

Ejercicio 1

Demostrar que, si conociéramos exactamente el valor de cada acción, es decir si $Q_t(a) = E[R_t | A_t = a]$, entonces la acción *greedy* $A_t = \operatorname{argmax}_a Q_t(a)$ es la acción óptima en el sentido de que permite maximizar las recompensas totales

Ejercicio 2

Lorem Ipsum