

## Praktikum Selbstlernende Systeme | Aufgabenblatt 2

### Aufgabe 1.2.a: Welche Explorationsstrategie?

- Boltzmann Approach

### Aufgabe 1.2.b: Wie funktioniert sie?

Bei der Exploration möchten wir idealerweise alle Informationen nutzen, die in den Q-Werten unseres Netzwerks enthalten sind. Die Boltzmann Exploration tut genau das. Anstatt immer die optimale oder zufällige Aktion zu nehmen, beinhaltet dieser Ansatz die Wahl einer Aktion mit gewichteten Wahrscheinlichkeiten. Um dies zu erreichen, wird ein Softmax verwendet. Dabei wird dann am ehesten (aber nicht garantiert) die Richtung gewählt, die der Agent für optimal hält.

$$P_t(a) = \frac{\exp(q_t(a)/\tau)}{\sum_{i=1}^n \exp(q_t(i)/\tau)},$$