Sparse non-linear denoising of fMRI: Performance and pattern reproducibility

Trine J. Abrahamsen
Lars K. Hansen

Non-linear embedding improves mental state decoding in fMRI. We investigate non-linear denoising by kernel Principal Component Analysis (PCA)

The mapping of denoised feature space points back into input space, also referred to as the "pre-image problem", is inherently illposed

We investigate sparse pre-image reconstruction by Lasso regularization.

Sparse estimation provides better predictability and a more reproducible pre-image. The combined prediction/reproducibility metric provides means for optimizing the degree of sparsity and the non-linearity of the kernel embedding.

Introduction

The basic idea of denoising by kernel PCA is to implement a projection onto a signal manifold in three steps

1. Map the original input space data into a feature space in which the manifold is linearized
2. Use a conventional linear algorithm, like PCA, to identify the signal manifold by linear projection in feature space
3. Estimate the denoised input space point that best represents the projected feature space point

The latter step is referred to as the pre-image problem. Unfortunately, finding a reliable pre-image is challenging and has given rise to several algorithms [1, 2, 3, 5, 6, 7].

Kernel PCA

Let \( \mathcal{F} \) be the Reproducing Kernel Hilbert Space (RKHS) associated with the kernel function \( k(x, x') = \langle \psi(x) \mid \psi(x') \rangle \), where \( \psi \) maps the D-dimensional input space \( \mathcal{X} \) to the high dimensional feature space \( \mathcal{F} \). Let \( \Psi \) denote the kernel matrix with element \( \Psi_{ij} = k(x_i, x_j) \), then kernel PCA can be performed by solving the eigenvalue problem

\[
\Psi \omega_i = \lambda_i \omega_i
\]

where \( \Psi \) is the centered kernel matrix.

The projection of a \( \psi \)-mapped test point onto the \( i \) th principal component is \( \beta_i = \sum_{n=1}^{N} o_{n,i} \psi(x_n) \), and the projection of \( \psi \) onto the subspace spanned by the first \( q \) eigenvectors can be found as

\[
P_\psi(\psi) = \sum_{i=1}^{q} \sum_{n=1}^{N} o_{n,i} \psi(x_n) + \hat{\psi} = \sum_{i=1}^{q} \sum_{n=1}^{N} y_n \psi(x_n) + \hat{\psi}
\]

where \( y_n = \sum_{i=1}^{q} \lambda_i o_{n,i} \).

Kernel PCA satisfies properties similar to those for linear PCA, namely that the squared reconstruction error is minimal and the retained variance is maximal. However, these properties hold in \( \mathcal{F} \) not in \( \mathcal{X} \).

We focus on the Gaussian kernel of the form \( k(x, x') = \exp(-\| x - x' \|^2) \), where \( c \) controls the width of the kernel and thereby the non-linearity of the associated feature map

Data

The fMRI data was acquired by Dr. Egill Rostrup at Hvidovre Hospital on a 1.5 T Magnetom Vision MR scan. The scanning sequence was a 2D gradient echo EPI (T2*-weighted) with 66 ms echo time and 50° RF flip angle. The images were acquired with a matrix of 128 x 128 pixels, with FOV of 230 mm, and 10 mm slice thickness, in a para-axial orientation parallel to the calcaneous sulcus, hence capturing possible activation in visual cortices. The visual stimulus paradigm consisted of a rest period of 20 sec of darkness using a light fixation dot, followed by 10 sec of full-field checkerboard reversing at 8 Hz, and ending with 20 sec of rest (darkness). In total, 150 images were acquired in 70 sec, corresponding to a period of approximately 330 msec per image. The experiment was repeated in 10 separate runs containing 150 images each. In order to reduce saturation effects, the first 29 images were discarded, leaving 121 images for each run.

Results

![Figure 2: Top panel: raw data. Middle panel: dense reconstruction. Lower panel: sparse reconstruction.](image)

Visual examples of the raw data as well as both dense and sparse reconstructions. It is easily seen how both the dense and sparse reconstructions capture the signal in the visual cortex. However, for the sparse pre-image the signal is more distinctive and the baseline scans in general suffer from less noise.

Conclusions

In this contribution we addressed the reliability of sparse pre-image estimation. Using experimental results on fMRI it was shown that sparse reconstruction not only leads to visually appealing pre-images but the estimates are also highly reproducible. We thus recommend to augment the cost function for pre-image estimation in Equation (5), with a \( \ell_1 \)-norm penalty term in order to impose sparsity on the sought pre-image. Using the \( \ell_1 \) estimation metric points to useful values of the two important denoising kernel PCA control parameters, namely the smoothing scale, \( c \), and the sparsity level control, \( \lambda \).

Acknowledgments

The research is funded in part by the Danish Lundbeck Foundation through CBMR Center for Integrated Molecular Brain Imaging.

Sparse pre-image estimation

Since we cannot expect an exact pre-image, we follow [6] and relax the quest to find an approximate pre-image, i.e., a point in input space which maps into a point in feature space as close as possible to \( \Psi(x) \). The basic quadratic objective function is

\[
R(x) = ||\psi(x) - P_\psi(\psi)||^2 + \gamma (k(x) - \tilde{k}(x)) + \sum_{i=1}^{N} \lambda_i \omega_i + \Omega
\]

where \( \gamma \)-independent terms are collected in \( \Omega \).

Regularization is commonly used to stabilize estimates of high variability. We focus on a special case from the power penalty family, namely the Lasso [9] where

\[
\hat{x} = \arg \min_{x} R(x) + \lambda ||x||_1
\]

For RBF kernels the cost function is given by

\[
\rho(x) = -2 \sum_{i=1}^{N} y_n k(x, x_n) + \lambda \sum_{i=1}^{N} \| y_n \|_2^2 + ||\tilde{\psi}||_2^2
\]

We apply the generalized path seeking (GPS) framework as introduced by Friedman in [4] to estimate the pre-image. The GPS method starts from a large \( \lambda \) and reduces it through a series of estimates of decreasing sparsity, providing solutions with multiple degrees of sparsity along the path.

Measuring predictability and robustness

In many application domains, including functional neuroimaging, statistical learning has two equally important objectives:

1. An accurate predictive model
2. A robust interpretation of the underlying physical mechanism that allows prediction

We follow [8] and use split-half resampling to produce unbiased estimates of the variability of the denoised pre-images. Brain state decoding performance is evaluated in cross-validation using a simple linear discriminant.

Three values of the smoothing scale of the Gaussian kernel are investigated ranging from a very nonlinear feature space map to a near-linear case equivalent to conventional PCA, viz., the scale is chosen as 1) the 5th percentile, 2) the median and 3) the maximum of the mutual distances in the data set. Furthermore, we run the experiments with a varying degree of sparsity imposed on the reconstruction (points along the GPS path).

![Figure 3: The top panel shows the prediction/reproducibility plot (pr-plot) using all scans. The lower left panel shows the pr-plot for the baseline scans while the lower right panel shows the pr-plot for the active scans. Locations in the upper right of the plot are preferred. The GPS estimate when using a non-linear kernel are seen to outperform all other estimates in terms of combined prediction and reproducibility measures.](image)

References