Reinforcement Learning

Anmerkung: Die Grenzen zu Dynamic Programming verschwimmen.

Aus Zusammenfassung

- Hier betrachten wir die Umwelt nicht nur propabilistisch, wir kennen die Umwelt nicht einmal.
- Das heißt, der Agent kennt die Umwelt gar nicht, das heißt die einzige Chance, mit der Umwelt umzugehen, ist die Umwelt zu explorieren. - Das nennt sich **Reinforcement Learning**
- Methodisch passiert hierbei folgendes: Aus dynamischem Programmieren mit einem Modell der Welt (Sie nehmen eine Wertefunktion, die Ihnen die optimale Strategie beschreibt. Die Wertefuntion ist rekursiv aufgebaut und sie können zurückiterieren und so die optimale Entscheidung treffen) wird **Q-Learning** oder ältere Formen wie TD-Learning oder SARSA.
- Es ist wichtig zu verstehen, dass wenn ich ein Modell, das ich kenne, stochastisch mache indem ich es ersetze durch Daten, hierbei Q-Learning herauskommt.

Der Agent, der eine Entscheidung treffen soll hat keine Ahnung von der Umwelt, er kann nur mit der Blackbox interagieren. Die Blackbox gibt nur einen Reward für die Aktion aus.

Es ist eine stochastische Variante von dynamischem Programmieren.

Klausurrelevant ist

- "Konzeptionell ist das Wichtigste überhaupt bei Reinforcement Learning, dass man die Grundlage von Reinforcement Learning versteht: Dass man Q-Learning ableiten kann durch eine stochastische Varianz und dynamisches Normieren, und das impliziert auch den Konvergenzbeweis von Q-Learning" --> kann aber in der Klausur nicht getestet werden.
- "Ein Agent läuft in der Welt rum, wie oft müsste er zufällig bis zum Ziel kommen bis er am Start eine Value != 0 hat (einfacher als in Übung)
- Epsilon-greedy Exploration sollte man wissen