

Mustererkennung Tutorium 3

Dr. Ronald Böck,
Olga Egorow

12.06.2018

Worum geht es heute?

- Kenntnisse aus der Mustererkennung in die Tat umsetzen
- Wie geht man vor, wenn man etwas erkennen/klassifizieren will?
- Welche Tools gibt es?
- Umsetzung für ein konkretes Beispiel

Gliederung

- Einzelne Schritte der Mustererkennung
 - Vorverarbeitung Daten
 - Merkmalsextraktion
 - Vorverarbeitung Merkmale
 - Eigentliche Erkennung
 - Auswertung
- Beispielaufgabe

Vorverarbeitung der Daten

- Bereinigung, Klassenzuweisung, etc.
- Welche Daten sollen klassifiziert werden?
- Wie viele Instanzen, wie viele Klassen, wie sind die Instanzen auf die Klassen verteilt?

Merkmalsextraktion

- Welche Merkmale kommen überhaupt in Frage?
- Welche Merkmale sind klassenbestimmend → entscheidend für Klassifizierung?
- Wie ist der Wertebereich der Merkmale definiert → binär/numerisch, symbolisch?

Vorverarbeitung der Merkmale

- Standardisierung, Aufteilung, etc.
- Müssen die Merkmale diskretisiert / eliminiert / angepasst werden?
- Muss der Wertebereich der Daten vergleichbar gemacht werden → Ist eine Datennormierung (Standardisierung / Normalisierung) notwendig
- Müssen die Daten bereinigt werden (Fehlwerte, verrauschte Daten, Ausreißer, Redundanzen, etc.)?

Eigentliche Erkennung I

- Überwachtes Lernen / Klassifikation
 - Lernphase
 - Auswahl von zufälligen Instanzen aus der Datenbasis → Trainingsmenge
 - Diese Instanzen müssen über Klassenzuweisung (Labels) verfügen
 - Erstellen eines Modells (z.B. Stützvektoren bei SVM) → dieses kann zu einem vorgegebenen Merkmalsvektor die Klasse angeben ("Klassifikator")
 - Klassifikationsphase
 - restliche Instanzen (Testmenge) werden mit dem gelernten Modell den Klassen zugeordnet

Eigentliche Erkennung II

- Unüberwachtes Lernen / Clustering
 - Cluster automatisch aus den Daten gebildet, keine vorausgehende Klassenzuweisung erforderlich
 - dazu werden aus den Rohdaten die vorhandenen Strukturen extrahiert
 - ähnliche Objekte in die gleiche Klasse, unähnliche Objekte in unterschiedliche Klassen (Objekte innerhalb einer Klasse möglichst homogen, Cluster möglichst stabil)

Auswertung

- Ziel von Mustererkennung: das gelernte Modell auf ungesehene Daten anzuwenden, diese klassifizieren
→ deswegen Daten vorher aufteilen in Trainings-/Testmenge
- Welche Fehler sind möglich?
 - Kontingenztafel
- Wie wird die Performanz gemessen?
 - Recall: $\text{TruePositives}/(\text{TruePositives}+\text{FalseNegatives})$
 - Precision: $\text{TruePositives}/(\text{TruePositives}+\text{FalsePositives})$

		Wirkliche Klasse	
		A	B
Klassifizierung	A	TP	FP
	B	FN	TN

Beispielaufgabe



Datenmenge:

- 3 Tomaten:
 - 1 grüne, 2 rote (normal & Datteltomate)
- 3 Birnen:
 - 1 rote, 1 grüne, 1 gelbe Nashi-Birne
- 3 Äpfel:
 - 1 roter, 1 gelber, 1 grüner

Beispielaufgabe

- Welche Merkmale kommen in Frage?
 - Farbe: rot, gelb, grün
 - Form: länglich, rund
- Sind diese Merkmale ausreichend für eine Klassifikation?
- Welche Merkmale könnte man noch hinzunehmen?

Farbe	Form	Klasse
Grün	Rund	Tomate
Rot	Rund	Tomate
Rot	Länglich	Tomate
Rot	Länglich	Birne
Grün	Länglich	Birne
Gelb	Rund	Birne
Rot	Rund	Apfel
Gelb	Rund	Apfel
Grün	Rund	Apfel

Umsetzung

- Demo / Klassifikation mit Hilfe von KNIME

