

Maschinelles Lernen 2

Sommersemester 2007

Blatt 5

Abgabe 12. Juni 2007 in der Vorlesung, oder bis 23:59:59 Uhr bei mikio@cs.tu-berlin.de

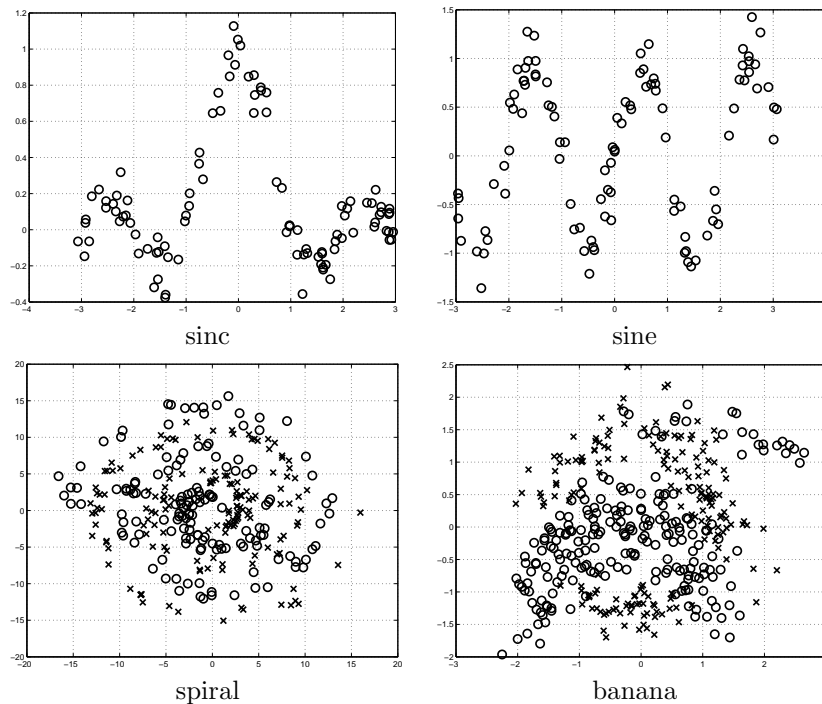
Die Aufgaben können auch in Gruppen bearbeitet werden, allerdings sollte die Gruppenzusammensetzung über des Semester stabil bleiben. Für praktische Aufgaben bitte ebenfalls Code und Ausgabe (d.h. Ergebnisse, Plots) abgeben. Verwende Matlab oder Octave. Lehrversionen von Matlab sind im cs-Netz (IRB) installiert. Zum Start `ml/bin/matlab` aufrufen. Octave ist ein freie Matlab-clone, der unter www.octave.org verfügbar ist. Für die Abgabe von praktischen Aufgaben bitte die Coding-Richtlinien beachten, die für das Praktikum gelten (siehe hierzu http://ml.cs.tu-berlin.de/de/ss07_ml_praktikum.html, Hinweise auf dem ersten Übungsblatt).

Aufgabe

In der Vorlesung wurde die “empirical kernel map” vorgestellt. Auf diesem Übungszettel sollen einige Datensätze auf ihre Komplexität hin untersucht werden.

Die Datensätze `sinc`, `sine`, `spiral` und `banana` befinden sich auf der Praktikums homepage <https://ml01.zrz.tu-berlin.de/cgi-bin/twiki/view/Main/TheLectS07MachineLearning>

Es handelt sich hierbei um eindimensionale Regressionsdatensätze und zweidimensionale Klassifikationsdatensätze.



- Berechne für jede der Datensätze die Kernel-PCA-Koeffizienten für einen RBF-Kern mit Kernbereiten $w \in \{0.1, 1, 10\}$. (10 Punkte)
- Fitte das Zweikomponentenmodell

$$s_i \sim \begin{cases} \mathcal{N}(0, \sigma_1^2) & 1 \leq i \leq d \\ \mathcal{N}(0, \sigma_2^2) & d+1 < i \leq n \end{cases}$$

und bestimme den optimalen Cut-off-Parameter d , sowie die erreichte Likelihood. Wähle hierdurch das optimale w aus der obigen Menge aus. (10 Punkte)

- Berechne die Projektion des Labelvektors Y auf die ersten Kernel-PCA-Komponenten bis zur Cut-off-Dimension für das ausgewählte w und vergleiche den entauschten Labelvektor mit den ursprünglichen Label. (10 Punkte)

