

Relatório do Projeto 2 de Inteligência Artificial

Mihail Brinza

83533

Ricardo Brancas

83557

6 de Dezembro de 2017

1 Métodos de Classificação

2 Métodos de Regressão

3 Aprendizagem por Reforço

3.1 Trajetórias Aprendidas

3.1.1 Exemplo 1

$5 \xrightarrow{0} 6 \xrightarrow{0} 6 \xrightarrow{0} 6 \xrightarrow{0} 6$

3.1.2 Exemplo 2

$5 \xrightarrow{0} 6 \xrightarrow{0} 1 \xrightarrow{1} 0 \xrightarrow{1} 0$

3.2 Modelo do Mundo

3.2.1 Exemplo 1

Neste primeiro exemplo o ambiente consiste numa série de quadrículas sequenciais, tal como demonstrado na figura 1, em que a ação 1 corresponde a dar um passo para a esquerda e a ação 0 corresponde a dar um passo para a direita. Tentar andar para fora dos limites não tem qualquer efeito.



Figura 1: Ambiente 1

Por fim determinámos que o agente é recompensado sempre que o estado inicial (antes da ação) é um estado limite. (i.e. 0 ou 6).

Como tal, quando começamos no estado 5 o melhor curso de ação é andar rapidamente para o estado limite mais próximo, o 6, e depois mantermo-nos lá.

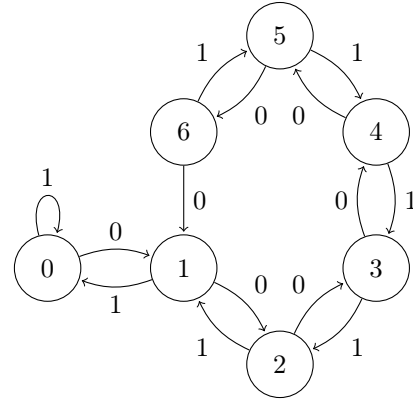


Figura 2: Ambiente 2

3.2.2 Exemplo 2

No segundo exemplo o ambiente é muito semelhante ao primeiro com exceção do estado 6. Agora quando tentamos andar para a direita (ação 1) nesse estado voltamos para o estado 1, tal como representado na figura 2. A função de recompensa mantém-se também inalterada, sendo 1 quando o estado inicial é o zero ou o seis, e 0 caso contrário.

Como tal [quando começamos no estado 5] já não é possível usar a estratégia anterior de ficar parado no estado 6. Assim, o melhor a fazer nesta situação é tentarmos dirigir para o estado 0, de modo a maximizar a recompensa a longo termo, que é exatamente o que obtemos com o algoritmo Q-learning.