

## Übungsblatt 5: Support Vector Machines

**Abgabeschluss:** Montag, der 7. Juli 2008 um 10:00 Uhr.

Für dieses Aufgabenblatt sind sowohl Code als auch eine schriftliche Ausarbeitung abzugeben. Der Code und eine elektronische Version der Ausarbeitung (als PDF) muss über PASS abgegeben werden (siehe link auf der Website). Zusätzlich ist ein Ausdruck der Ausarbeitung abzugeben, entweder in der Veranstaltung oder über unser Sekretariat.

### Aufgaben

#### Teil 1: Implementation

##### Aufgabe 1 (20 Punkte)

Implementiere den SMO Algorithmus für Support-Vektor-Maschinen. Die Signatur orientiert sich wieder an Kernel-Ridge-Regression:

```
C = svm_smo(X, y, kernel, kernelparameter, regularization)
```

mit den folgenden Kernen (und Parametern)

Name	Kern	Parameter
linear	$k(x, z) = \langle x, z \rangle$	(keiner)
polynomial	$k(x, z) = (\langle x, z \rangle + 1)^p$	Grad $p \in \{1, 2, 3, \dots\}$
gaussian	$k(x, z) = \exp(-\ x - z\ ^2 / 2dw^2)$	Kernbreite $w$

Hier ist  $d$  die Dimension der  $X$ .

Verwende die Debug-Datasets (siehe unten), um Deine Implementation zu testen.

##### Aufgabe 2 (10 Punkte)

Implementiere eine Support-Vektor-Maschine mit Hilfe der Funktion `quadprog` aus der Matlab Optimization Toolbox. Signatur und Aufbau ist identisch mit `svm_smo`.

```
C = svm_quadprog(X, y, kernel, kernelparameter, regularization)
```

Verwende die Debug-Datasets (siehe unten), um Deine Implementation zu testen.

#### Teil 2: Anwendung

Wende beide SVM Implementationen auf die folgenden Datensätze an. Wenn sich die gefundenen Lösungen unterscheiden, stelle sie beide dar und weise auf die Unterschiede hin.

##### Aufgabe 3 (10 Punkte)

1. Zum Debuggen verwende die Datensätze `U05_easy_1d` und `U05_easy_2d` mit einem linearen Kern. Plote die resultierenden Entscheidungsebenen und markiere die Supportvektoren.
2. Wenden den Algorithmus auf den `U05_spiral` Datensatz mit einem Gaussschen Kern an. Finde manuell gute Einstellungen für die Kernbreite und Regularisierungskonstante. Plote die Entscheidungsebene und gib dabei die Werte für die Parameter an.

#### Aufgabe 4 (10 Punkte)

1. (USPS-Datensatz) Wir möchten jetzt die SVM auf einem “richtigen” Datensatz anwenden, dem USPS Datensatz, den wir auf Blatt 1 und 3 bereits betrachtet wurden. Hierbei wird jeweils eine Ziffer gegen den Rest trainiert.
  - (a) Wende 5-fold Cross-Validation an (siehe Blatt 4), um den besten Kern und Kernparameter zu finden. Gib den besten Kern und Kernparameter sowie den geschätzten Testfehler an.
  - (b) Plote für jeden Kern jeweils 5 zufällig gewählte Support-Vektoren aus der positiven und der negativen Klasse.