

## Bestärkendes oder verstärkendes Lernen (engl. *reinforcement learning*)

Als sequenzielles Entscheiden wird jeder dynamische Prozess verstanden, der eine zeitlich geordnete Folge von Entscheidungen umfasst, die sich durch positive oder negative Verstärkung auf das Erreichen eines bestimmten kumulierten Gewinns auswirken. Auf diese Weise soll die Entscheidungssequenz ermittelt werden, die den erwarteten Gesamtgewinn maximiert.

Neben dem überwachten und dem unüberwachten Lernen stellt das bestärkende Lernen ein drittes Paradigma im Bereich Machine Learning dar. Als solches weist es zwar Gemeinsamkeiten mit den beiden benachbarten Modellen auf, wird jedoch gezielt bei der sequenziellen Entscheidungsfindung angewendet und stützt sich daher auf eine andere Art von Trainingsdaten. Das bestärkende Lernen basiert zu einem großen Teil auf der Theorie der **dynamischen Programmierung**, die Richard Bellman in den 1950er-Jahren formulierte (Bellman, 1957). Wie weiter unten ausgeführt, legt diese Theorie zwar die mathematische Grundlage für das bestärkende Lernen, stößt aber angesichts der erforderlichen Rechenkapazität an ihre Grenzen. In der Praxis bedeutet das, dass die dynamische Programmierung zumindest in ihrer Reinform keine geeignete Lösung für Datenanwendungen mit hohen Durchsatzraten darstellt, wie sie heute typisch sind.

Als besonders bemerkenswerte frühe Anwendung des bestärkenden Lernens gilt das sequenzielle Entscheidungsmodell TD-Gammon, das auf konkurrenzfähigem Niveau Backgammon spielen kann. Ein Beispiel aus der jüngeren Vergangenheit ist das von DeepMind Technologies entwickelte Schachprogramm AlphaZero. Bestärkendes Lernen eignet sich insbesondere für Spielanwendungen und profitiert dabei vor allem auch von der Kombination mit tiefen neuronalen Netzen (engl. *deep reinforcement learning*). So wird das Modell derzeit etwa erfolgreich im Bereich der Videospieldentwicklung eingesetzt.