

Aprendizado por Reforço

Prof. Domingos Napolitano

Aula 0: Apresentação do Curso

Aprendizado por Reforço

Disciplina: IBM8919 - 6 Semestre 2025.2

Graduação em Ciência de Dados e Inteligência Artificial

Prof. Dr. Domingos M R Napolitano

Quartas Feiras : 9:40 – 11:40

Sextas Feiras : 7:40 – 9:40 h

DOMINGOS.NAPOLITANO@professores.ibmec.edu.br



Bem-vindos à Disciplina

Nesta disciplina, exploraremos o fascinante mundo do Aprendizado por Reforço, uma das áreas mais promissoras e dinâmicas da Inteligência Artificial moderna.

Quem somos

Engenheiro Mecânico com 30 anos de experiência em gestão de projetos em Engenharia, Construção e TI. Mestre e Doutor em Informática.

Nossa jornada

Ao longo do semestre, construiremos uma base sólida desde os conceitos fundamentais até aplicações avançadas, capacitando-os a desenvolver soluções práticas utilizando algoritmos de Aprendizado por Reforço.

Objetivo principal

Capacitar você a compreender e aplicar os fundamentos do Aprendizado por Reforço, modelando problemas sequenciais com agentes autônomos que aprendem a interagir com o ambiente.



O que é Aprendizado por Reforço?

O Aprendizado por Reforço (RL) é um paradigma de aprendizado de máquina onde agentes autônomos aprendem a tomar decisões sequenciais em ambientes incertos, através de interações baseadas em recompensas.

Diferentemente de outros paradigmas de aprendizado de máquina:

- **Aprendizado Supervisionado:** Aprende com exemplos rotulados
- **Aprendizado Não-Supervisionado:** Encontra padrões em dados não rotulados
- **Aprendizado por Reforço:** Aprende por tentativa e erro com feedback do ambiente

O agente desenvolve uma política para maximizar a recompensa cumulativa ao longo do tempo, sem supervisão explícita.



O que é Aprendizado por Reforço?

O Aprendizado por Reforço (RL) é um paradigma de aprendizado de máquina onde agentes autônomos aprendem a tomar decisões sequenciais em ambientes incertos, através de interações baseadas em recompensas.



O agente desenvolve uma política para maximizar a recompensa cumulativa ao longo do tempo, sem supervisão explícita.

Por que estudar Aprendizado por Reforço?



Inspirado na natureza

O RL se inspira em como humanos e animais aprendem naturalmente, tomando decisões baseadas em experiências passadas e adaptando comportamentos para maximizar recompensas.



Autonomia

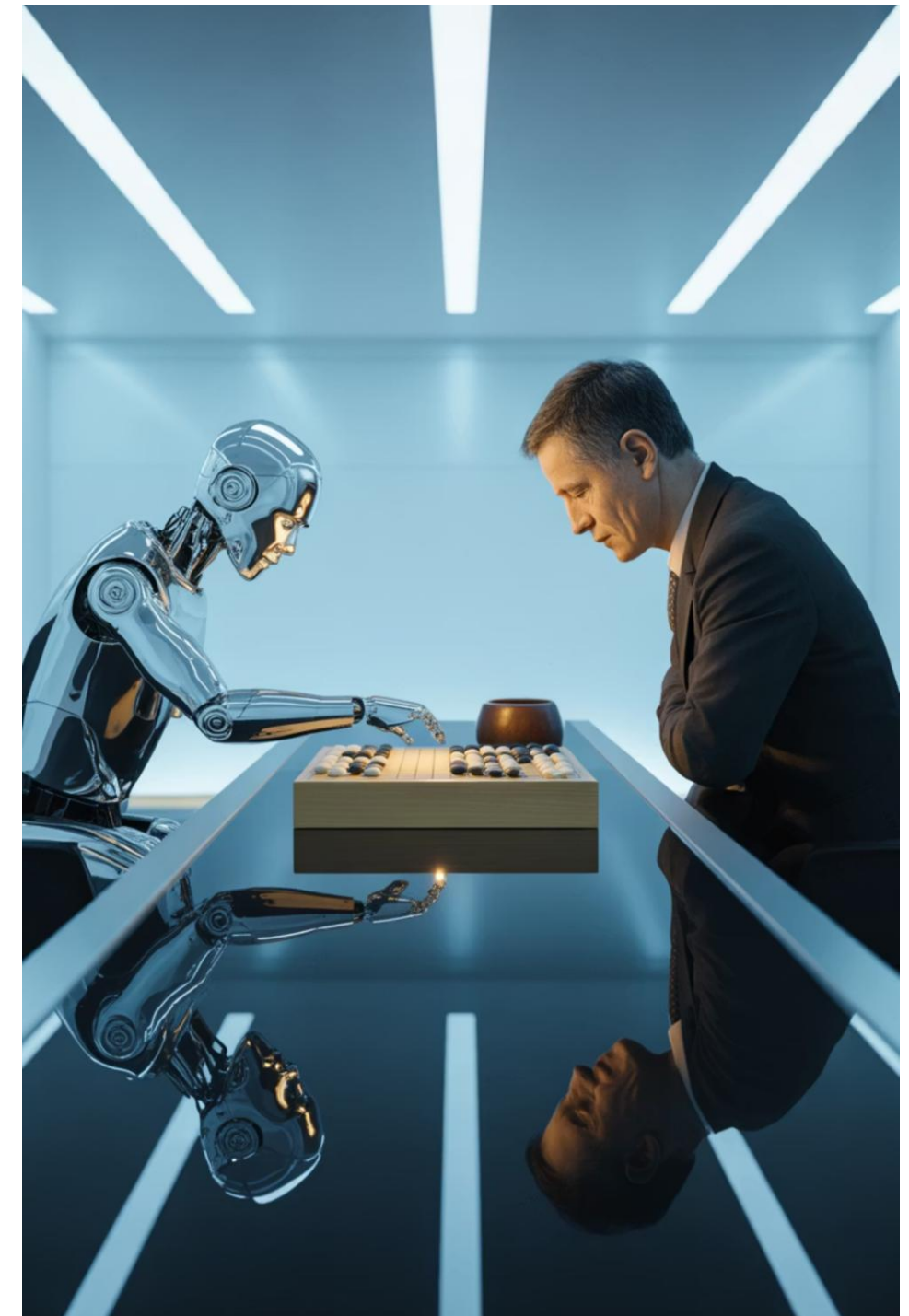
Capacita sistemas a tomarem decisões independentes em ambientes complexos e dinâmicos, essencial para robótica avançada e sistemas autônomos.



Superando humanos

Algoritmos de RL têm alcançado resultados impressionantes, superando campeões humanos em jogos como xadrez, Go e StarCraft II.

O aprendizado por reforço representa uma fronteira fundamental da IA, permitindo que sistemas aprendam a realizar tarefas complexas que seriam difíceis de programar explicitamente. As habilidades que você desenvolverá nesta disciplina serão aplicáveis em diversos domínios de ponta na indústria e pesquisa.



Aplicações do Aprendizado por Reforço



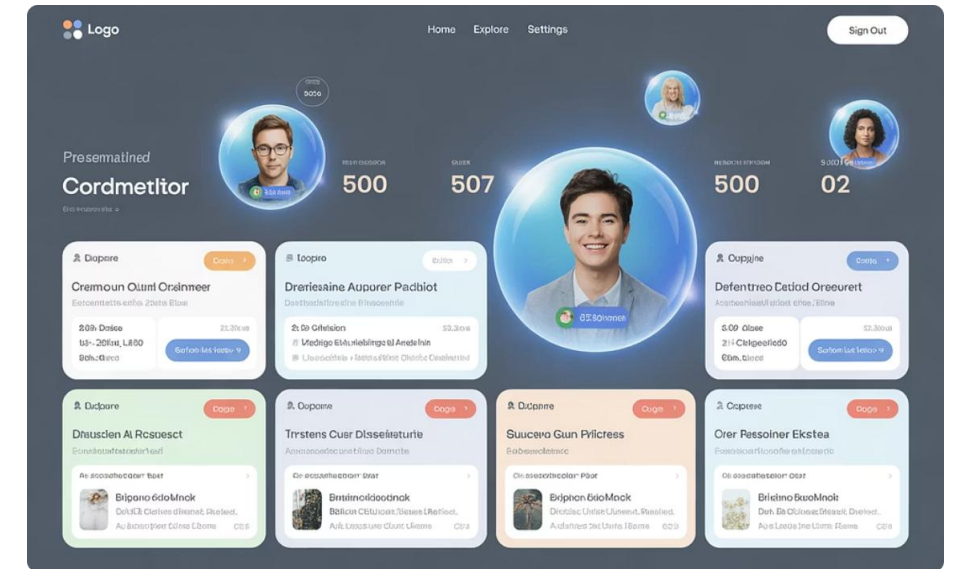
Veículos Autônomos

Desenvolvimento de políticas de navegação, estacionamento e resposta a situações imprevistas em tempo real.



Robótica Industrial

Controle de braços robóticos para manipulação de objetos, montagem de peças e otimização de processos fabris.



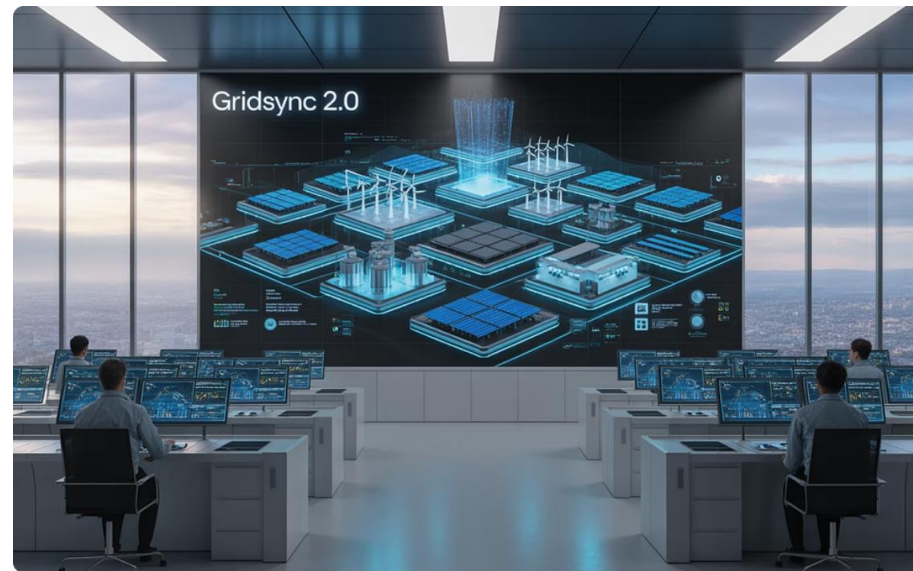
Sistemas de Recomendação

Personalização de conteúdo, otimização de interfaces e estratégias de engajamento adaptativas.



Mercado Financeiro

Desenvolvimento de estratégias de negociação, otimização de portfólios e gerenciamento de riscos.



Gestão Energética

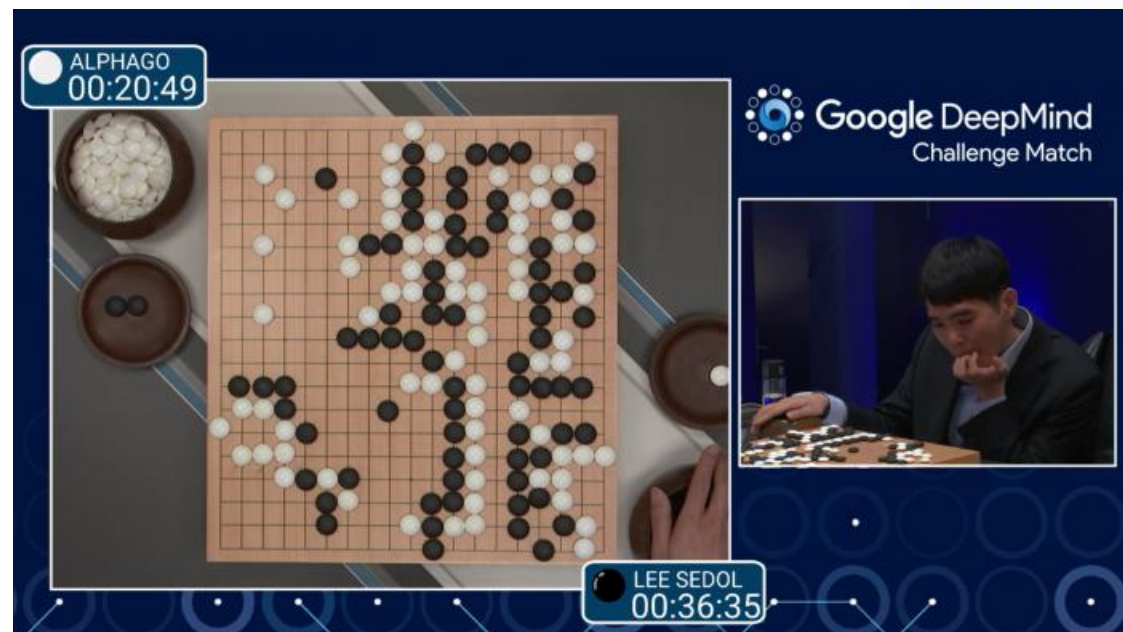
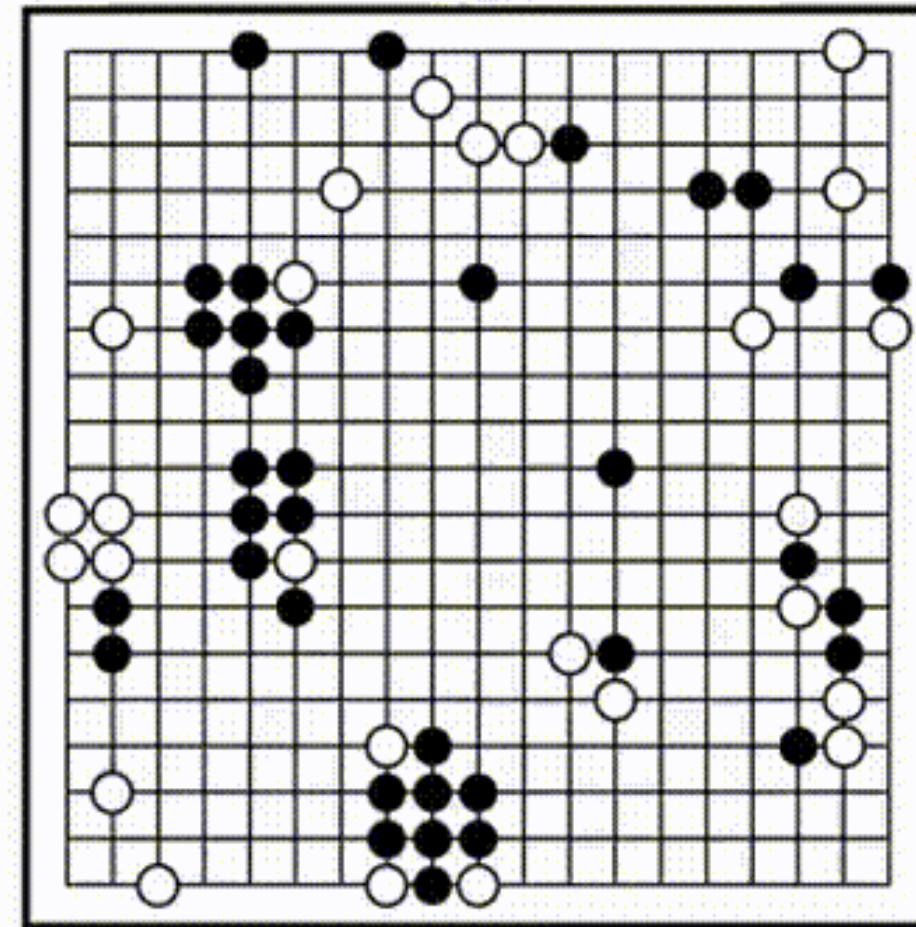
Otimização do consumo de energia em redes inteligentes e controle de sistemas de geração distribuída.



Saúde

Diagnóstico médico, personalização de tratamentos e assistência em procedimentos cirúrgicos.

Aplicações do Aprendizado por Reforço



Objetivos da Disciplina

01

Compreender os fundamentos

Introduzir os conceitos básicos do Aprendizado por Reforço, incluindo a formulação de agentes, recompensas, estados e políticas que formam a base deste paradigma.

02

Dominar os modelos matemáticos

Apresentar os fundamentos matemáticos de Modelos de Decisão de Markov (MDPs) e sua aplicação na modelagem de ambientes estocásticos.

03

Implementar algoritmos clássicos

Explorar e implementar algoritmos tradicionais como Programação Dinâmica, Métodos de Monte Carlo e Diferenças Temporais para solução de problemas sequenciais.

04

Aplicar métodos baseados em valor

Utilizar algoritmos como Q-learning e SARSA para tomada de decisão em ambientes simulados e reais, comparando suas características e desempenho.

05

Explorar tópicos avançados

Discutir métodos baseados em política, ambientes contínuos, aprendizado por reforço profundo e aplicações práticas em diversas áreas.

Conteúdo Programático



Fundamentos do RL

Introdução, conceitos básicos (agente, ambiente, política, função valor, recompensa) e aplicações práticas.



Processos de Decisão

Cadeias de Markov, Modelos de Decisão de Markov (MDPs), definição formal e propriedades do processo de decisão.



Métodos de Solução

Programação Dinâmica, Equações de Bellman, Value Iteration, Policy Iteration, convergência e complexidade.



Métodos de Amostragem

Monte Carlo, estratégias de exploração vs. aproveitamento, Diferenças Temporais (TD Learning), TD(0), TD(λ).



Algoritmos de Valor

Aprendizado off-policy e on-policy, Q-Learning, SARSA e variações destes algoritmos.



Tópicos Avançados

Métodos baseados em política, espaços contínuos, aprendizado por reforço profundo, múltiplos agentes e aplicações.

A disciplina está estruturada de forma progressiva, começando com os fundamentos teóricos e avançando gradualmente para técnicas mais sofisticadas e aplicações práticas.

Metodologia de Ensino

Aulas Teórico-Práticas

Exposições dialogadas combinadas com resolução de exercícios e implementações práticas, utilizando slides, quadros digitais e notebooks interativos.

Exercícios e Projetos

Listas de exercícios teóricos e práticos, atividades individuais e em grupo, com desenvolvimento incremental de agentes e análise de desempenho.

Ambientes de Simulação

Uso de plataformas como Gymnasium (antigo OpenAI Gym), RLGlue e Google Colab para experimentação e implementação de algoritmos em ambientes controlados.

Discussões e Aprofundamento

Análise de artigos científicos e discussão de tópicos avançados, promovendo o pensamento crítico e a exploração de subáreas emergentes.

O tempo estimado para estudo extraclasse é de no mínimo 3 horas semanais para revisão de conteúdo e realização de tarefas. É fundamental manter a regularidade nas atividades práticas para consolidar os conceitos teóricos.

Pré-requisitos e Conhecimentos Recomendados

Conhecimentos Prévios Importantes



Programação em Python

Estruturas de dados, funções, classes, bibliotecas científicas (NumPy, Pandas) e manipulação de arquivos.



Estatística

Probabilidade, variáveis aleatórias, distribuições, esperança matemática e processos estocásticos.



Inferência Estatística

Estimação, testes de hipóteses, intervalos de confiança e métodos de amostragem.

Disciplinas Relacionadas

Conhecimentos de outras disciplinas do mesmo semestre que compõem o grupo de Aprendizado de Máquina:

- **Aprendizado de Máquina:** Técnicas supervisionadas e não-supervisionadas que complementam o Aprendizado por Reforço
- **Projeto de Machine Learning:** Metodologias de desenvolvimento que serão aplicadas nos projetos da disciplina
- **Processamento de Linguagem Natural:** Técnicas que podem ser integradas com RL para sistemas mais complexos

Estas disciplinas fornecem uma visão holística do campo de Inteligência Artificial e como o Aprendizado por Reforço se integra ao ecossistema de técnicas de IA.

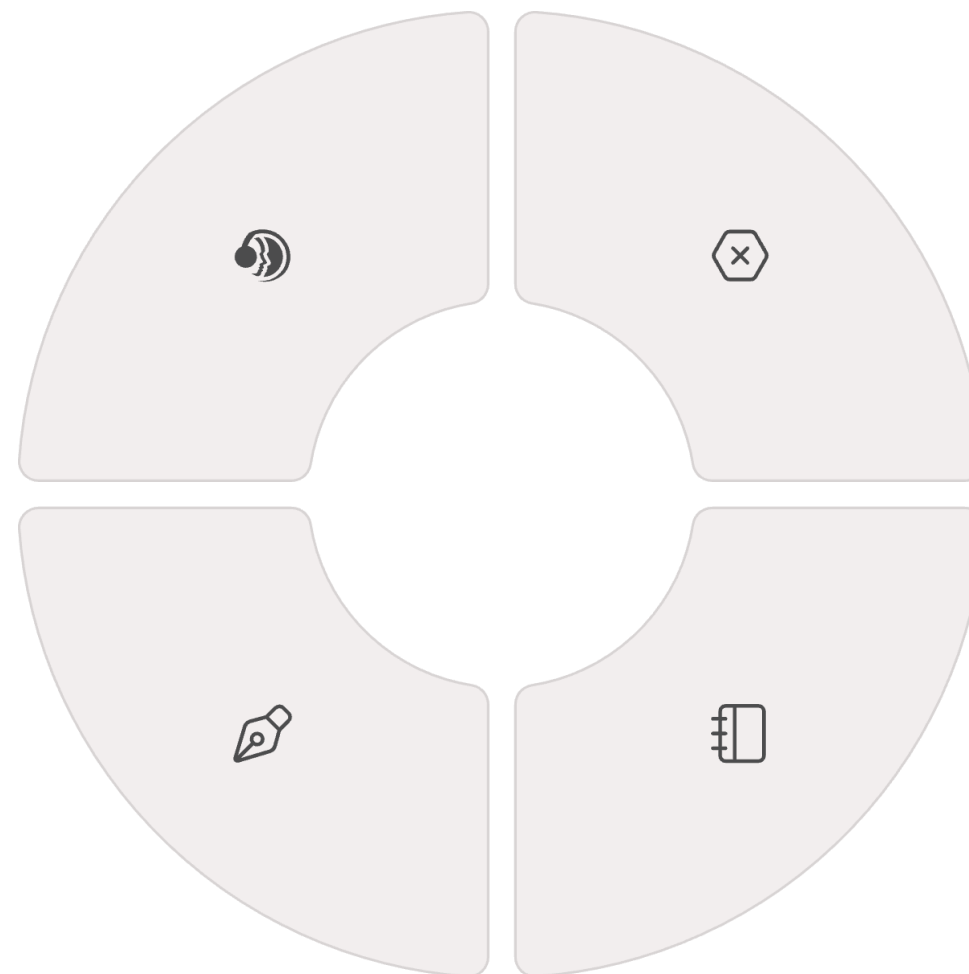
Sistema de Avaliação

AP1 (40%)

Prova individual teórico-prática abrangendo todo o conteúdo ministrado até a semana anterior à avaliação.

AS (Opcional)

Avaliação suplementar facultativa que substituirá a menor nota entre AP1 e AP2, abrangendo todo o conteúdo do semestre.



AP2 (40%)

Prova individual teórico-prática abrangendo todo o conteúdo ministrado até a semana anterior à avaliação.

AC (20%)

Trabalhos individuais ou em grupo realizados com auxílio de Jupyter Notebooks, envolvendo conceitos teóricos e práticos.

Fórmula para cálculo da Média Final

$$\text{Média Final} = (0,4 \times \text{AP1}) + (0,4 \times \text{AP2}) + (0,2 \times \text{AC})$$

Será considerado aprovado o aluno que obtiver Média Final igual ou superior a 7 (sete) e frequência mínima de 75% nas aulas.

Ferramentas e Ambientes de Aprendizado

Gymnasium

Plataforma para desenvolvimento e avaliação de algoritmos de Aprendizado por Reforço, com ambientes padronizados para testar e comparar diferentes abordagens.



Google Colab

Ambiente de notebook Jupyter baseado em nuvem que permite escrever e executar código Python, ideal para implementação e experimentação com algoritmos de RL.



Bibliografia

Bibliografia Básica

FACELI, Katti et al. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. Rio de Janeiro: LTC, 2021.

NETTO, Amilcar; MACIEL, Francisco. Python para Data Science e Machine Learning Descomplicado. Rio de Janeiro: Alta Books, 2021.

SÁ, Y.V.A. Desenvolvimento de aplicações IA: Robótica, Imagem e Visão Computacional. São Paulo: Platos, 2021.

Bibliografia Complementar

SUTTON, Richard; BARTO, Andrew. Reinforcement Learning: An Introduction. Cambridge: MIT Press, 2018.

RUSSELL, Stuart J.; NORVIG, Pete. Inteligência Artificial: Uma Abordagem Moderna. Rio de Janeiro: GEN LTC, 2022.

SILVA, Fabrício M. et al. Inteligência artificial. Porto Alegre: SAGAH, 2018.

Além dos livros listados, compartilharemos ao longo do curso artigos científicos recentes, tutoriais online e recursos complementares para aprofundamento em tópicos específicos.



O Poder do Aprendizado por Reforço Profundo



Atari Games (2013)

A DeepMind revolucionou o campo com o DQN (Deep Q-Network), que aprendeu a jogar jogos Atari apenas a partir dos pixels da tela, superando humanos em vários jogos.



AlphaGo (2016)

Derrotou o campeão mundial de Go, Lee Sedol, um feito considerado impossível na época devido à complexidade do jogo (mais configurações de tabuleiro que átomos no universo).



Dexterous Manipulation (2018)

Sistemas que aprendem a manipular objetos com destreza semelhante à humana, desenvolvidos pela OpenAI e outros grupos de pesquisa.



AlphaFold (2020)

Revolucionou a biologia estrutural, prevendo com precisão a estrutura 3D de proteínas a partir da sequência de aminoácidos, um problema fundamental da ciência.

Estes avanços exemplificam como o Aprendizado por Reforço, especialmente quando combinado com redes neurais profundas, tem transformado diferentes áreas e resolvido problemas anteriormente considerados intratáveis.

Desafios Atuais no Aprendizado por Reforço



Eficiência de Amostra

Algoritmos tradicionais de RL necessitam de muitas interações com o ambiente para aprender políticas efetivas, tornando o treinamento demorado e computacionalmente custoso.



Exploração vs. Aproveitamento

Encontrar o equilíbrio ideal entre explorar novos estados e explorar conhecimento adquirido continua sendo um desafio fundamental.



Transferência de Conhecimento

Desenvolver agentes capazes de transferir aprendizado entre tarefas relacionadas, reduzindo a necessidade de treinamento do zero para cada novo problema.



Aplicações no Mundo Real

Transpor o sucesso dos algoritmos em ambientes simulados para aplicações práticas em sistemas reais, com suas incertezas e restrições físicas.



Interpretabilidade

Compreender e explicar as decisões tomadas pelos agentes, especialmente quando utilizam modelos complexos como redes neurais profundas.



Segurança e Robustez

Garantir que os agentes de RL atuem de forma segura e previsível, evitando comportamentos indesejados ou exploração de falhas nos sistemas.

Plano de Aula para o semestre 2025.2 - Agosto

ibmec.br

AULAS		Disiplina: IBM8919Aprendizado por Reforço
PLANO DE AULAS		
Dia/Mês	Conteúdo da Aula	Atividades de Apoio
	Apresentação da disciplina, plano de ensino, material de aula e modelos avaliativos	
06/ago	Revisão Parte 1 : Revisão de Python para Aprendizado por Reforço e Ciência de Dados (Demonstração prática de Códigos em Jupyter Notebook a ser executada pelos alunos como tarefa para casa) Revisão Parte 2 : Revisão de Probabilidades para Aprendizado por Reforço (Demonstração prática de Códigos em Jupyter Notebook a ser executada pelos alunos como tarefa para casa)	Slides Aula 0 Jupyter Notebooks para Revisão
08/ago	Introdução ao Aprendizado por Reforço aplicações e ferramentas. Conceitos básicos de aprendizado por reforço: agente, ambiente, política, função valor, função de recompensa.	Pré Work : Aula 1 Slides Aula 1
13/ago	O problema do k armed Bandits: Intuição e Conceituação (Dinâmica: de Estudo Médico)	Pré Work : Aula 2 Slides Aula 2
15/ago	O problema do k armed Bandits: Implementação em Python	Jupyter Notebook Aula 3 Slides Aula 3
20/ago	O problema da Exploração e Aproveitamento O problema do k armed Bandits: ϵ-greedy (Implementação em Python)	Jupyter Notebook Aula 4 Slides Aula 4 Lista de Exercícios 1
22/ago	Cadeias de markov: conceitos básicos e definição.	Pré Work : Aula 5 Slides Aula 5
27/ago	Cadeias de markov: Implementação em Python Proposição AC1	Pré Work Aula 6 Jupyter Notebook Aula 6 Slides Aula 6
29/ago	Processo de Decisão de Markov (MDP) Intuição e Conceituação (Dinâmica Gridworld)	Slide Aula 7

Plano de Aula para o semestre 2025.2 - Setembro

ibmec.br

AULAS	PLANO DE AULAS	Disciplina: IBM8919 Aprendizado por Reforço
Dia/Mês	Conteúdo da Aula	Atividades de Apoio
03/set	Processo de Decisão de Markov (MDP) Equações de Bellman	Slides Aula 8
05/set	Processo de Decisão de Markov (MDP) Implementação de Soluções Analíticas	Slides Aula 9 Jupyter Notebook Aula 9 Lista de Exercícios 2
10/set	Apresnetação AC1 Programação Dinâmica: Intuição e Conceituação	Slides Aula 10
12/set	Programação Dinâmica: Implementação em Python	Slides Aula 11 Jupyter Notebook Aula 11 Lista de Exercícios 3
17/set	Programação Dinâmica: Implementação em Python	Slides Aula 12 Jupyter Notebook Aula 12
19/set	Correção Listas de Exercícios 1, 2 e 3	Gabaritos Lista de Exercícios 1, 2, e 3
24/set	Revisão para AP 1 Resolução AC 2	Todo o material apresnetado até o Momento
26/set	Período Avaliação Parcial AP 1	
01/out	Período Avaliação Parcial AP 1	

Plano de Aula para o semestre 2025.2 - Setembro

ibmec.br

AULAS		PLANO DE AULAS	Disciplina: IBM8919 Aprendizado por Reforço
Dia/Mês	Conteúdo da Aula		Atividades de Apoio
03/set	Processo de Decisão de Markov (MDP)	Equações de Bellman	Slides Aula 8
05/set	Processo de Decisão de Markov (MDP)	Implementação de Soluções Análíticas	Slides Aula 9 Jupyter Notebook Aula 9 Lista de Exercícios 2
10/set	Apresnetação AC1 Programação Dinâmica: Intuição e Conceituação		Slides Aula 10
12/set	Programação Dinâmica: Implementação em Python		Slides Aula 11 Jupyter Notebook Aula 11 Lista de Exercícios 3
17/set	Programação Dinâmica: Implementação em Python		Slides Aula 12 Jupyter Notebook Aula 12
19/set	Correção Listas de Exercícios 1, 2 e 3		Gabaritos Lista de Exercícios 1, 2, e 3
24/set	Revisão para AP 1 Resolução AC 2		Todo o material apresnetado até o Momento
26/set	Período Avaliação Parcial AP 1		
01/out	Período Avaliação Parcial AP 1		

Plano de Aula para o semestre 2025.2 - Outubro

ibmec.br

AULAS	PLANO DE AULAS	Disciplina: IBM8919Aprendizado por Reforço
Dia/Mês	Conteúdo da Aula	Atividades de Apoio
03/out	Introdução ao Método de Monte Carlo	Slides Aula 13 Jupyter Notebook Aula 13
08/out	Métodos de Monte Carlo em Aprendizado por Reforço: Predição, Controle, On Policy OFF Policy	Slides Aula 14
10/out	Métodos de Monte Carlo em Aprendizado por Reforço: Implementação em Python	Slides Aula 15 Jupyter Notebook Aula 15 Lista de Exercícios 4
15/out	Métodos de Diferença Temporal Introdução e Conceitos	Slides Aula 16
17/out	Métodos de Diferença Temporal : Q Learning	Slides Aula 17 Jupyter Notebook Aula 17
22/out	Métodos de Diferença Temporal : SARSA	Slides Aula 18 Jupyter Notebook Aula 18
24/out	Métodos de Diferença Temporal : Expected SARSA Proposição AC 3	Slides Aula 19 Jupyter Notebook Aula 19 Lista de Exercícios 5
29/out	Tópicos Avançados: <ul style="list-style-type: none">• Métodos baseados em política (Policy Gradient);• Aprendizado em espaço contínuo;• Aprendizado com múltiplos agentes;• Aplicações em jogos, robótica e sistemas autônomos..	Slides Aula 20
31/out	Tópicos Avançados: <ul style="list-style-type: none">• Aprendizado por reforço profundo (DQN);	Slides Aula 21

Plano de Aula para o semestre 2025.2 - Novembro

AULAS		PLANO DE AULAS	Disciplina: IBM8919 Aprendizado por Reforço
Dia/Mês		Conteúdo da Aula	Atividades de Apoio
05/nov		Apresnetação AC3	Atividade Avaliativa
07/nov		Resolução Lista de Exercícios 4 e 5	Lista de Exercícios 4 e 5
12/nov		Revisão para AP2 - Resolução AC4	Todo o material apresnetado até o Momento
14/nov		Período de Aplicação a AP2	
19/nov		Período de Aplicação a AP2	
21/nov		Recesso	
26/nov		Período de Aplicação da AS	
28/nov		Período de Aplicação da AS	
03/dez		Período de Aplicação da AS	

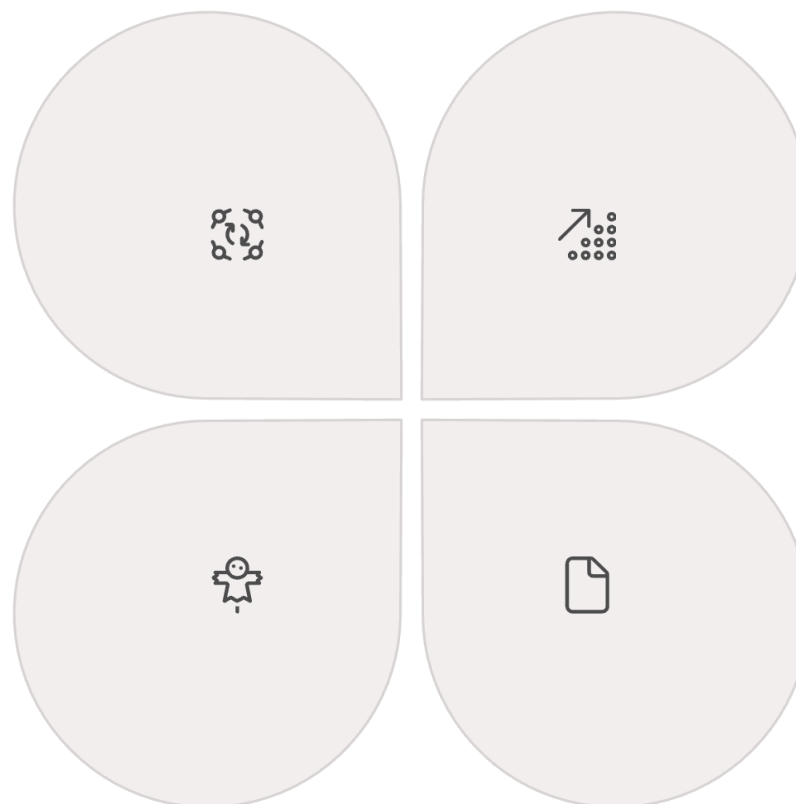
Vamos Começar Nossa Jornada!

Comunidade Ativa

O campo do Aprendizado por Reforço possui uma comunidade vibrante de pesquisadores e praticantes, com conferências, workshops e fóruns dedicados.

Oportunidades Profissionais

Habilidades em Aprendizado por Reforço são altamente valorizadas no mercado, abrindo portas para carreiras em pesquisa, desenvolvimento e inovação.



Crescimento Acelerado

A área está em rápida evolução, com novos algoritmos e aplicações surgindo constantemente, oferecendo inúmeras oportunidades para contribuições.

Interdisciplinaridade

O RL conecta-se com diversas áreas como neurociência, psicologia, economia, controle ótimo e matemática aplicada.

"A inteligência é a capacidade de se adaptar à mudança." - Stephen Hawking

Estamos animados para explorar com vocês os fundamentos e as fronteiras do Aprendizado por Reforço neste semestre! Vamos juntos desenvolver as habilidades necessárias para criar os sistemas inteligentes do futuro.

Repositório para acesso aos notebooks e slides!

https://github.com/DeepFluxion/IBMEC_Aprendizado_Refor-o/

IBMEC_Aprendizado_Refor-o / Aula_0_Revisão de Probabilidades.ipynb

DeepFluxion Criado usando o Colab 97f07e2 · 8 minutes ago History

Preview Code Blame 774 lines (774 loc) · 57.5 KB Raw Copy Download Edit

Open in Colab

Quebra-Gelo: Probabilidades para Aprendizado por Reforço

Objetivo da Aula

Relembrar conceitos básicos de probabilidade de forma prática e divertida, preparando o terreno para entender como a **incerteza** e a **aleatoriedade** são fundamentais no Aprendizado por Reforço.

Warm-up: O Jogo dos Dados

Pergunta Inicial: "Qual resultado é mais provável?"

Vamos começar com uma pergunta simples para ativar o pensamento probabilístico:

Ao lançar dois dados, qual soma tem maior probabilidade de ocorrer?



IBMEC.BR

 /IBMEC

 IBMEC

 @IBMEC_OFICIAL

 @IBMEC

 **ibmec**