**Universität Bielefeld**

CITEC / Faculty of Technology
Ambient Intelligence Group
Dr. Thomas Hermann

Sekretariat: S. Strunk
Office: CITEC-3.311
106-6891

**Exercises of the lecture "Introduction to Data Mining"**

**winter term 2018/2019 Excercise sheet 5**

# - Principal curves and principal surfaces -

## Problem 5.1, Principal curves

In the file `polynom.dat` in the e-Learning space (Lernraum) of the ekVV you can find a list of pairs of numbers. Each pair represents a point in $\mathbb{R}^2$.

Let $\vec{f}(y)$ be a curve, whose two components are polynomials.

$$\vec{f} : \mathbb{R} \to \mathbb{R}^2$$

$$\vec{f}(y) = \begin{pmatrix} f_1(y) \\ f_2(y) \end{pmatrix}, \quad f_j(y) = \sum_{i=0}^{3} a_{ij} y^i$$

with

$$a_{ij} \in \mathbb{R}.$$

Learn a principal curve that is fit a polynomial to the data set. For this, optimize the $a_{ij}$ with regard to the cost function

$$E = \left\langle \min_y \|\vec{x} - \vec{f}(y)\|^2 \right\rangle_{\{\vec{x}\}}.$$

Initialize the $a_{ij}$ randomly and use stochastical exploration (informally spoken: Shake each $a_{ij}$ a little and if this reduces the cost, keep the jittered $a_{ij}$) to gradually converge to the optimum. What coefficients yield an optimal curve? Visualize the resulting curve together with the data set.

In [ ]:

## Problem 5.2, Mean value as Optimum of a cost function

In the lecture we have learned that the centroid (i.e. mean) vector minimizes the squared distance to all data points. Discuss the following cost functions qualitatively! How do they differ from the quadratic cost function for example in regard to outliers?

$$E(\vec{w}) = \sum_{\alpha=1}^{N} f(\|\vec{x}^{\alpha} - \vec{w}\|)$$

- (a) with $f(r) = r^4$
- (b) with $f(r) = r$
- (c) with $f(r) = r^4 - r^2$
- (d) with $f(r) = 1 - \exp\left(-\frac{r^2}{\sigma}\right)$


| | | | |
|---|---|---|---|
| Thomas Hermann (thermann@techfak) | Lecture Wed 12-14, CITEC lecture hall | office building | public transportation |
| Ferdinand Schlatt (fschlatt@techfak) | Tutorial Tue 12:15-13:45, H14 | Universität Bielefeld | light rail 4 |
| | | Universitätsstraße 25 | from Bahnhof and Jahnplatz |
| | | 33615 Bielefeld | |


In [ ]: