

Enunciados dos Exercícios – Cap. 21 – Russell & Norvig

1. **(21.4)** O método da estimativa de utilidade direta da Seção 21.2 utiliza estados terminais distintos para indicar o fim de um teste. De que maneira ele poderia ser modificado no caso de ambientes com recompensas descontadas e sem estados terminais?
2. **(21.6)** Adapte o mundo do aspirador de pó (Capítulo 2) para aprendizagem por reforço, incluindo recompensas pela captura de cada fragmento de sujeira e por voltar para a posição inicial e se desligar. Torne o mundo acessível, fornecendo percepções adequadas. Agora experimente diferentes agentes de aprendizagem por reforço.