

Maschinelles Lernen und Data Mining

Übungsblatt für den 11.1.2005

Aufgabe 1

Gegeben sei folgende Version der Wetter-Daten, die 2 numerische Attribute enthält.

```
@relation weather
```

```
@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
```

```
sunny,80,90,TRUE,no
```

```
overcast,83,86,FALSE,yes
```

```
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 ursprünglichen Werte erhalten können (drei Werte für Temperature, zwei für Humidity), aber betrachten Sie auch die um eins höher oder niedriger liegenden Werte.

Aufgabe 2

Führen Sie auf den diskretisierten Daten (oder den ursprünglichen Daten) eine Feature Subset Selection mit dem Relief-Verfahren durch (i.e., berechnen Sie die Feature-Gewichte für alle 4 Attribute).