

**Information Mining - Wintersemester 2014**

Dipl.-Inform. Vu Tran, LF 139

vtran@is.inf.uni-due.de

**Übungsblatt 5**

---

**Aufgabe 10: Numerische Vorhersage mit *RapidMiner***

Es soll die Geschwindigkeit von CPUs<sup>1</sup> vorhergesagt werden.

- Erstelle mit *RapidMiner* einen Prozess, der eine lineare Regressionsfunktion lernt.
- Wie sieht die gelernte Regressionsfunktion aus?
- Welcher Wert wird für eine CPU mit den folgenden Daten vorhergesagt?  
MYCT=270, MMIN=4000, MMAX=8000, CACH=120, CHMIN=12, CHMAX=32

**Aufgabe 11: Naiver Bayes**

Beim Lernverfahren *Naiver Bayes* wird geschätzt, mit welcher Wahrscheinlichkeit eine Instanz  $E$  (Evidenz) zu einer Klasse  $H$  (Hypothese) gehört:

$$Pr(H|E) = \frac{Pr(E|H) \cdot Pr(H)}{Pr(E)}$$

- Welche Annahmen liegen dem Naiven Bayes zugrunde?
- Was geben die folgenden Wahrscheinlichkeiten an:  
 $Pr(E|H)$ ,  $Pr(H|E)$ ,  $Pr(H)$ ,  $Pr(E)$ ?

Es seien die Wetterdaten 1 aus der Vorlesung gegeben. `play` ist das Klassenattribut, `Tag` dient nur zur Identifizierung von Tagen und ist kein reguläres Attribut.

Gegeben seien ferner die beiden folgenden Instanzen:

- `outlook = rainy, temperature = hot, humidity = normal, windy = false`
- `outlook = overcast, temperature = hot, humidity = normal, windy = true`

- Berechne die Wahrscheinlichkeiten  $Pr(H|E)$  für die beiden Instanzen.  
Benutze einen modifizierten Wahrscheinlichkeitsschätzer (siehe Vorlesung) mit  $\mu = 1$  (Attribut `outlook`), wenn das Problem mit der Häufigkeit 0 auftritt.
- Zu welchen Klassen werden diese beiden Instanzen zugewiesen?

---

<sup>1</sup>[http://www.is.inf.uni-due.de/courses/im\\_ws14/uebung/data\\_a10.arff](http://www.is.inf.uni-due.de/courses/im_ws14/uebung/data_a10.arff)

Tag	outlook	temperature	humidity	windy	play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rainy	mild	high	false	yes
5	rainy	cool	normal	false	yes
6	rainy	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rainy	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rainy	mild	high	true	no

Abbildung 1: Die Wetterdaten