

Maschinelles Lernen und Data Mining

Übungsblatt für den 26.1.2006

Aufgabe 1

Gegeben sind folgende 12 Beispiele der Wetter-Daten:

```
@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}
```

```
@data
sunny,hot,high,FALSE,no
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
```

Berechnen Sie die Relief Feature-Gewichte für alle 4 Attribute. Berechnen Sie den Nearest Hit und Nearest Miss für jedes Beispiel ($m = 12$). Als Distanz-Funktion nehmen Sie einfach die Anzahl der verschiedenen Attribute.

Aufgabe 2

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

```
@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}
```

```
@data
sunny,85,85,FALSE,no
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
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 (2 Beispiele wurden entfernt, siehe Aufgabe 1).