

# Data Mining und Maschinelles Lernen



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Wintersemester 2015/2016  
12. Übungsblatt

## Aufgabe 1 Relief

Gegeben sind folgende 12 Beispiele der Wetter-Daten:

ID	outlook	temperature	humidity	windy	play
1	sunny	hot	high	FALSE	no
2	rainy	mild	high	FALSE	yes
3	rainy	cool	normal	FALSE	yes
4	rainy	cool	normal	TRUE	no
5	overcast	cool	normal	TRUE	yes
6	sunny	mild	high	FALSE	no
7	sunny	cool	normal	FALSE	yes
8	rainy	mild	normal	FALSE	yes
9	sunny	mild	normal	TRUE	yes
10	overcast	mild	high	TRUE	yes
11	overcast	hot	normal	FALSE	yes
12	rainy	mild	high	TRUE	no

Berechnen Sie die Relief Feature-Gewichte (Foliensatz “Instanzenbasiertes Lernen”, Folie “RELIEF (2)”) für alle 4 Attribute (die ID ist nur zur leichten Identifizierung eines Beispiels vorhanden). Berechnen Sie den Nearest Hit und Nearest Miss für jedes Beispiel ( $r = 12$ ). Gehen Sie davon aus, dass jedes Beispiel genau einmal gewählt wird (Es kann aber natürlich mehrfach als Nearest-Neighbor auftauchen.). Als Distanz-Funktion nehmen Sie einfach die Anzahl der verschiedenen Attribute.

## Aufgabe 2 Diskretisierungsmethoden

Gegeben sei folgende Version der Wetter-Daten mit 12 Trainings-Beispielen und 2 numerischen Attributen.

ID	outlook	temperature	humidity	windy	play
1	sunny	85	85	FALSE	no
2	rainy	70	96	FALSE	yes
3	rainy	68	80	FALSE	yes
4	rainy	65	70	TRUE	no
5	overcast	64	65	TRUE	yes
6	sunny	72	95	FALSE	no
7	sunny	69	70	FALSE	yes
8	rainy	75	80	FALSE	yes
9	sunny	75	70	TRUE	yes
10	overcast	72	90	TRUE	yes
11	overcast	81	75	FALSE	yes
12	rainy	71	91	TRUE	no

Diskretisieren Sie die beiden numerischen Attribute mit den Verfahren, die Sie in der Vorlesung kennen gelernt haben:

- equal-width
- equal-frequency
- chi-merge
- info-split

Wählen Sie die Anzahl der Intervalle so, daß Sie die bekannten Daten erhalten könnten (drei Werte für Temperature, zwei für Humidity). Vergleichen Sie die Resultate miteinander und mit den bekannten Daten (die aus der vorherigen Aufgabe).