Kernel parameter dependence in spatial factor analysis

Nielsen, Allan Aasbjerg

Published in:
IGARSS

Link to article, DOI:
10.1109/IGARSS.2010.5653545

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
1. INTRODUCTION


In this paper we shall apply a kernel version of maximum autocorrelation factor (MAF) [7, 8] analysis to irregularly sampled stream sediment geochemistry data from South Greenland and illustrate the dependence of the kernel width. The 2,097 samples each covering on average 5 km² are analyzed chemically for the content of 41 elements.

2. DATA AND GEOLOGY

In 1979-80 the GGU, the Geological Survey of Greenland (now GEUS, the Geological Survey of Denmark and Greenland), collected stream sediment samples from a 10,000 km² area in South Greenland and illustrate the dependence of the kernel width. The 2,097 samples each covering on average 5 km² are analyzed chemically for the content of 41 elements.

2.1. Geological Setting

The study area is underlain by a Palaeoproterozoic orogen, the Ketilidian orogen, which consists of three major tectono-stratigraphic units: (1) a northern Border zone of tectonically reworked Archaean gneissic basement overlain by Palaeo-proterozoic metasediments and metavolcanics in the northeast, (2) a central zone occupied by a calc-alkaline batholith, and (3) a southern migmatite complex of predominantly Palaeoproterozoic metasediments and metavolcanics intruded by post-tectonic rapakivi type granites, see Figure 1 (top) and [11]. The plate-tectonic setting of the orogen has been interpreted in [12]. In Mesoproterozoic times the boundary region between the border and the granite zones was subjected to rifting and intrusions of numerous dykes of basaltic to trachytic compositions as well as of felsic alkaline complexes including carbonatites. The region affected by the alkaline magmas is termed the Gardar province, [13].

3. KERNEL PCA AND MAF

A kernel formulation of principal component analysis (PCA) [1] may be obtained from Q-mode or dual formulation of the problem combined with the so-called kernel trick [3].

Let us consider a data set with n observations and p variables organized as a matrix X with n rows and p columns; each column contains measurements over all observations from one variable and each row consists of a vector of measurements x_i from p variables for a particular observation

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}.$$  (1)

The superscript $T$ denotes the transpose. X is sometimes called the data matrix or the design matrix. Without loss of generality we assume that the variables in the columns of X have mean value zero.

3.1. R-mode PCA

In ordinary (primal also known as R-mode) PCA we analyze the variance-covariance matrix $S = X^T X/(n-1) = 1/(n-$
\[ \sum_{i=1}^{n} x_i x_i^T \] which is \( p \) by \( p \). If \( X^T X \) is full rank \( r = \min(n,p) \) this will lead to \( r \) non-zero eigenvalues \( \lambda_i \) and \( r \) orthogonal or mutually conjugate unit length eigenvectors \( u_i \) \((u_i^T u_i = 1)\) from the eigenvalue problem

\[
\frac{1}{n-1} X^T X u_i = \lambda_i u_i. \tag{2}
\]

We see that the sign of \( u_i \) is arbitrary. To find the principal component scores for an observation \( x \) we project \( x \) onto the eigenvectors, \( x^T u_i \). The variance of these scores is \( u_i^T S u_i = \lambda_i u_i^T u_i = \lambda_i \) which is maximized by solving the eigenvalue problem.

### 3.2. Q-mode PCA

In the dual formulation (also known as Q-mode analysis) we analyze \( X X^T / (n-1) \) which is \( n \) by \( n \) and which may be very large. \( X X^T \) is called the Gram matrix and its elements are the inner products \( x_i^T x_j \) between the rows of the data matrix \( X \). Multiply both sides of Equation 2 from the left with \( X^T \)

\[
\frac{1}{n-1} X X^T (X u_i) = \lambda_i (X u_i) \tag{3}
\]

or

\[
\frac{1}{n-1} X X^T v_i = \lambda_i v_i \tag{4}
\]

with \( v_i \) proportional to \( X u_i \), \( v_i \propto X u_i \), which is normally not unit length if \( u_i \) is. Now multiply both sides of Equation 4 from the left with \( X^T \)

\[
\frac{1}{n-1} X X^T (X^T v_i) = \lambda_i (X^T v_i) \tag{5}
\]

to show that \( u_i \propto X^T v_i \) is an eigenvector of \( S \) with eigenvalue \( \lambda_i \). We scale these eigenvectors to unit length assuming that \( v_i \) are unit vectors \((1 = v_i^T v_i \propto u_i^T X^T X u_i = (n-1)\lambda_i u_i^T u_i = 1)\)

\[
u_i = \frac{1}{\sqrt{(n-1)\lambda_i}} X^T v_i. \tag{6}
\]

We see that if \( X^T X \) is full rank \( r = \min(n,p) \), \( X X^T / (n-1) \) and \( X X^T / (n-1) \) have the same \( r \) non-zero eigenvalues \( \lambda_i \) and that their eigenvectors are related by \( u_i = X^T v_i / \sqrt{(n-1)\lambda_i} \) and \( v_i = X u_i / \sqrt{(n-1)\lambda_i} \).

### 3.3. Kernel Formulation of PCA

We now replace \( x \) by \( \phi(x) \) which maps \( x \) nonlinearly into a typically higher dimensional feature space. As an example consider a two-dimensional vector \([z_1 \ z_2]^T\) being mapped into \([\phi^2(z_1) \ \phi(z_2)]^T\). This maps the original two-dimensional vector into a five-dimensional feature space

1) named after Danish mathematician Jørgen Pedersen Gram (1850-1916) so that for example a linear decision rule becomes general enough to differentiate between all linear and quadratic forms including ellipsoids.

The mapping by \( \phi \) takes \( X \) into \( \Phi \) which is an \( n \) by \( q \) \((q \geq p)\) matrix

\[
\Phi = \begin{bmatrix}
\phi(x_1) & \phi(x_2) & \cdots & \phi(x_n)
\end{bmatrix}.	ag{7}
\]

For the moment we assume that the mappings in the columns of \( \Phi \) have zero mean. In this higher dimensional feature space \( C = \Phi^T \Phi / (n-1) = 1 / (n-1) \sum_{i=1}^{n} \phi(x_i) \phi(x_i)^T \) is the variance-covariance matrix and for PCA we get the primal formulation

\[
\frac{1}{n-1} \Phi^T \Phi u_i = \lambda_i u_i \tag{8}
\]

where we have re-used the symbols \( \lambda_i \) and \( u_i \) from above.

For the corresponding dual formulation we get

\[
\frac{1}{n-1} \Phi \Phi^T v_i = \lambda_i v_i \tag{9}
\]

where we have re-used the symbol \( v_i \) from above. As above the non-zero eigenvalues for the primal and the dual formulations are the same and the eigenvectors are related by

\[
u_i = \frac{1}{\sqrt{(n-1)\lambda_i}} \Phi^T v_i \tag{10}
\]

and \( v_i = \Phi u_i / \sqrt{(n-1)\lambda_i} \).

### 3.3.1. Kernel Substitution

Applying kernel substitution also known as the kernel trick we replace the inner products \( \phi(x_i) \phi(x_j) \) in \( \Phi \phi^T \) with a kernel function \( \kappa(x_i, x_j) = \kappa_{ij} \) which could have come from some unspecified mapping \( \phi \). In this way we avoid the explicit mapping \( \phi \) of the original variables. We obtain

\[
K v_i = (n-1)\lambda_i v_i \tag{11}
\]

where \( K = \Phi \Phi^T \) is an \( n \) by \( n \) matrix with elements \( \kappa(x_i, x_j) \). \( K \) is symmetric and must be positive semi-definite, i.e., its eigenvalues are non-negative; we say that \( \kappa \) is a Mercer kernel. Normally we let the eigenvalues subsume the factor \( n-1 \)

\[
K v_i = \lambda_i v_i. \tag{12}
\]

In this case \( u_i = \Phi^T v_i / \sqrt{\lambda_i} \) and \( v_i = \Phi u_i / \sqrt{\lambda_i} \).

It is easy to show that both centering to zero mean of the mappings in the columns of \( \Phi \) as well as the projection of observations \( x \) onto the primal eigenvectors \( u_i \) may be expressed by means of the kernel function \( \kappa(x_i, x_j) \) without explicit use of the nonlinear mapping.
3.4. Kernel MAF

In a similar fashion maximum autocorrelation factor (MAF) analysis [7, 8, 14] which may be considered as a form of spatial factor analysis may be kernelized, for details see [15]. In this context a popular kernel is the Gaussian

\[ \kappa(x_i, x_j) = \exp\left(-\frac{1}{2} \left( \frac{\|x_i - x_j\|}{\sigma} \right)^2 \right) \]

where the kernel width is given by the scale parameter \( \sigma \), and \( x_i \) and \( x_j \) (here) are 41-dimensional vectors of concentrations. Below we give results of the kernel MAF analysis with different choices of \( \sigma \).

4. RESULTS AND DISCUSSION

Figure 1 (bottom) shows the 2,097 sample sites in Southern Greenland in red. The study area is approximately 320 km east-west and 210 km north-south. The Delaunay triangulation is shown in blue. The analyses shown below are based on concentrations standardized to unit variance, see also [10, 9, 16].

For \( \sigma \) equal to the mean distance between observations in 41-dimensional feature space kMAFs 1, 2 and 3 in Figure 2 top focus on extreme observations associated with the intrusions marked with dense plus signs “+” in the Granite zone (Figure 1 top). Also they neatly adapt to an even strongly varying multivariate background. Although other samples have high scores, this is true also for kMAFs with \( \sigma \) equal to ten times the mean, Figure 2 bottom. In spite of a tendency to highlight more samples in the so-called Gardar intrusion, the same overall impression is true for kMAFs with \( \sigma \) equal to a hundred times the mean, Figure 3 top. For kMAFs with \( \sigma \) equal to a thousand times the mean (Figure 3 bottom) we see a depiction of the three major geological units named “Border Zone”, “Granite Zone” and “Migmatite Complex” in the geological map, Figure 1 top.

In conclusion we see that by varying the kernel width \( \sigma \) we may analyse the phenomenon under study at different scales which highlight different relevant geological features.

5. REFERENCES


Fig. 2. Kernel MAFs 1, 2 and 3 as RGB, kernel width $\sigma$ is mean of distances in feature space (top), kernel MAFs 1, 2 3 as RGB, kernel width $\sigma$ is 10 times mean of distances in feature space (bottom).


Fig. 3. Kernel MAFs 1, 2 and 3 as RGB, kernel width $\sigma$ is 100 times mean of distances in feature space (top), kernel MAFs 1, 2 3 as RGB, kernel width $\sigma$ is 1,000 times mean of distances in feature space (bottom).


