Approaches to dimension reduction for multidimensional complex patterns based on localised principal components

Project C4

Jochen Einbeck, Ludger Evers, Gerhard Tutz

Abstract

Highdimensional data structures arise frequently in a variety of disciplines. The treatment of multidimensional data is difficult, in particular when the relation between the variables is non-functional, as it is the case e.g. for spirals and other complex patterns. Here we present two approaches to this problem based on a localised version of principal component analysis. The first concept is a new approach to find principal curves, which are descriptively defined as smooth curves which pass through the middle of a (curved) multidimensional data cloud. The idea is to calculate a series of local centers of mass, where the movement along the data cloud is performed by follwing the first local eigenvector, each starting at the last local gravity point. We call the resulting curve *local principal curve*. Localised principal components however can also be used to construct a *global* divisive algorithm: Starting from the first principal component of the whole dataset, one successively divides each partition orthogonally to the extracted principal component. In between two split steps the extracted directions are updated in a manner similar to the k−means clustering algorithm. This approach does not primarily aim to find a closed principal curve, but to find "best local projections" on it. These best local projections can e.g. be useful in highdimensional regression problems with very complex (e.g. not additive) structure.

PRINCIPAL CURVES

Existing Principal Curve Algorithms

There exist a variety of approaches how to find the "middle" of a multidimensional data cloud:

Hastie & Stützle (HS, 1989) define a point on the principal curve as the average of all data points that project there ("self-consistency"). Self-consistent curves are critical points of the distance function

$$\Delta(m) = E\left(\inf_t ||X - m(t)||^2\right) \qquad (1)$$

and genaralize linear principal components in a natural way. A bias corrected version was provided by Banfield & Raftery (BR, 1992).

Tibshirani (1992) defines principal curves using an probabilistic approach, such that for data

$$X = m(t) + \epsilon \quad \text{mit} \quad E(\epsilon) = 0$$

the curve m is also principal curve for X.

Kégl, Krzyzak, Linder & Zeger (KKLZ, 2000) calculate the (piecewise linear) principal curve, that minimzes (1) over all curves with bounded length L.

Delicado (2001) defines principal curves as a sequence of fix points of function

$$\mu^*(x) = E(X|X \in H),$$

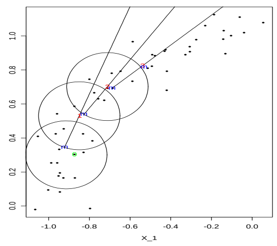
where H denotes the hyperplane which locally minimises the variance of the data points projected to it.

Local Principal Curves

Assume a data cloud $X = (X_1, \ldots, X_n)$, with $X_i = (X_{i1}, \ldots, X_{id})$. The local principal curve (LPC) is constructed as follows:

1. Choose a suitable starting point $x_{(0)}$. Set $x = x_{(0)}$.
2. Calculate the local center of mass μ^x around x in virtue of
$$\mu^x = \sum_{i=1}^n k_i X_i$$
with $k_i = K_H(X_i - x)X_i / \sum_{i=1}^n K_H(X_i - x)$.
3. Perform a PCA locally at x, i.e. calculate the first eigenvector of the local covariance matrix $\Sigma^x = (\sigma^x_{jk})_{(1\leq j,k \leq d)}$, where
$$\sigma^x_{jk} = \sum_{i=1}^n k_i (X_{ij} - \mu^x_j)(X_{ik} - \mu^x_k).$$
4. Go from μ^x in direction of the first local principal component γ^x to $x := \mu^x + t_0\gamma^x$.
5. Repeat 2. to 4. until the μ^x stay constant.

The series of μ^x make up the local principal curve. The following picture demonstrates the algorithm.



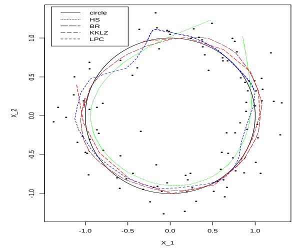
0: starting point; m: points on the local principal curve;
$1, 2, 3$: position of x after 1,2, and 3 steps.

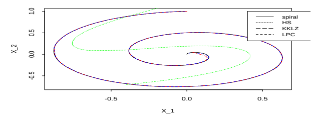
Technical Details

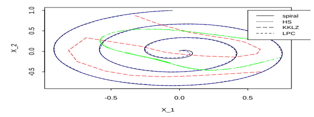
- "Signum flipping": If
$$\gamma^x_{(i-1)} \circ \gamma^x_{(i)} < 0$$
set $\gamma^x_{(i)} := -\gamma^x_{(i)}$.
- Go from $x_{(0)}$ in directions $\gamma^x_{(0)}$ and $-\gamma^x_{(0)}$.
- Angle penalization to handle crossings.
- Multiple initializiations for data consisting of several branches.

Examples

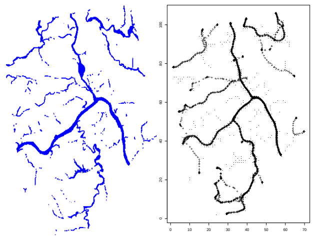
- Simulations







- Floodplains in Pennsylvania



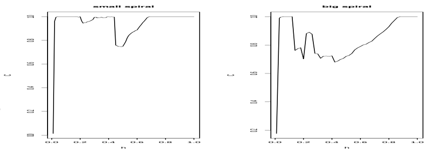
Parameter Selection

Given a bandwidth matrix $H = h^2 \cdot \mathbf{I}$, we set $t_0 = h$. It remains to select h. For a principal curve $m(h)$ consisting of a set of points $P_{m(h)}$ calculated with bandwidth h we define the self-coverage

$$C_{m(h)}(h) = \frac{\#\{x \in X\,|\exists p \in P_{m(h)} \text{ with } ||x - p|| \leq h\}}{n}$$

and choose

$$\hat{h} = \text{first local maximum of } \quad C_{m(h)}(h).$$

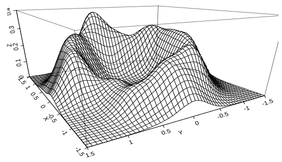


Theoretical Background

The local principal curve approximates the crest line of a kernel density estimate f_K of the data, since

$$\mu^x - x \sim f_K'(x)$$

holds (Comaniciu & Meer, 2002), provided that the center of mass and the density are estimated with the same bandwidth h. The following picture shows a kernel density estimate for the circle data:



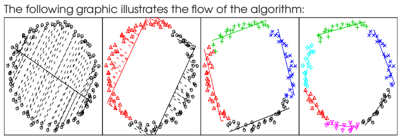
ADAPTIVE PARTITIONING

Localised principal components can also be used to define a hierarchic partitioning principal curve algorithm. This especially fast algorithm does not lead to a closed curve, but finds line segments which approximate the principal curve. Starting with the first principal component, a divisive approach is used:

- Each partition is split up, whenever this leads to a significant reduction of the average distance between the line segment and the points in that partition.
- Between the splitting cycles two supplementary steps are executed:
  - Update the membership by assigning each observation to the nearest line segment.
  - Update the line segments by calculating the first principal component line for each partition.

  These two steps, which can be iterated several times, are very close in mind to the k-means algorithm used in cluster analysis.
- Furthermore one can check after each cycle whether two neighbouring partitions can be combined.

The following graphic illustrates the flow of the algorithm:



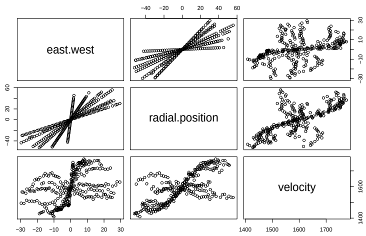
This approach can be useful when being applied to the covariate space in a regression model:

- Instead of using the covariates itself for the model building, their projections onto the different segments are used for estimating local models.
- These local models can — as in the example below — simply be weighted regression models using the projection index as covariate. Despite the partitioning done beforehand, all the data are used for these models, but are weighted according to their distance to the corresponding line segment.
- These models are connected by using a distance weighted sum of the predictions in order to obtain a smooth prediction.

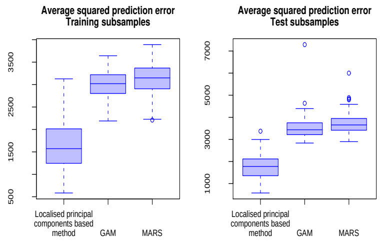
This technique permits the exploitation of the information in the structure of X which can often not be accessed when using additive models or other techniques like MARS. Besides its flexibility, a decisive advantage of this approach is that the model for the projection indices is only one-dimensional and needs not to be as complex as a global model. From this point of view the approach proposed here can be seen as a some sort of local dimension reduction technique for the covariate space.

Example

S-Plus galaxy dataset, prediction of the velocity of galaxies using one of its coordinates (east-west) and its radial position.



PAIRS PLOT
Test and training errors of the proposed method compared to those obtained using a generalised additive model (GAM) or multivariate adaptive regression splines (MARS):



References

(1) Banfield, J. D. & Raftery, A. E. *Ice flow identification in stallite images using mathematical morphology and clustering about principal curves*, **JASA** Vol. 87, 1996; p. 7-16.

(2) Comaniciu, D. & Meer, P. *Mean shift: A robust approach toward feature space analysis*, **IEEE Trans. Pattern Anal. Mach. Intell.** Vol. 24, 2002; p. 603-619.

(3) Delicado, P.: *Another Look at Principal Curves and Surfaces*, **Journal of Multivariate Analysis** Vol. 77, 2001; p. 84-116

(4) Hastie, T., Stuetzle, W. : *Principal Curves*, **JASA** Vol. 84, 1989; p. 502–516.

(5) Kégl, B., Krzyzak, A., Linder, T. & Zeger, K. *Learning and Design of Principal Curves*, **IEEE Trans. Pattern Anal. Mach. Intell.** Vol. 22, 2000; p. 281-297

(6) Tibshirani, R.: *Principal Curves Revisited*, **Statistics and Computation** Vol.2, 1992; p. 183-190