

11. Übungsblatt

(zum 21., 22., 26. bzw. 27.01.2015)

Aufgabe 38 Maschinelles Lernen: Entscheidungsbaum

Entlassungswelle in einer großen Bank: Aus (nicht ganz) heiterem Himmel bekommen Angestellte ihre Kündigung oder werden an andere Arbeitsplätze versetzt. Im Betriebsrat grübelt man, welche Empfehlungen die kürzlich angeheuerte Unternehmensberatung der Unternehmensführung gegeben haben mag. Folgende Daten über die Entlassungen wurden bisher gesammelt:

Pers.-Nr.	Abteilung	Firmenzugehörigkeit	Alter	Tätigkeit	Klassifikation
1	EDV	kurz	jung	Sachbearbeiter	+
2	EDV	kurz	jung	Führungsposition	-
3	EDV	kurz	älter	Führungsposition	+
4	EDV	lang	älter	Sachbearbeiter	-
5	Kundenbetreuung	kurz	jung	Führungsposition	+
6	Kundenbetreuung	lang	älter	Führungsposition	+
7	Kundenbetreuung	kurz	jung	Sachbearbeiter	-
8	Marketing	kurz	jung	Sachbearbeiter	-
9	Marketing	lang	jung	Führungsposition	-
10	Marketing	lang	älter	Sachbearbeiter	+

Kurze Firmenzugehörigkeit heißt hier *weniger als fünf Jahre*, *jung* bedeutet *jünger als 35*. Die Klassifizierung + bedeutet, dass eine Kündigung oder Versetzung ausgesprochen wurde.

- Bestimmen Sie einen Entscheidungsbaum zur Klassifizierung der Kündigungen bzw. Versetzungen. Gehen Sie bei der Attributauswahl wie folgt vor: Das Attribut, das die meisten Beispiele eindeutig klassifiziert, wird für die nächste Verzweigung ausgewählt. Gibt es kein solches Attribut oder mehrere Attribute mit derselben Anzahl eindeutig klassifizierter Beispiele, gilt die Reihenfolge der Attribute in der Tabelle (von links nach rechts).
- Bestimmen Sie die entsprechenden Entscheidungsregeln.
- Interpretieren Sie das Ergebnis: Wie könnte der Rat der Unternehmensberatung gelaute haben?
- Jemand behauptet, der Unternehmensberatung käme es auf die Abteilungen gar nicht an, sie wolle sowieso nur *gewisse Leute loswerden* oder sich profilieren. Können Sie diese Behauptung widerlegen?

Aufgabe 39 Maschinelles Lernen: Assoziationsregeln

Gegeben seien folgende Transaktionen:

TID	Itemset
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C
10	A, B, C, F

- a) Berechnen Sie auf dieser Grundlage die *frequent itemsets* (minimum support = 0.2). Verwenden Sie hierzu den apriori-Algorithmus.
- b) Bestimmen Sie aus den *frequent itemsets* aus Teilaufgabe a) alle Regeln, die einen minimalen Konfidenzwert von 0.8 überschreiten.

Aufgabe 40 Maschinelles Lernen: Entscheidungsbaum

Bestimmen Sie analog zur zweiten Aufgabe des letzten Übungsblattes einen Entscheidungsbaum der anhand des folgenden Datensatzes das Attribut „Klassifizierung“ bestimmt, welches aussagt, ob es bei den derzeitigen Wetterbedingungen ratsam ist, Tennis zu spielen. Wenn zwei oder mehr Attribute die gleiche Qualität liefern, wählen Sie jenes aus, welches weiter links in der Tabelle steht.

Ausblick	Temp. ($^{\circ}F$)	Luftfeuchtigkeit (%)	Windig?	Klassifizierung
sonnig	85	85	nein	Nicht spielen
sonnig	80	90	ja	Nicht spielen
bewölkt	83	78	nein	Spielen
regnerisch	70	96	nein	Spielen
regnerisch	68	80	nein	Spielen
regnerisch	65	70	ja	Nicht spielen
bewölkt	64	65	ja	Spielen
sonnig	72	95	nein	Nicht spielen
sonnig	69	70	nein	Spielen
regnerisch	75	80	nein	Spielen
sonnig	75	70	ja	Spielen
bewölkt	72	90	ja	Spielen
bewölkt	81	75	nein	Spielen
regnerisch	71	80	ja	Nicht spielen