

Übungen zur Vorlesung
Wissensentdeckung in Datenbanken
Sommersemester 2007
Blatt 10

Aufgabe 10.1

In der Vorlesung wurde zur Optimierung der SVM-Parameter das Verfahren SMO vorgestellt.

- (a) Warum optimiert man hierbei nicht alle Parameter gleichzeitig, und wie viele Parameter werden statt dessen gleichzeitig optimiert und wieso?
- (b) Gehen Sie davon aus, dass $C = 0, 1$ ist. Welchen Wert kann α_2 mindestens und welchen Wert maximal annehmen? Lösen Sie diese Aufgabe zeichnerisch, und leiten Sie Ihre Lösung daraus ab. Gehen Sie einmal von $y_1 = y_2$ und dann von $y_1 \neq y_2$ aus.

Aufgabe 10.2

Sie haben außerdem die strukturelle Risikominimierung kennengelernt (Folien-SVM2, Folie 31 ff.), die dazu dient, die Ausdrucksstärke von Modellklassen zu messen. Sie basiert auf der in Folien als η notierten VC Dimension.

- (a) Das empirische Risiko zweier jeweils bester Modelle aus unterschiedlichen Modellklassen sei identisch, die erste Modellklasse besitze endliche, die zweite unendliche VC Dimension. Welches Modell empfiehlt sich hinsichtlich der strukturellen Risikominimierung und warum?
- (b) Wie groß ist die VC Dimension der Modellklasse unbeschränkt tiefer Entscheidungsbäume für binäre Klassifikationsprobleme bei mindestens einem kontinuierlichen Merkmal? Begründen Sie bitte Ihre Ansicht.
- (c) Wir betrachten die Modellklasse der Kreise im \mathbb{R}^2 (für Zweiklassenprobleme). Die Klasse besitzt die Parameter "Mittelpunkt" und "Radius". Jedes Modell klassifiziert genau die Punkte innerhalb des Kreises als positiv. Beweisen Sie, zum Beispiel graphisch, dass diese Hypothesenklasse eine VC Dimension von mindestens 3 besitzt.

Aufgabe 10.3

Die Regressions-SVM wurde für Zeitreihen vorgestellt. Bei Zeitreihen kann es zu Trends oder Zyklen kommen, die zu beachten sind. Geben Sie zusätzlich zu den in der Vorlesung und auf den Folien vorgestellten Zeitreihen drei Beispiele für Datenreihen, die Zyklen oder Trends enthalten.