

Maschinelles Lernen

Zusammenfassung

Thomas Mohr

Contents

1	Grundlagen	3
1.1	(Un)-überwachtes Lernen	3
1.2	Inkrementelles Lernen	3
1.3	Aktives Lernen	3
1.4	Data cleansing	3
1.5	Datensatz	3
2	Deskriptive Statistik	5
3	Regression	5
4	Klassifikation	5
5	Clustering	5
6	Warenkorbanalyse	5
7	Analyse von Graphdaten	5

1 Grundlagen

1.1 (Un)-überwachtes Lernen

- Eine **überwachte** Lernaufgabe liegt vor, wenn wir Beispiele haben, die das zu lernende Attribut bereits tragen (Zielvariable).
 - **Regression** im Fall von kontinuierlichen Werten (z.B. \mathbb{R})
 - **Klassifikation** im Fall von diskreten Labels (z.B. *TRUE*, *FALSE*; ausgezeichnet, durchschnittlich, schlecht)
- Eine **unüberwachte** Lernaufgabe liegt vor, wenn es kein Attribut gibt, das wir lernen wollen und für das wir bereits Beispiele haben.
 - Clustering, also die Unterteilung der Daten in eine Menge von Gruppen
 - Finden von Ausreißern

1.2 Inkrementelles Lernen

- Anstatt das Modell stets von Null an zu lernen, wird das alte Modell mit neuen Beispielen erweitert.
-

1.3 Aktives Lernen

- Aktive Lernverfahren erzeugen die Beispiele selbst, d.h., sie sagen dem Benutzer, welches Tupel benötigt wird.

1.4 Data cleansing

- Fehlende Werte auffüllen
- Rauschen aus den Daten entfernen
- Daten glätten
- Ausreißer entfernen
- Identische Tupel identifizieren
- Daten komprimieren

1.5 Datensatz

- Ein Datensatz ist eine Tabelle
- Eine Instanz (auch Objekt) ist eine Zeile in dieser Tabelle

- Ein Attribut ist ein Feld, das ein Merkmal des Objekts repräsentiert. Mögliche Arten von Attributen sind
 - nominal (kategorisch)
 - * Keine sinnvolle Ordnung
 - * Wir können nicht rechnen (z.B. Mittelwert, Median, Abstände)
 - ordinal (sortierte Kategorien)
 - * Sinnvolle Ordnung
 - * Der Unterschied zwischen zwei Ausprägungen ist i.d.R. unbekannt
 - binär
 - * Können nur zwei Werte annehmen
 - numerisch
 - * Messbare Quantitäten
 - * Abstand zwischen zwei Werten kann quantifiziert werden
 - * Auf den Attributen kann gerechnet werden
 - * Wir unterscheiden
 - diskrete Attribute (endliche oder abzählbar unendliche Menge von womöglichen Ausprägungen)
 - kontinuierliche Werte, reelle Zahlen
 - Attribute mit echtem Nullpunkt (Gewicht, Größe)
 - Attribute ohne echten Nullpunkt (Jahresangaben, Temperatur in °C)
- Ein Datensatz besitzt N Instanzen und d Attribute
 - x_i beschreibt die i -te Instanz
 - x_{ij} beschreibt das j -te Attribut der i -ten Instanz
 - x beschreibt einen d -dimensionalen Vektor
 - Liegt eine überwachte Lernaufgabe vor, so ist das Label der i -ten Instanz t_i

2 Deskriptive Statistik

3 Regression

4 Klassifikation

5 Clustering

6 Warenkorbanalyse

7 Analyse von Graphdaten