

Data Mining und Maschinelles Lernen



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Wintersemester 2015/16

8. Übungsblatt

Aufgabe 1 RISE

Gegeben sei folgende Beispielmenge:

Outlook	Temperature	Humidity	Wind	PlayTennis
Overcast	Hot	High	False	Yes
Rainy	Cool	Normal	True	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	High	False	No
Overcast	Mild	High	True	Yes
Sunny	Cool	Normal	False	Yes
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	True	No
Sunny	Hot	High	False	No
Sunny	Mild	Normal	True	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Hot	High	True	No
Rainy	Mild	High	False	Yes
Overcast	Hot	Normal	False	Yes

- a) Wenden Sie den Algorithmus RISE (Foliensatz “*Instance-based Learning*”, Folie “*RISE (Domingos, 1996)*”) auf den obigen Datensatz an. Berechnen Sie zur Vereinfachung nur die *erste Iteration* und diese nur für die *negativen Beispiele* und nur für die Regeln aus der Theorie mit *negativem Head*.

Bedenken Sie, dass bei der Berechnung der Accuracy für ein Beispiel die Regel, die aus diesem Beispiel selbst erzeugt wurde, nicht einbezogen wird, da dies bereits zu einer perfekten Klassifikation führen würde und der Algorithmus somit bereits beendet wäre. Beachten Sie auch, daß die Accuracy für die gesamte Theorie berechnet wird, obwohl wir in der Iteration nur die negativen Regeln daraus betrachten.

Benutzen Sie zur Berechnung der Distanz der Attribute die *Value Difference Metric* und nehmen Sie hierbei an, dass $k = 1$ gilt. Als Vereinfachung müssen Sie die *VDM* nicht normieren.

Zur Bestimmung des Abstandes eines Beispiels zu einer Regel verwenden Sie die euklidische Distanz, wobei Sie das Ziehen der Wurzel weglassen können.

Müssen Sie 2 Regeln zusammenfassen, so ist es Voraussetzung, dass diese die gleiche Klasse vorhersagen.

Aufgabe 2 KD-Trees

- a) Bauen Sie einen KD-Tree aus der folgenden 2D Punktmenge auf und zeichnen Sie sowohl den Baum als auch die grafische Lösung im 2D-Raum:
 $\{\{4, 8\}, \{7, 4\}, \{5, 10\}, \{1, 3\}, \{2, 2\}, \{9, 1\}, \{10, 7\}, \{8, 10\}, \{6, 6\}\}.$

Verwenden Sie bei gerader Anzahl Punkte den größeren Wert bei der Bestimmung des Medians.

- b) Wenden Sie 1-NN für die folgenden beiden Queries $\{7, 9\}$ und $\{1, 1\}$ auf den Baum an und geben Sie die genaue Traversierung des Baumes an.