

2. Recherchieren Sie eine weitere Explorationsstrategie (nicht ϵ -greedy).

a) Welche haben Sie gewählt?

semi-uniform distributed exploration strategy

Literatur: <https://www.cs.mcgill.ca/~cs526/roger.pdf>

b) Wie funktioniert Sie?

Selects an action according to the semi-uniform distributed exploration strategy. If several actions have the same highest utility estimate, select a random one amongst them.

Description:

"Rather than generating actions randomly with uniform probability, an alternative would be to generate actions with a probability distribution that is based on the utility estimates that are currently available to the agent. One such approach selects the action having the highest current utility estimate with some predefined probability P_{best} . Each of the other actions is selected with probability $1 - P_{best}$ regardless of its currently utility estimate.

$$P(a) = \begin{cases} P_{best} + \frac{1 - P_{best}}{\# \text{ of actions}} & , \text{ if } a \text{ maximizes } f(a) \\ \frac{1 - P_{best}}{\# \text{ of actions}} & , \text{ otherwise} \end{cases}$$

The P_{best} parameter facilitates a continuous blend of exploration and exploitation that ranges from purely random exploration ($P_{best} = 0$) to pure exploitation ($P_{best} = 1$)."

Literatur: <https://www.cs.mcgill.ca/~cs526/roger.pdf>