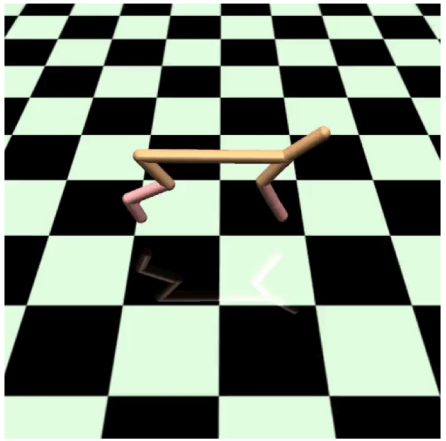
Bem vindo(a) ao Curso

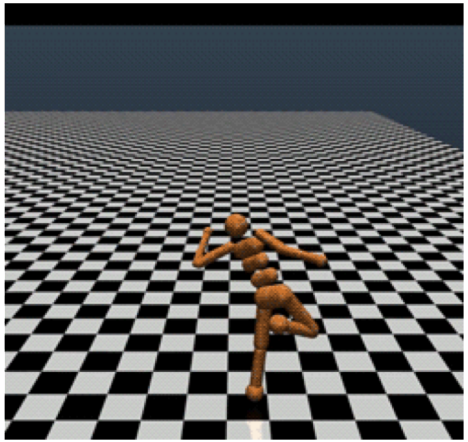
Bem-vindo(a) ao curso!

Half-Cheetah



Fonte: https://github.com/alexis-jacq/numpy_ARS

Humanoid



Fonte: www.argmin.net

Bem-vindo(a) ao curso!

Entender
os conceitos
principais

Intuição
+
Prática

Aprender
como aplicar
na prática

Pré-requisitos

- Lógica de programação
- Programação básica em Python
- Instalação de softwares
- Básico sobre orientação a objetos

- Nível do curso: todos os níveis
- Dica: aumentar a velocidade do player
- Avaliação do curso!

Conteúdo

Conteúdo

O que você aprenderá nesta seção:

- Visão geral da ARS (Augmented Random Search)
- Como um perceptron funciona?
- Maximização de Recompensas
- Método de Diferenças Finitas
- Métodos básicos x ARS
- ARS x Outras IAs

Visão Geral da ARS (Augmented Random Search)

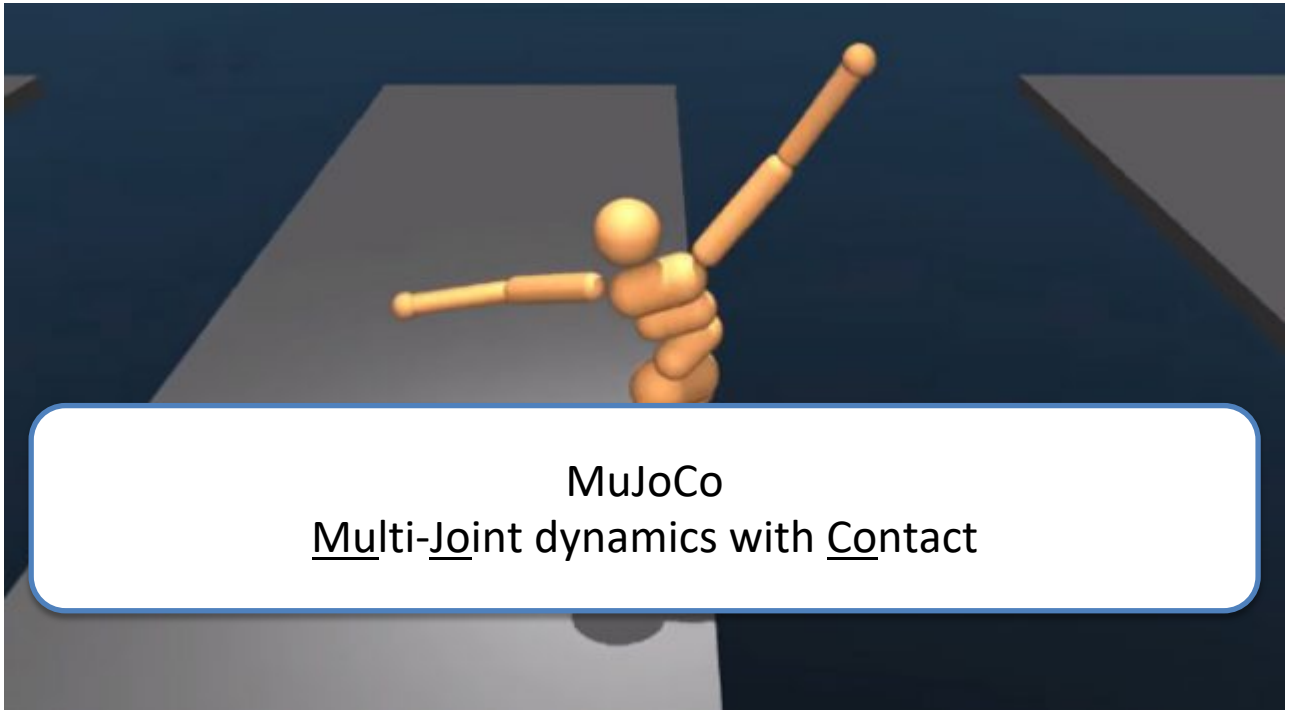
Visão Geral da ARS

Horia Mania

Aurelia Guy

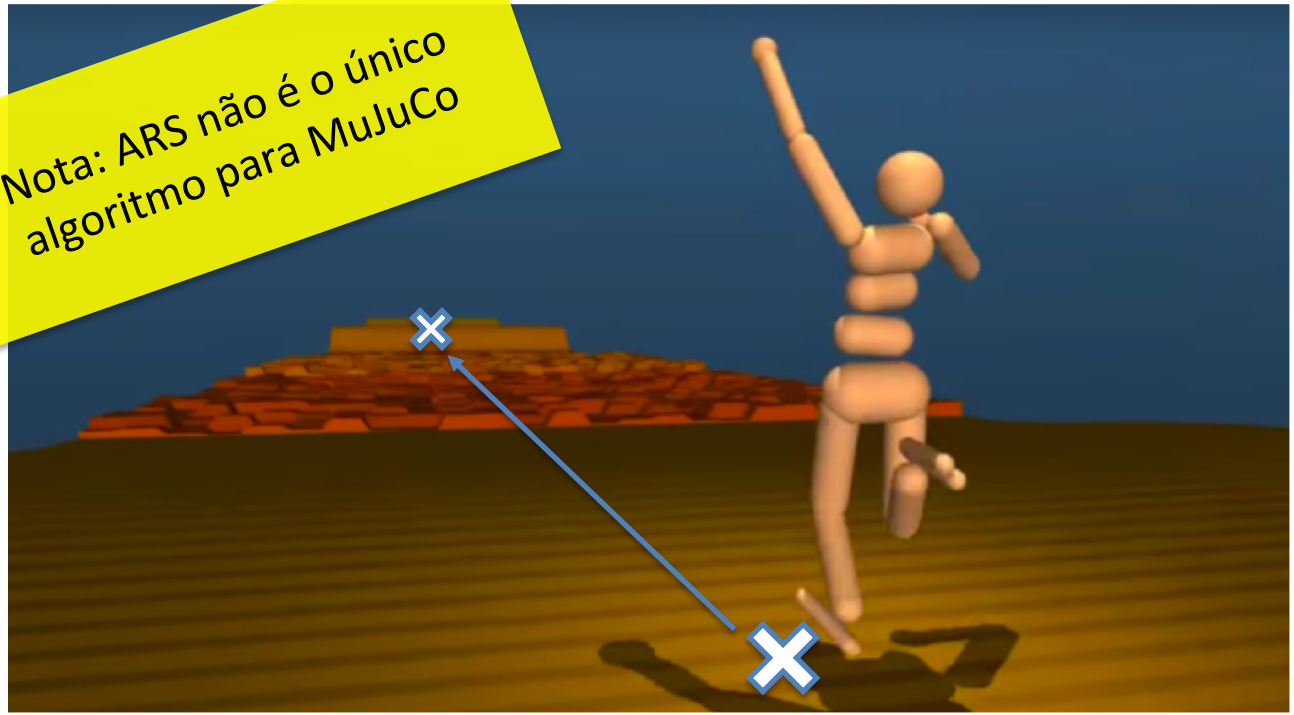
Fontes: people.eecs.berkeley.edu & linkedin.com

Visão Geral da ARS

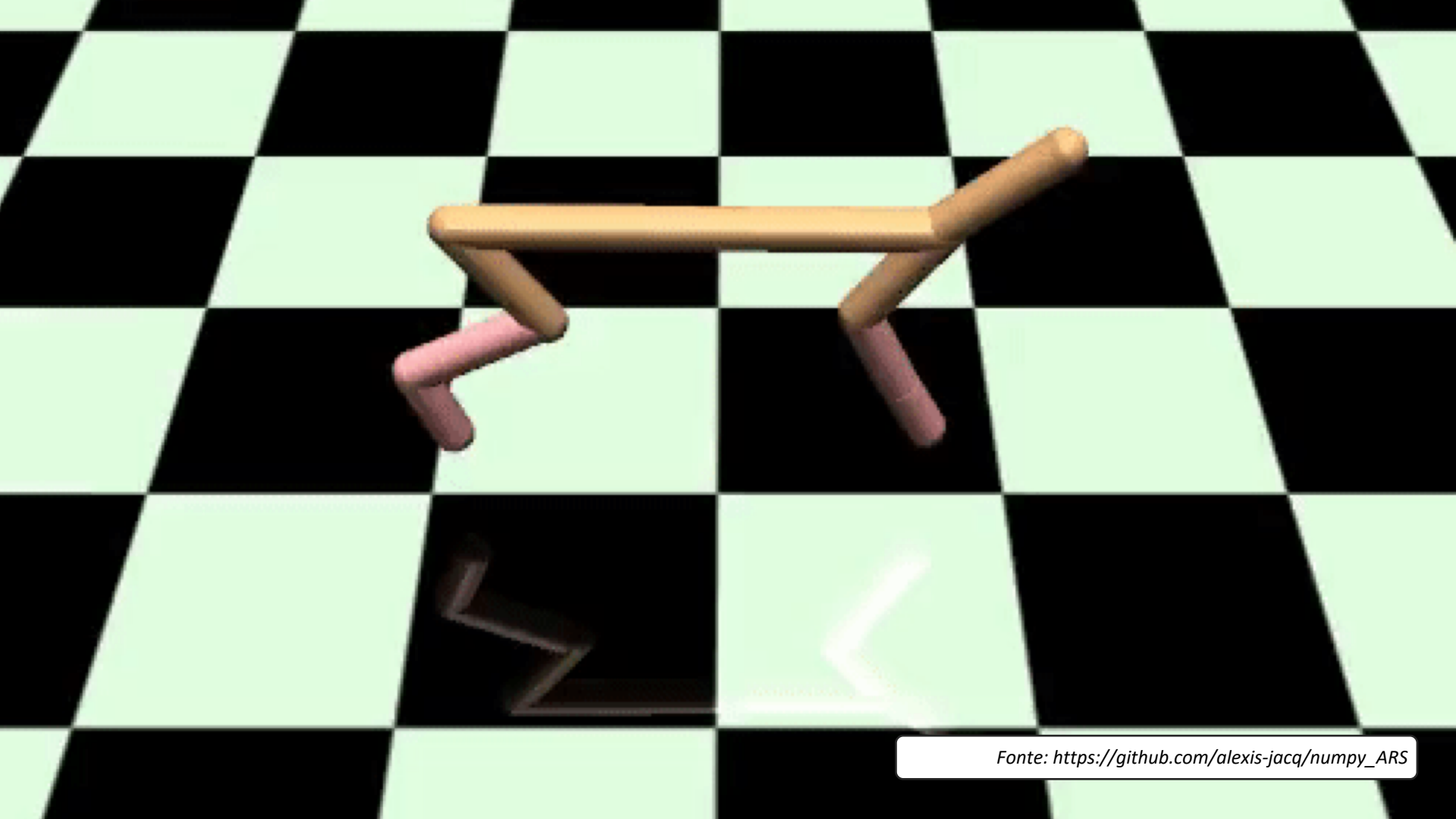


Source: www.extremetech.com

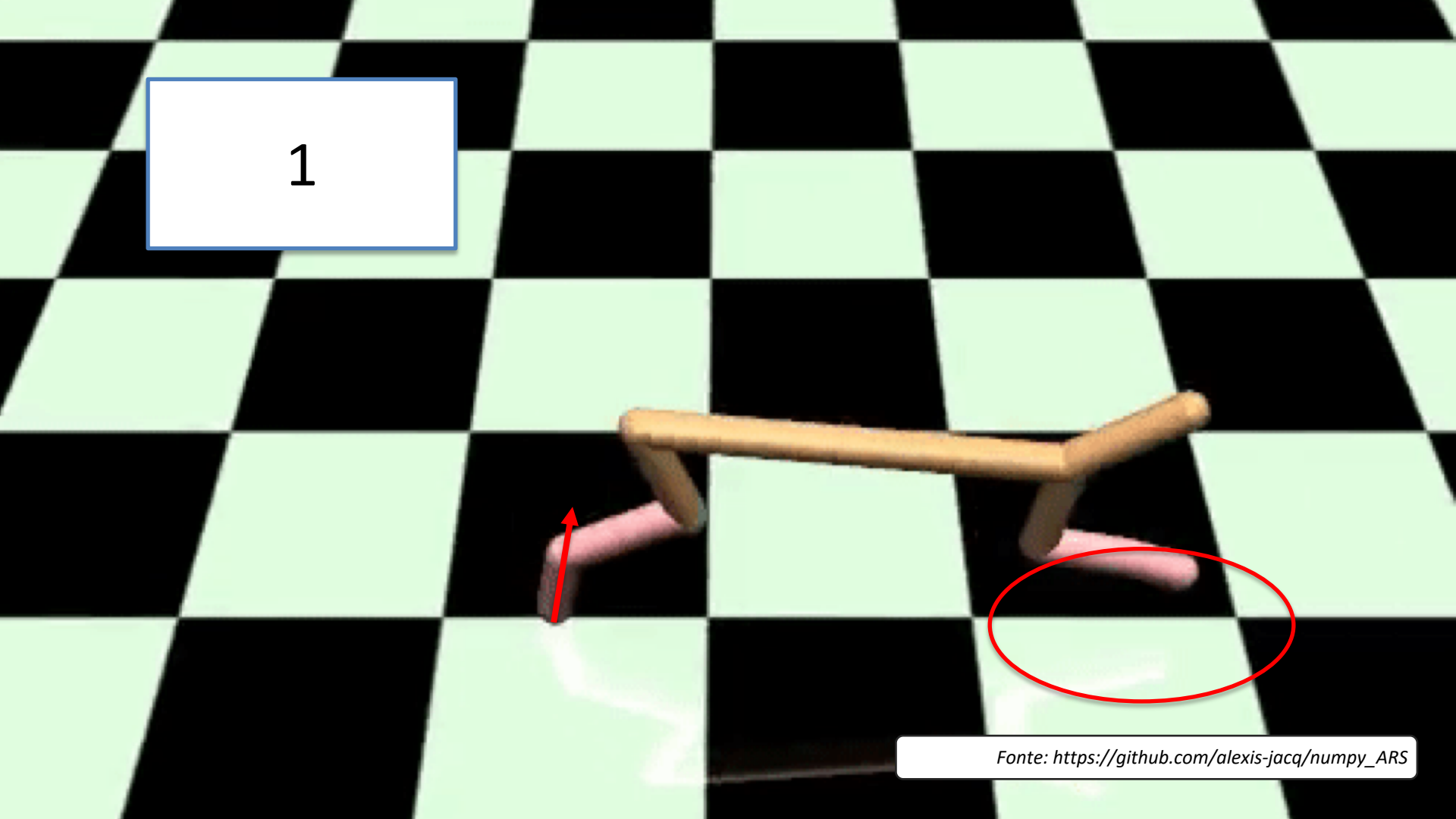
Visão Geral da ARS



Fonte: www.bt.com

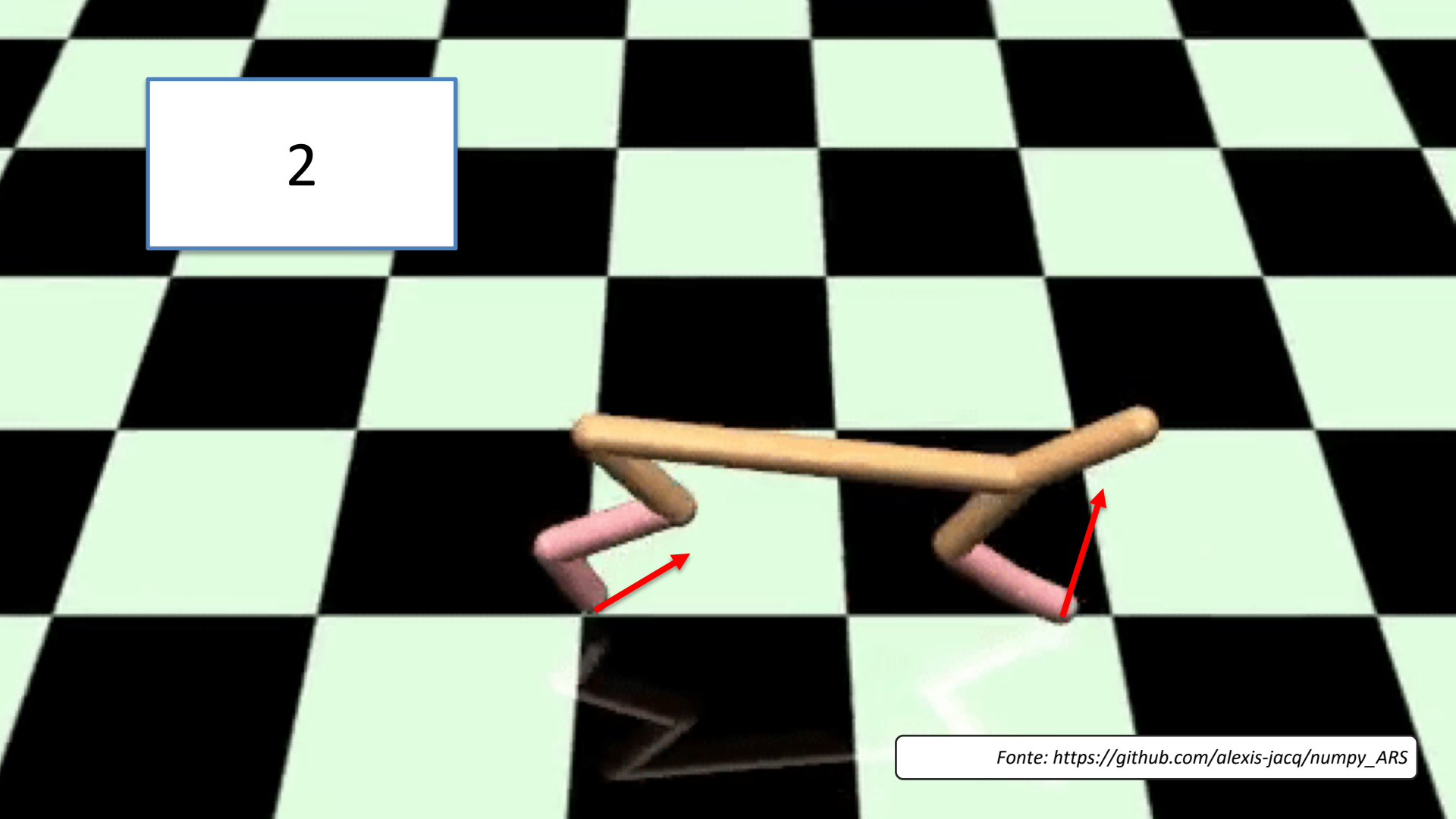


Fonte: https://github.com/alexis-jacq/numpy_ARS



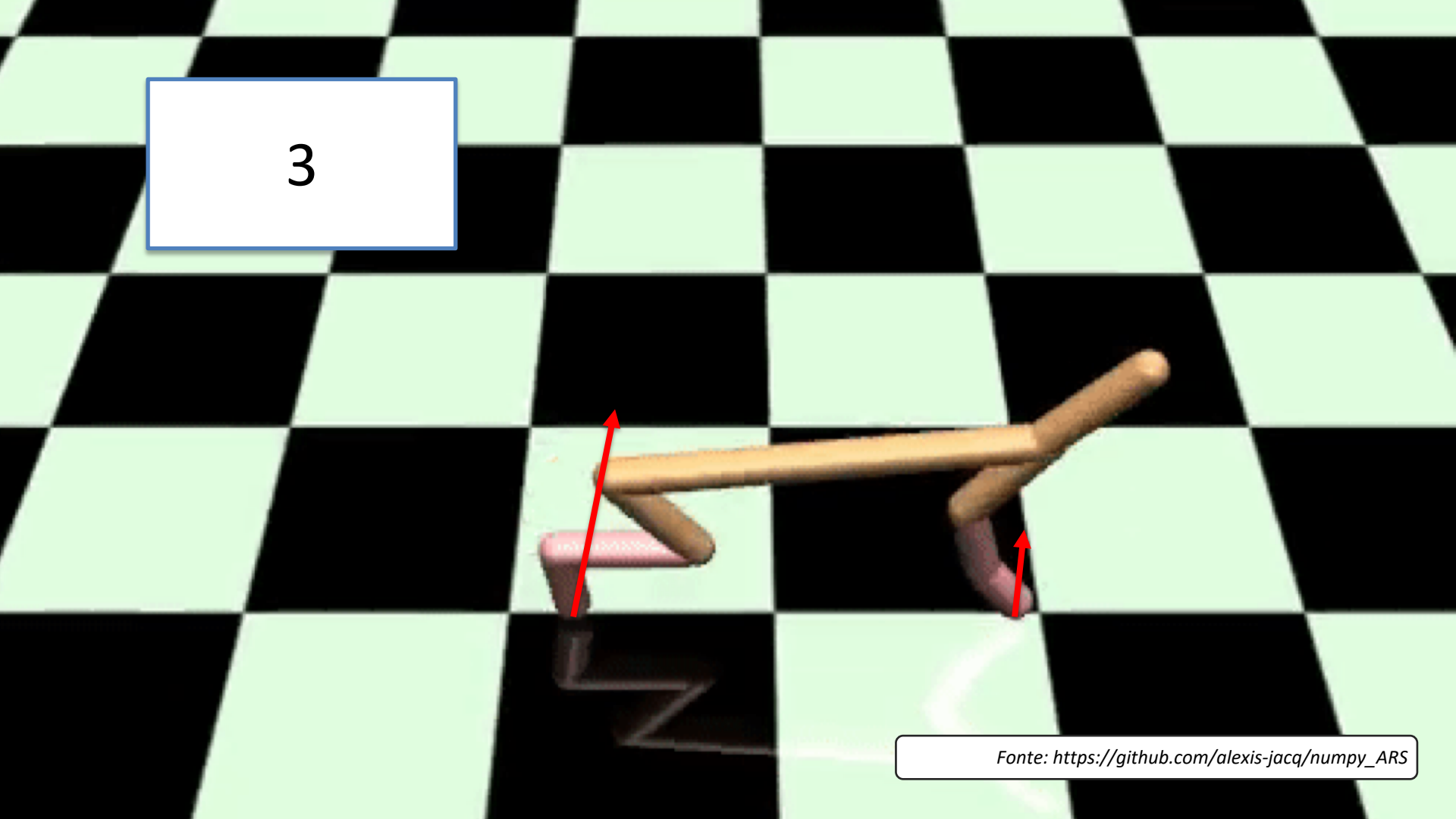
Fonte: https://github.com/alexis-jacq/numpy_ARS

2



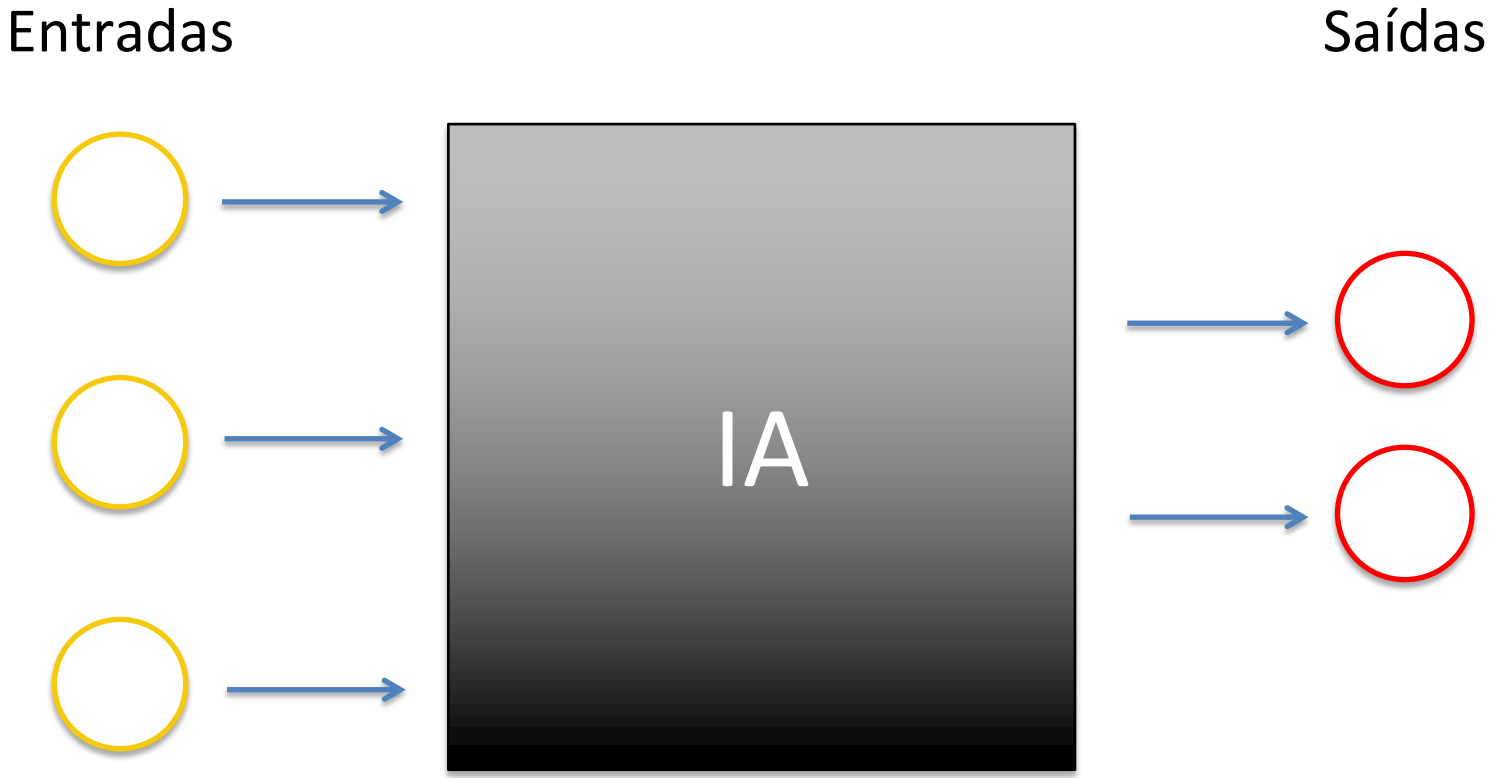
Fonte: https://github.com/alexis-jacq/numpy_ARS

3



Fonte: https://github.com/alexis-jacq/numpy_ARS

Visão Geral da ARS

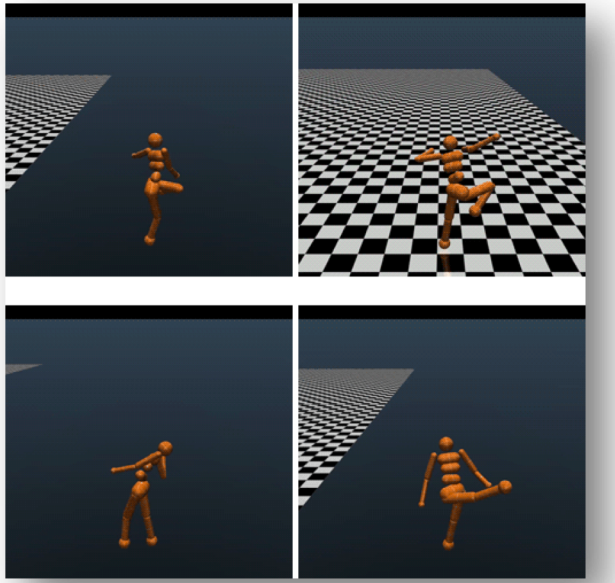


Leitura Adicional

Leitura Adicional:

Clues for Which I Search and Choose

Ben Recht (2018)

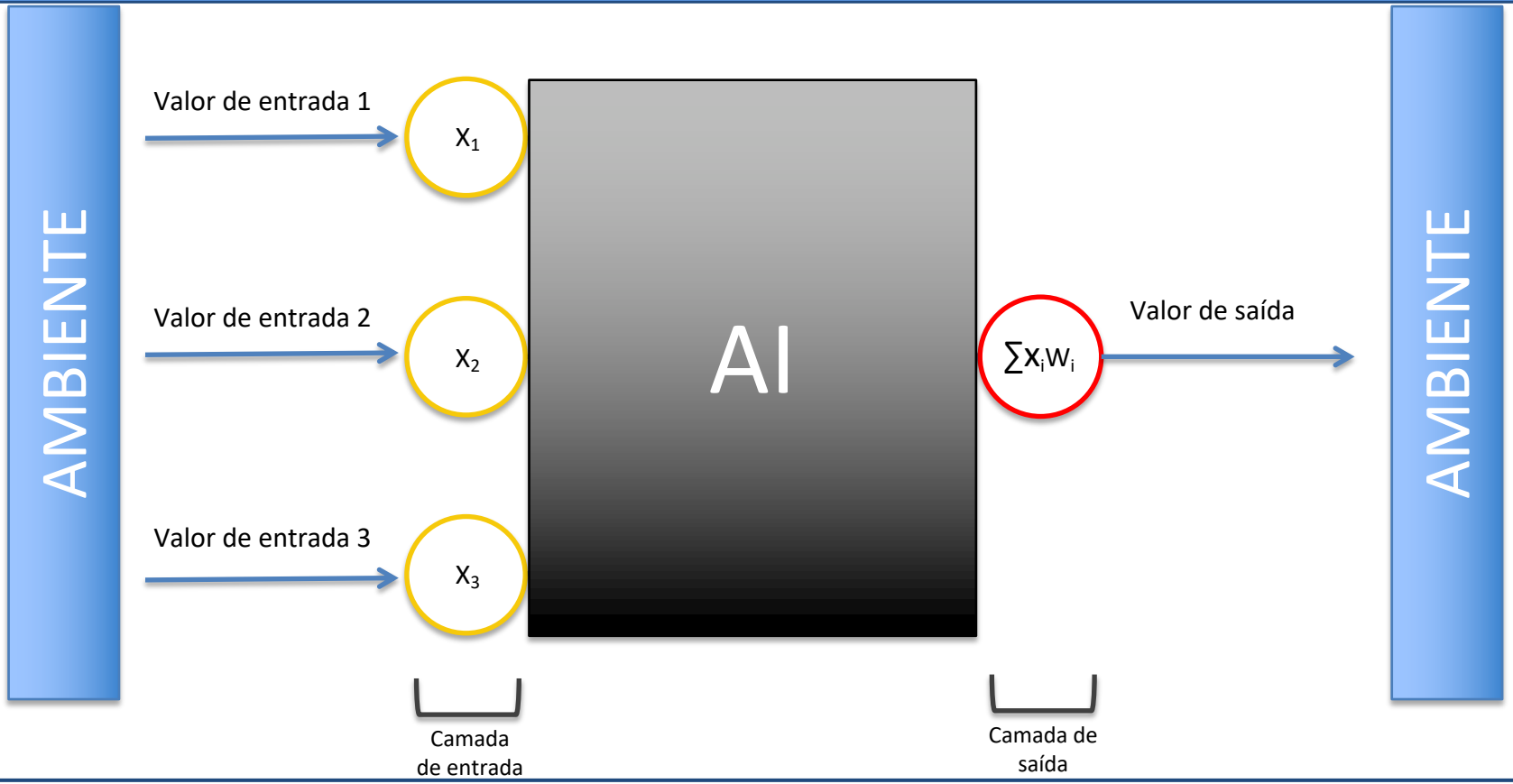


Link:

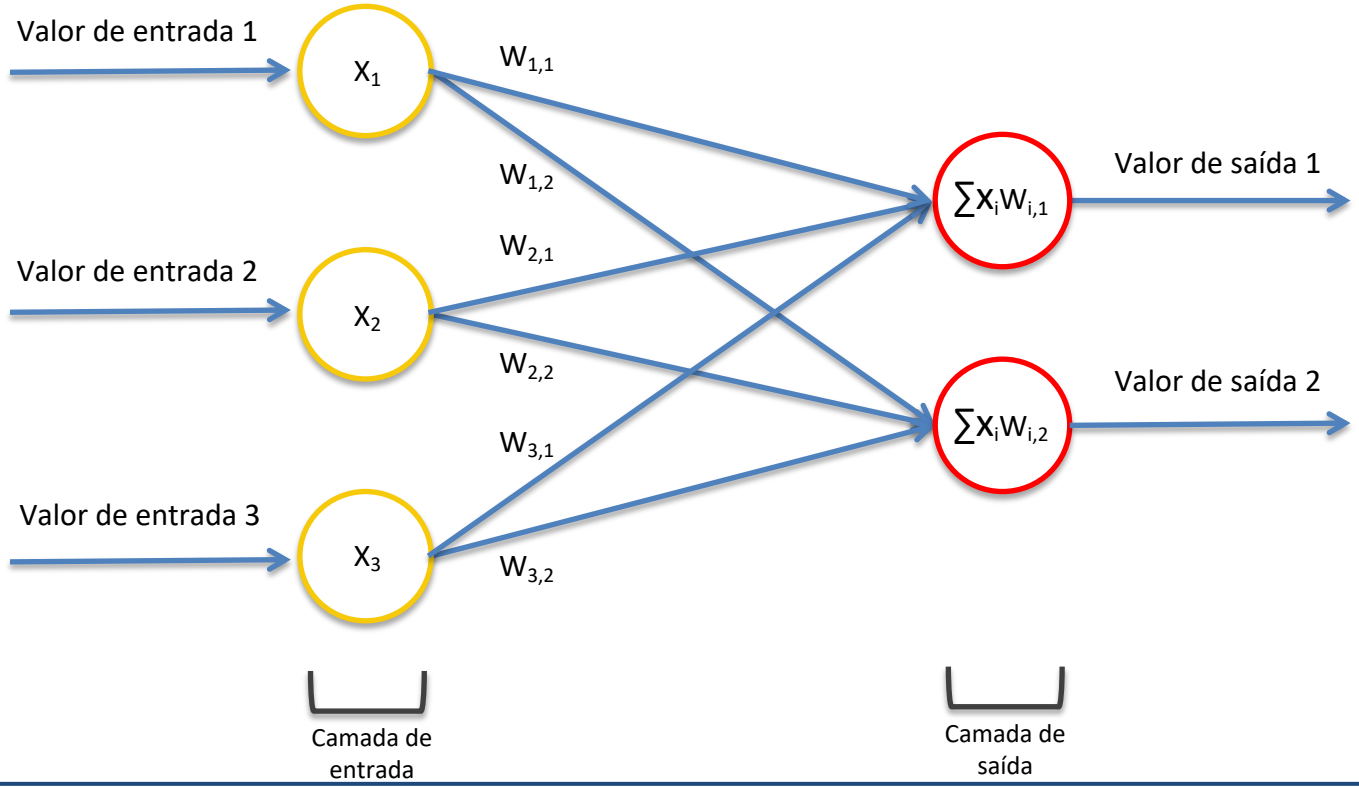
http://www.argmin.net/2018/03/20/mujocoloco/

Como um perceptron funciona?

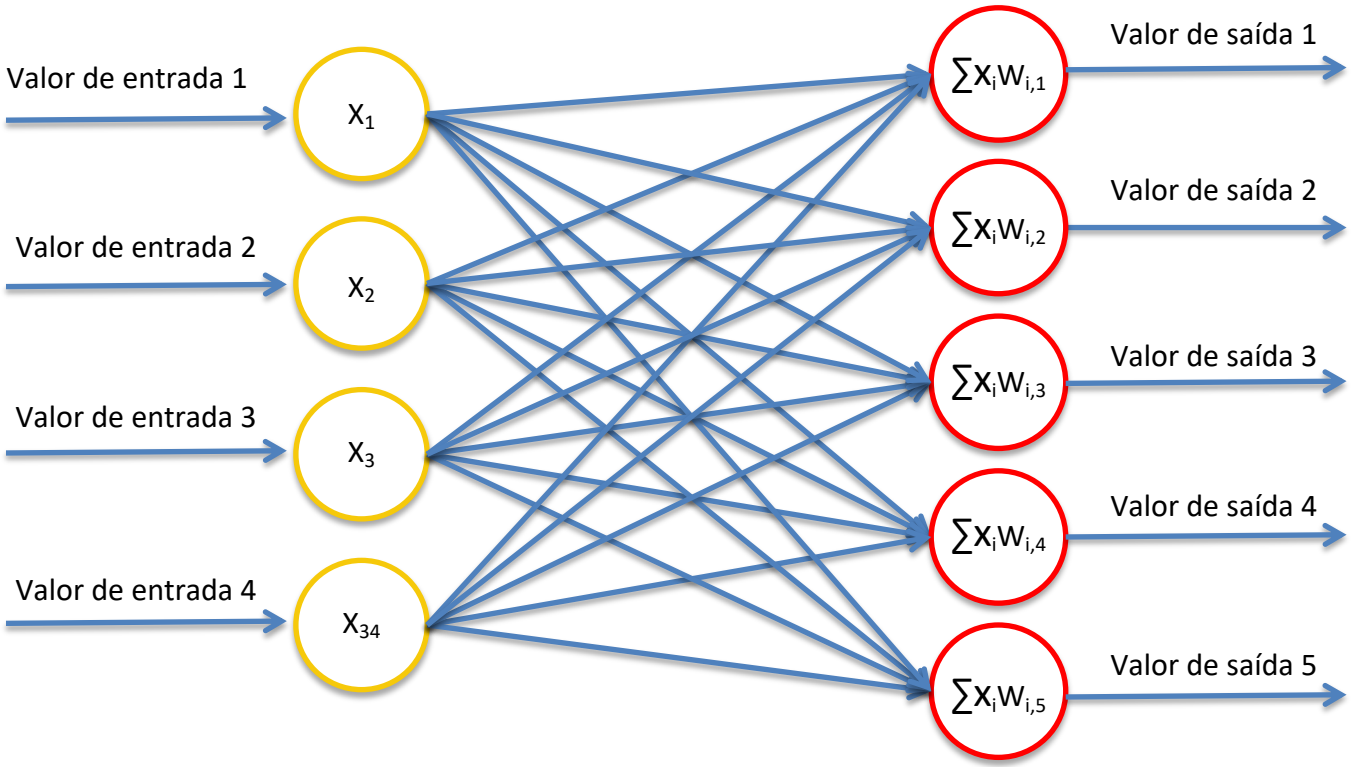
Como um perceptron funciona?



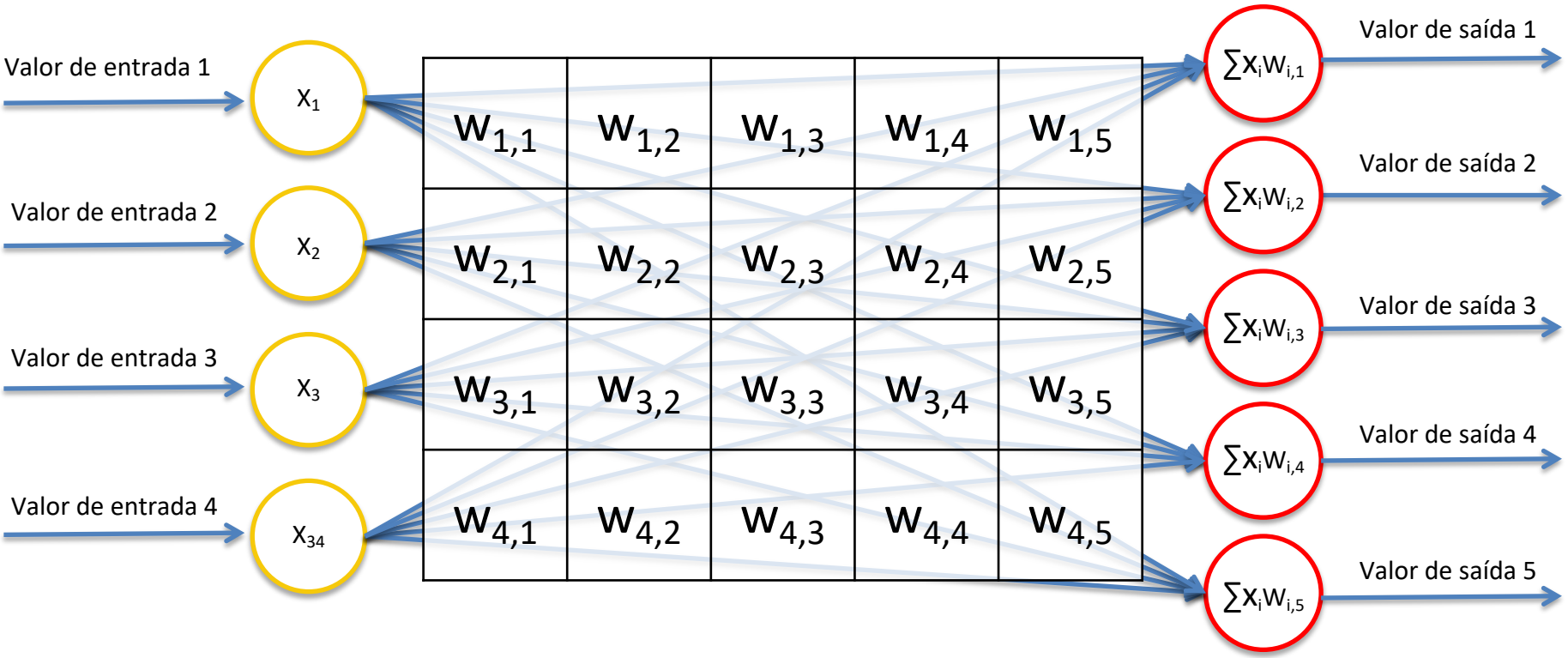
Como um perceptron funciona?



Como um perceptron funciona?



Como um perceptron funciona?



Como um perceptron funciona?

```
44 # Building the AI
45
46 class Policy():
47
48     def __init__(self, input_size, output_size):
49         self.theta = np.zeros((output_size, input_size))
50
51     def evaluate(self, input, delta = None, direction = None):
52         if direction is None:
53             return self.theta.dot(input)
54
```

Perceptron

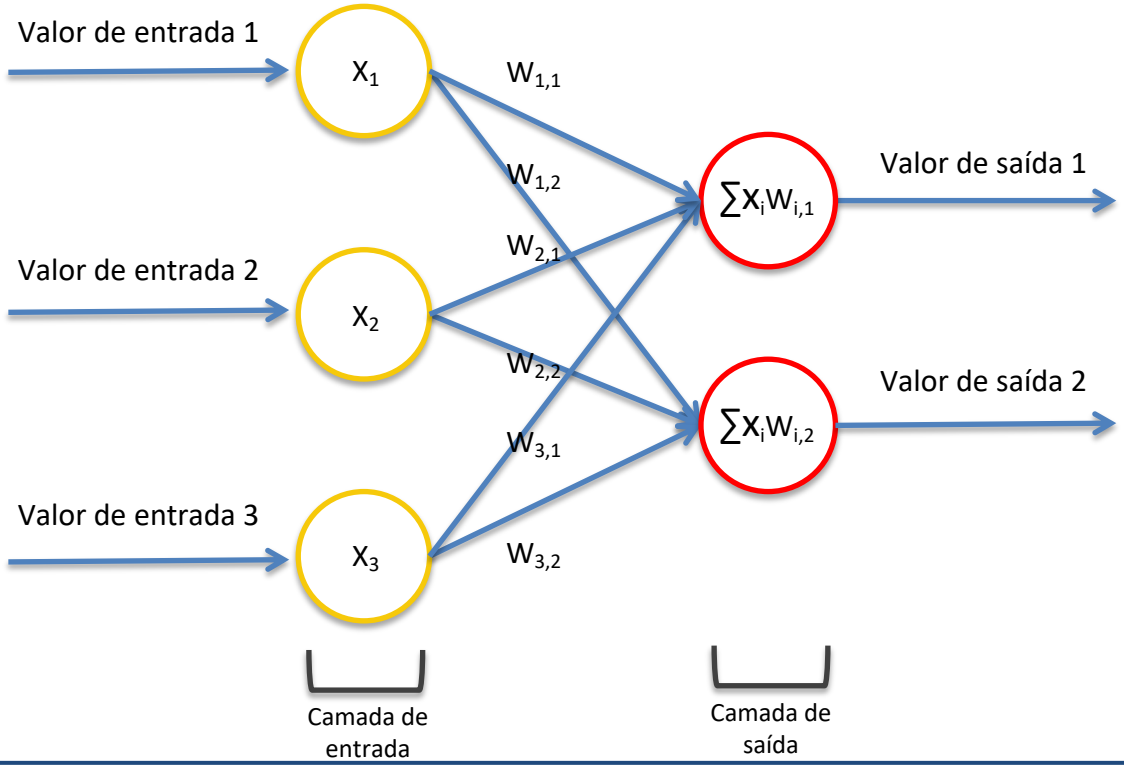
Matriz de Pesos

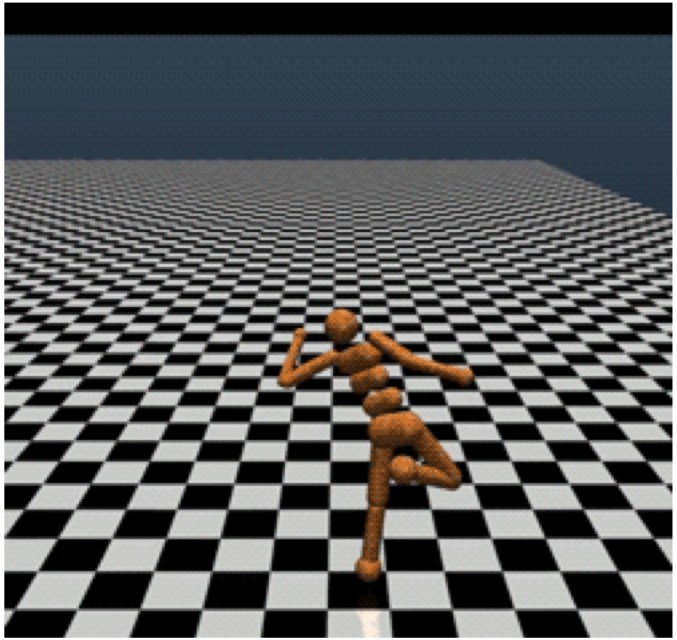
Multiplicação de Matrizes

Entradas

Maximização de Recompensas

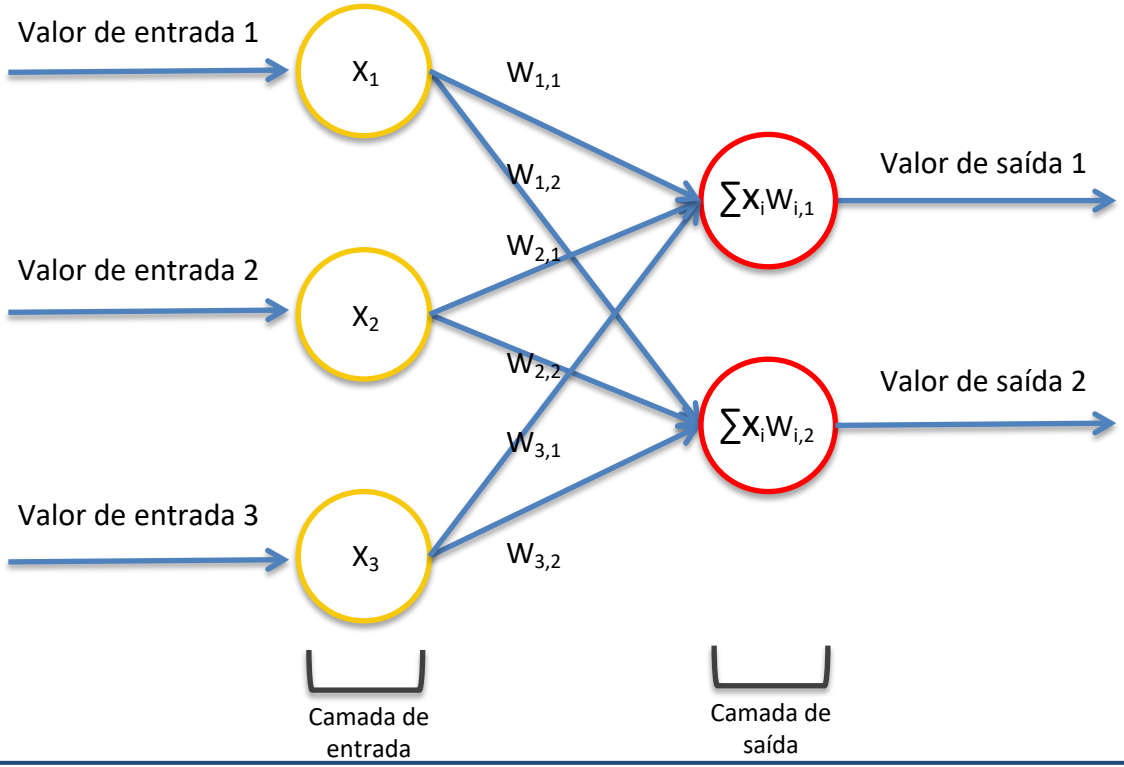
Maximização de Recompensas

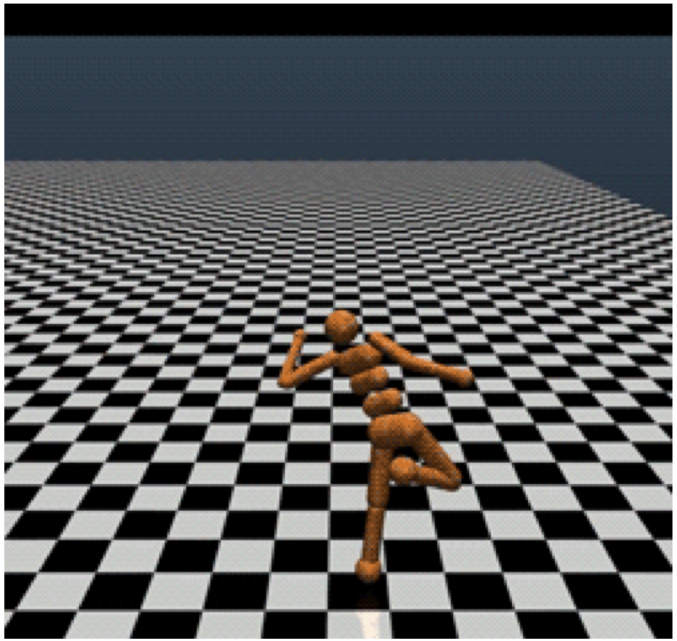




Fonte: www.argmin.net

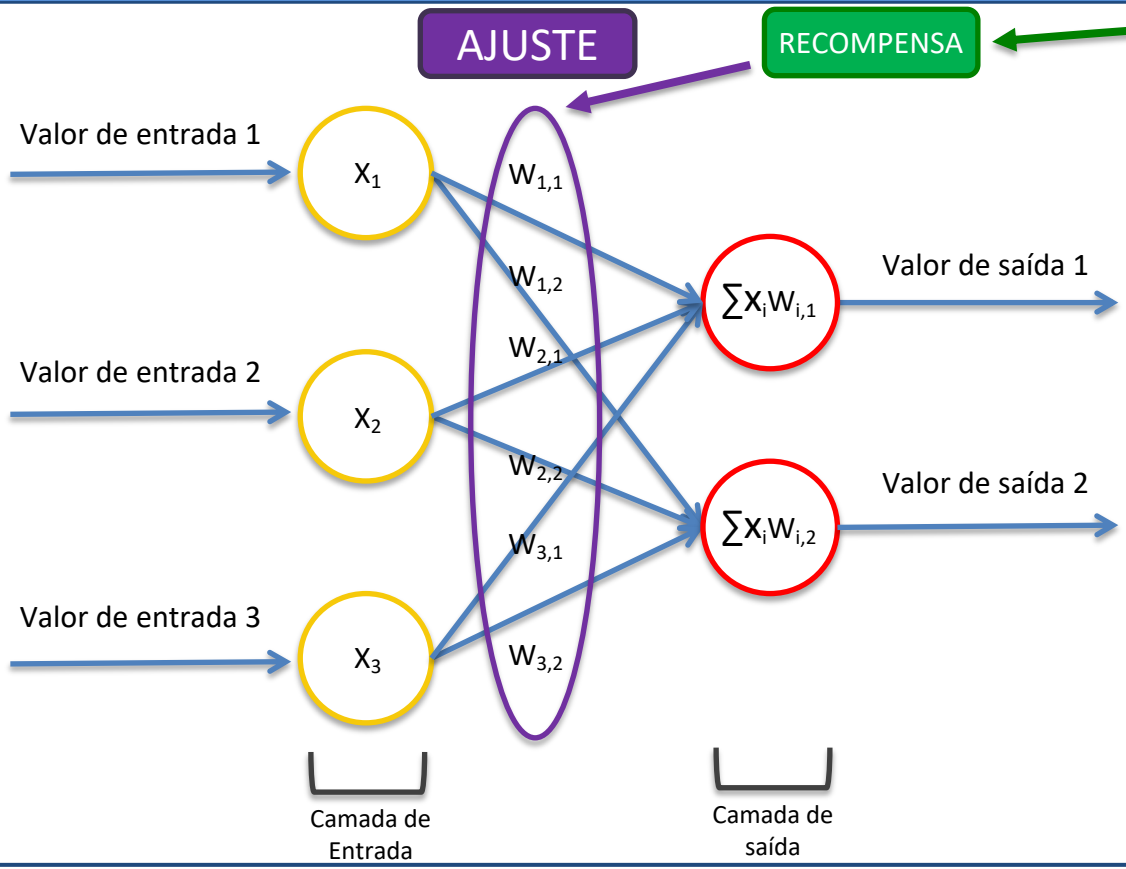
Maximização de Recompensas

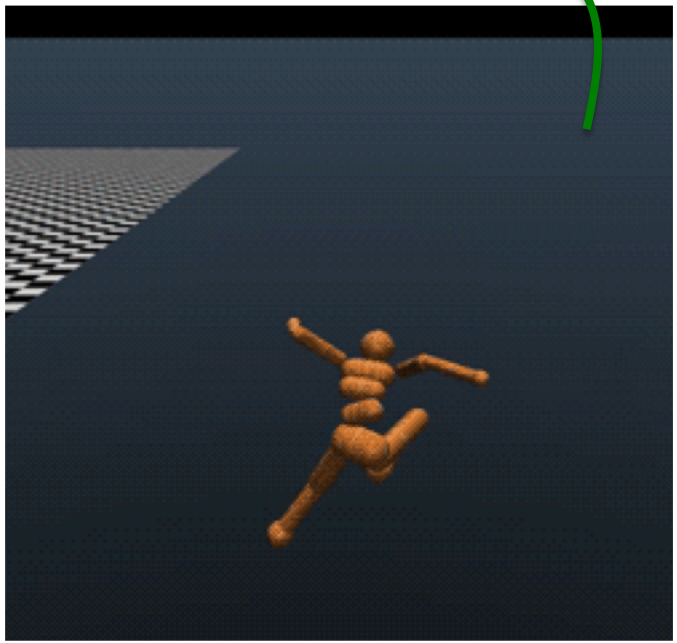




Fonte: www.argmin.net

Maximização de Recompensas

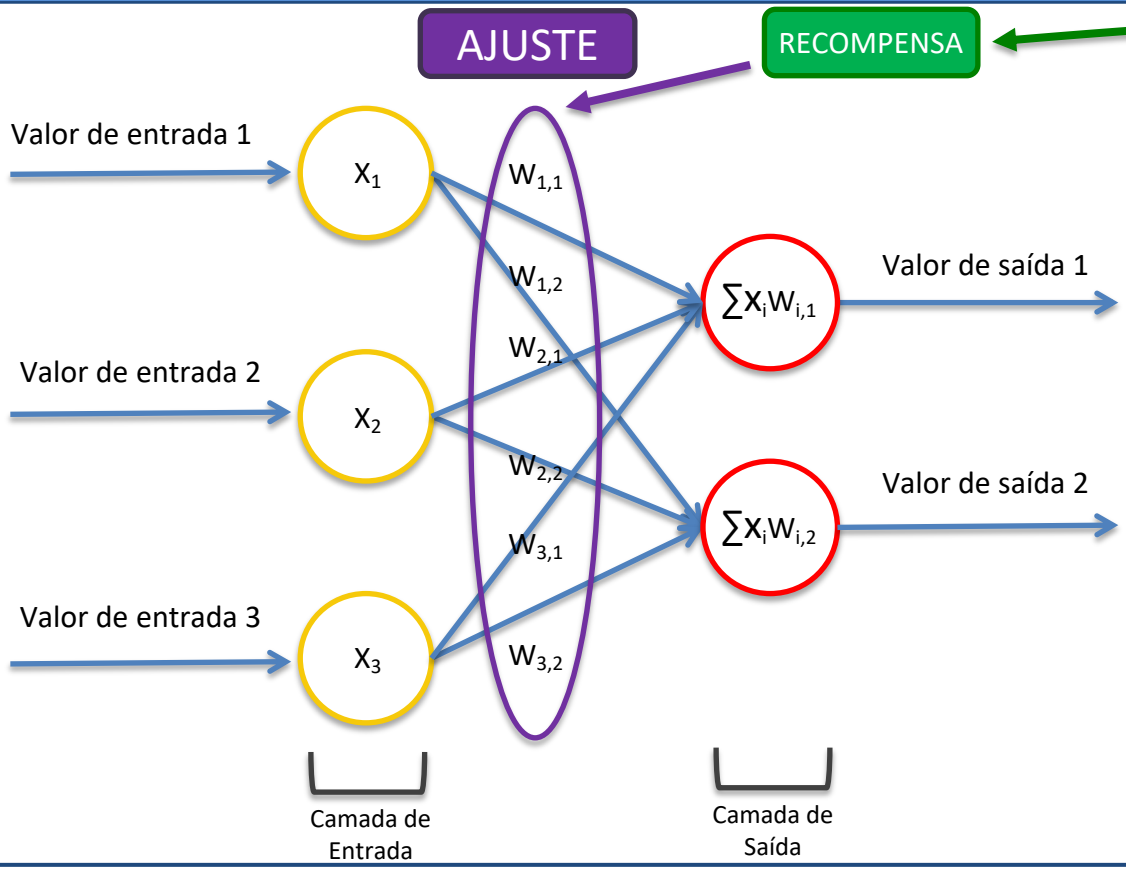


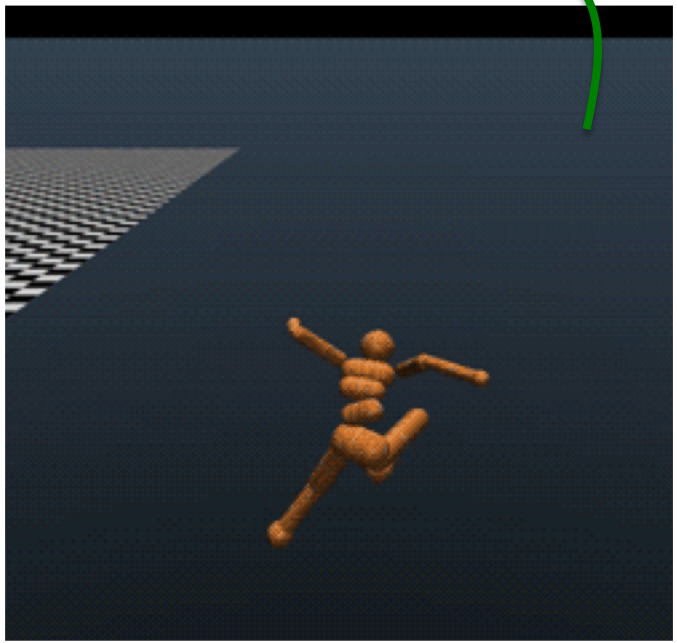


Fonte: www.argmin.net

Método de Diferenças Finitas (Method of Finite Differences)

Método de Diferenças Finitas





Fonte: www.argmin.net

Método de Diferenças Finitas

Usualmente em IA:

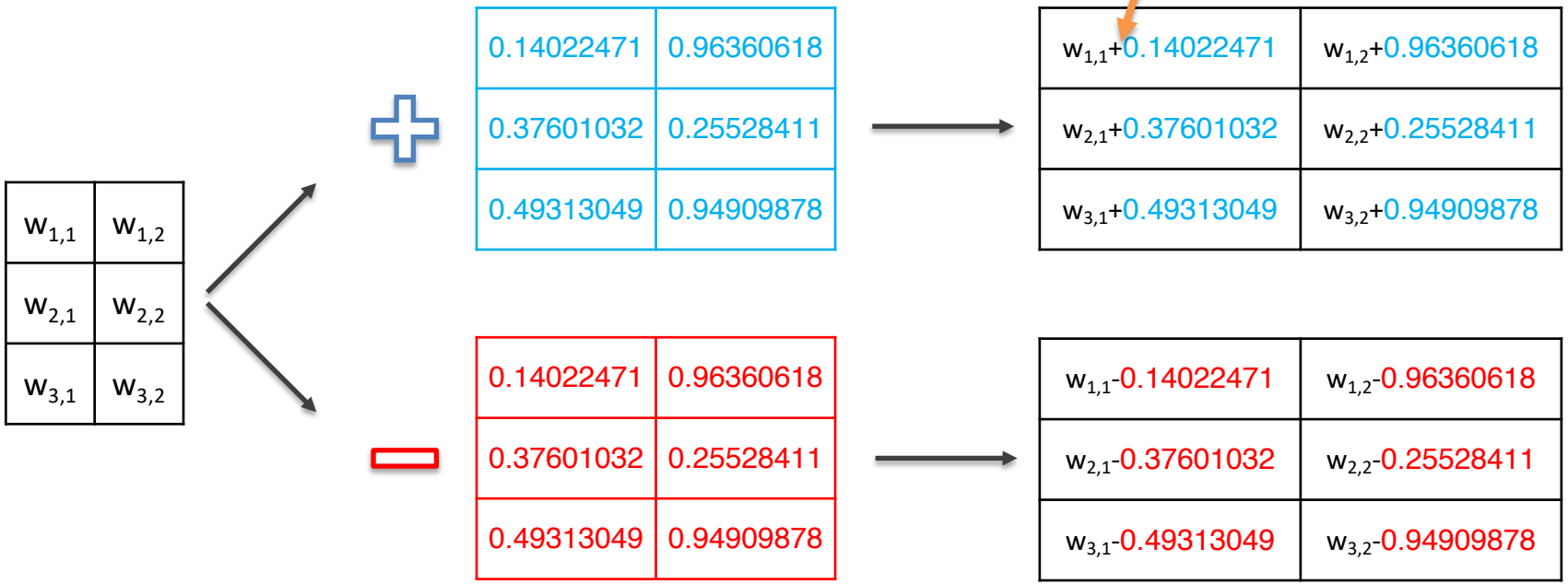
$$f'(x) = \frac{df}{dx}$$

Em ARS:

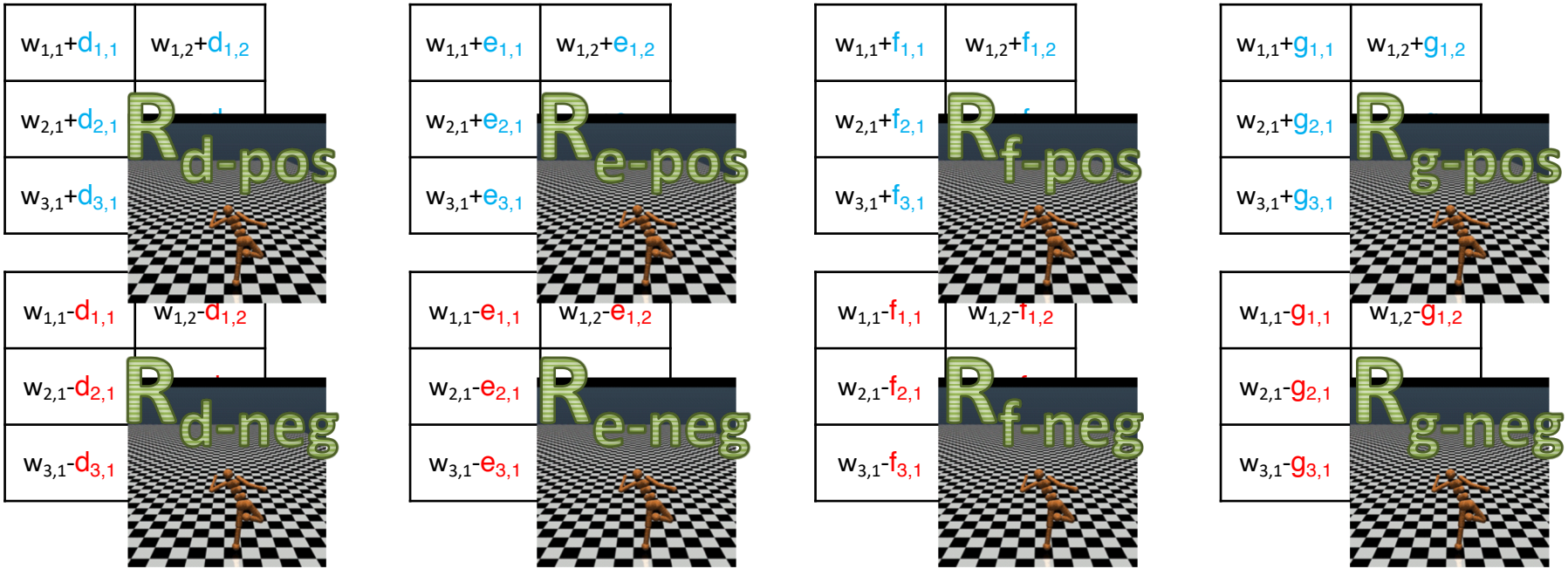
$$f'(a) \approx \frac{f(a+h) - f(a)}{h}$$

Método de Diferenças Finitas

Coeficiente



Método de Diferenças Finitas



Método de Diferenças Finitas

Coeficiente

$$\begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} + (R_{d\text{-}pos} - R_{d\text{-}neg}) * \begin{bmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \\ d_{3,1} & d_{3,2} \end{bmatrix} + (R_{e\text{-}pos} - R_{e\text{-}neg}) * \begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \end{bmatrix}$$

$$+ (R_{f\text{-}pos} - R_{f\text{-}neg}) * \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{bmatrix} + (R_{g\text{-}pos} - R_{g\text{-}neg}) * \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \\ g_{3,1} & g_{3,2} \end{bmatrix}$$

Leitura Adicional

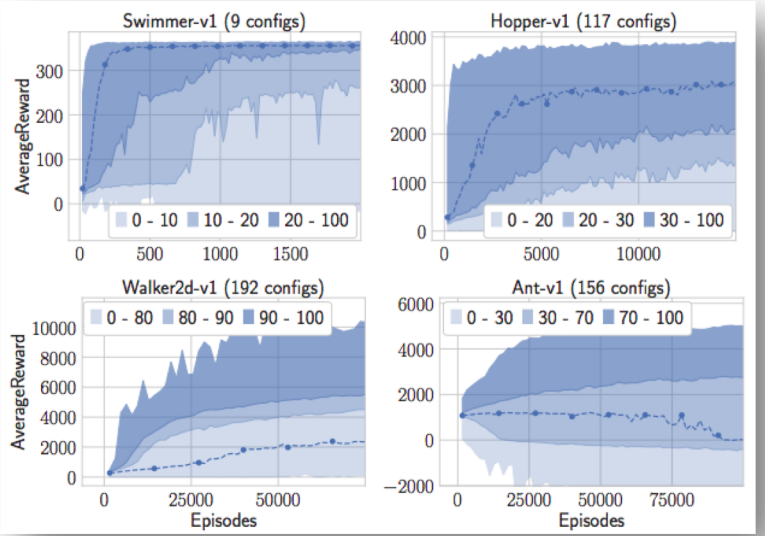
Leitura Adicional:

Simple random search provides a competitive approach to reinforcement learning

Horia Mania et al. (2018)

Link:

https://arxiv.org/pdf/1803.07055.pdf



Métodos Básicos x ARS

Métodos Básicos x ARS

Três atualizações principais:

- Etapa de atualização de escala usando o desvio padrão das recompensas
- Normalização em tempo real dos estados
- Descartar direções que levam para as recompensas mais baixas

Método de Diferenças Finitas

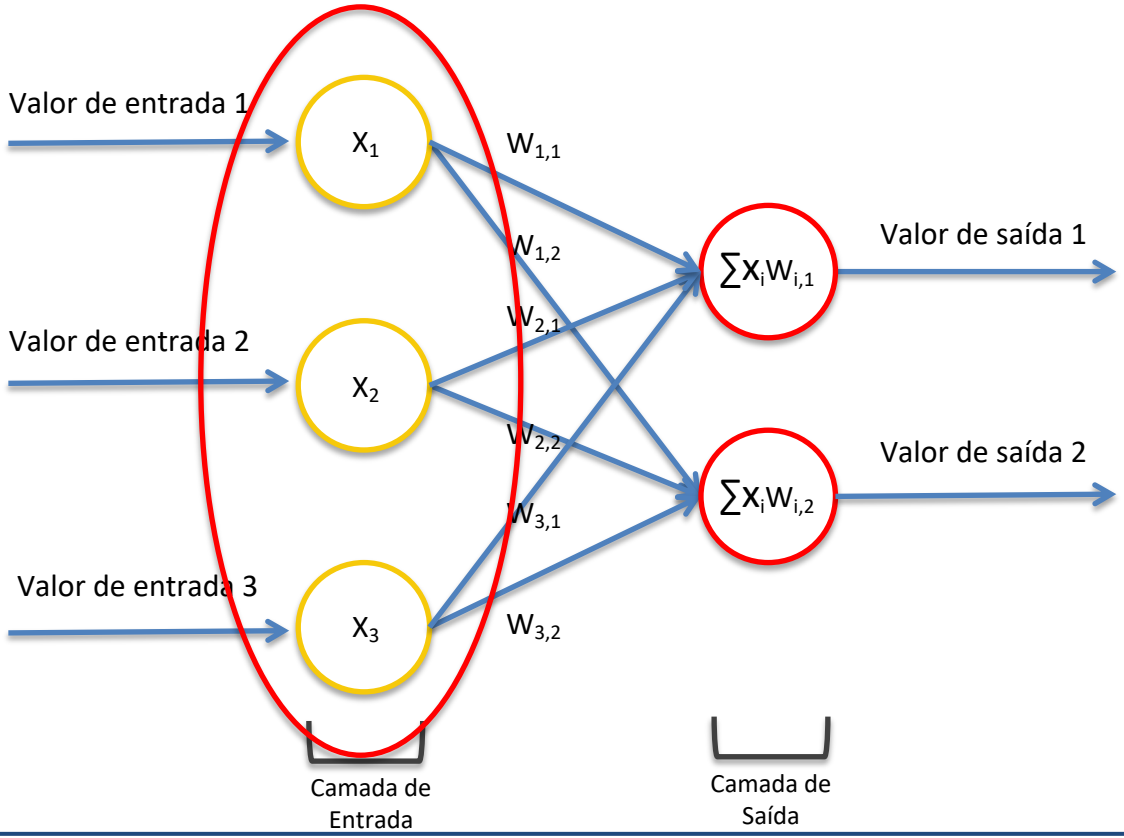
Coeficiente

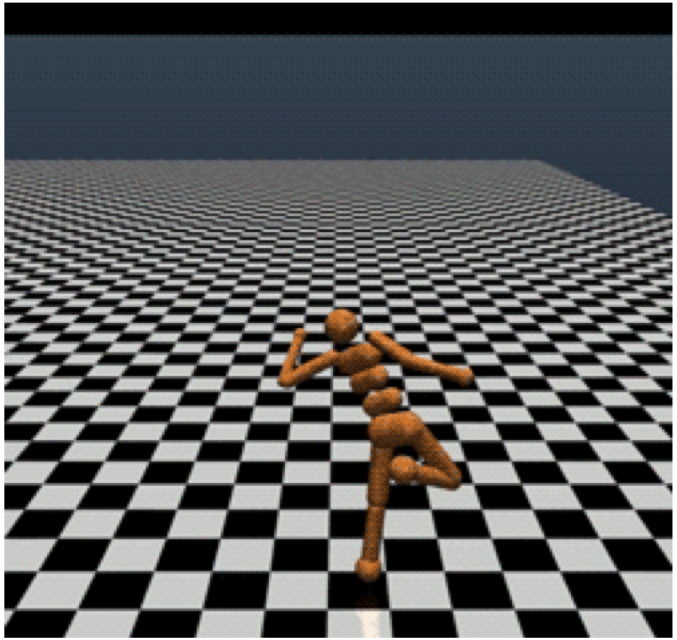
/ Desvio padrão das recompensas

$$\begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} + \Big[(R_{d\text{-}pos} - R_{d\text{-}neg}) * \begin{bmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \\ d_{3,1} & d_{3,2} \end{bmatrix} + (R_{e\text{-}pos} - R_{e\text{-}neg}) * \begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \end{bmatrix}$$

$$+ (R_{f\text{-}pos} - R_{f\text{-}neg}) * \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{bmatrix} + (R_{g\text{-}pos} - R_{g\text{-}neg}) * \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \\ g_{3,1} & g_{3,2} \end{bmatrix} \Big]$$

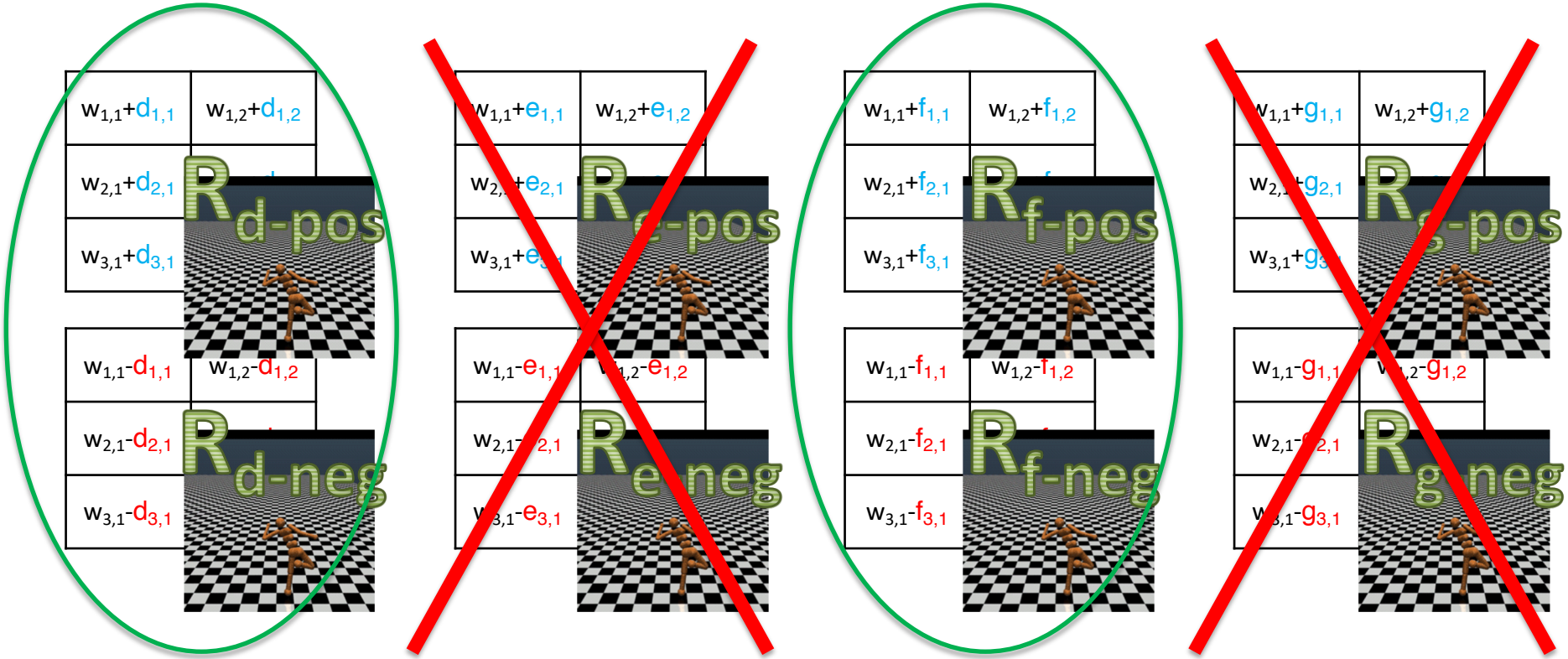
Método de Diferenças Finitas





Fonte: www.argmin.net

Método de Diferenças Finitas



Método de Diferenças Finitas

Coeficiente

/ Desvio padrão das recompensas

$$\begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} + (R_{d\text{-}pos} - R_{d\text{-}neg}) * \begin{bmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \\ d_{3,1} & d_{3,2} \end{bmatrix} + (R_{e\text{-}pos} - R_{e\text{-}neg}) * \begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \end{bmatrix}$$

$$+ (R_{f\text{-}pos} - R_{f\text{-}neg}) * \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{bmatrix} + (R_{g\text{-}pos} - R_{g\text{-}neg}) * \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \\ g_{3,1} & g_{3,2} \end{bmatrix}$$

Leitura Adicional

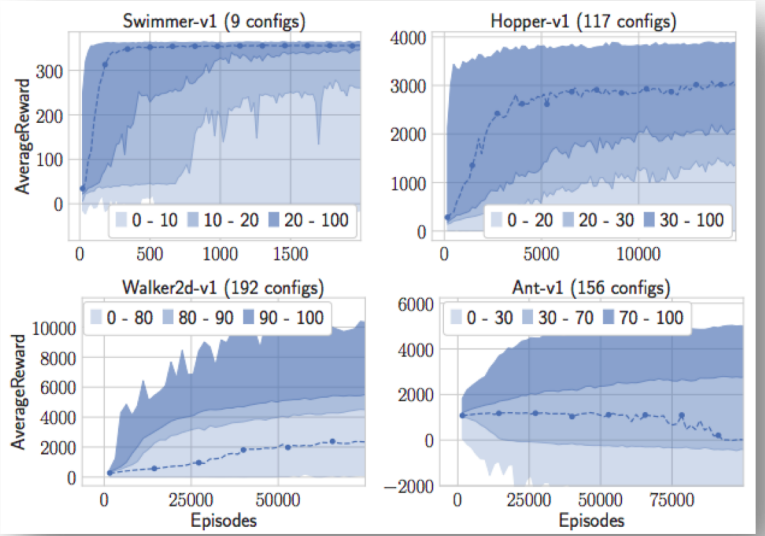
Leitura Adicional:

Simple random search provides a competitive approach to reinforcement learning

Horia Mania et al. (2018)

Link:

https://arxiv.org/pdf/1803.07055.pdf



ARS x Outras IAs

ARS x Outras IAs

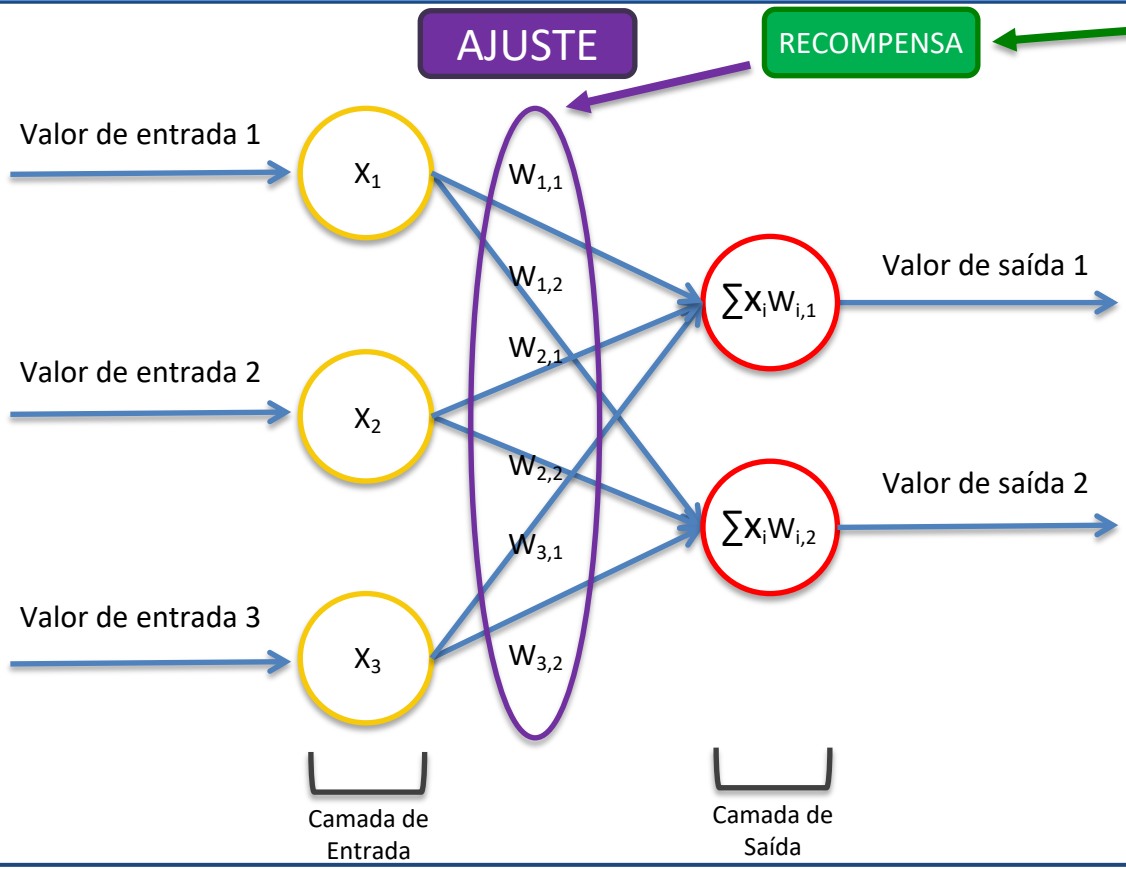
ARS

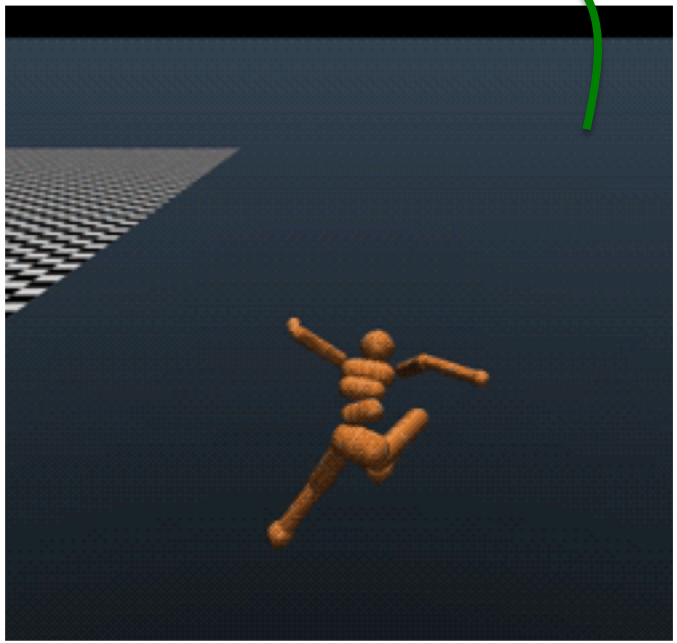
1. Exploração no espaço de Políticas

Outras IAs

1. Exploração no espaço de Ações

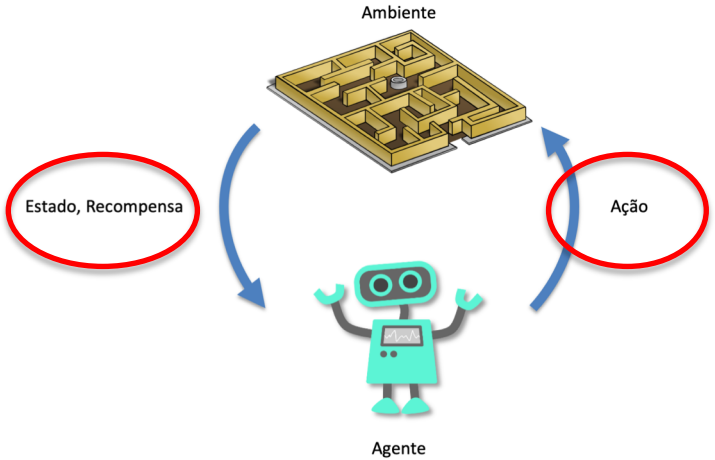
ARS x Outras IAs



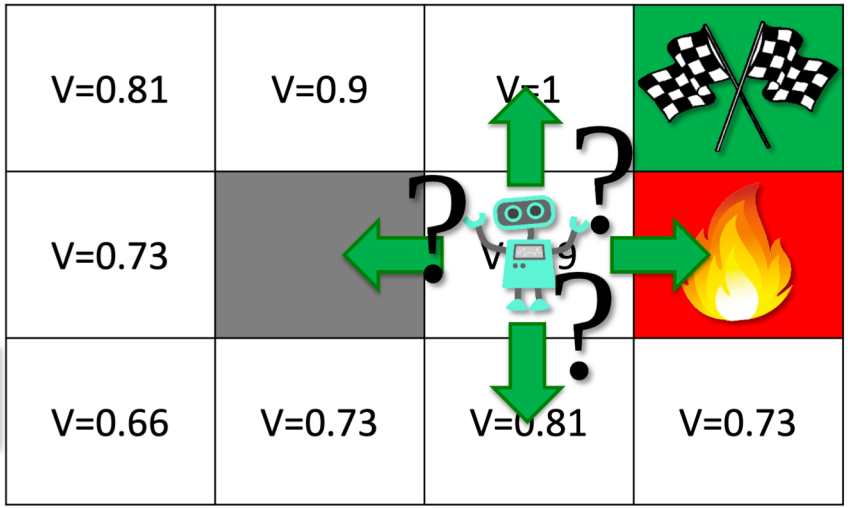


Fonte: www.argmin.net

ARS x Outras IAs



$$V(s) = \max_{a}(R(s,a) + \gamma V(s'))$$



ARS x Outras IAs

ARS

1. Exploração no espaço de Políticas
2. Método de Diferenças Finitas

Outras IAs

1. Exploração no espaço de Ações
2. Descida do Gradiente

ARS x Outras IAs

Coeficiente

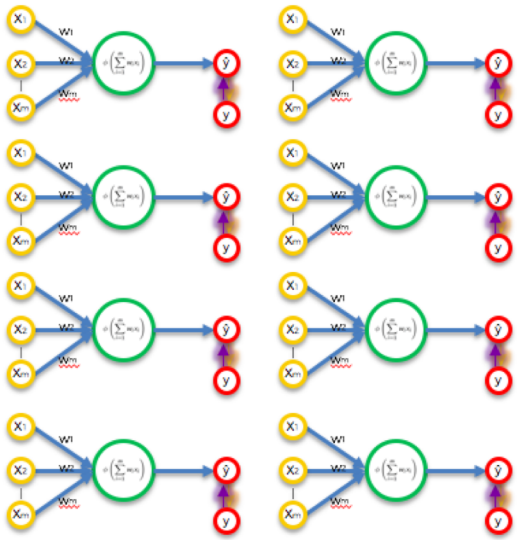
/ Desvio padrão das recompensas

$$\begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} + \Big[(R_{d\text{-}pos} - R_{d\text{-}neg}) * \begin{bmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \\ d_{3,1} & d_{3,2} \end{bmatrix} + (R_{e\text{-}pos} - R_{e\text{-}neg}) * \begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \end{bmatrix}$$

$$+ (R_{f\text{-}pos} - R_{f\text{-}neg}) * \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{bmatrix} + (R_{g\text{-}pos} - R_{g\text{-}neg}) * \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \\ g_{3,1} & g_{3,2} \end{bmatrix} \Big]$$

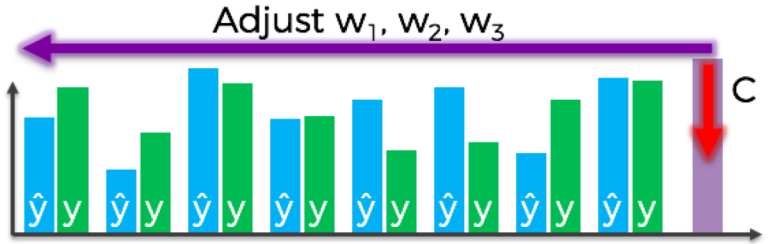
ARS x Outras IAs

Gradient Descent



Row ID	Study Hrs	Sleep Hrs	Quiz	Exam
1	12	6	78%	93%
2	22	6.5	24%	68%
3	115	4	100%	95%
4	31	9	67%	75%
5	0	10	58%	51%
6	5	8	78%	60%
7	92	6	82%	89%
8	57	8	91%	97%

$$C = \sum \tfrac{1}{2}(\hat{y} - y)^2$$



Deep Learning A-Z © SuperDataScience

ARS x Outras IAs

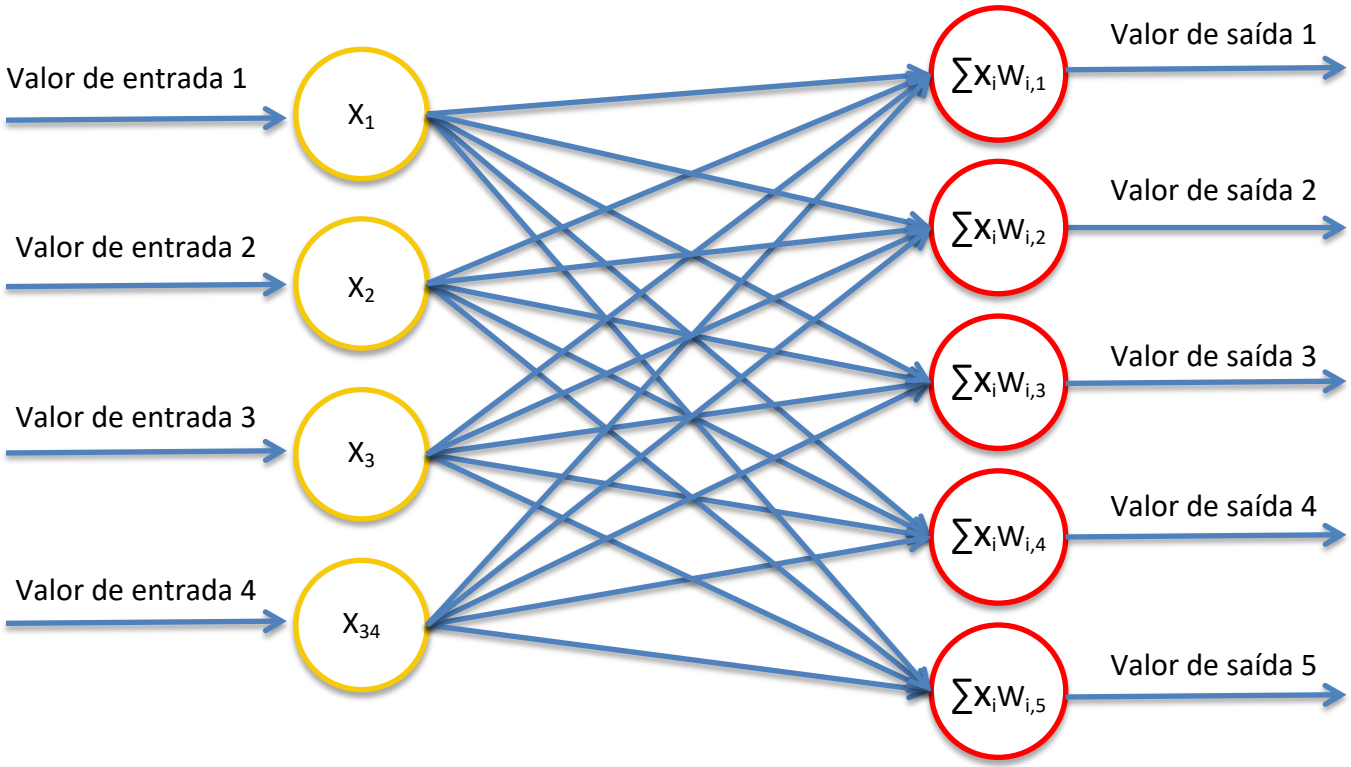
ARS

1. Exploração no espaço de Políticas
2. Método de Diferenças Finitas
3. Aprendizagem Superficial (Shallow)

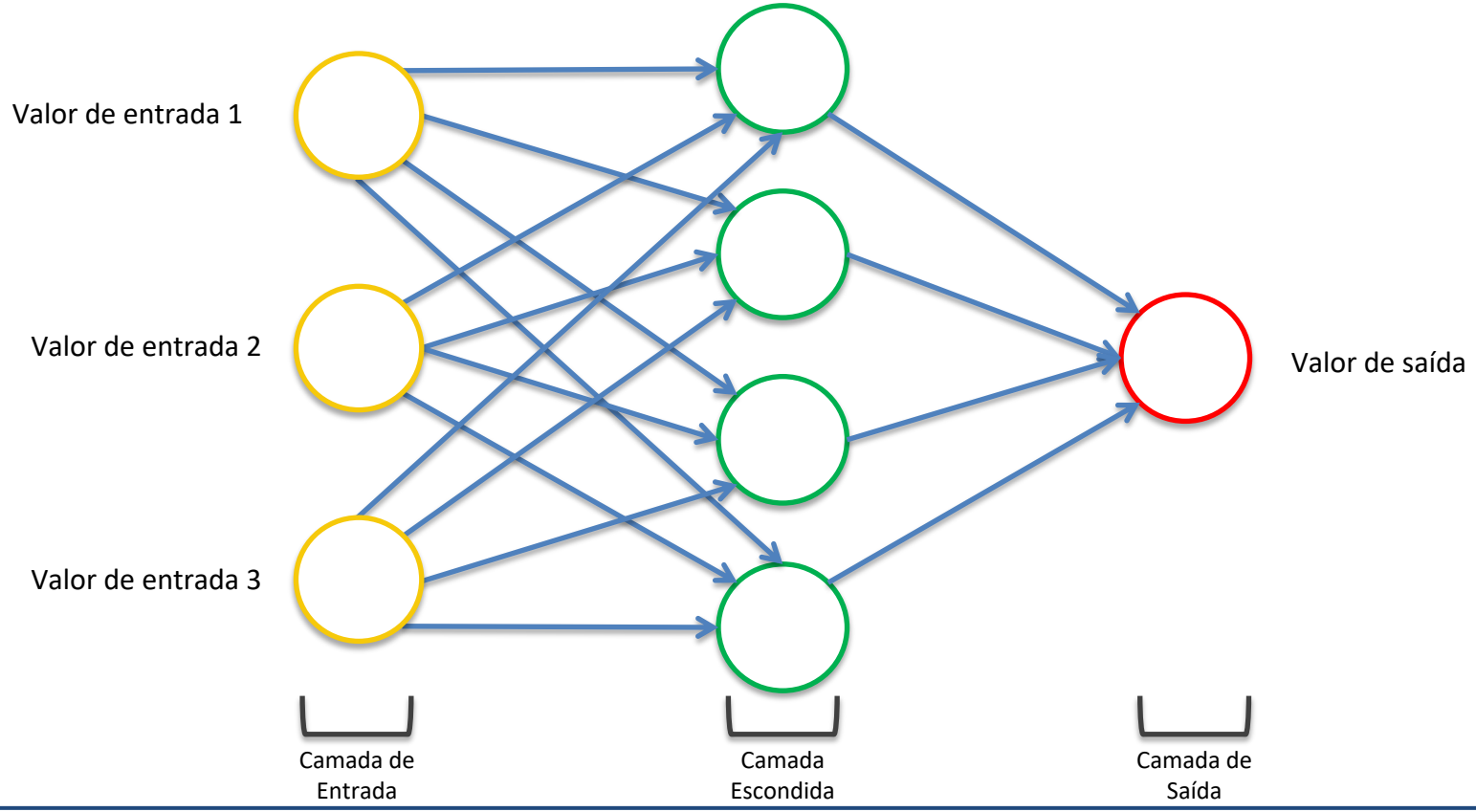
Outras IAs

1. Exploração no espaço de Ações
2. Descida do Gradiente
3. Em geral Deep Learning

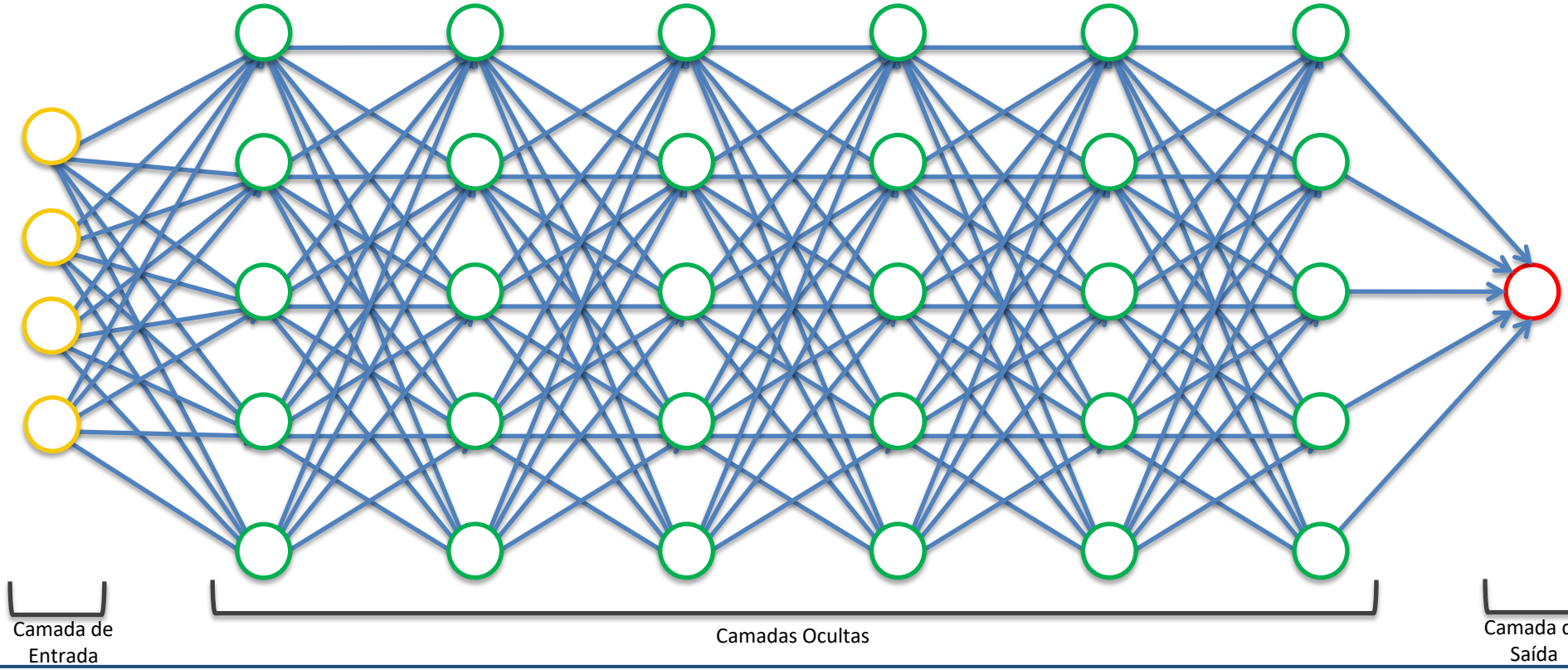
ARS x Outras IAs



ARS x Outras IAs



ARS x Outras IAs



ARS x Outras IAs

ARS

1. Exploração no espaço de Políticas
2. Método de Diferenças Finitas
3. Aprendizagem Superficial (Shallow)

Outras IAs

1. Exploração no espaço de Ações
2. Descida do Gradiente
3. Em geral Deep Learning

ARS é em torno de 15x mais rápido e leva a recompensas mais altas em aplicações específicas

Leitura Adicional

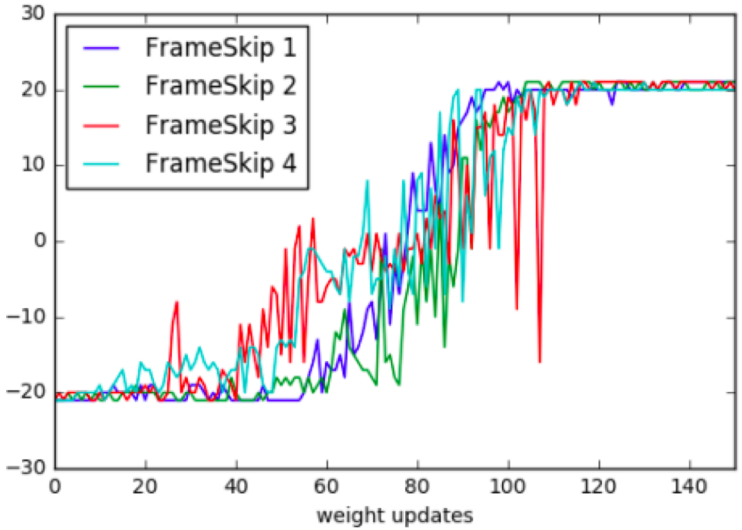
Leitura Adicional:

Evolution Strategies as a Scalable Alternative to Reinforcement Learning

Tim Salimans et al., Open AI (2017)

Link:

https://arxiv.org/pdf/1703.03864.pdf



Normalização (Normalization)

$$x = \frac{x\ - mínimo(x)}{máximo(x) - mínimo(x)}$$

Idade	Renda anual
60	30.000
35	45.000
20	29.500

Idade	Renda anual
1,00	0,03
0,37	1,00
0,00	0,00

$$x = \frac{60\ - 20}{60 - 20} = 1{,}00 \quad x = \frac{30.000\ - 29.500}{45.000 - 29.500} = 0{,}03$$

$$x = \frac{35\ - 20}{60 - 20} = 0{,}37 \quad x = \frac{45.000\ - 29.500}{45.000 - 29.500} = 1{,}00$$

$$x = \frac{20\ - 20}{60 - 20} = 0{,}00 \quad x = \frac{29.500\ - 29.500}{45.000 - 29.500} = 0{,}00$$

Padronização (Standardization)

$$x = \frac{x\ - média(x)}{desvio\ padrão(x)}$$

Idade	Renda anual	Idade	Renda anual
60	30.000	1,07	-0,54
35	45.000	-0,16	1,15
20	29.500	-0,90	-0,60

Idade
Média = 38,33
Desvio padrão = 20,20

Renda
Média = 34.833,33
Desvio padrão = 8.808,14

$$x = \frac{60\ - 38{,}33}{20{,}20} = 1{,}07 \qquad x = \frac{30.000\ - 34.833{,}33}{8.808{,}14} = -0{,}54$$

$$x = \frac{35\ - 38{,}33}{20{,}20} = -0{,}16 \qquad x = \frac{45.000\ - 34.833{,}33}{8.808{,}14} = 1{,}15$$

$$x = \frac{20\ - 38{,}33}{20{,}20} = -0{,}90 \qquad x = \frac{29.500\ - 34.833{,}33}{8.808{,}14} = -0{,}60$$