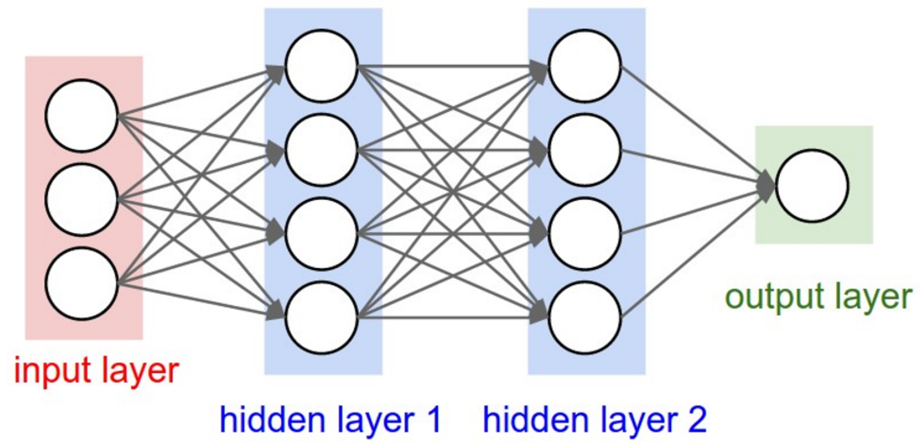


Naive Bayes

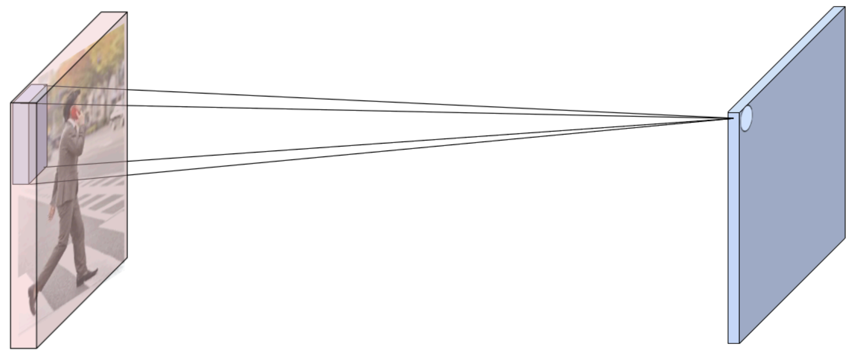


KNN

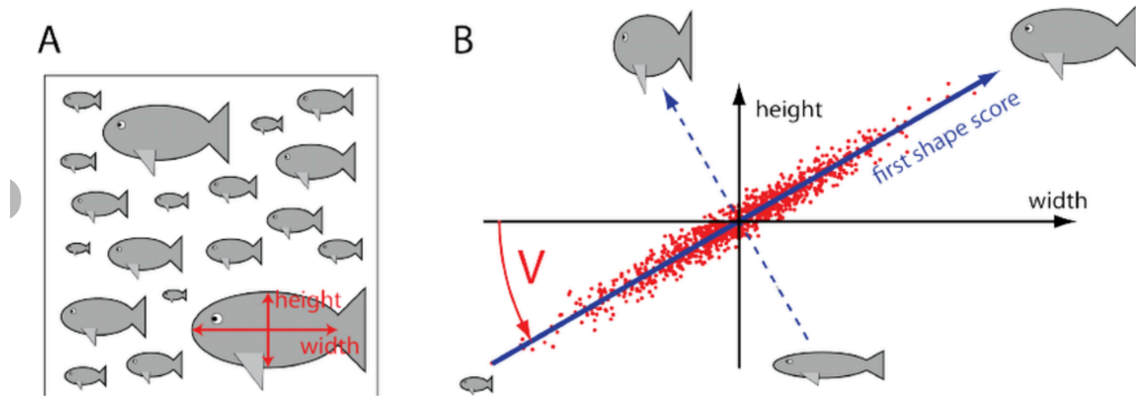
Neural networks



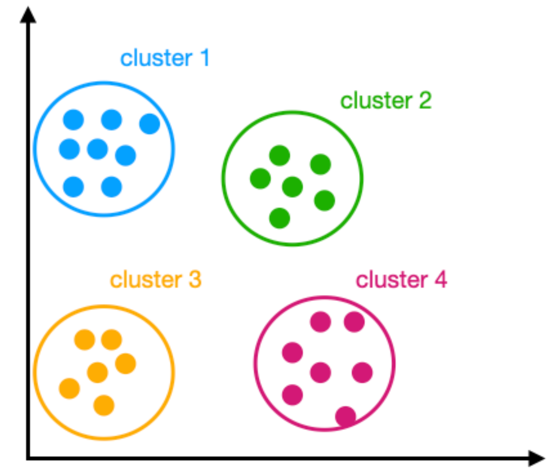
Convolutional neural networks



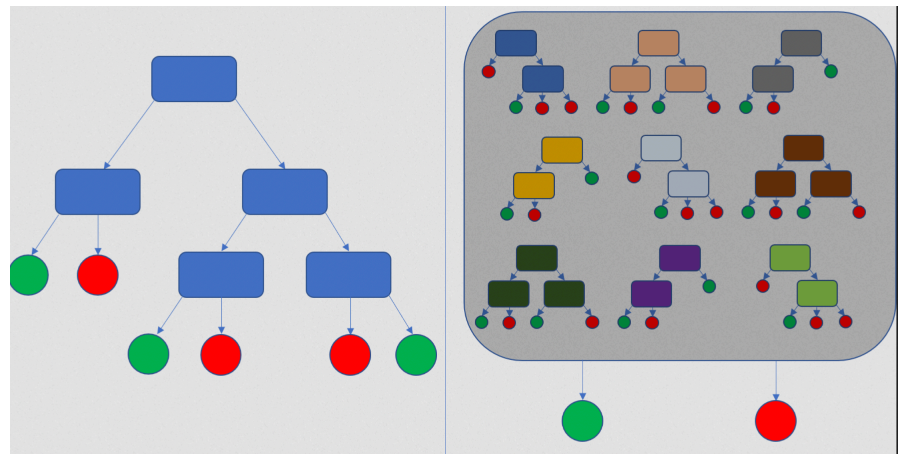
Dimensionality reduction



Clustering

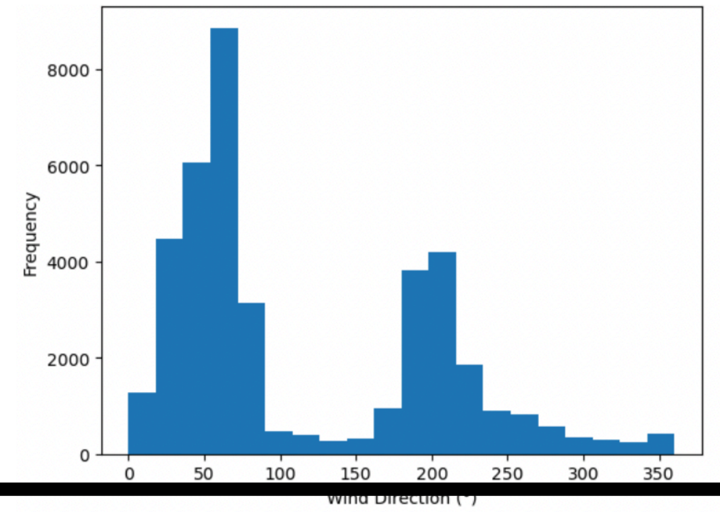


Decision-trees



AI ethics

Reinforcement learning



Schedule for the last week

- Wednesday Dec 11 Quiz, PCA, k-means, decision trees
- Monday Dec 16 given by my doctoral student (I am traveling)
- Wednesday Dec 18 last exercise hour.
- Review Q&A on Wednesday Jan 29, 3:30-4:30 pm. Room to be determined.

AI Ethics: reflections
Discussion outcome

What I like you to take away?
Critical thinking of what we are doing, Who is benefiting from it, and how this is aligned with the challenges in our societies we face.
Our choices in what approach we take, who we work for, and how we work Matter!

Some points students raised:
- Potential to use AI to solve some big problems: biodiversity monitoring, energy optimization, developing medicine
- Just as big risk to exacerbate other big problems: undermine trust, accelerate consumerism, overuse energy and resources
- Motivations of main actors are usually opaque and may differ from public declarations
- Today's international governance is not conducive to proper oversight if AI, just as it is ineffective in solving other big social or environmental problems

AI Ethics: assignment

EPFL

Based on the videos posted on website in Week of Dec. 2

- Now, suppose we are using/developing AI tools, what should be aware of?
- 5 brief videos and one slide set for a total of one hour, prepared by Johan Rochel
- Answer the question in the bonus assignment based on this video

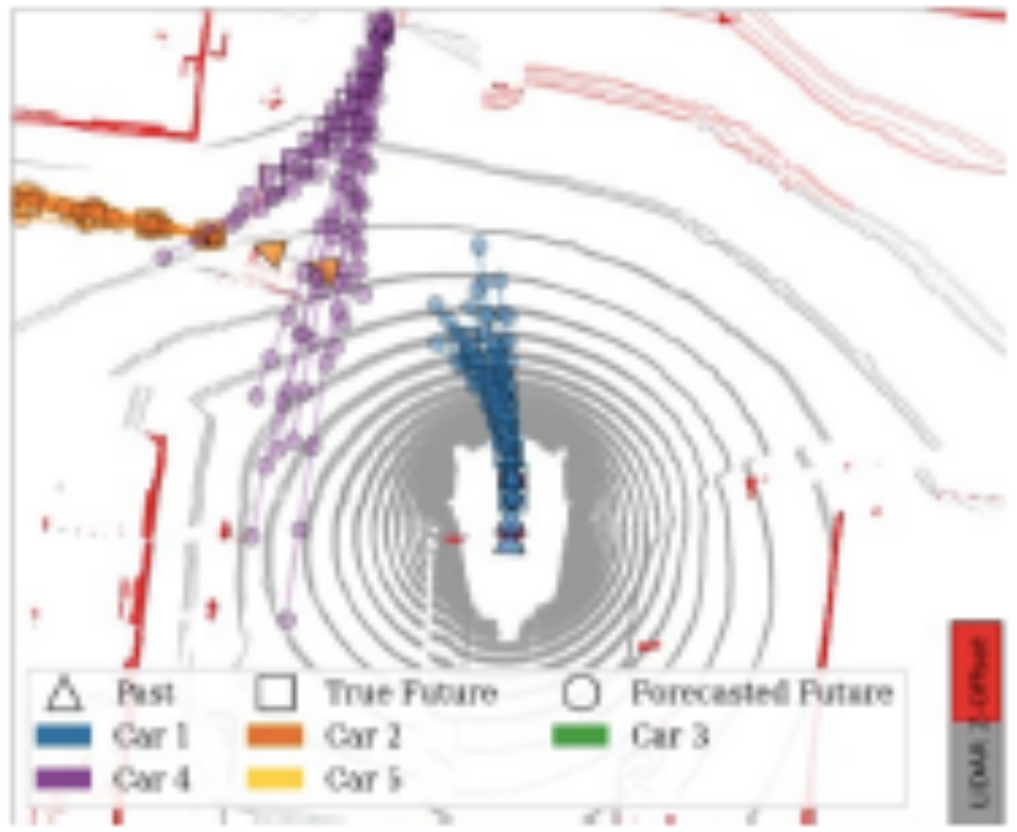
Reinforcement learning

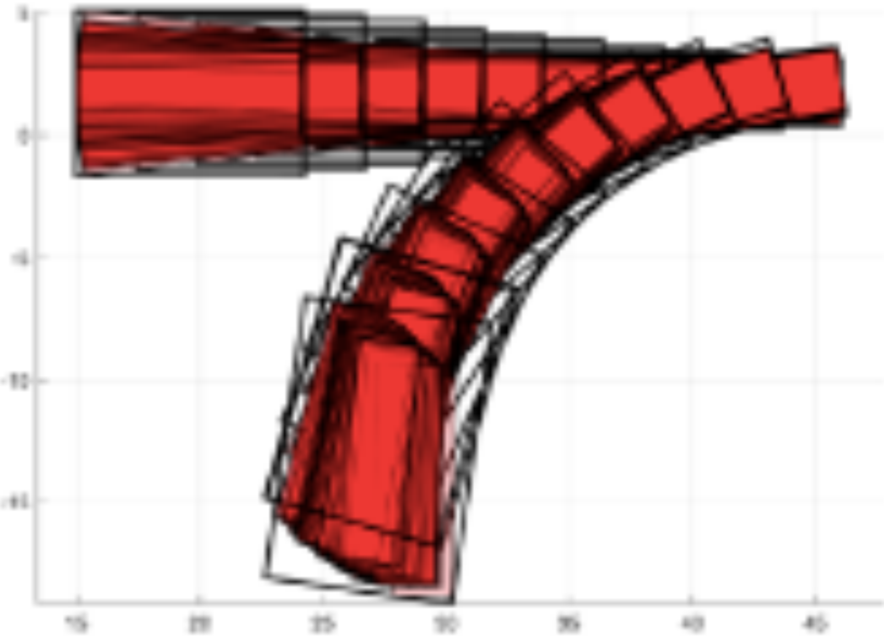
Introduction: discrete-time stochastic optimal control

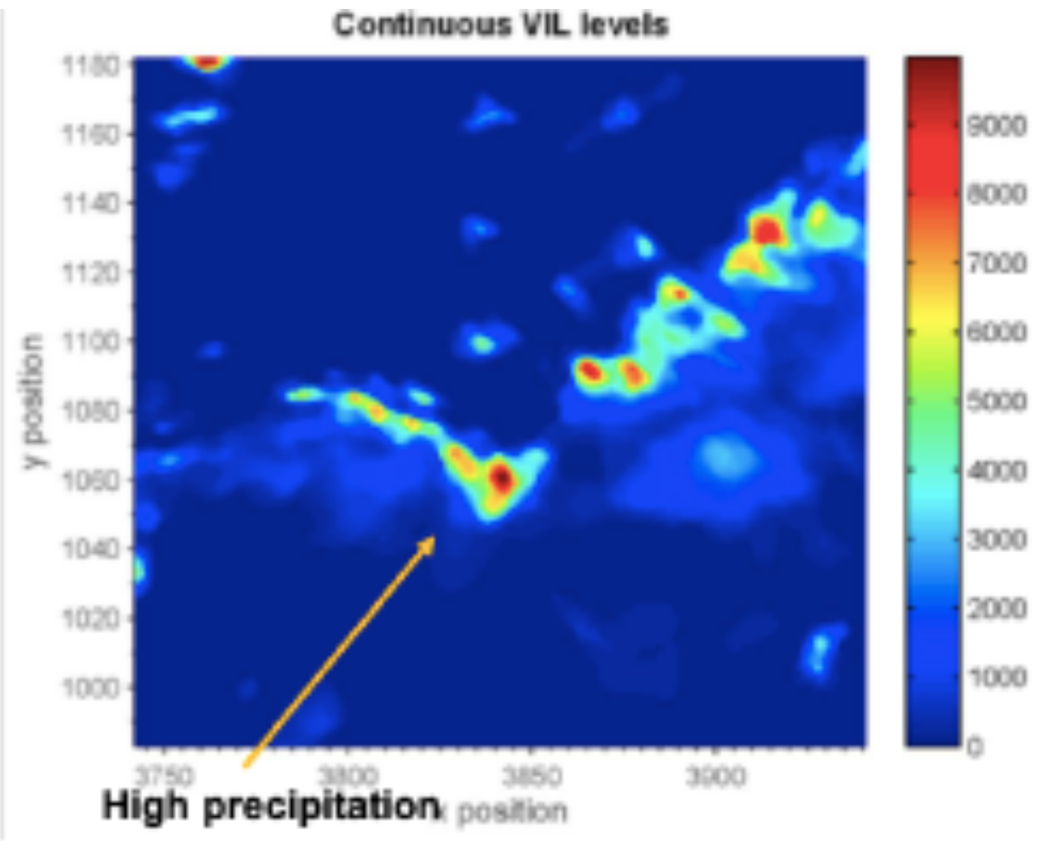
Work from my research group (Sycamore)

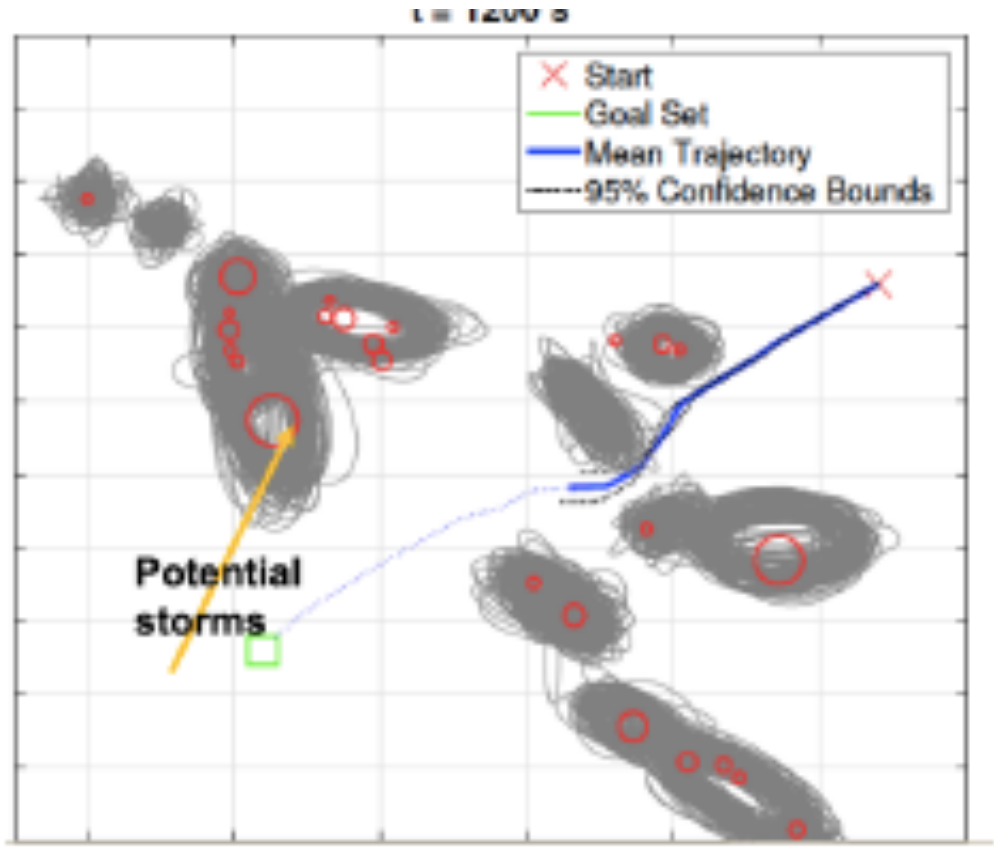
EPFL

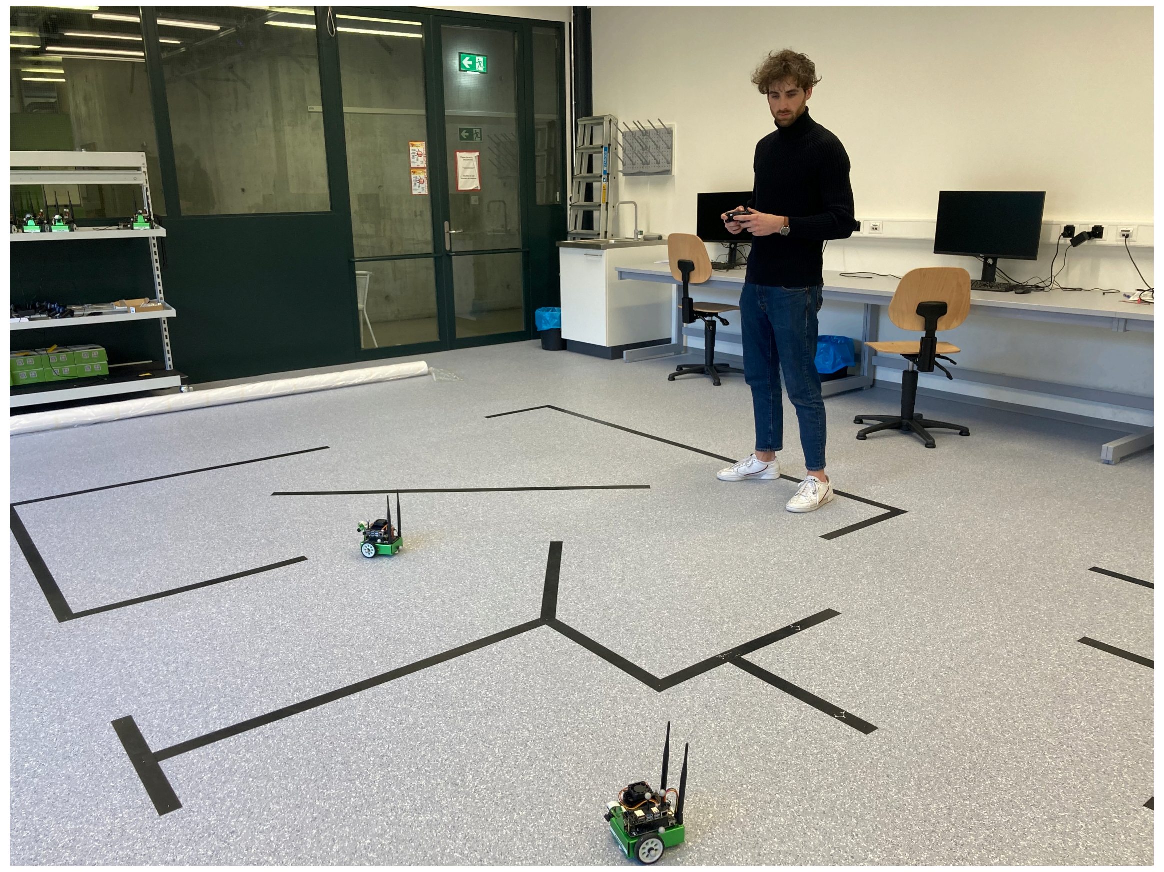
- Aircraft trajectory planning, autonomous driving, robotic construction, power grid control

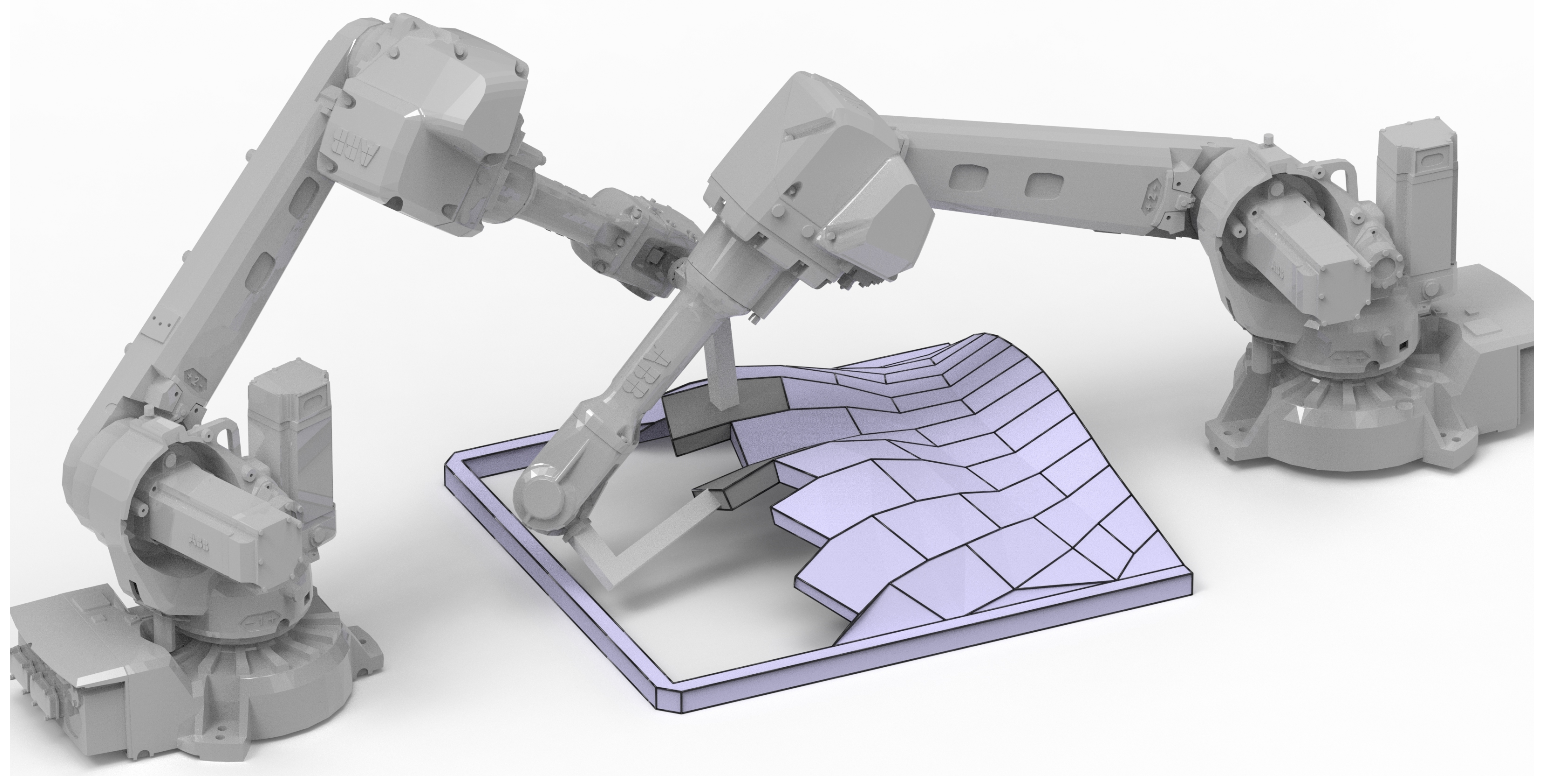


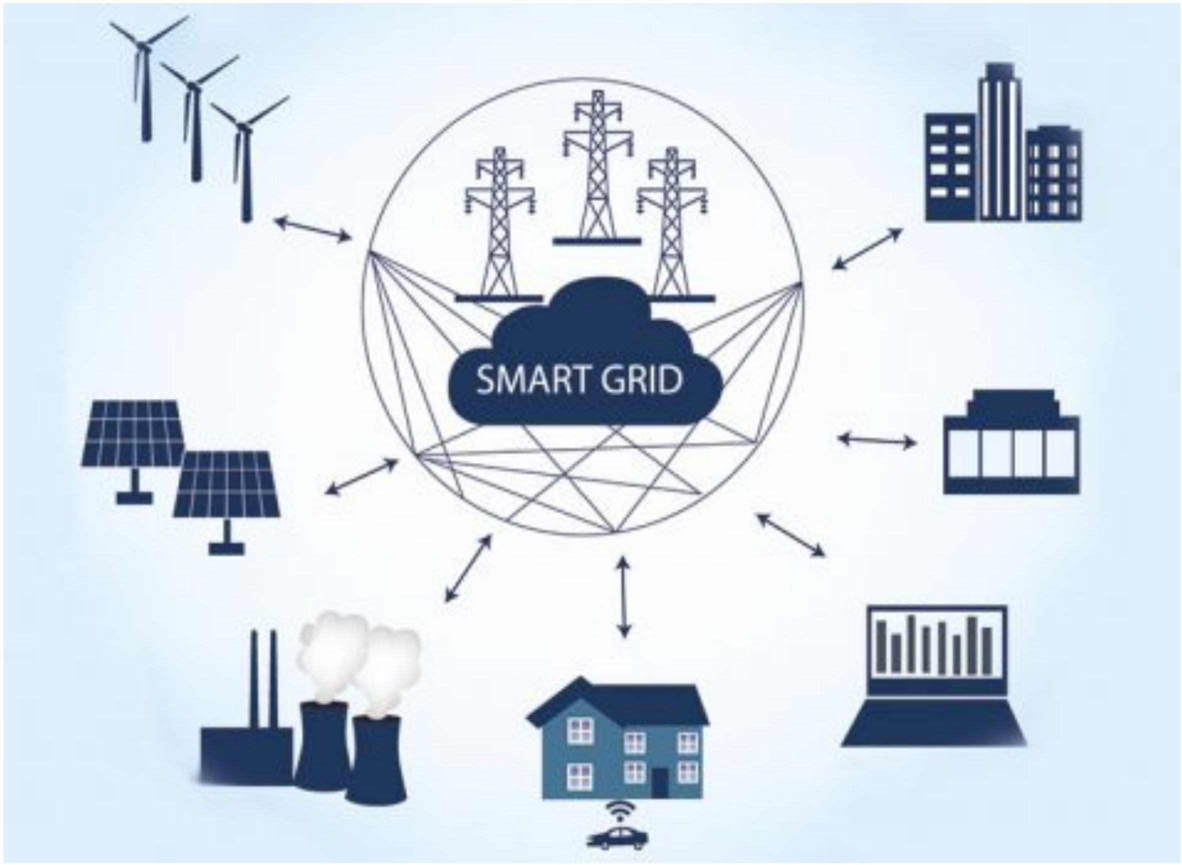












Optimal control of dynamical systems

Dynamics, cost function, objective

Dynamics (2) $\dot{s}(t) = f(s(t), a(t))$, $s(t) \in \mathbb{R}^n$ state

$a(t) \in \mathbb{R}^m$ control input

cost function $\min_{\substack{a(t) \\ t\in[0,T]}} \int_0^T c(s(t), a(t))\,dt$ (1)

objective : to solve (1) subject to (2) and

initial condition $s(0)$

Example application

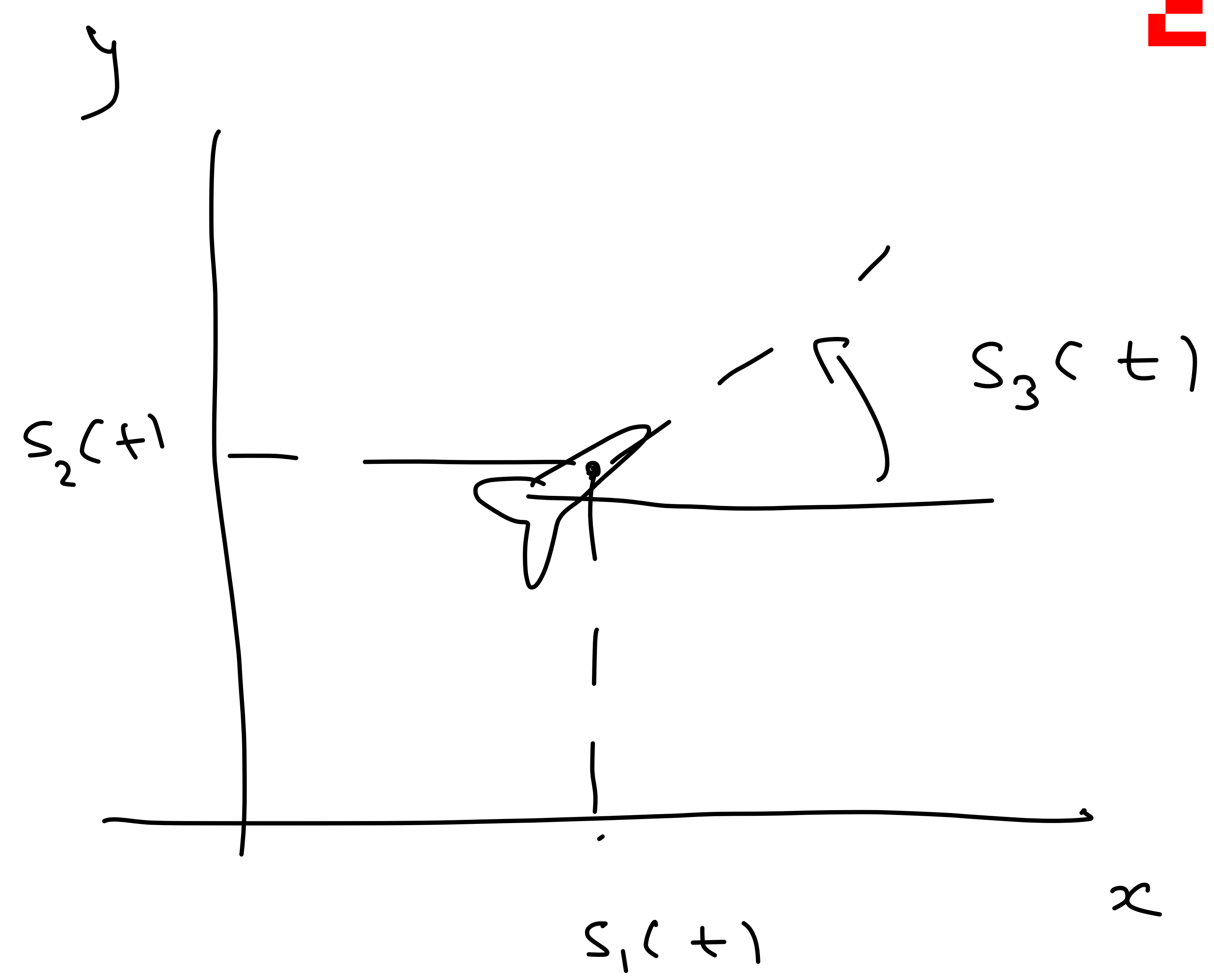
Dynamics, cost function, objective

unicycle model of an aircraft

$s^1(t)$ x position

$s^2(t)$ y position

$s^3(t)$ heading angle



$$\dot{s}^1(t) = a^1(t)\cos(s^3(t))$$

$$\dot{s}^2(t) = a^1(t)\sin(s^3(t))$$

$$\dot{s}^3(t) = a^2(t)$$

$a^1(t)$ velocity

$a^2(t)$ turn rate

$$\min \left\| (s^1(T), s^2(T)) - (s^1_{desired}, s^2_{desired}) \right\|_2^2$$

Dynamical system in discrete-time

Euler discretization

Motivation: 1. finding optimal $a(t)$ $t \in [0,T]$ challenging

2. in practice we use computers / digital controllers

$$\frac{ds}{dt} = \dot{s}(t) = f(s(t), a(t))$$

$$\frac{ds}{dt} \approx \frac{s(t+\delta) - s(t)}{\delta}, \quad f(s(t), a(t)) = \frac{s(t+\delta) - s(t)}{\delta} \Rightarrow$$

$$s(t+\delta) = s(t) + \delta f(s(t), a(t))$$: Euler discretization

Define $s_k := s(t + k\delta)$, $a_k = a(t + k\delta)$

$$s_{k+1} = s_k + \delta f(s_k, a_k)$$, difference equation

Discrete-time optimal control

Finite dimensional optimization

$$\min_{\{a_\ell\}_{\ell=0}^{K}} \sum_{k=0}^{K} c(s_k, a_k)$$

$$\text{s.t.} \quad s_{k+1} = \tilde{f}(s_k, a_k)$$

$$s_0 \in \mathbb{R}^n \quad \text{given.}$$

Decision variables $a_\ell \in \mathbb{R}^m, \quad \ell = 0, \dots, K$

in contrast to $a(t), \quad t \in [0, T]$ in the ODE

Dynamical system control

EPFL

Uncertainty - stochastic model

Consider the aircraft dynamics. What happens in the presence of wind?

$$s^1_{k+1} = s^1_k + \delta a^1_k \cos s^3_k + w^1_k$$

$$s^2_{k+1} = s^2_k + \delta a^1_k \cos s^3_k + w^2_k$$

$$s^3_{k+1} = s^3_k + \delta a^2_k$$

} wind component in x, y coordinates

$w := \begin{bmatrix} w^1 \\ w^2 \end{bmatrix} \sim D$ ~~of~~ probability distribution, example $w_k \sim N(0, \Sigma_w)$,

$\Sigma_w \in \mathbb{R}^{2\times 2}$

Dynamical system control - effects of uncertainties

EPFL

Open-loop versus state feedback policy

open - loop $\quad a_0, a_1, \dots, a_K$

state feedback policy : $\quad a_k = \pi_k(s_k)$, $\quad s_k \in \mathbb{R}^n$ is state of the system at time k.

$$\pi_k : \mathbb{R}^n \to \mathbb{R}^m, \quad k = 0, \dots, K.$$

Stochastic dynamical system control

EPFL

Probabilistic transition model

$s_{k+1} \sim P(\cdot \mid s_k, a_k)$ in open loop

$\hookrightarrow$ given s_k, a_k,

$P(\cdot \mid s_k, a_k)$ is a probability distribution on $\mathbb{R}^n$ (state space)

$s_{k+1} \sim P(\cdot \mid s_k, \pi_k(s_k))$ in closed loop

$P(\cdot \mid s_k)$ is a probability distribution on the state space

Example application - continued

Probabilistic transition

$$\begin{bmatrix} w^1_k \\ w^2_k \end{bmatrix} \sim N(0, \Sigma_w), \quad \Sigma_w \in \mathbb{R}^{2\times 2}, \quad \text{then}$$

$$\left.\begin{aligned} s^1_{k+1} &= s^1_k + \delta V \cos(s^3_k) + w^1_k, \\ s^2_{k+1} &= s^2_k + \delta V \sin(s^3_k) + w^2_k \\ s^3_{k+1} &= s^3_k + \delta a^2_k \end{aligned}\right\} \Longrightarrow$$

V = velocity $= a^1_{\cancel{k}}$

heading rate a^2_k

$$\begin{bmatrix} s^1_{k+1} \\ s^2_{k+1} \\ s^3_{k+1} \end{bmatrix} \sim N\left(\begin{bmatrix} s^1_k + \delta V \cos s^3_k \\ s^2_k + \delta V \sin s^3_k \\ s^3_k + \delta a^2_k \end{bmatrix}, \underbrace{\begin{bmatrix} \Sigma_w & 0 \\ 0 & 0 \end{bmatrix}}_{\in \mathbb{R}^{3\times 3}} \right),$$

Optimal control of stochastic dynamical system

EPFL

Approaches

$$\min_{\{\pi_\ell\}_{\ell=0}^{K}} \mathbb{E} \sum_{k=0}^{K} c(s_k, \pi_k(s_k))$$

$$\text{s.t.} \quad s_{k+1} \sim P(\cdot \mid s_k, \pi_k(s_k))$$

- model predictive control
  solve a sequence of optimization problems;
  receding horizon control

- Dynamic programming approach to stochastic control.

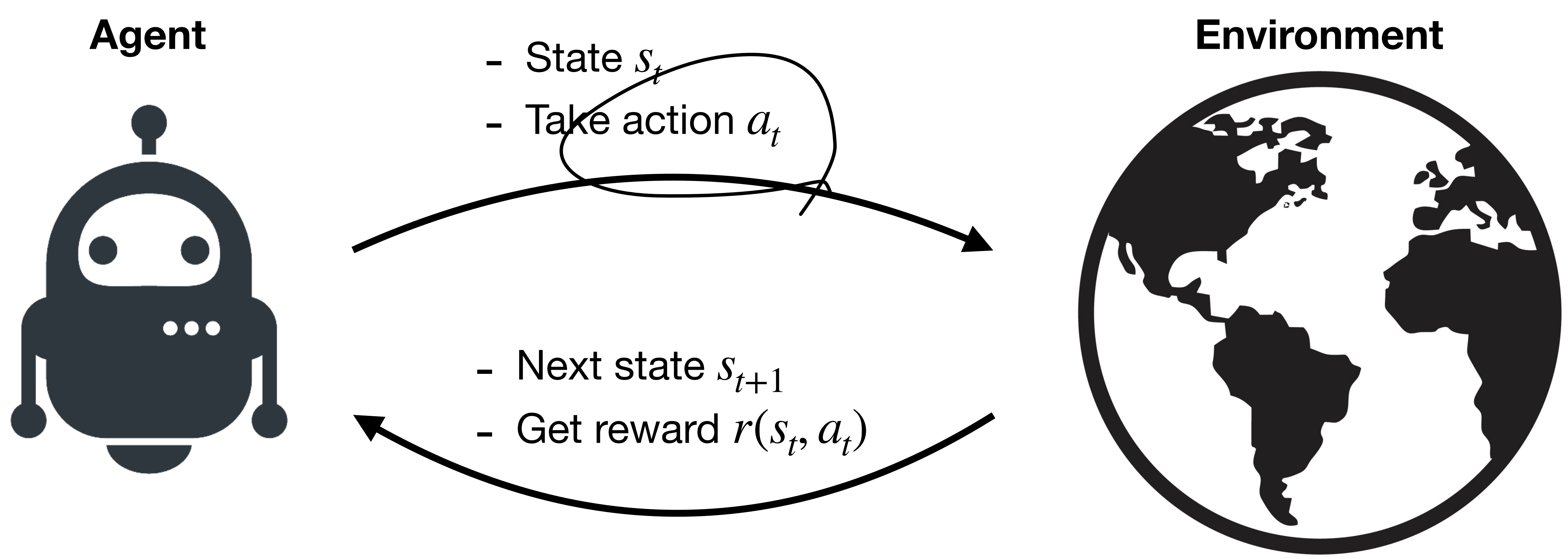
- Reinforcement learning approach

What is reinforcement learning?

EPFL

- **Sutton and Barto, 1998:** *"Reinforcement learning is learning what to do - how to map situations to actions - so as to maximize a numerical reward signal".*

- **ChatGPT, 2022:** *"Reinforcement learning is a type of machine learning in which an* agent *learns to interact with its environment in order to maximize a reward signal".*



Recent advances

EPFL

2013
Atari
Deep Q-learning for Atari games [1].

2016
Energy saving
DeepMind AI reduces Google data centre cooling bill by 40% [3].

2017
AlphaGo/ AlphaZero
AI achieving grand master level in **chess, go,** and **shogi** [4,5].

2018
OpenAI Five
Training five artificial intelligence agents to play the **Dota 2** [6].

2019
Alpha Star
AI achieving grand master level in **StarCraft II** game [7].

Rubik's Cube
Solving **Rubik's Cube** with a human-like robot hand [8].

2022
AlphaTensor
Discovering faster **matrix multiplication** algorithms [9].

ChatGPT
A **language model** trained to generate human-like responses to text input [10].

(Potential) real world applications

- **Robotics**

- **Autonomous driving**

- **Control of power grids**