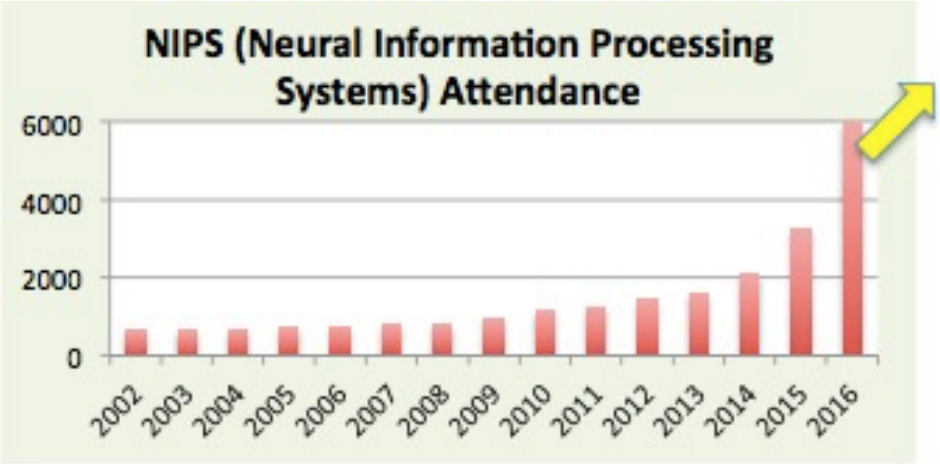
Master Reinforcement Learning 2022 Lecture 1: Intro

Aske Plaat

DRL is Hot

- After Supervised Learning, Reinforcement Learning is now very much in fashion
  - Attendance AI conferences
  - Attendance RL workshops
  - Startups

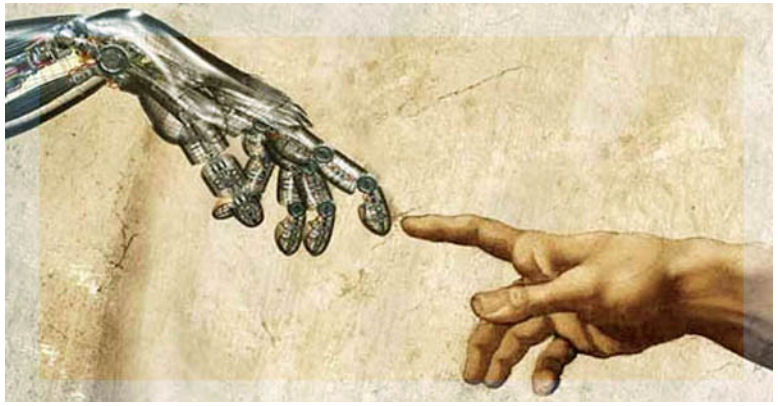


- Everybody wants to understand how it works and what it can do
- Many blogs. Some quite good


The Economist
JUNE 25TH–JULY 1ST 2016
Inside China's Ministry of Truth
Trump in trouble
Who are the Niger Delta Avengers?
The flaws in executive pay
Motorcycles that fly
March of the machines
A SPECIAL REPORT ON ARTIFICIAL INTELLIGENCE


Google
amazon
Alibaba.com
ebay
腾讯
Tencent
DeepMind
D E Shaw & Co
新浪微博
weibo.com
NATIONAL SECURITY AGENCY
UNITED STATES OF AMERICA
Booking
IBM
Microsoft
Baidu 百度
NETFLIX

Quiz: Why is DRL Hot?*

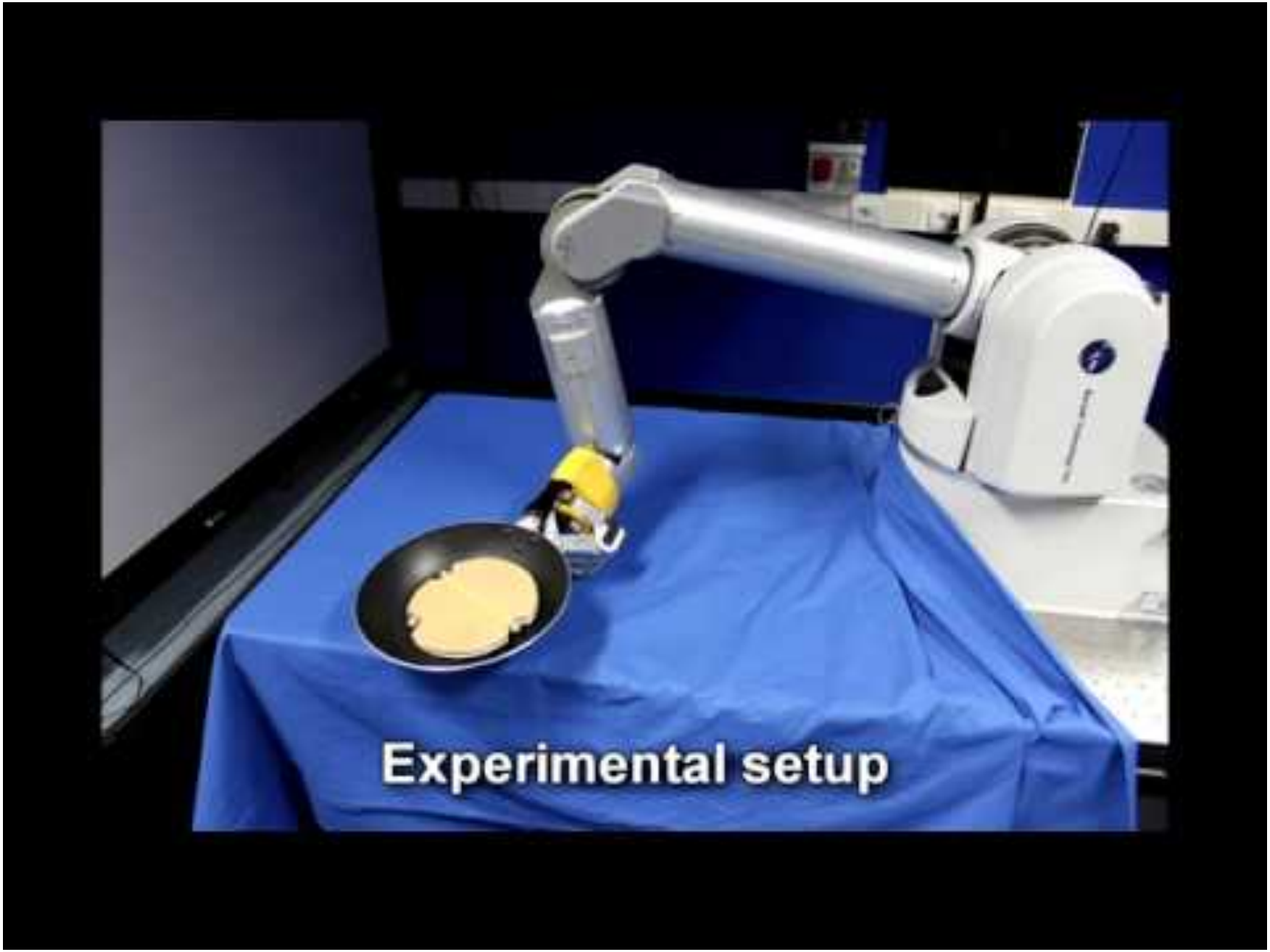


Applications


柯洁
00:4


SPAROV VS
DEEP BLU
the rematch

movie



movie

What is RL?

Neurological Learning

Unconditioned Response
(Salivation)

Unconditioned Stimulus
(Food)

No Response

Neutral Stimulus
(Bell Ringing)

- conditioning
- learning by trial and error

Unconditioned Response
(Salivation)

Neutral Stimulus
(Bell Ringing)

Unconditioned Stimulus
(Food)

Conditioned Response
(Salivation)

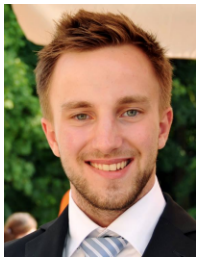
Conditioned Stimulus
(Bell Ringing)

Motivation for this course

- Intelligence is fascinating
- How can computers learn as humans?

Your Teachers





- Aske Plaat, Thomas Moerland, Matthias Müller-Brockhausen, Mike Huisman, Zhao Yang
- Members of LIACS Reinforcement Learning Group
- email: rl@liacs.leidenuniv.nl
- **Brightspace**: register, course documents, assignments
- Question Hour Fridays after course

Prerequisites

- Bachelor level Python
- Bachelor level Artificial Intelligence

Lectures

- **Theory lectures Mandatory**
- Explanations & Interaction
- Fridays 09:15-11:00.

- Practice lectures (Q&A) for reflection & assignment help
- Fridays 11:15-13:00.
- **See SCHEDULE! on Brightspace**

2022: New!

- This year's course is about **Deep** Reinforcement Learning
- **Deep Reinforcement** Learning =
  Deep Learning + **Reinforcement** Learning

- *Old assignments no longer valid (they never are)*
- *New Book (still free)* *https://deep-reinforcement-learning.net*
- *Old books are still useful but only as additional book*
- *Programming! Every three weeks an assignment*

Book

- **Book: https://deep-reinforcement-learning.net**
- **Mandatory material**
- Follow Book, Slides are extra
- **Preprint: Comments are Greatly Appreciated**
- [Optional Book Learning to Play, RL + Games]
- [Optional Book Sutton & Barto, the Bible of RL]

Different Aspects

- Deep Learning [B]
- Model-free
  - Value-based [2,3]
  - Policy-based [4]
- Model-based
  - Learned Transition model [5]
  - Perfect; Two-Agent [6]
- Multi-agent [7]
- Hierarchical Reinforcement Learning (Sub-goals) [8]
- Meta Learning [9]

Teaching Philosophy

- Read Chapter
  - Attend Lecture
    - Try **Assignment**
      - Practicals Ask Help

Field of RL

- Teach latest scientific achievements
- Learning by doing: Assignments are where you learn the most
- Work hard learn much.
  LOTS OF PROGRAMMING!
- Results in research papers are based on much tuning, which they do not tell with so many words.
- Computational Demands. Long Training
- Rough ride in friendly atmosphere

Assignments

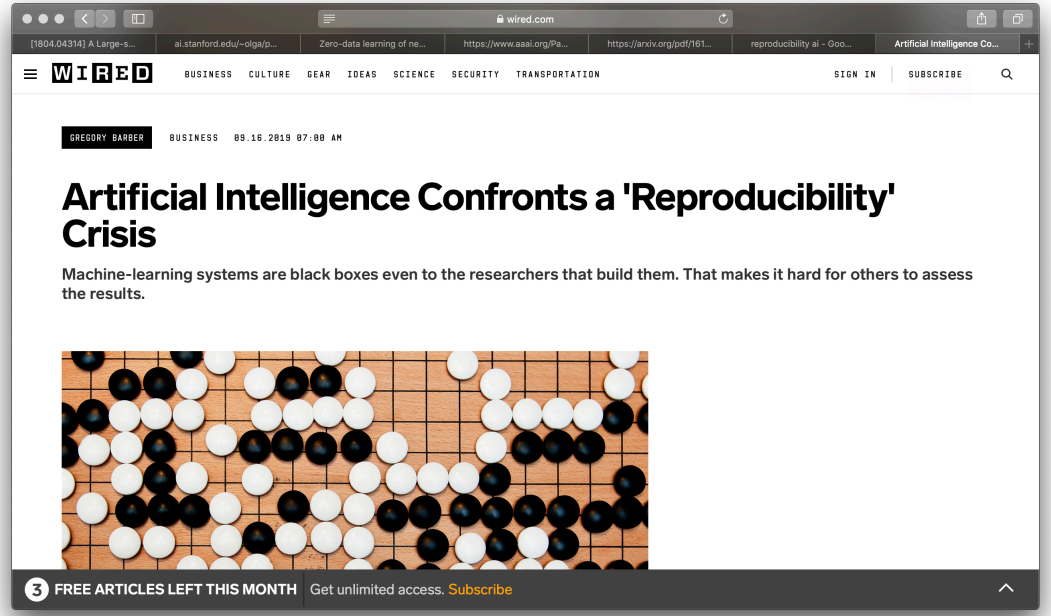
- 3 Assignments, each 25%
- 1 Theory exam 25%
  (Quizzes: see Book. Sample Exam see Blackboard)
- Final grade is the average of the 5 individual elements:
  4 assignments and 1 theory exam. No pass thresholds for individual items. You pass when the average >5.5
  9-9-9-9-1 is fine: 37/5=7.4
- There is no retake for the assignments, but there is a late-hand-in policy: one point off for each day late.
- All assignments are valid for only one year. If you do this course a second time you have to do everything over.

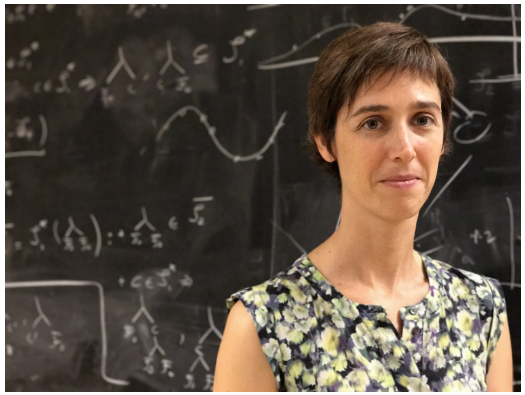
Assignment Groups

- First assignment: individual
- Second & Third assignment: Groups of 3
  Put all names on each report
- If from Math, Astro, Physics,
  choose a CS in your group (programming!)
- Register on Brightspace
- Submit assignments to 1 course (not also to Neural Networks or to Game AI)
- Working code + short report
- Well written, well documented,
  clear report, reproducible, insights

Grading Criteria

- See Brightspace
- Clear Explanation of experiments
- Good measurements, graphs
- Try explanations of findings
- Fair grading: all who seriously try should be able to pass





Joelle Pineau

Exam

- Theory exam and re-take
- 25%
- **Multiple Choice**
- Closed book, No SmartWatches, SmartPhones, Laptops (only human intelligence and mechanical time piece allowed)
- Inspired by “Questions” in Exercise Sections in Book
- Sample Exam is available on website and Brightspace

No Fraud

- Fraud is absolutely forbidden
- All copying of other's work should be properly referenced in customary scientific fashion, e.g., [Sutton, 2018]
- All cases of fraud will be directed to the exam committee. You may be expelled from the course or from your master program. Crying will not help you
- No using part of the code of last year, another group, etc.
- No looking at your neighbours exam
- No electronic equipment during exam
- Cheating is fraud

No Fraud

- A consequence of fraud is likely failing this course
- If you follow a track in which this course is mandatory, you waste a full year by cheating
- Your work should be your own, or that of your group
- Take precautions against others copying your work

Recommended

- Deep Learning and Neural Networks, Wojtek Kowalczyk
- Modern Game AI, Mike Preuss
- Elective Seminar Advanced Deep RL Sep-Dec 2022
  Limited enrolment

Resources

- Book:
  Deep Reinforcement Learning: [Plaat 2022, Preprint] https://deep-reinforcement-learning.net

- Additional material
  - Reinforcement Learning: An Introduction [Sutton & Barto 2018] http://incompleteideas.net/book/RLbook2018.pdf

  - Deep Learning [Goodfellow et al. 2016] http://www.deeplearningbook.org

  - RL course David Silver UCL http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching.html

  - RL course Emma Brunskill Stanford http://web.stanford.edu/class/cs234/index.html
  - Deep RL Sergey Levine Berkeley http://rail.eecs.berkeley.edu/deeprlcourse/

  - ML course Nando de Freitas https://www.cs.ox.ac.uk/people/nando.defreitas/machinelearning/

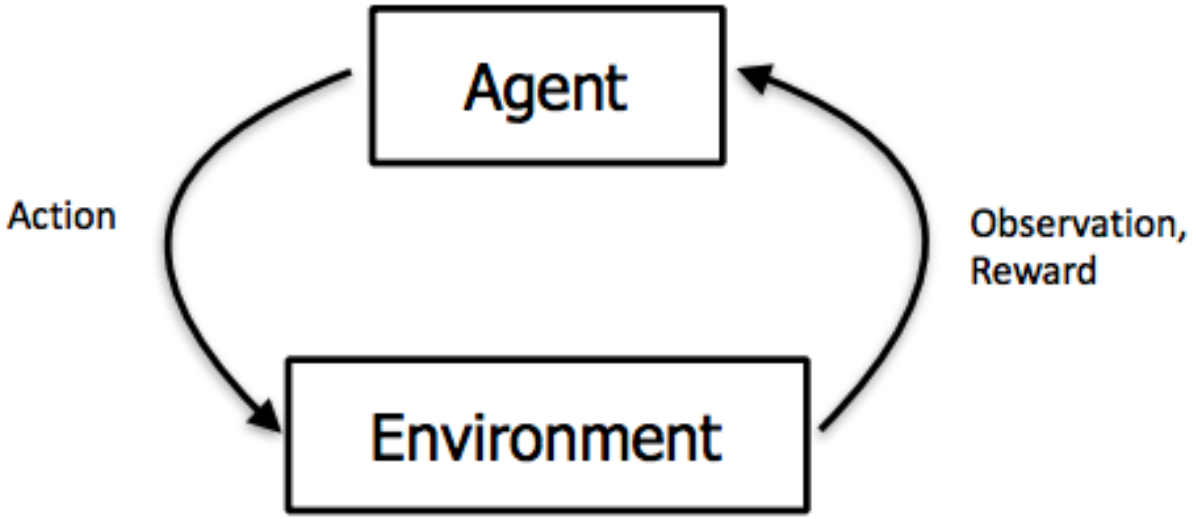
Assignments due

- **6 Mar Assignment 1 (2: Tabular) Individual**
- 3 Apr Assignment 2 (3: DQN) Group of 3
- 8 May Assignment 3 (4: Policy) Group of 3
- 9 june Exam Individual
- 14 july Retake exam Individual

Questions?

Overview of Course

- 1: Machine Learning
- 2: MDP + Tabular Value Based
- 3: Deep Value Based
- 4: Policy Based
- 5: Model Based
- 6: Two-Agent (AlphaZero)
- 7: Multi-Agent (Poker, StarCraft)
- 8: Hierarchical
- 9: Meta Learning
- [B: Deep Supervised Learning]



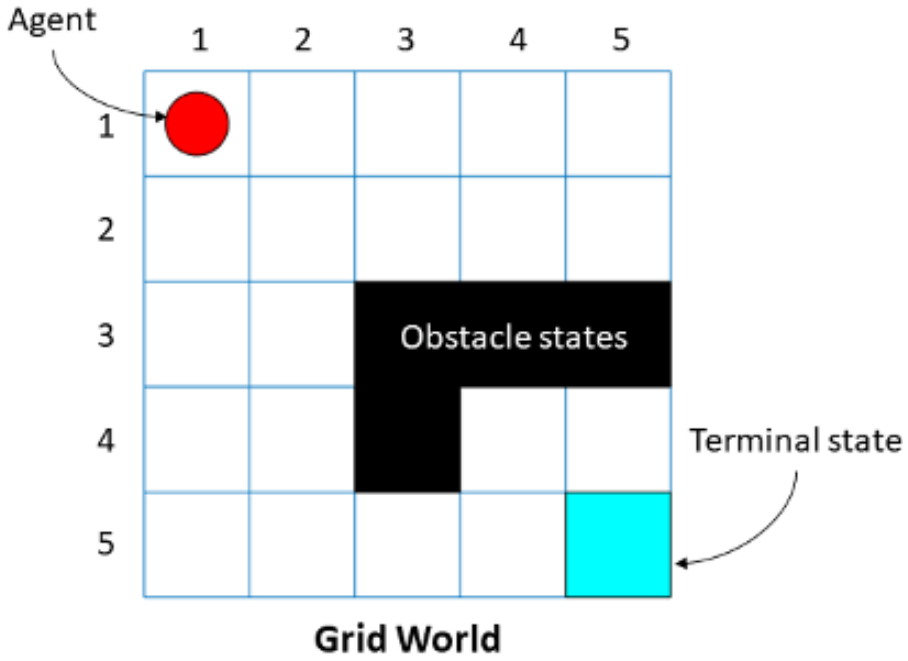
1. Machine Learning

- Reverse Engineering a function from data

- 1. Supervised: labeled dataset
- 2. Reinforcement: environment that return a number
- 3. Unsupervised: no labels, inherent clustering

2. MDP + Tabular

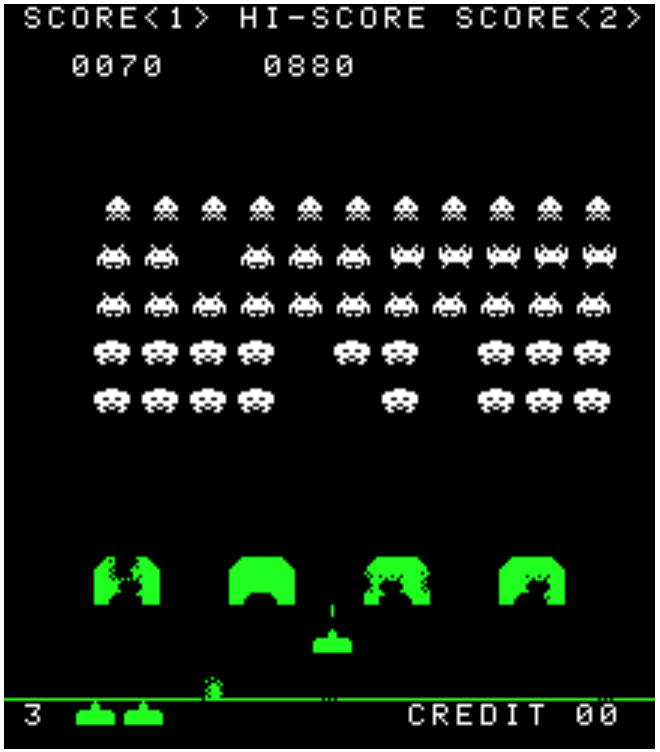
- State
- Action
- Transition
- Reward
- Discount



- Learn the **Policy** Function: what action to take in what state
- “Tabular” means array
- Exact algorithms for small problems that fit fully in memory, **Q-learning**

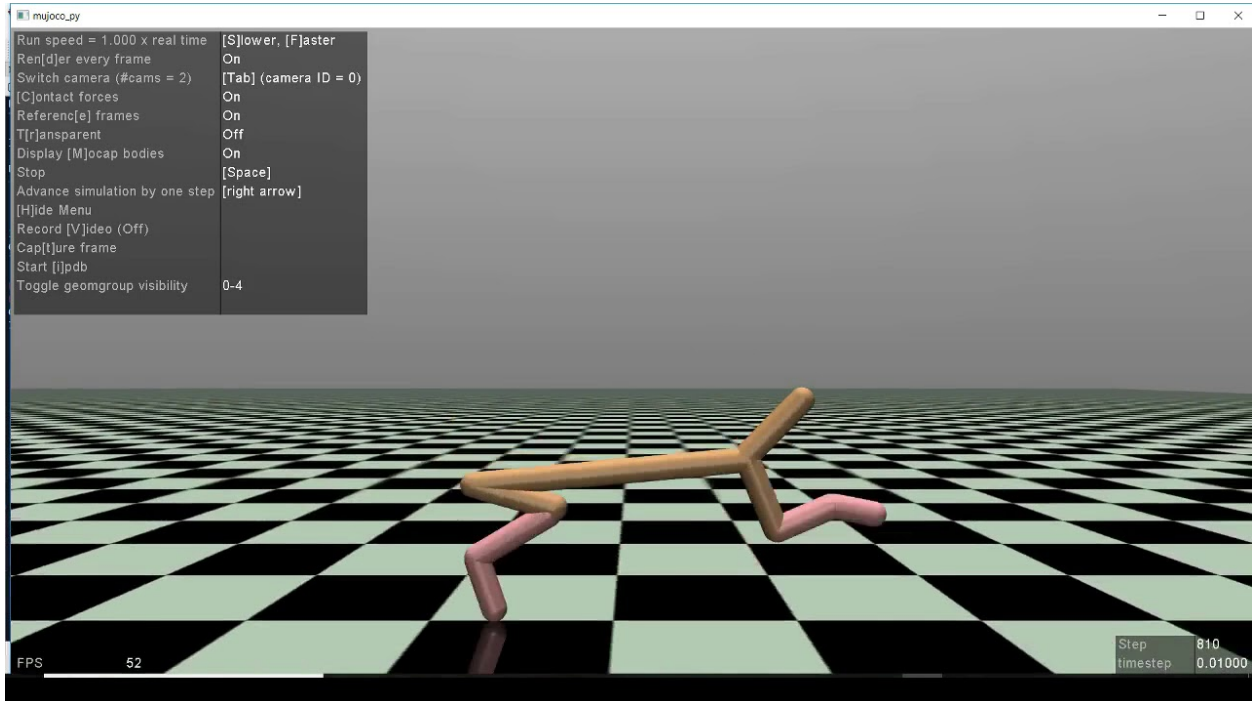
3. Deep Value-Based

- Large problem that are “solved” with function approximation
- Deep Learning of the policy function
- **DQN**
- 1980’s Atari video games, policy from pixel to joystick
- Space invaders as topic of serious scientific breakthrough



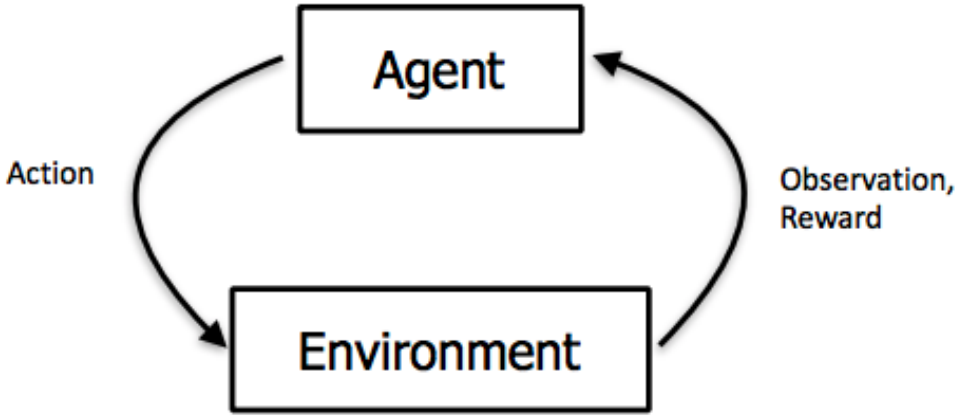
4. Policy-Based

- Direct Policy for Continuous Action Space
- PPO
- (Simulated) Robots that learn themselves to walk



5. Model-Based

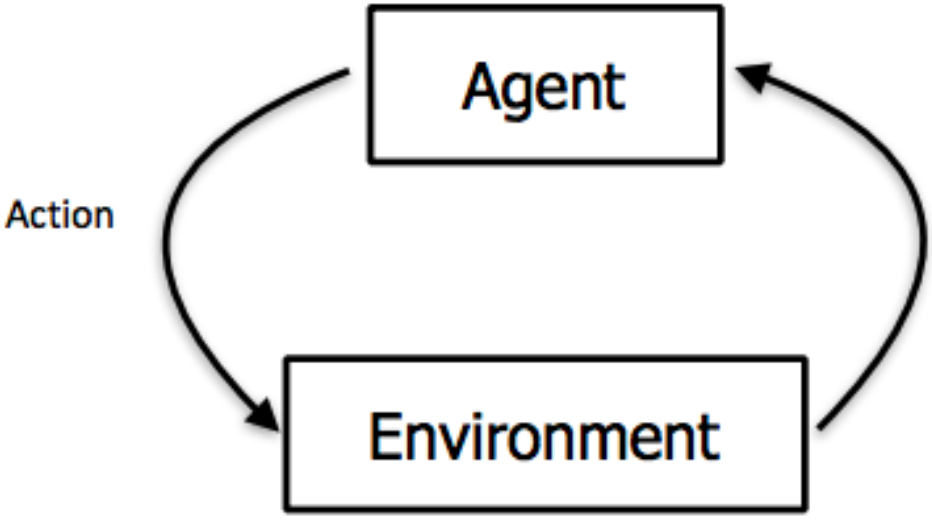
- s1->a1->s2->a2->s3->a3->s4->a4->s5
- **Policy** function is s->a
- **Transition** function is a->s

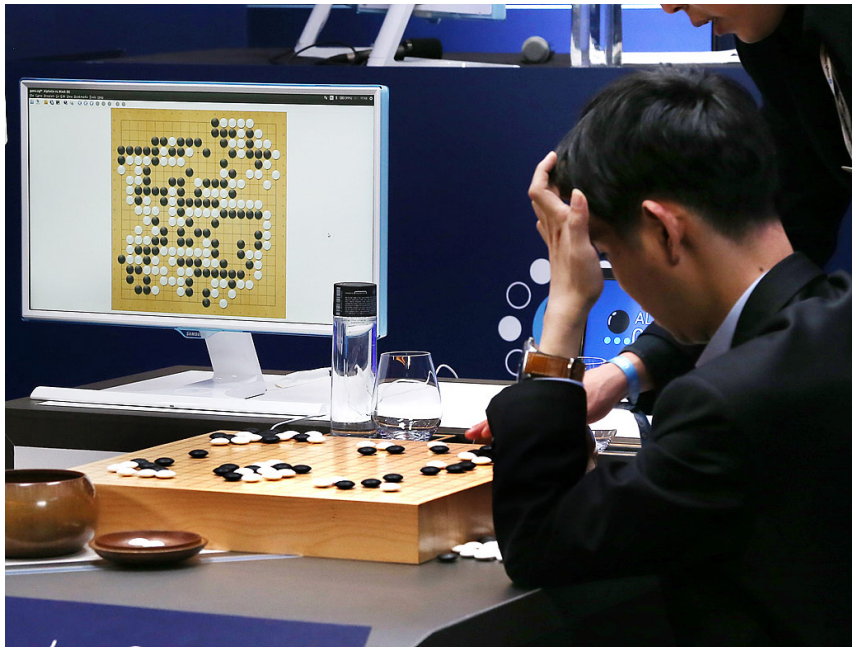


- Model-free: Agent learn **policy** from environment
- Model-based: Agent first learns **Transition** from environment and then uses planning to learn **policy**

6. Two Agent Self Play

- What if agent has PERFECT transition function? (Because you are your own opponent)
- Then we can learn “perfect” policy by self-play
- Board games: TD-Gammon, AlphaGo



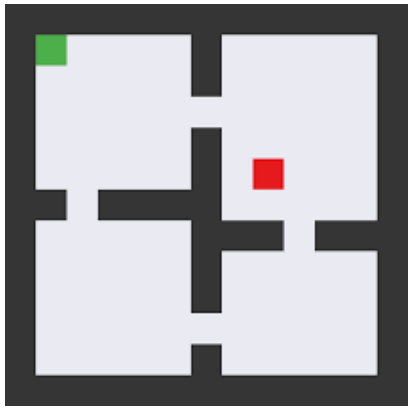


7. Multi Agent

- Model many Agents
- Competition, Cooperation
- Nash, Pareto, Prisoner's Dilemma
- Non-stationarity, Partial Observability
- Population-based approaches

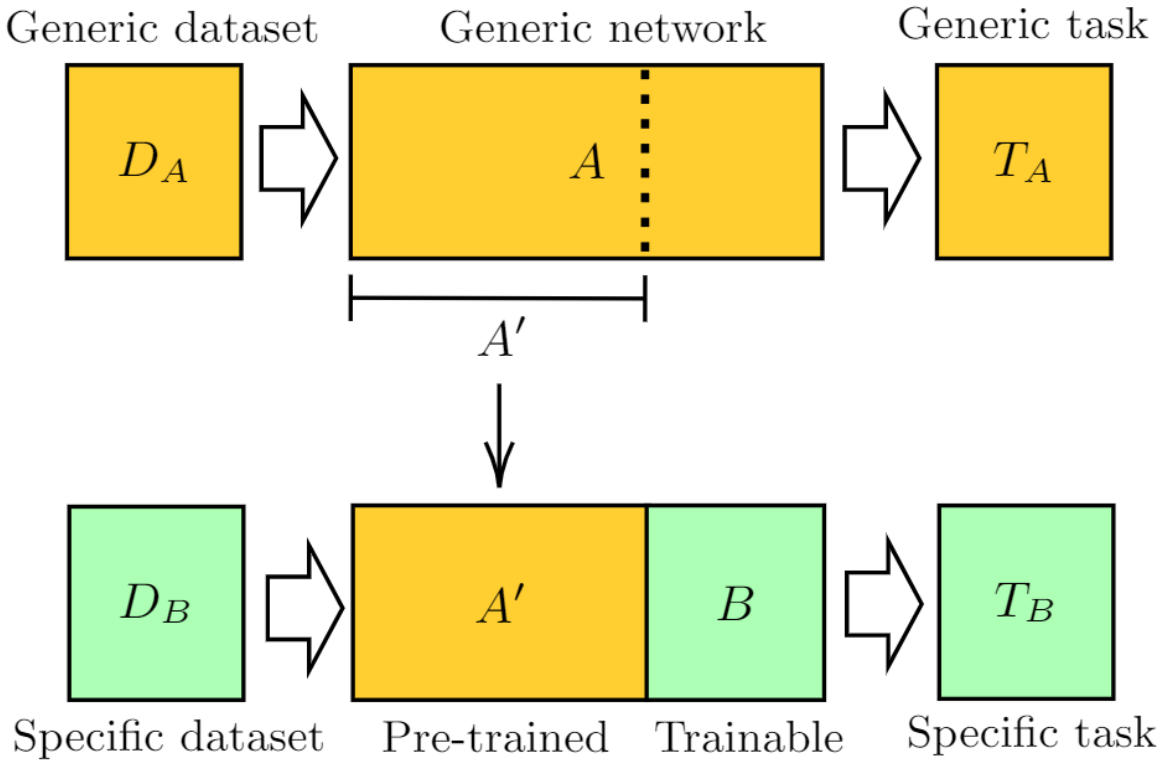
8. Hierarchical RL

- Travel intuition
- Subgoals, subpolicies
- Goal conditioned functions
- Intrinsic motivation



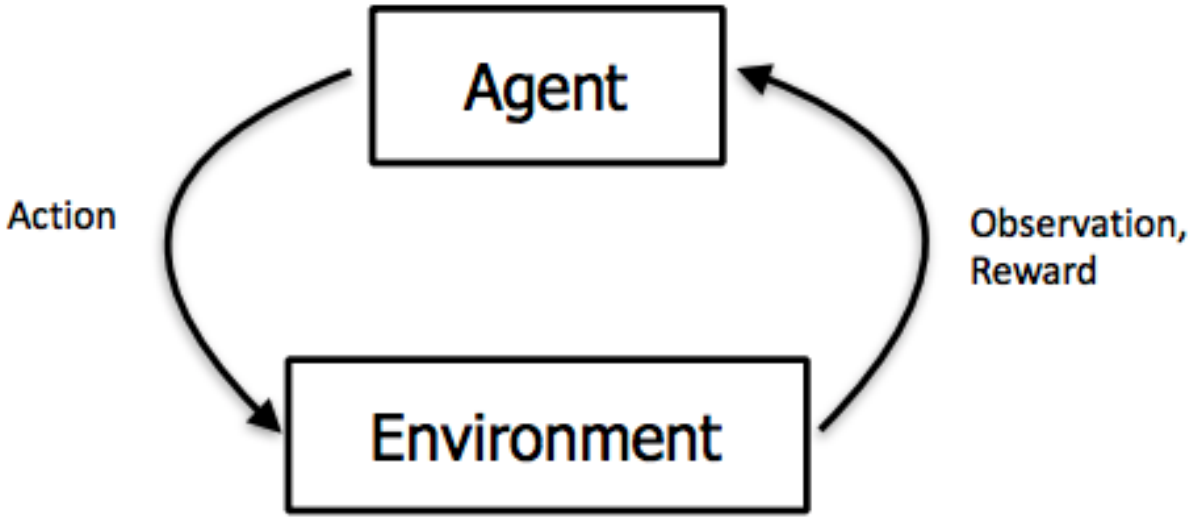
9. Meta Learning

- Learn faster by using knowledge from related tasks
- Lower layers contain more general knowledge
- Upper layers more task specific
- Transfer lower layers
- Learning to learn from a sequence



Overview of Course

- 1: Machine Learning
- 2: MDP + Tabular Value Based
- 3: Deep Value Based
- 4: Policy Based
- 5: Model Based
- 6: Two-Agent (AlphaZero)
- 7: Multi-Agent (Poker, StarCraft)
- 8: Hierarchical
- 9: Meta Learning
- [B: Deep Supervised Learning]



Questions?