

Reinforcement Learning

M.Sc. in AI & Robotics

Roberto Capobianco



SAPIENZA
UNIVERSITÀ DI ROMA

About Me

— — —



SAPIENZA
UNIVERSITÀ DI ROMA



Name: Roberto Capobianco

Current Positions: External Professor @ Sapienza |
Senior Research Scientist @ Sony AI

Research Interests: Reinforcement Learning,
Explainable AI, Robot Learning, Knowledge
Representation (<https://krlgroup.github.io>)

Hobbies: Playing drums, drawing, hiking



About your 'Corso di Laurea'

SONY

— — —
Coordinator: Prof. Domenico Lembo

Google Classroom:

<https://classroom.google.com/c/NDAxOTU1NTQ0Mjc4?cjc=y2ae7sp>

Useful for news, general information and upcoming seminars (including a *welcome* seminar soon)



SAPIENZA
UNIVERSITÀ DI ROMA

About the Course

SONY

— — —
Teaching Assistants: Andrea Fanti, Federico di Valerio



Google Classroom: [Link](#) (or code kbrncgq)

Schedule:

- Thursday 09:00 - 11:00 (theory), Room B2, Via Ariosto 25
- Friday 11:00 - 15:00 (theory + practical), Room 201, Building D, Viale Regina Elena 295

Covered Topics: Sequential Decision Making, Dynamic Programming, Linear Quadratic Regulators, Model-Free Reinforcement Learning, Model-Based Reinforcement Learning, Deep Reinforcement Learning, Exploration in RL, Multi-Agent RL



SAPIENZA
UNIVERSITÀ DI ROMA

About the Course

Teaching Assistants: Andrea Fanti, Federico di Valerio



Math-Heavy Course

Google Classroom: [Link](#) (or code kbrncgq)

Schedule:

- Thursday 09:00 - 11:00 (theory), Room B2, Via Ariosto 25
- Friday 11:00 - 15:00 (theory + practical), Room 201, Building D, Viale Regina Elena 295

Covered Topics: Sequential Decision Making, Dynamic Programming, Linear Quadratic Regulators, Model-Free Reinforcement Learning, Model-Based Reinforcement Learning, Deep Reinforcement Learning, Exploration in RL, Multi-Agent RL



SAPIENZA
UNIVERSITÀ DI ROMA

About the Course



Math-Heavy Course

$$\begin{aligned} Q^{\pi^{t+1}}(s, a) - Q^{\pi^t}(s, a) &= \gamma \mathbb{E}_{s' \sim P(s, a)} \left[Q^{\pi^{t+1}}(s', \pi^{t+1}(s')) - Q^{\pi^t}(s', \pi^t(s')) \right] \\ &= \gamma \mathbb{E}_{s' \sim P(s, a)} \left[Q^{\pi^{t+1}}(s', \pi^{t+1}(s')) - Q^{\pi^t}(s', \pi^{t+1}(s')) + Q^{\pi^t}(s', \pi^{t+1}(s')) - Q^{\pi^t}(s', \pi^t(s')) \right] \\ &\geq \gamma \mathbb{E}_{s' \sim P(s, a)} \left[Q^{\pi^{t+1}}(s', \pi^{t+1}(s')) - Q^{\pi^t}(s', \pi^{t+1}(s')) \right] \geq \dots, \geq -\gamma^\infty / (1 - \gamma) = 0 \end{aligned}$$

$$\begin{aligned} \nabla_\theta J(\theta) &= \nabla_\theta \sum P(\tau; \theta) R(\tau) \\ &= \sum_\tau \nabla_\theta P(\tau; \theta) R(\tau) \\ &= \sum_\tau \frac{P(\tau; \theta)}{P(\tau; \theta)} \nabla_\theta P(\tau; \theta) R(\tau) \\ &= \sum_\tau P(\tau; \theta) \frac{\nabla_\theta P(\tau; \theta)}{P(\tau; \theta)} R(\tau) \\ &= \sum_\tau P(\tau; \theta) \nabla_\theta \log P(\tau; \theta) R(\tau) \end{aligned}$$

$$\begin{aligned} \|V^t(s) - V^\pi(s)\| &\leq \gamma^t \|V^0 - V^\pi\| \\ \forall s, \left| V^{t+1}(s) - V^\pi(s) \right| &= \left| r(s, \pi(s)) + \gamma \mathbb{E}_{s' \sim P(\cdot | s, \pi(s))} V^t(s') - \left(r(s, \pi(s)) + \gamma \mathbb{E}_{s' \sim P(\cdot | s, \pi(s))} V^\pi(s') \right) \right| \\ &= \gamma \left| \mathbb{E}_{s' \sim P(\cdot | s, \pi(s))} V^t(s') - \mathbb{E}_{s' \sim P(\cdot | s, \pi(s))} V^\pi(s') \right| \\ &\leq \gamma \mathbb{E}_{s' \sim P(\cdot | s, \pi(s))} \left| V^t(s') - V^\pi(s') \right| \\ &\leq \gamma \|V^t - V^\pi\|_\infty \end{aligned}$$













About the Course

Zoom & Streaming:

Zoom Information:

This year, classes won't be streamed or recorded. However, all the slides, assignments, practicals and recorded material from last year is already available on the Classroom web-page. New material that is covered this year will not be recorded.

Recordings

	Class (theory + practical), Sept 28th, 2022	Publicazione: 28 set 2022
	Class (theory), Sept 30th, 2022	Publicazione: 4 ott 2022
	Class (theory + practical), Oct 05th, 2022	Publicazione: 6 ott 2022
	Class (theory), Oct 07th, 2022	Publicazione: 10 ott 2022
	Class (theory + practical), Oct 12th, 2022	Publicazione: 13 ott 2022
	Class (theory), Oct 14th, 2022	Publicazione: 14 ott 2022
	Class (theory + practical), Oct 19th, 2022	Publicazione: 20 ott 2022
	Class (theory), Oct 21st, 2022	Publicazione: 26 ott 2022
	Class (theory + practical), Oct 26th, 2022	Publicazione: 26 ott 2022
	Class (assignment & project description), O...	Publicazione: 3 nov 2022

Mostra altro



SAPIENZA
UNIVERSITÀ DI ROMA

About the Course

SONY

Exam:

Grading

The exam will be composed by 3 assignments (or a written exam) plus a final project for a total of 95% of the maximum score (which is 30 cum laude - or 33 - according to the Italian system). The missing 5% can be obtained through the attendance bonus or through a valid contribution to the course repository, to be agreed with the TAs.

Grade breakdown

- Assignments (3 homeworks: 15% each) or written exam 45% ;
- Final project 50%;
- Attendance bonus or repository contribution 5%.

Assignment rules

Students may discuss assignments, but each student must code up and write up their solutions independently. Students must also indicate on each homework the names of the colleagues they collaborated with and what online resources they used.

Late days policy

Assignments must be submitted by the posted due date. However, we recognize that students may face unusual circumstances and require some flexibility during the semester. For this reason, each student will have a total of 6 free late (calendar) days to use as (s)he prefers, that are counted and tracked at the granularity of days: e.g., 3 hours late is one late day. If an assignment is submitted beyond the late-day budget, the student will receive a 33.33% penalty per day over budget (up to zero). Late days can only be used for assignments, not for the project.

Participation/extra effort bonus

We encourage participation including asking/answering questions in lectures. For this reason, active participation in class will be tracked and considered in the final score. Similarly, contributions to the code repository of the course will be considered, previous agreement with the TAs.



SAPIENZA
UNIVERSITÀ DI ROMA

About the Course

SONY

Exam:

Grading

The exam will be composed by 3 assignments (or a written exam) plus a final project for a total of 95% of the maximum score (which is 30 cum laude - or 33 - according to the Italian system). The missing 5% can be obtained through the attendance bonus or through a valid contribution to the course repository, to be agreed with the TAs.

Grade breakdown

- Assignments (3 homeworks: 15% each) or written exam 45% ;
- Final project 50%;
- Attendance bonus or repository contribution 5%.

Assignment rules

Students may discuss assignments, but each student must code up and write up their solutions independently. Students must also indicate on each homework the names of the colleagues they collaborated with and what online resources they used.

Late days policy

Assignments must be submitted by the posted due date. However, we recognize that students may face unusual circumstances and require some flexibility during the semester. For this reason, each student will have a total of 6 free late (calendar) days to use as (s)he prefers, that are counted and tracked at the granularity of days: e.g., 3 hours late is one late day. If an assignment is submitted beyond the late-day budget, the student will receive a 33.33% penalty per day over budget (up to zero). Late days can only be used for assignments, not for the project.

Participation/extra effort bonus

We encourage participation including asking/answering questions in lectures. For this reason, **active participation** in class will be tracked and considered in the final score. Similarly, contributions to the code repository of the course will be considered, previous agreement with the TAs.

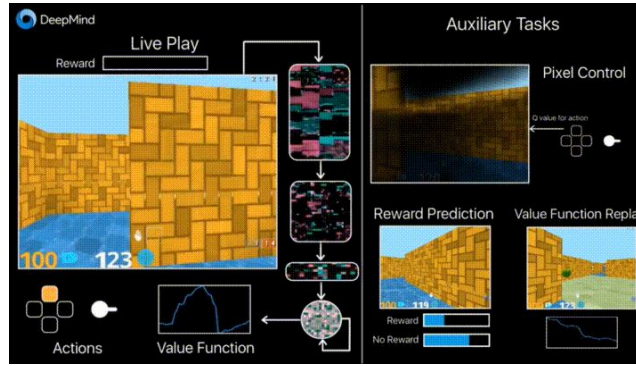
5 meaningful
questions/ans
wers to my
questions



SAPIENZA
UNIVERSITÀ DI ROMA

Reinforcement Learning

- Problem of interest in psychology, AI, neuroscience, etc.
- (Trending) research field
- Set of algorithms and approaches



Reinforcement Learning

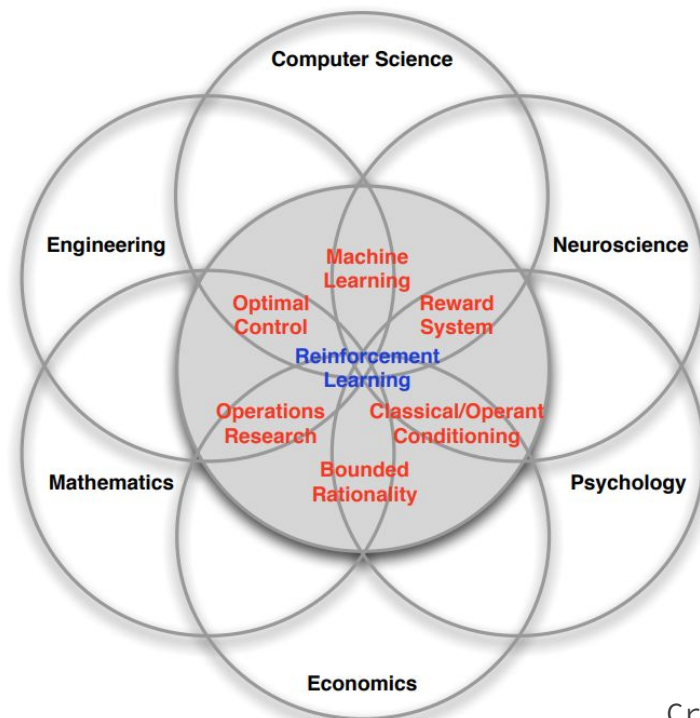
Reinforcement Learning benefits from fruitful interactions with other scientific disciplines

- General machine learning
- Use of function approximators to address the curse-of-dimensionality
- Work towards simple general principles for AI
- Psychology
- Neuroscience



Reinforcement Learning

— — —



Credits: David Silver

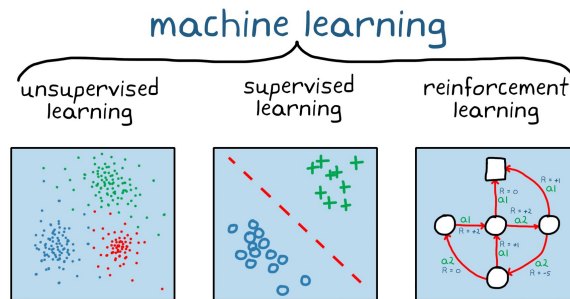


SAPIENZA
UNIVERSITÀ DI ROMA

Reinforcement Learning

Reinforcement Learning is not:

- Supervised learning (learning from a training set of labeled examples)
 - Important but not adequate for interaction
 - Impractical to obtain correct and representative examples of desired behavior
- Unsupervised learning (finding structure hidden in unlabeled data)
 - RL uses a reward signal
 - Useful but does not solve the problem



Credits: MathWorks



Characteristics of Reinforcement Learning

Reinforcement Learning is learning what to do to **maximize a numerical reward signal**

- Learner must discover which actions yield the most reward by trying them
- Actions might affect not only immediate reward, but also subsequent ones
- Trial-and-error
- Delayed rewards



Characteristics of Reinforcement Learning

— — —

Agent has to:

- **Exploit** what it has already been experienced to obtain a reward
- **Explore** in order to make better action selections in the future

Challenges:

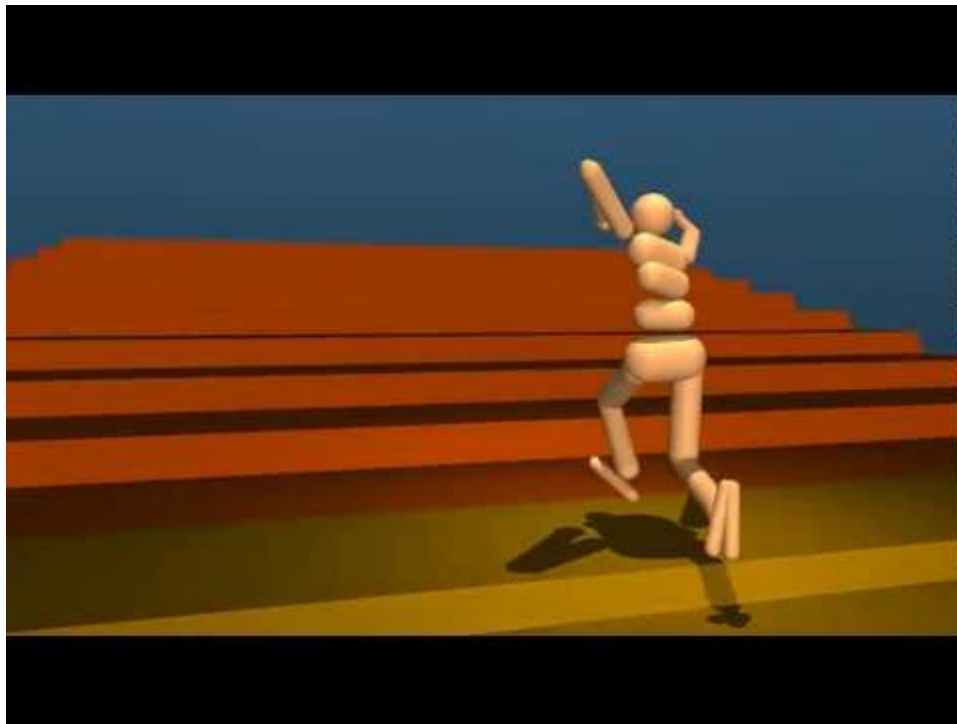
- Exploration-exploitation trade-off (both have to be achieved)
- On stochastic tasks, each action must be tried many times (to correctly estimate expected reward)
- Considers whole problem of a goal-directed agent (not subproblems)
- Significant uncertainty about the environment







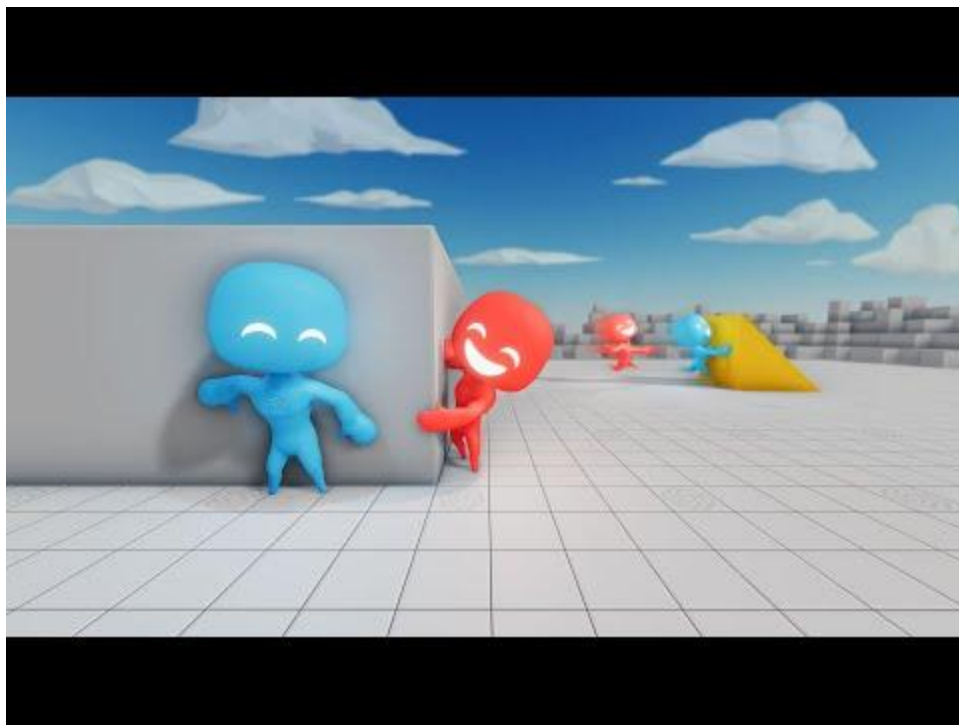
SAPIENZA
UNIVERSITÀ DI ROMA



SAPIENZA
UNIVERSITÀ DI ROMA



SAPIENZA
UNIVERSITÀ DI ROMA



SAPIENZA
UNIVERSITÀ DI ROMA



SAPIENZA
UNIVERSITÀ DI ROMA

THE MAKING OF GRAN TURISMO SOPHY

AI AND THE QUEST FOR SUPERHUMAN SPEED Sony AI



SAPIENZA
UNIVERSITÀ DI ROMA

MARL Example

— — —

Pistonball, a cooperative environment from PettingZoo and its observation space

