Remote Sensing and Image Analysis – Principle component analysis

Edzer Pebesma

November 19, 2008

ifgi
Institute for Geoinformatics
University of Münster

PCA – what is it?

- who has used it?
- what for?
- what does it do?
- what does it *not* do?

What does it do?

- ▶ Goal: exploration, data reduction
- ▶ Data reprojection
- ▶ Look at projections on the first n eigenvalues of the (i) correlation matrix or (ii) covariance matrix
- ▶ Searches for main patterns, important messages
- ▶ ignores spatial organization of data
- ▶ finds degree of (multivariate) correlation
- ▶ size and shape effect

It's place in multivariate statistics

- unsupervised (no distinction between dependent and independent)
- ordination (search for directions, assuming continuous variability)

It's place in unsupervised techniques

- Alternatives: cluster analysis where one tries to find some form of grouping (class = categorical outcome)
- PCA: continous outcome

It's place in ordination techniques

Alternatives:

- ► linear regression predicts one variable based on others
- ► correspondence analysis CA ordinates where data are counts, rather than continuous varying quantities
- ► Max/min autocorrelation factors (MAF) or maximum nois fraction (MNF) transformation, try to capture both variability/correlation **AND** spatial correlation (A transformation for ordering multispectral data in terms of imagequality with implications for noise removal Green, A.A.; Berman, M.; Switzer, P.; Craig, M.D. Geoscience and Remote Sensing, IEEE Transactions on Volume 26, Issue 1, Jan 1988 Page(s):65 - 74)
- ► independent component analysis ICA (not only uncorrelated, but independent)