

Projecto Inteligência Artificial

(LEIC 3º Ano, 1º Semestre 2019/2020)

Aprendizagem por Reforço

1 Agentes que aprendem

Nesta segunda entrega vamos deixar um agente aprender qual o melhor comportamento num dado ambiente. Por exemplo o ambiente pode ser uma cidade, e um ladrão quer ver que localizações evitar.

1.1 Bibliografia e ambiente de desenvolvimento

A matéria teórica necessária ao desenvolvimento do projecto pode ser encontrada no livro de texto adoptado. O projecto deverá ser implementado em Python.

2 2a Fase

Nesta entrega iremos estudar métodos de aprendizagem por reforço e de exploração. Haverá 1 classe com pelo menos 3 funções a implementar. O protótipo é:

```
# LearningAgent to implement
# no knowledge about the environment can be used
# the code should work even with another environment
class LearningAgent:

    # init
    # nS maximum number of states
    # nA maximum number of action per state
    def __init__(self,nS,nA):

        # define this function
        self.nS = nS
        self.nA = nA
        # define this function

    # Select one action, used when learning
    # st - is the current state
    # aa - is the set of possible actions
    # for a given state they are always given in the same order
    # returns
    # a - the index to the action in aa
    def selectactiontolearn(self,st,aa):
        # define this function
        # print("select one action to learn better")

        a = 0
        # define this function
        return a

    # Select one action, used when evaluating
    # st - is the current state
    # aa - is the set of possible actions
    # for a given state they are always given in the same order
    # returns
```

```

# a - the index to the action in aa
def selectactiontoexecute(self,st,aa):
    # define this function
    a = 0
    # print("select one action to see if I learned")
    return a

# this function is called after every action
# st - original state
# nst - next state
# a - the index to the action taken
# r - reward obtained
def learn(self,ost,nst,a,r):
    # define this function
    #print("learn something from this data")

    return

```

Durante a aprendizagem há um ciclo que chama a função `SELECTACTIONTOLEARN(SELF,ST,AA)` para que seja escolhida uma acção. O resultado desta função é uma acção que após ser executada no mundo dá origem a uma recompensa. Esses dados serão usados para aprender quais são as melhores acções em cada estado na função `LEARN(SELF,OST,NST,A,R)`.

Após um tempo limitado de aprendizagem o agente é testado. Ai já não há aprendizagem e o agente vai executar as acções escolhidas na função `SELECTACTIONTOEXECUTE(SELF,ST,AA)`.

Para avaliar a rapidez de aprendizagem, em cada exemplo o agente irá ser testado 2 vezes. A primeira com alguns passos e depois a aprendizagem irá ser chamada novamente com mais passos. Não alterar o número de passos no ficheiro de testes.

A cotação será de 14 valores para o final da aprendizagem (muitos passos) e de 6 valores para a aprendizagem inicial (poucos passos).

Devido à aleatoriedade na aprendizagem, os testes irão ser corridos 5 vezes e a média será usada.

3 Informação a ser usada

Nos ficheiros para teste é trivial ver qual o resultado. Esses dados não podem ser usados. O agente só pode aprender através dos dados passados para `LEARN(SELF,OST,NST,A,R)`. O agente irá ser testado em outros ambientes usando uma interface equivalente.

4 Entregas e Prazos

Deverá ser submetido um só ficheiro contendo o código do seu projecto (com o nome LGGG em que L é A - Alameda ou T - Tagus, e GGG é o número do grupo (usar 3 caracteres). O ficheiro de código deve conter em comentário, na primeira linha, os números e os nomes dos alunos do grupo, bem como o número do grupo. Não é necessário incluir os ficheiros disponibilizados pelo corpo docente.

As entregas têm que ser feitas até ao limite definido a seguir, data e hora, não sendo aceites projectos fora de prazo sob pretexto algum.

- 2ª Entrega - até até às 23:59 do dia 06/12/2019

4.1 Condições de realização e discussão dos projectos

Projectos muito semelhantes serão considerados cópia e rejeitados. A detecção de semelhanças entre projectos será realizada utilizando software especializado e caberá exclusivamente ao corpo docente a decisão do que considera ou não cópia. Em caso de cópia, todos os alunos envolvidos terão 0 no projecto, serão reprovados na cadeira e referenciados para o conselho pedagógico.

Os trabalhos serão realizados em grupos de 2 pessoas mas cada pessoa deverá ser capaz de explicar todo o trabalho.

Alguns alunos serão chamados, de forma aleatória ou caso seja necessário confirmar a aquisição a competências, **individualmente** para uma discussão oral do trabalho e/ou uma demonstração do funcionamento do programa.