Universidad de los Andes
Facultad de Ingeniería
Escuela de Sistemas
Departamento de Computación
Nombres: Diego Alejandro Martínez Rívas y José Gregorio Contreras Suárez
Profesor: Jesús Peréz
Asignatura: Sistemas Computacionales A2023

Lectura 1: Multi-Armed Bandits

Comparación de los métodos de selección de acciones en un entorno Two-Armed Bandit. ¿Cúal método de selección es mejor en 100 pasos?.

Método	α	ϵ	R(T) Acción 0	R(T) Acción 1	Mejor Acción
Random	0.1		32	200	1
Greedy	0.1		55	0	0
Epsilon-Greddy	0.1	0.1	34	300	1
Epsilon-Greddy	0.1	0.2	46	200	1
Epsilon-Greddy	0.1	0.3	18	500	1
Epsilon-Greddy	0.1	0.4	29	300	1

Con el método aleatorio se logró observar que cualquiera de las dos acciones podía ser la mejor, había una misma probabilidad de elegir la acción 0 que la acción 1, y cuando se obtenía una recompensa por la acción 1 ya era claramente que sería la mejor.

Con el método greedy (codicisoso) a diferencia de con el método aleatorio, siempre la mejor acción fue la acción 0. Porque principalmente seleccionaba la opción 0, dado que es la más probable de ganar una recompensa. Lo que conlleva a que tuviería siempre (o casi) el valor mayor de las dos acciones, cuando escogía la acción 1 era muy poco probable que ganara, por lo tanto escogería la acción 0 de nuevo en el siguiente turno.

Con el método epsilon-greedy se observaron similitudes con el método greedy, pero la diferencia es que daba más oportunidades de escoger la acción que no necesariamente era la mejor en ese momento, dependiendo del valor de epsilon. Mientras más grande fuera epsilon era más probable de hacerlo.

En los metodos aleatorio y epsilon-greedy la recompensa total mayor fue dada por la acción 1 por gran diferencia, mientras que en con el método greedy fue la acción 0.

De todos los métodos pienso que el mejor es el epsilon-greedy por lo explicado anteriormente. Con un valor de epsilon lo suficiente pequeño para minimizar la exploraración y máximizar la explotación.