

Übung zur Vorlesung **Maschinelles Lernen**

Wintersemester 2008/2009
Blatt 6

Wiederholung

1. Wie genau fließt das Skalarprodukt in die Klassifikation mittels separierender Hyperebene ein? Wie wird durch das Skalarprodukt die Ähnlichkeit von Vektoren gemessen?
2. Gehen Sie die Entwicklung des dualen Problems aus der Formulierung der Optimierungsaufgabe nach Lagrange noch einmal durch!

Aufgabe 1

3 Punkte

1. Die meisten SVM-Implementierungen halten die *Kernelmatrix* K über den Beispielen mit

$$K_{i,j} = \langle \vec{x}_i, \vec{x}_j \rangle$$

im Speicher. Überlegen Sie sich, wieviel Speicher das Training eines SVM-Modells mit 100 (1000) Beispielen braucht, wenn jede Vektorkomponente als 64-Bit `double` gespeichert wird.

2. Zeigen Sie die Äquivalenz der Nebenbedingungen

$$\left. \begin{array}{l} \langle \vec{x}, \vec{\beta} \rangle + \beta_0 \geq +1 \quad \forall \vec{x} \in C_+ \\ \langle \vec{x}, \vec{\beta} \rangle + \beta_0 \leq -1 \quad \forall \vec{x} \in C_- \end{array} \right\} \Leftrightarrow y \left(\langle \vec{x}, \vec{\beta} \rangle + \beta_0 \right) - 1 \geq 0 \quad \forall \vec{x} \in \mathbf{X}.$$

Aufgabe 2

7 Punkte

Langsam sollte sich Ihre Implementierung des ID3-Algorithmus vervollständigen. Nachdem auf dem letzten Übungsblatt die Auswahl des besten Split-Attributes implementiert werden sollte, soll Ihr Operator in diesem Blatt um die Möglichkeit erweitert werden, numerische Attribute zu behandeln.

1. Erweitern Sie das ID3-Modell Ihrer Implementierung so, daß numerische Attribute innerhalb der Methode

```
public int[] split( ExampleSet examples, Attribute attribute )
```

entsprechend des Vorgehens aus der Vorlesung behandelt werden.

2. Ergänzen Sie Ihren zugehörigen Lern-Operator so, daß innerhalb von RapidMiner auch numerische Attribute in den ExampleSets möglich sind.