

Data Mining und Maschinelles Lernen



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Wintersemester 2015/16
11. Übungsblatt

Aufgabe 1 AdaBoost

Rechnen Sie das AdaBoost-Beispiel aus der Vorlesung (Ensemble-Methoden, Folien 15ff) nach. Verwenden Sie für die einzelnen Datenpunkte die folgenden Koordinaten (x, y, Klasse):

1, 5, +	3, 1, -
2, 2, +	4, 6, -
5, 8, +	7, 4, -
6, 10, +	9, 3, -
8, 7, +	10, 9, -

Als Basis-Lerner sollen Decision Stumps (also waagrechte bzw. senkrechte Splits, z.B. $x > 4 \rightarrow +$) verwendet werden. Der Basislerner wählt unter allen möglichen Splits jenen aus, bei dem die Gesamtsumme der Gewichte der falsch klassifizierten Beispiele minimiert wird. Wählen Sie bei Gleichstand den zuerst gefundenen Test beginnend mit vertikalen Splits mit aufsteigenden Thresholds.

- Berechnen Sie die ersten 3 AdaBoost-Iterationen.
- Generieren Sie aus den eben berechneten Decision Stumps einen Entscheidungsbaum.

Aufgabe 2 Stacking

In dieser Aufgabe sollen Sie unter Verwendung mehrerer Basislerner und der Ensemble-Methode Stacking einen Entscheidungsbaum lernen. Verwenden Sie hierfür den Datensatz und die drei Decision Stumps aus der vorherigen Aufgabe.

- Konvertieren Sie diesen Datensatz, d.h. ersetzen sie die Attribute durch eine neue Attributmenge, die jeweils ein Attribut für jeden Decision Stump beinhaltet. Als Attributwerte werden die Vorhersagen des entsprechenden Klassifizierers verwendet.
- Bestimmen Sie nun auf dem konvertierten Datensatz einen Entscheidungsbaum mittels des Verfahrens ID3 und Maß Information Gains. Entscheiden Sie sich bei Gleichstand z.B. für den als erstes gefundenen Test.
- Zeichnen Sie diesen Baum und vergleichen Sie ihn mit dem Entscheidungsbaum aus Aufgabe 1.