

**Knowledge Discovery in Databases II**  
SoSe 2008

**Übungsblatt 7: Multirepräsentiertes Data Mining**

Besprechung am Freitag, 7.6.2008

**Aufgabe 7-1** *Kombination von zwei Ähnlichkeitsmaßen*

Gegeben seien zwei Kernels  $k_1$  und  $k_2$ . Wir kombinieren sie in einen gemeinsamen Kernel  $k_{com}$

$$k_{com} = \alpha k_1 + (1 - \alpha) k_2, \quad (1)$$

wobei  $\alpha \in [0; 1]$ .

Wir wenden den Kernel  $k_{com}$  auf zwei Klassifikationsprobleme an, wobei wir das Experiment für verschiedene Werte von  $\alpha$  wiederholen. In der folgenden Abbildung stellen wir die Klassifikationsgenauigkeit auf dem ersten Datensatz (R1) und auf dem zweiten Datensatz (R2) in Abhängigkeit von  $\alpha$  dar:

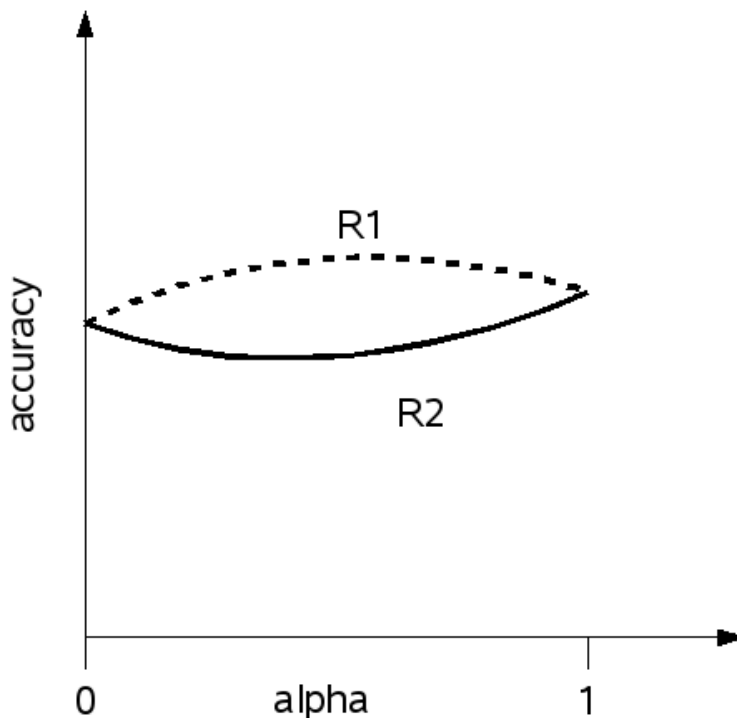


Abbildung 1: Classification accuracy vs.  $\alpha$

Beantworten Sie folgende Fragen anhand von Abbildung 1:

- Auf welchem der beiden Datensätze lohnt sich die Kombination der beiden Kernels?
- Wann funktionieren die Kernels  $k_1$  und  $k_2$  alleine besser als kombiniert?

**Aufgabe 7-2**      *Komplementarität von Klassifikatoren*

Gegeben seien zwei binäre Klassifikatoren  $f_1$  und  $f_2$ , die auf je einer Repräsentation der Objekte eines Datensatzes  $D$  mit Klassen  $\{0, 1\}$  arbeiten. Entscheiden Sie, ob in den folgenden Fällen eine Kombination der Klassifikatoren sinnvoll ist:

- (a)  $f_1(x) = f_2(x)$  für alle  $x \in D$
- (b)  $f_1(x) = 1 - f_2(x)$  für alle  $x \in D$

**Aufgabe 7-3**      *Abhängigkeitsmaß*

Gegeben sei ein Maß  $h$ , welches die Abhängigkeit zwischen zwei Kernelmatrizen  $K$  und  $K'$  misst. Anschaulich heißt das, dass  $h(K, K')$  groß ist, wenn die zugehörigen Kernels  $k$  und  $k'$  dieselben Objekte als ähnlich und als unähnlich betrachten. Wenn sie die Ähnlichkeit derselben Objekte unterschiedlich bewerten, sei  $h(K, K')$  niedrig.

Seien nun ein Datensatz  $D$  mit einem Klassenlabel und  $r$  Repräsentationen pro Objekt gegeben. Wir berechnen eine Kernelmatrix  $K_i$  für jede der  $r$  Repräsentationen und eine Kernelmatrix  $L$  auf den Klassenlabels. Überlegen Sie sich, wie man mittels  $h$  eine Linearkombination der  $K_i$  bestimmen kann, die die Ähnlichkeit der Klassenlabels möglichst gut widerspiegelt.