

Prof. Dr. Katharina Morik,  
Prof. Dr. Claus Weihs,  
Dr. Wouter Duivesteijn,  
M.Sc. Sarah Schnackenberg,  
B.Sc. Melanie Dagge

Dortmund, 04.06.14  
Abgabe: bis Do, 11.06.2015,  
10 Uhr, an  
[wouter.duivesteijn@tu-dortmund.de](mailto:wouter.duivesteijn@tu-dortmund.de)  
und/oder in den Briefkasten "Duivesteijn"  
im OH12, R4.005

Übungen zur Vorlesung  
**Wissensentdeckung in Datenbanken**  
Sommersemester 2015

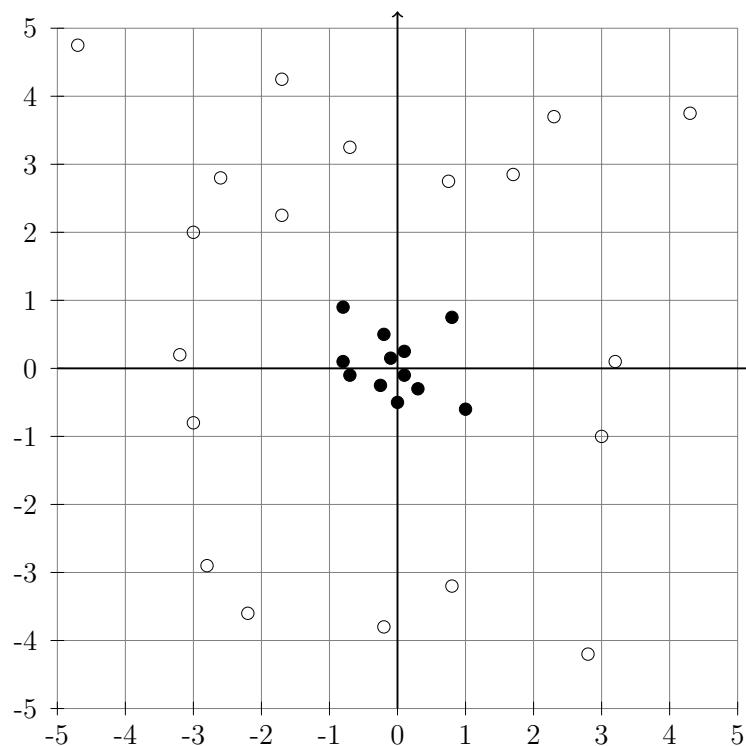
Blatt 8

**Aufgabe 8.1 (4 Punkte)**

Betrachten Sie den folgenden Datensatz! Fällt Ihnen eine Transformation ein, die zu einer linearen Trennung der Daten führen könnte?

Den Datensatz finden Sie unter:

<http://www-ai.cs.uni-dortmund.de/LEHRE/VORLESUNGEN/KDD/SS15/DATA/svm-data.csv>



1. (2 Punkte) Beschreiben Sie, wie Sie die gegebenen Daten trennen würden! Geben Sie dazu in kurzen Stichpunkten an, welche Eigenschaften der Daten Sie ausnutzen würden.  
*Hinweis:* Denken Sie daran, dass die  $\Phi$ -Transformation nicht notwendigerweise in einen Raum gleicher Dimensionszahl abbilden muss!
2. (2 Punkte) Denken Sie sich eine Transformationsfunktion  $\Phi$  aus und transformieren Sie die Daten in einen Raum, der eine lineare Trennung ermöglicht!

### Aufgabe 8.2 (3 Punkte)

Nicht alle Kernelfunktionen sind gültig. Denken Sie sich 5 Beispiele von ungültigen Kernelfunktionen und beweisen Sie es mit einem Beispiel.

*Hinweis:* Gültige Kernelfunktionen müssen symmetrisch sein.  $k(x, y) = k(y, x)$

### Aufgabe 8.3 (3 Punkte)

Wenn man ein Mehrklassenproblem mit SVM lösen möchte, muss man es in mehrere 2-Klassen-Probleme umwandeln. Zwei bekannte Strategien sind One-vs-All (OVA) und One-vs-One (OVO). Stellen Sie sich vor, dass sie ein Problem mit 3 klassen haben, bei der jede Klasse je 100 Datenpunkte hat.

1. (1 Punkt) Beschreiben Sie in wenigen Worten beide Strategien (OVA), (OVO).
2. (1 Punkt) Wie viele 2-Klassen-SVM müssen Sie für beide Strategien trainieren?
3. (1 Punkt) Wie viele Datenpunkte benutzen Sie für jede 2-Klassen-SVM, um beide Strategien zu trainieren?