

Übungen zur Vorlesung  
**Wissensentdeckung in Datenbanken**  
Sommersemester 2007  
Blatt 9

**Aufgabe 9.1**

In der Vorlesung haben Sie die Stützvektormethode (SVM) kennen gelernt. In der ersten Aufgabe sollen Sie einleitend grundlegende Begriffe der SVM erläutern.

- (a) Was versteht man unter 'empirical risk minimization', und warum wird dieses Verfahren benutzt?
- (b) Warum benutzt man für die SVM eine Darstellung in Vektorform?
- (c) Was versteht man unter der optimal separierenden Hyperebene? Erläutern Sie in diesem Zusammenhang auch den Begriff Margin!

**Aufgabe 9.2**

Benutzen Sie nun Yale, um mittels SVM auf einem Trainingsdatensatz zu lernen und schließlich das gelernte Modell auf einem Testdatensatz anzuwenden.

- (a) Erzeugen Sie vorerst einen Trainingsdatensatz aus folgenden Daten:

Label	Größe	Gewicht
little	1	1
little	3	4
little	3	6
big	4	8
little	5	2
big	7	8
little	9	4
big	9	8
big	10	6
big	12	8

Erzeugen Sie hierfür eine 'Leerzeichen'-separierte Textdatei, die Sie mit dem ExampleSource-Operator einlesen können. Hierfür müssen Sie im Operator den 'Configuration- Wizard' benutzen. Dieser hilft Ihnen, Schritt für Schritt aus der txt-Datei eine aml-Datei zu erstellen, die Sie in Yale benutzen können. Lassen Sie Ihren Versuch laufen und geben Sie den Scatter-plot (Größe u. Gewicht gegen das Label geplottet) mit ab.

- (b) Benutzen Sie den JMySVMClassifier (unter Kernel-Learner) zum Lernen.
- (c) Erstellen Sie wie in der vorangegangenen Aufgabe einen Testdatensatz, der zwei Daten mit (Größe, Gewicht) (5,3) und (8,7) enthält.
- (d) Lesen Sie den Testdatensatz ein, und benutzen Sie den ModelApplier, um das gelernte Modell anzuwenden. Geben Sie die resultierende 'data-view' an, und zeigen Sie, welcher Klasse die beiden Daten jeweils zugeordnet wurden.

### **Aufgabe 9.3**

Benutzen Sie den vorangegangenen Versuch, um die separierende Hyperebene zu skizzieren. Beachten Sie dabei, beim JMySVMClassifier 'scale' und 'calculate-weights' auszuwählen. Nun wird zusätzlich zum Model noch der Normalenvektor der separierenden Hyperebene ausgegeben. Diesen, sowie den Bias  $b$  und die normalisierten Werte (im Reiter 'Model') benötigen Sie zur Skizzierung der Hyperebene. Geben Sie die Skizze der Werte und der separierenden Hyperebene mit ab. Achtung: Die Werte entsprechen nun nicht mehr den originalen sondern den normalisierten Werten!

Geben Sie alle erstellten Experiment- sowie Datensatz-Dateien mit ab!