

**Knowledge Discovery in Databases II**  
WiSe 2011/12

**Übungsblatt 7: Multi-Repräsentiertes Data Mining**  
Besprechung am Freitag, 16.12.2011

**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:

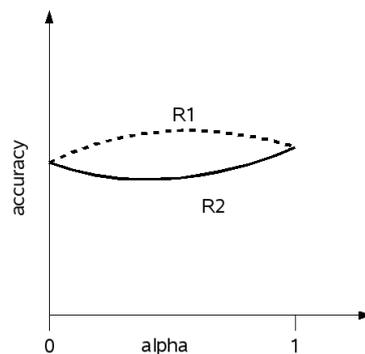


Abbildung 1: Classification accuracy vs.  $\alpha$

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** *Multi-Repräsentierte Klassifikation*

Gegeben sei ein Datensatz mit multiplen Repräsentationen jedes Datenobjekts. Wir möchten Klassenzugehörigkeiten mittels dieser multiplen Repräsentationen ermitteln.

- In welcher Phase des Klassifikationsprozesses können wir die verschiedenen Repräsentationen integrieren?
- Wie können wir die multiplen Repräsentationen beim Trainieren integrieren?
- Wie können wir die multiplen Repräsentationen beim Vorhersagen integrieren?
- Ist in beiden Fällen zuvor eine Normalisierung erforderlich?