

Übung zur Vorlesung **Maschinelles Lernen**

Wintersemester 2008/2009

Blatt 10

Wiederholung

- Erläutern Sie den Unterschied zwischen den Lernaufgaben *Klassifikation*, *Clustering* und *Regression*!
- Welche Aufgabe löst die *SVMstruct*? Erläutern Sie die grundlegende Idee des *SVMstruct*-Ansatzes!

Aufgabe 1

7 Punkte

In der Vorlesung wurden verschiedene Clustering Verfahren vorgestellt. Ein einfacher Clusteralgorithmus für numerische Daten ist das *k*-Means Verfahren, welches in dieser Aufgabe zu implementieren ist. Als Abstandsmaß soll der euklidische Abstand benutzt werden.

Dazu soll ein Operator implementiert werden, der ein *Prediction-Modell* zurückliefert das als Ergebnis für jeden Punkt das zugehörige Cluster vorhersagt. Ist das vorhergesagte Label nominal, kann das Ergebnis auf einfache Weise mit dem Scatter-Plotter dargestellt werden.

Beispiel-Daten dazu finden sich im Übungsprojekt im Verzeichnis `samples/clustering`.

Aufgabe 2

3 Punkte

Erweitern Sie ihren Operator um die Möglichkeit auch nominale Attribute zu behandeln. Versuchen Sie dann mit Hilfe des Word-Vector Tools eine kleine Sammlung von Texten zu clustern. Ergeben sich sinnvolle Gruppierungen auf ihren Daten? Stellen Sie ihre Ergebnisse vor!