

Projeto #2

TÓPICOS DE MATEMÁTICA EM APRENDIZAGEM AUTOMÁTICA
Mestrado em Inteligência Artificial & Mestrado em Ciência de Dados
Departamento de Matemática – ISCTE-IUL
Ano académico: 2025/2026

Data de entrega: 12 de dezembro de 2025

Este projeto deverá ser feito em grupos de 3 ou 4 alunos, salvo exceções previamente acordadas. O trabalho deverá ser desenvolvido na linguagem Python, e o código (um único Jupyter notebook) enviado por email para `jorge.miguel.rocha@iscte-iul.pt` no formato `.ipynb`, devidamente comentado.

1. Nas últimas aulas do curso trabalhamos com um código (Lab3) que avalia uma política aleatória equiprovável π num “gridworld” 4×4 e que a melhora (“policy improvement”). Obtivemos resultados concordantes com a Figura 4.1 da referência Sutton & Barto.

- (a) Considere agora que a política original é diferente (chamemos-lhe a política $\tilde{\pi}$): em cada estado, em vez de selecionar uma das quatro ações possíveis com a mesma probabilidade ($1/4$), a política pré-melhoria $\tilde{\pi}$ é definida pela seguinte tabela (a mesma para todos os estados):

ação	probabilidade
left	0.4
right	0.1
up	0.3
down	0.2

Adapte o código do Lab3 (portanto, sem escrever um novo código de raiz) para obter a função de valor de cada um dos 14 estados deste sistema.

- (b) No caso da política aleatória equiprovável, π , a função de valor antes da melhoria, $v_{\pi}(s)$, era, no máximo, -14 , enquanto que após melhoria para uma nova política π' se pode verificar por inspeção imediata (ver Figura 4.1 do Sutton & Barto) que a função de valor $v_{\pi'}(s)$ vale -1 , -2 ou -3 , dependendo do estado. Faça uma análise semelhante para a política $\tilde{\pi}$ e para a sua versão melhorada $\tilde{\pi}'$.