

## Maschinelles Lernen und Data Mining

Sommersemester 2011

### Übungsblatt 6

Besprechung des Übungsblattes am 30.06./01.07.2011

#### Aufgabe 6-1 Minimale Oberfläche

Ein geschlossener Karton soll ein Fassungsvermögen von  $36 \text{ cm}^3$  haben. Zusätzlich soll die Breite seiner Grundfläche genau die dreifache Länge der Grundfläche betragen.

Berechnen Sie Länge, Breite und Höhe des Kartons mit der kleinsten Oberfläche.

#### Aufgabe 6-2 Bestimmung der optimal-trennenden Hyperebene

Bestimmen Sie die optimale, separierende Hyperebene des zwei-Klassen-Datensatzes  $(A, B)$  mit:

$$A = \left\{ p_1 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, p_2 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, p_3 = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}, p_4 = \begin{pmatrix} 2.5 \\ 3 \end{pmatrix}, p_5 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\},$$
$$B = \left\{ p_6 = \begin{pmatrix} 0.5 \\ 1.5 \end{pmatrix}, p_7 = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, p_8 = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix} \right\}$$

Die Klassenlabels  $y$  der Klasse  $A$  sind deklariert mit 1, die von  $B$  mit  $-1$ .

Visualisieren Sie das Ergebnis und geben Sie die Supportvektoren an. Wie breit ist die Margin?

#### Aufgabe 6-3 Modellvergleich

Vergleichen Sie Modelle der Regression mit Basisfunktionen. Die Vorhersage für einen Datenpunkt  $\mathbf{x}_i \in \mathbb{R}$  sei gegeben durch:

$$f(\mathbf{x}_i, \mathbf{w}) = \sum_{j=1}^{M_\Phi} w_j \phi_j(\mathbf{x}_i)$$

Verwenden Sie die PLS-Lösung  $\hat{\mathbf{w}} = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T \mathbf{y}$  mit  $\Phi_{i,j} = \phi_j(\mathbf{x}_i) = \mathbf{x}_i^{j-1}$ . Gegeben sei der Datensatz  $\mathbf{X}, \mathbf{y}$  der Größe  $N = 10$ , basierend auf einer Rauschvarianz von  $\sigma^2 = 0.25$ :

$\mathbf{X}$	0.3	0.4	0.8	1.5	1.8	3.6	4	4.3	4.6	5
$\mathbf{y}$	7	4.7	0.6	-1.1	-0.3	4.6	5.5	5.7	3.1	-0.3

Es soll das beste Modell für Basisfunktionen mit  $M_\Phi \in \{1, \dots, 6\}$  bestimmt werden. Als Loss-Funktion nehmen Sie im folgenden den mittleren quadratischen Fehler (MSE).

- a) Bestimmen Sie das beste Modell durch Kreuzvalidierung (5-fach und 10-fach). Unterstützen die paarweisen Tests aus der Vorlesung die Entscheidung des MSE? Welchen Einfluss hat der  $\lambda$ -Parameter?

b.w.

- b) Kommen Sie mit den frequentistischen Verfahren ( $C_p$  Statistik und AIC) und dem Bayes'schen Verfahren (BIC) zu den gleichen Schlüssen?
- c) Welchen Einfluss hat die Datengröße  $N$ , wenn Sie einen vergleichbaren Datensatz für  $N = \{100, 1000\}$  simulieren?