

Arten des Machine Learnings - Supervised, Unsupervised und Reinforcement Learning

Laura Hartzheim

2018

Inhaltsverzeichnis

1	Einleitung	1
2	Supervised Learning	2
2.1	Klassifikation	2
2.2	Regression	2
3	Unsupervised Learning	4
3.1	Clustering	4
3.2	Dimensionsreduktion	4
3.3	Anomalie Erkennung	4
3.4	Association rule-mining	4
4	Reinforcement Learning	5
5	Schluss	6

1 Einleitung

2 Supervised Learning

Supervised Learning gehört zu den erfolgreichsten und meist verbreiteten Arten des Machine Learnings. (Müller, S.25) Beim Supervised Learning werden bekannte Daten und Ausgaben während dem Trainieren und Prüfen des Modells genutzt, welche auch Training-Daten und Label genannt werden.(Sarkar, S.35) Diese optimieren das Modell, auf Basis der Vorhandenen Daten, durch anpassen der Parameter.(Suthaharan, S.140) Ein Modell besteht aus den in- und output-Paaren des Training Datensatzes. (Müller, S.25) Das Hauptziel ist es die eingehenden Daten x auf die ausgehenden y Abzubilden($f(x) = y$), um später für neue Daten x' die zugehörigen y' Daten zu bestimmen.(Sarkar, S.35) Durch eine größere Menge an Training-Daten ist eine bessere Abdeckung von Verschiedenen Fällen möglich, dies kann aber auch zu Overfitting führen.Um das zu verhindern muss das Training früh genug beendet werden.(Suthaharan, S.140) Es gibt zwei Methoden für Supervised Learning, Klassifikation und Regression. Die Wahl der Methode hängt von der zu erfüllenden Aufgabe ab.(Sarkar, S.35)

2.1 Klassifikation

Das Ziel der Klassifikation ist es ein Klassenlabel für die eingehenden Daten voraus zusagen. Die verschiedenen Label sind Teil einer vorgegebenen Liste. (Müller, S.25) Die Klassifikation kann in binäre und multiklassen Klassifikation aufgeteilt werden. Bei binärer Klassifikation sind nur zwei Klassen verfügbar, die Problemstellung lässt sich also auf eine Ja/Nein-Frage ableiten. In der multiklassen Klassifikation sind mehrere Klassen möglich.(Müller, S.25) Während der Trainings-Phase werden Regeln für das Zuteilen von Labels erstellt, die später dabei helfen Test-Daten Labels zu zuweisen. (Suthaharan, S.8)

2.2 Regression

In Regressions-Problemen sollen oft Zahlen oder Werte ermittelt werden. Im Gegensatz zur Klassifikation gibt es keine Klassen oder Labels denen Daten zugeordnet werden können. Regressions-Modelle lernen den Zusammenhang aus Eingangs- und Ausgangsdaten, um für neue Daten den passenden Output vorherzusagen.(Sarkar, S.37)

Simple linear regression-Modelle versuchen mit nur einer Variable x eine Output-Variable y zu bestimmen und können somit lineare Probleme lösen.(Sarkar, S.37)

Multivariable Regressions Methoden werden für Probleme mit mehreren input-Variablen in Form eines Vektors und nur einer Output-Variable verwendet.(Sarkar, S.38)

Ein Sonderfall der Multivariablen Regression ist die Polynomiale Regression. Hier

ist die Ausgabevariable Polynom n-ten Grades der Eingangsvariable (Sarkar, S.38)
Nichtlineare Regressions Modelle stellen zwischen Ein- und Ausgehendendaten eine
Beziehung auf Basis einer Kombination aus nicht-linearen Funktionen her.(Sarkar,
S.38)

3 Unsupervised Learning

Bei unsupervised Learning haben die Trainings-Daten keine Label. (Sarkar, S.38) Es ist also kein erwarteter/gewünschter Output bekannt und auch schwer feststellbar ob das Modell korrekte Ergebnisse erzielt. (Müller, S.132) Der Algorithmus bekommt nur die Input-Daten und muss anhand dessen Entscheidungen treffen und kategorisieren. (Müller, S.131)

Das Modell lernt inherente Strukturen, Muster und Beziehungen aus dem Datensatz ohne dabei Hilfe von außen zubekommen. (Sarkar, S.38) Hierbei werden Besonderheiten von Daten gefunden. (Kirk, S.16) Die Ergebnisse sind unsicherer als die von supervised Learning Algorithmen, eignen sich aber um weitere Informationen zu den Daten zu finden. (Sarkar, S.38) Weshalb diese Art des Machine Learnings oft in explorativen Bereichen eingesetzt wird um Daten besser zu verstehen. (Müller, S.132) Ein Einsatzgebiet ist als vorverarbeitungs Schritt des supervised Learnings, um neue Representatoren für die Daten zu finden und somit Genauigkeit, Speichernutzung und Geschwindigkeit zu verbessern. (Müller, S.132) Unsupervised Learning kann mit verschiedenen Methoden angewendet werden, diese sind: Clustering, Dimensionsreduktion, Anomalie Erkennung und Association rule-mining. (Sarkar, S.38)

3.1 Clustering

3.2 Dimensionsreduktion

3.3 Anomalie Erkennung

3.4 Association rule-mining

4 Reinforcement Learning

5 Schluss