How many Principal Components?

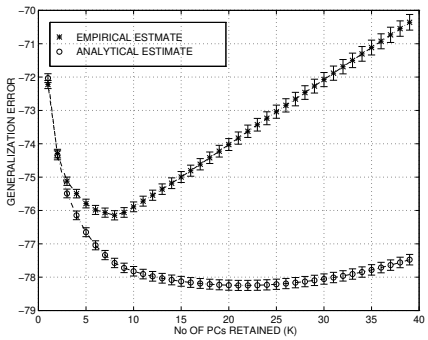
L.K. Hansen[1], F.Å. Nielsen[1], P. Toft[1], S.C. Strother[2], N. Lange[3], N. Mørch[1,4], C. Svarer[4], O.B. Paulson[4], R. Savoy[5], B. Rosen[5], E. Rostrup[6], P. Born[6]

[1] *Dept. of Mathematical Modelling, B321, Technical University of Denmark, Denmark*
[2] *PET Imaging Service, Minneapolis VA Medical Center, Minnesota, USA*
[3] *Brain Imaging Center, McLean Hospital & Harvard University, Boston, USA*
[4] *Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet, Denmark*
[5] *Massachussets General Hospital. Boston, USA* [6] *MR Unit, Hvidovre Hospital, Denmark*

Introduction. Principal Component Analysis (PCA) and Clustering are two prominent examples of unsupervised learning schemes that are gaining interest in brain mapping. Since learning proceeds from a finite sample of training data the estimated model parameters are stochastic variables depending on the particular (random) training set. Hence, there are questions of robustness and generalization: how robust are the estimated parameters to fluctuation and noise in the training set and how well will they perform on a new (test) datum? It is a common observation that good generalization is obtained when the model capacity is carefully optimized. Generalization aspects of unsupervised learning algorithms were recently discussed in (1). In particular the issues of selection of PCA subspace dimension and number of clusters in Clustering were addressed. PCA is often used as a preprocessing step in functional brain mapping and we here outline a generalization based procedure for selection of PCA subspace dimension in an fMRI context.

Methods. An fMRI activation image set of a single subject performing a left-handed finger-to-thumb opposition task was acquired. Multiple runs of 72 2.5-second (24 baseline, 24 activation, 24 baseline) whole-brain echo planar images where aligned, and an axial slice through primary motor cortex and SMA of 42 x 42 voxels (3.1 x 3.1 x 8 mm) extracted. A training set of size (N =) 300 was drawn at random from the pool of scans, and another *disjoint* independent set of 324 scans were drawn at random to form a test set. PCA analyses were carried out on the training set. The generalization error (1) was computed on the test set using the covariance structure estimated on the training sets and plotted versus size of the PCA subspace (K). In particular we assume that the slice x (total number of voxels L) can be modelled as a multivariate Gaussian variable whose main variation is confined to a subspace of dimension K, this component being degraded by additive, independent isotropic noise: $x = s + n$. Here the "signal" $s \sim \mathcal{N}(x_0, \Sigma_s)$, while the "noise" is distributed $n \sim \mathcal{N}(0, \Sigma_n)$. We assume that Σ_s is singular, i.e., of rank $K < L$, while $\Sigma_n = \sigma^2 I_L$, where I_L is an $L \times L$ unit matrix and σ^2 is a noise level. We estimate the parameters $\theta \equiv (x_0, \Sigma_s, \Sigma_n)$ on the training set by maximizing the likelihood. Furthermore, within maximum likelihood estimation we get a asymptotic estimate of the test error $\langle G(\widehat{\theta})\rangle_D \approx \langle E(\widehat{\theta})\rangle_D + \dim(\theta)/N$, where $\langle E(\widehat{\theta})\rangle_D$ is the average training error, and where the dimensionality of the parametrization (θ) grows quadratically with the number,K, of PCs retained in the PCA (1).

Results and Conclusion. In the figure is shown the development of the generalization error as we increase the model dimensionality (K), as estimated on the test set ($*$) and using the analytical asymptotic estimate (o). The error bars indicate the standard deviation observed in 10 replications of the experiment. Note that the minimal generalization error is obtained when projecting on $K = 8$ principal directions. It is apparent that the asymptotic theory overestimates the optimal value of K and underestimates the overall the generalization error . Hence we recommend the empirical test set-based scheme for estimating the proper dimension of the PCA subspace. It should be noted that principal component analysis is *unsupervised*, hence, does not use any information about the external stimulus (activation reference function). It can not be ruled out that interesting and generalizable structures are hidden in the rejected subspace, but such information has to be extracted using additional knowledge, e.g., about the stimulus.



Acknowledgments. This project has been funded by The Danish Research Councils and the Human Brain Project P20 MH57180 "Spatial and Temporal Patterns in Functional Neuroimaging".

Reference

1. L.K. Hansen, J. Larsen: Proc. Int. Conf. Neural Nets 1996