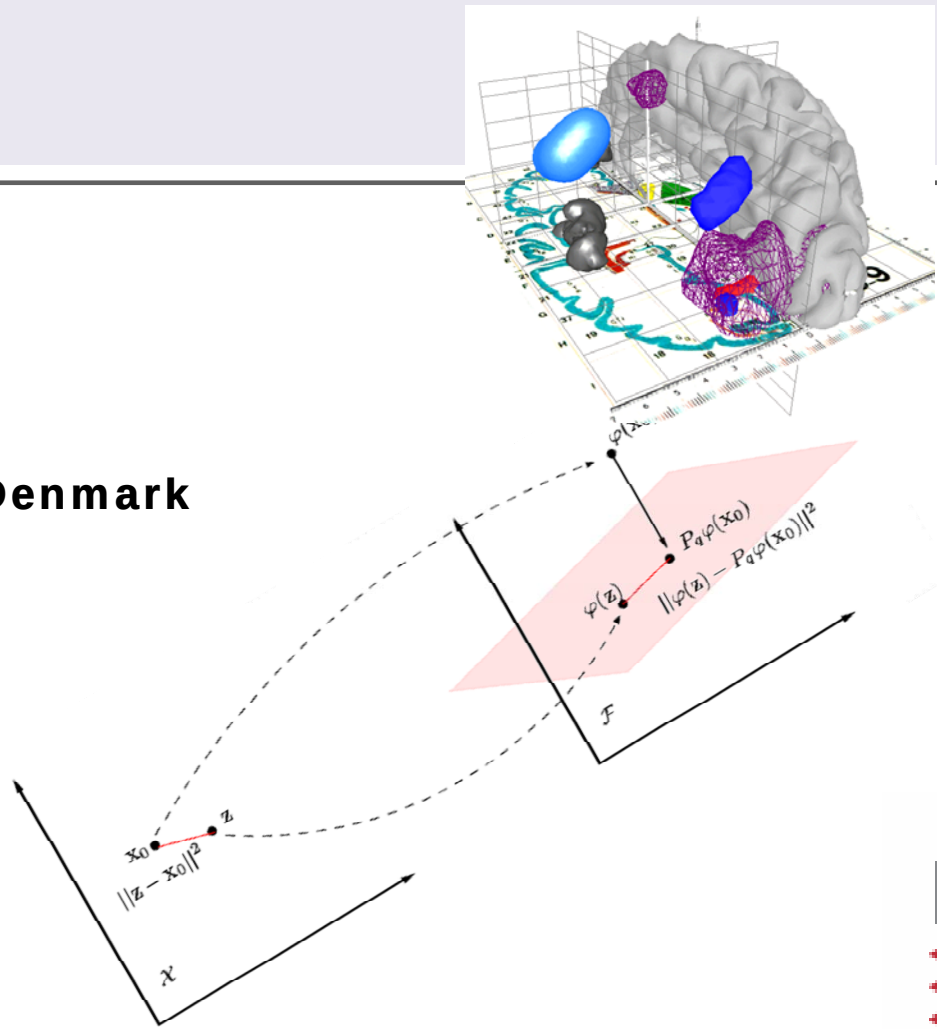
Scientific visualization with manifolds

Lars Kai Hansen

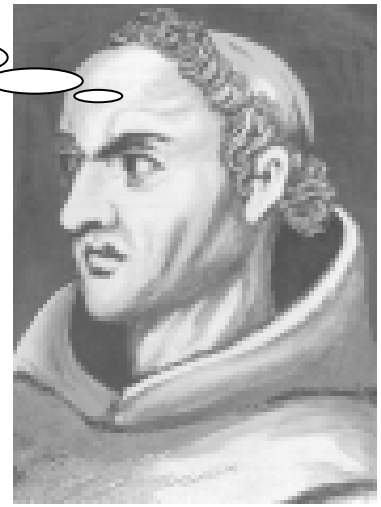
DTU Informatics
Technical University of Denmark



Cimbi
Center for integrated molecular brain imaging

DTU

Do not multiply causes!



OUTLINE

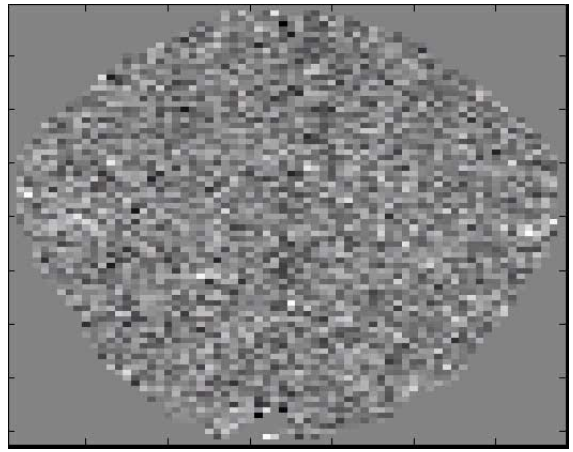
- Visualization is an important aspect of
  "scientific" applications of ML
  - To abstract generalizable relations from data
  - Robust visualization

- Unsupervised (explorative)
  - Factor models - Linear hidden variable representations
  - Independent component analysis (ICA)
  - Kernel representations: kPCA

- Supervised models (detection)
  - Visualization of non-linear kernel machines

Brain imaging with fMRI (functional magnetic resonance imaging)

- Indirect measure of neural activity - hemodynamics
- A cloudy window to the human brain
- Challenges:
  - Signals are multi-dimensional mixtures
  - No simple relation between measures and brain state - ”what is signal and what is noise”?



TR = 333 ms

Mind reading software could record your dreams…

Yukiyasu Kamitani: Y. Miyawaki et al. Neuron 60(5):915 - 929 (2008)

BOLD fMRI: Is hemodynamic de-convolution feasible?

LETTER Communicated by Karl Friston

Bayesian Model Comparison in Nonlinear BOLD fMRI Hemodynamics

Daniel J. Jacobsen
dj@decision3.com
Lars Kai Hansen
lkh@imm.dtu.dk
Intelligent Signal Processing, Informatics and Mathematical Modeling, Technical University of Denmark, Lyngby, N/A 2800, Denmark

Kristoffer Hougaard Madsen
khm@imm.dtu.dk
Intelligent Signal Processing, Informatics and Mathematical Modeling, Technical University of Denmark, Lyngby, N/A 2800, Denmark,

Neural Computation **20, 738–755** (2008)

Balloon model: Non-linear relations between stimulus and physiology – described by four non-linear differential eqs.

Bayesian averaging with resampling loop
to establish generalizability (predictive distribution)
and reproducibility of the posterior distribution

$$R(M) = -\frac{1}{K}\sum_{i=1}^{K}\int p(\theta \mid D_i^1, M)\log\frac{p(\theta \mid D_i^1, M)}{p(\theta \mid D_i^2, M)}d\theta,$$

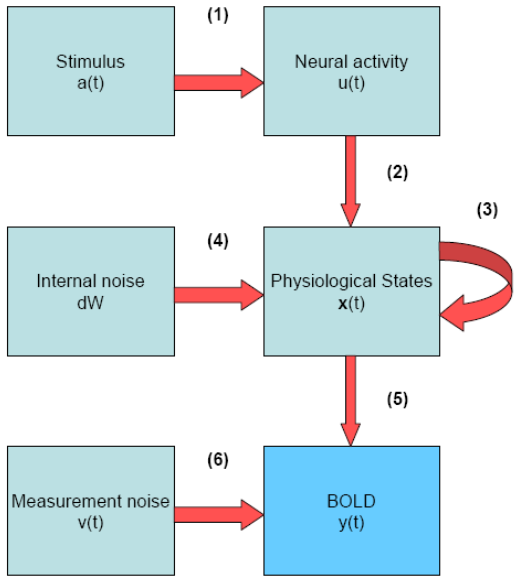


Figure 3.1: Overview diagram of hemodynamic models.

BOLD hemodynamics R-Bayes model selection

Model A: constant input vs Model B: Fading input

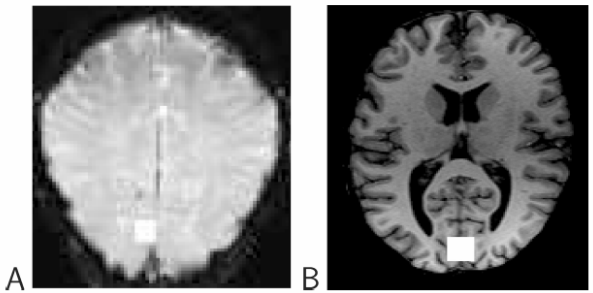


Figure 3: Regions of interest, marked with white squares. (A) Data set 1; T2* weighted image slice parallel to the calcarine sulcus. (B) Data set 2; MPRAGE (magnetization prepared rapid gradient echo) horizontal slice.

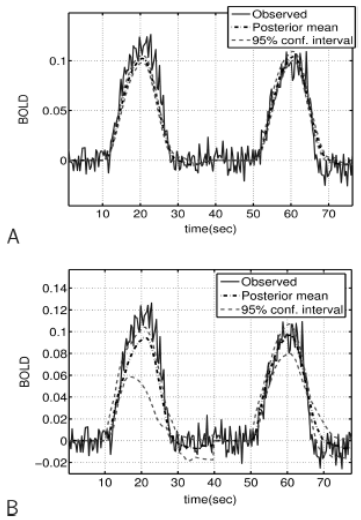


Figure 4: Prediction of data set 1. (A) Model A. (B) Model B. Note that the confidence interval is an empirical confidence interval for the mean prediction, based on the MCMC samples.

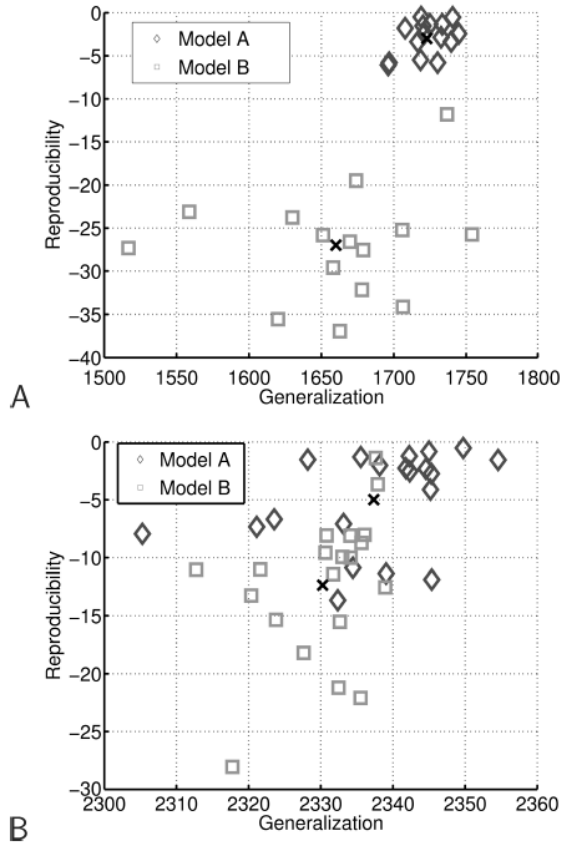


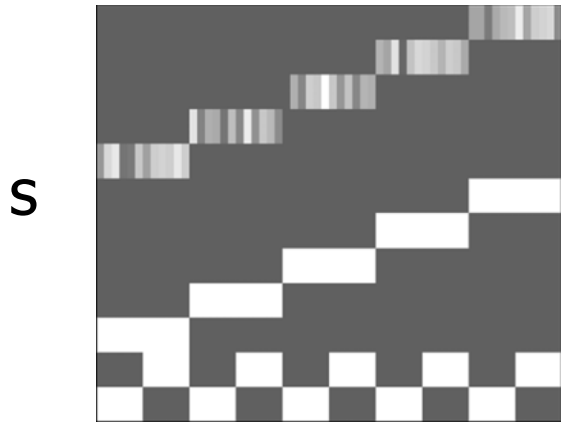
Figure 6: Generalization and reproducibility for real data; crosses mark the mean. (A) Data set 1. (B) Data set 2.

Multivariate neuroimaging models

Neuroimaging aims at extracting the mutual information between stimulus and response.

- Stimulus: Macroscopic variables, ”design matrix” ... *s(t)*
- Response: Micro/meso-scopic variables, the neuroimage ... *x(t)*
- Mutual information is stored in the joint distribution … *p(x,s).*

Often s(t) is assumed known….unsupervised methods consider s(t) or parts of s(t) ”hidden”…..



Multivariate neuroimaging models

- Univariate models -SPM, fMRI time series models etc.

$$p(x,s) = p(x \mid s)p(s) = \prod_j p(x_j \mid s) \cdot p(s)$$

- Multivariate models -PCA, ICA, SVM, ANN (Lautrup et al., 1994, Mørch et al. 1997)

$$p(x,s) = p(s \mid x)\,p(x)$$

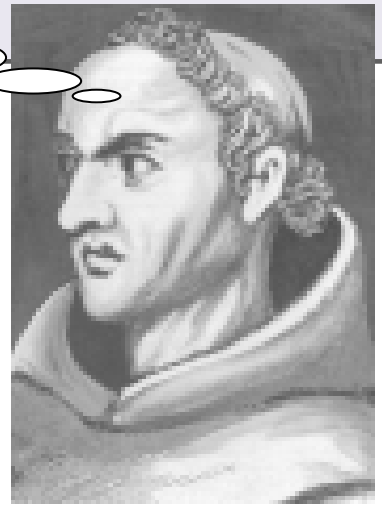
- Modeling from data (D) w. parameterized function families

$$p(s \mid x) \sim p(s \mid x, D) \sim p(s \mid x, \hat{\theta}), \qquad \hat{\theta} = \hat{\theta}(D)$$

$$p(x) \sim p(x \mid D) \sim p(x \mid \hat{\theta}),$$

Generalizability

Do not multiply causes!



- Generalizability is defined as *the expected performance on a random new sample*
  - A models mean performance on a ”fresh” data set is an unbiased estimate of generalization

- Typical loss functions:

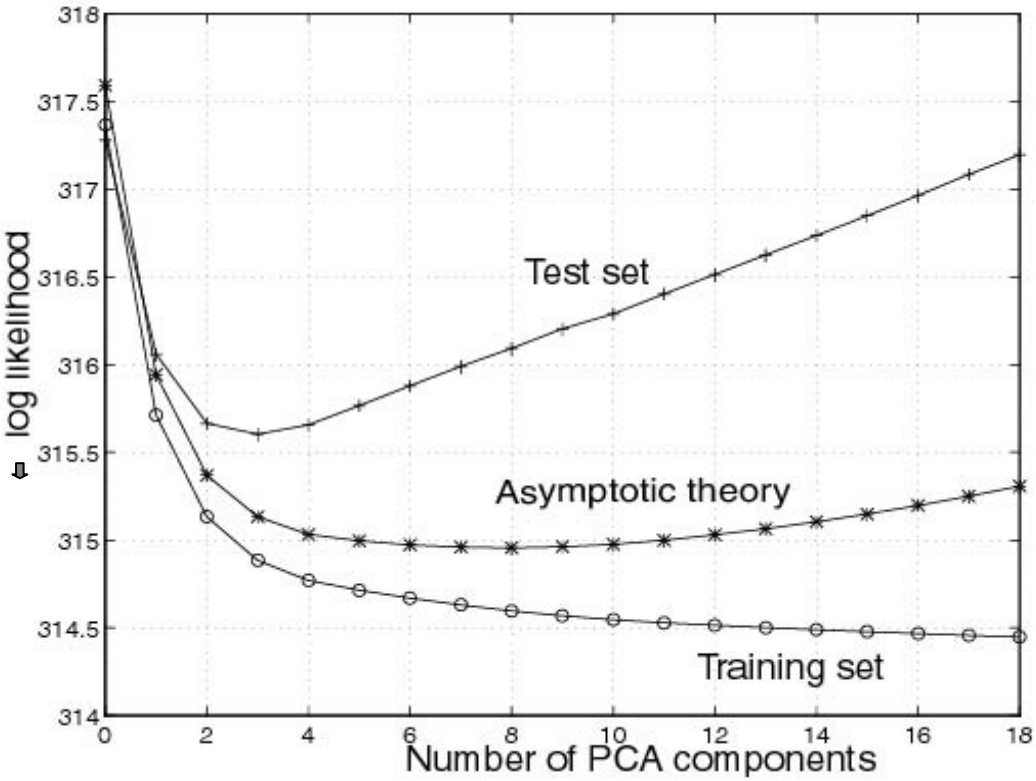
$$\left\langle -\log p(\mathbf{s} \mid \mathbf{x}, D) \right\rangle, \qquad \left\langle -\log p(\mathbf{x} \mid D) \right\rangle,$$

$$\left\langle (\mathbf{s} - \hat{\mathbf{s}}(D))^2 \right\rangle, \qquad \left\langle \log \frac{p(\mathbf{s}, \mathbf{x} \mid D)}{p(\mathbf{s} \mid D)\, p(\mathbf{x} \mid D)} \right\rangle$$

- Note: No problem to estimate generalization in hidden variable models!

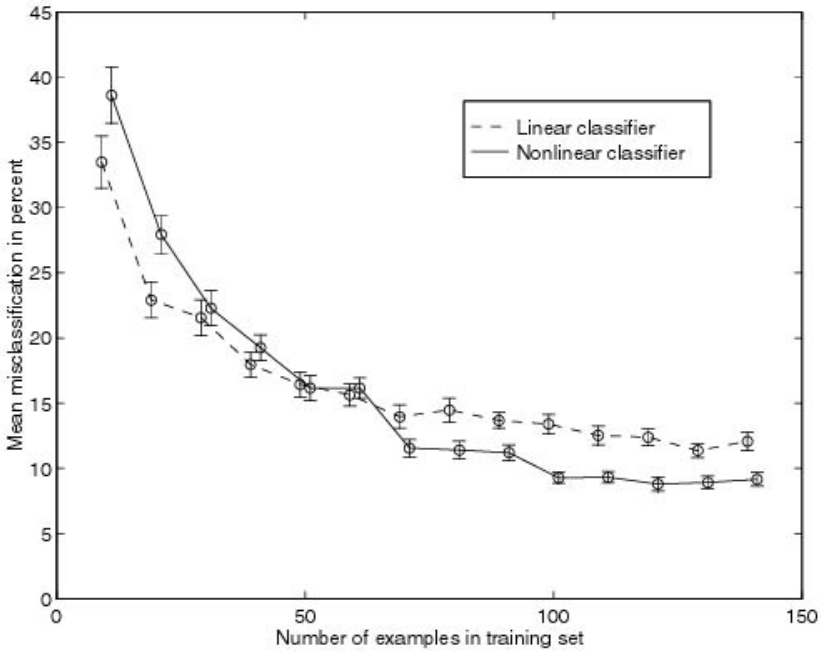
- Results can be presented as ”bias-variance trade-off curves” or ”learning curves”

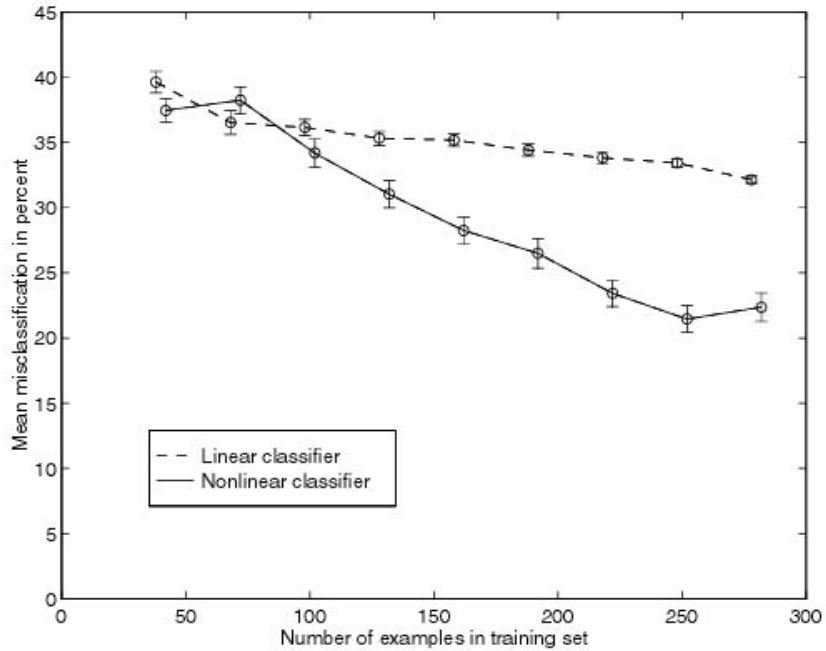
Bias-variance trade-off as function of PCA dimension



Hansen et al. *NeuroImage* (1999)

Learning curves for multivariate brain state decoding





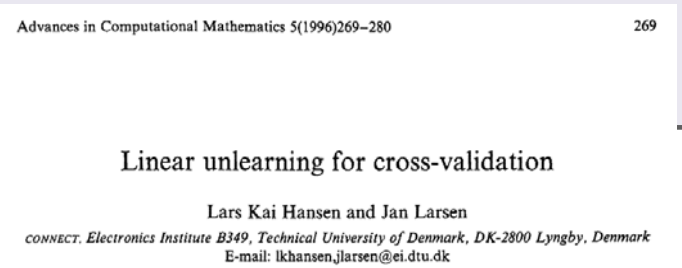
PET fMRI

Finger tapping, analysed by PCA dimensional reduction and Fisher LD / ML Perceptron. Mørch et al. *IPMI* (1997)…"first brain state decoding in fMRI"

Visualization of brain state decoder representations

- A brain map is a visualization of the information captured by the model:
  - The map should take on a high value in voxels/regions involved in the response and a low value in other regions…
- Statistical Parametric Maps
  - voxel based hypothesis test
  - Parameter maps for linear models (e.g. linear SVM's)
- The saliency map
- The sentivity map
- Consensus maps

...hints from asymptotic theory



Advances in Computational Mathematics 5(1996)269–280 269

Linear unlearning for cross-validation

Lars Kai Hansen and Jan Larsen
CONNECT, Electronics Institute B349, Technical University of Denmark, DK-2800 Lyngby, Denmark
E-mail: lkhansen,jlarsen@ei.dtu.dk

- Asymptotic theory investigates the sampling fluctuations in the limit N -> ∞
- Cross-validation good news: The ensemble average predictor is equivalent to training on all data (Hansen & Larsen, 1996)
- Simple asymptotics for parametric and semi-parametric models
- Some results for non-parametric e.g. kernel machines
- In general: Asymptotic predictive performance has bias and variance components, there is proportionality between parameter fluctuation and the variance component...

The sensitivity map

NeuroImage 15, 772–786 (2002)
doi:10.1006/nimg.2001.1033, available online at http://www.idealibrary.com on IDEAL®

The Quantitative Evaluation of Functional Neuroimaging Experiments: Mutual Information Learning Curves

U. Kjems,*,1 L. K. Hansen,* J. Anderson,†,‡ S. Frutiger,‡,§ S. Muley,§ J. Sidtis,§ D. Rottenberg,†,‡,§ and S. C. Strother†,‡,§,¶

*Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark; †Radiology Department, §Neurology Department, and ¶Biomedical Engineering, University of Minnesota, Minneapolis, Minnesota 55455; and ‡PET Imaging Center, VA Medical Center, Minneapolis, Minnesota 55417

$$m_j = \left\langle \left(\frac{\partial \log p(s|x)}{\partial x_j} \right)^2 \right\rangle$$

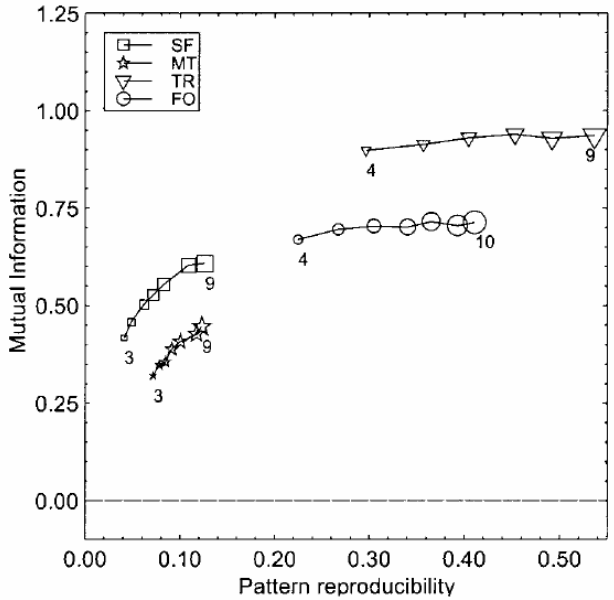
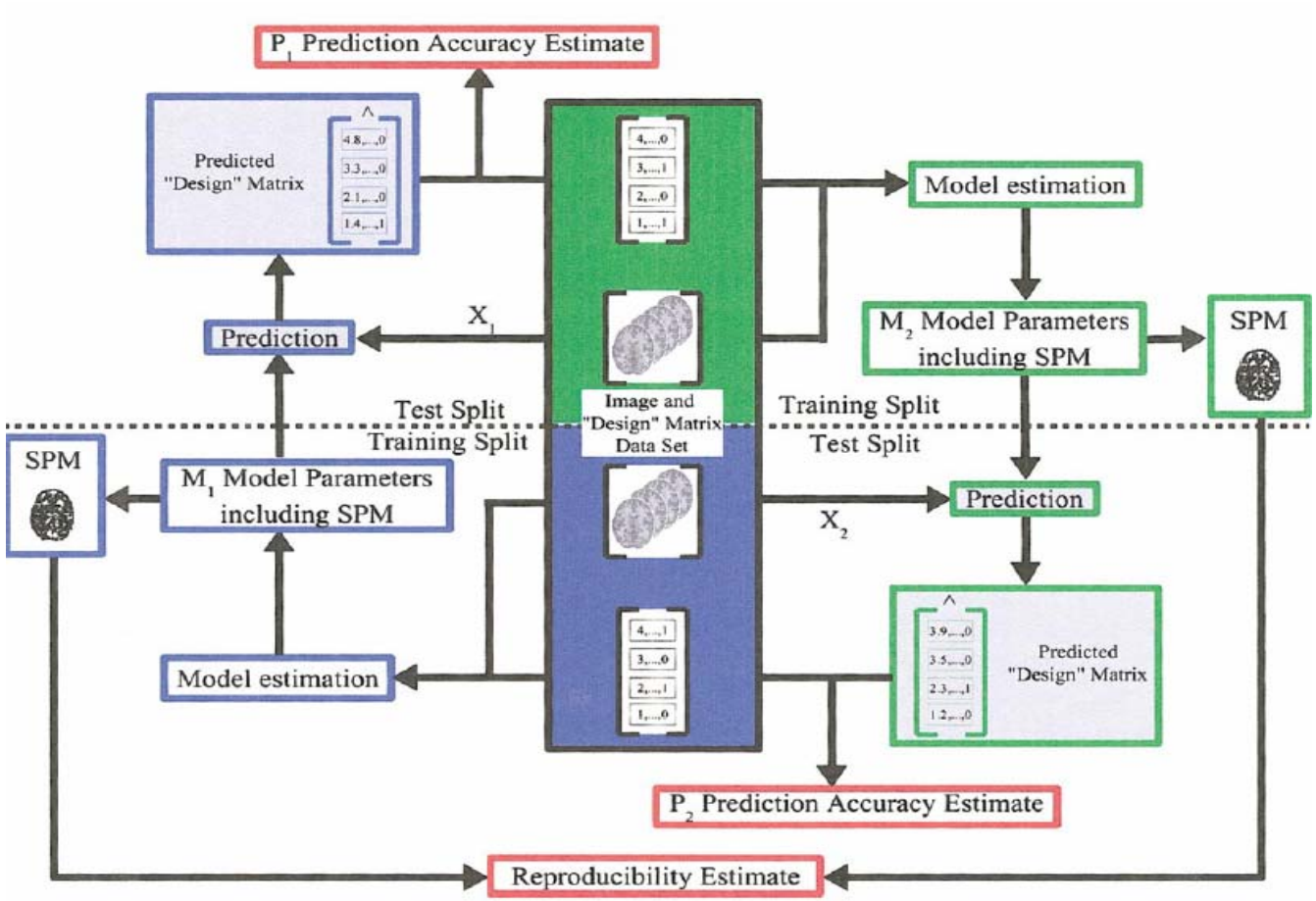


FIG. 3. Plot of scan/label mutual information versus reproducibility signal/noise for the four data sets, for varying numbers of subjects in the training set. There were 2 labels/4 scans per subject (balanced data set; Setup 1, Table 1) corresponding to the dashed solid line in Fig. 4. We see that both measures indicate improved performance of the model as the number of subjects increases.

- The sensitivity map measures the impact of a specific feature/location on the predictive distribution

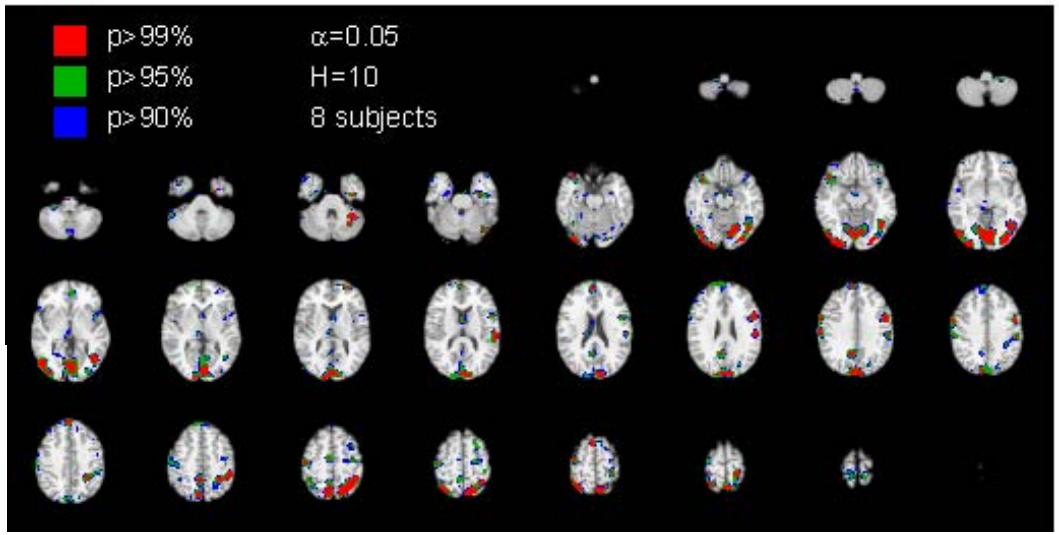
NPAIRS: Reproducibility of parameters

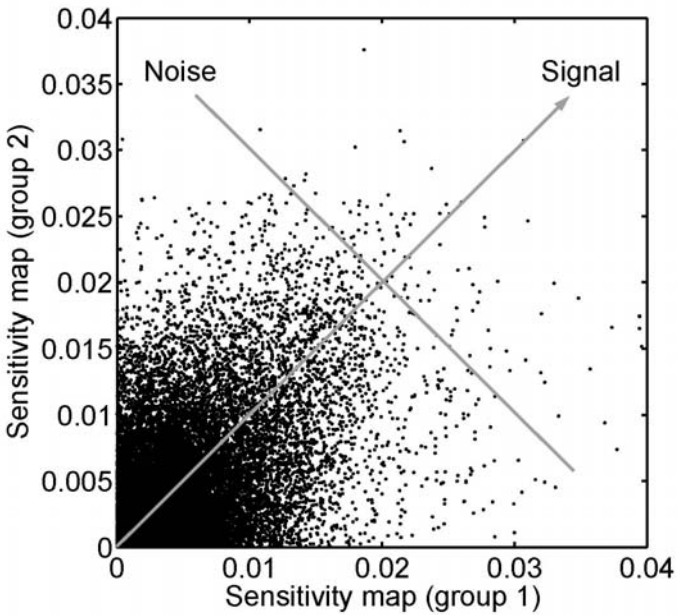
Joint work with Stephen Strother



NeuroImage: Hansen et al (1999), Hansen et al (2000), Strother et al (2002), Kjems et al. (2002), LaConte et al (2003), Strother et al (2004)

Reproducibility of internal representations





Split-half resampling provides unbiased estimate of reproducibility of SPMs (Strother et al. 2002, Kjems et al. 2002)

S. Sigurdsson et al. (2006): Sensitivity map of ANN discriminant

The sensitivity map

NeuroImage 15, 772–786 (2002)
doi:10.1006/nimg.2001.1033, available online at http://www.idealibrary.com on IDEAL®

The Quantitative Evaluation of Functional Neuroimaging Experiments: Mutual Information Learning Curves

U. Kjems,*,1 L. K. Hansen,* J. Anderson,†,‡ S. Frutiger,‡,§ S. Muley,§ J. Sidtis,§ D. Rottenberg,†,‡,§ and S. C. Strother†,‡,§,¶

*Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark; †Radiology Department, §Neurology Department, and ¶Biomedical Engineering, University of Minnesota, Minneapolis, Minnesota 55455; and ‡PET Imaging Center, VA Medical Center, Minneapolis, Minnesota 55417

$$m_j = \left\langle \left(\frac{\partial \log p(s|x)}{\partial x_j} \right)^2 \right\rangle$$

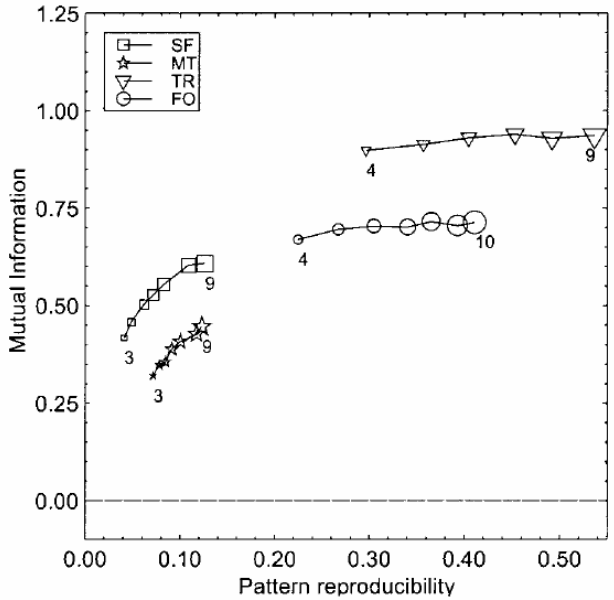


FIG. 3. Plot of scan/label mutual information versus reproducibility signal/noise for the four data sets, for varying numbers of subjects in the training set. There were 2 labels/4 scans per subject (balanced data set; Setup 1, Table 1) corresponding to the dashed solid line in Fig. 4. We see that both measures indicate improved performance of the model as the number of subjects increases.

- The sensitivity map measures the impact of a specific feature/location on the predictive distribution

Unsupervised learning:
Factor analysis generative model

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$$

$$p(\mathbf{x} \mid \mathbf{A}, \boldsymbol{\theta}) = \int p(\mathbf{x} \mid \mathbf{A}, \mathbf{s}, \boldsymbol{\Sigma}) p(\mathbf{s} \mid \boldsymbol{\theta}) d\mathbf{s}$$

$$p(\mathbf{x} \mid \mathbf{A}, \mathbf{s}, \boldsymbol{\Sigma}) = \mid 2\pi\boldsymbol{\Sigma} \mid^{-1/2} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{A}\mathbf{s})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{A}\mathbf{s})}$$

Source distribution:
PCA: … normal
ICA: … other
IFA: … Gauss. Mixt.
k-Means: .. binary

PCA: $\boldsymbol{\Sigma} = \sigma^2 \cdot \mathbf{1},$

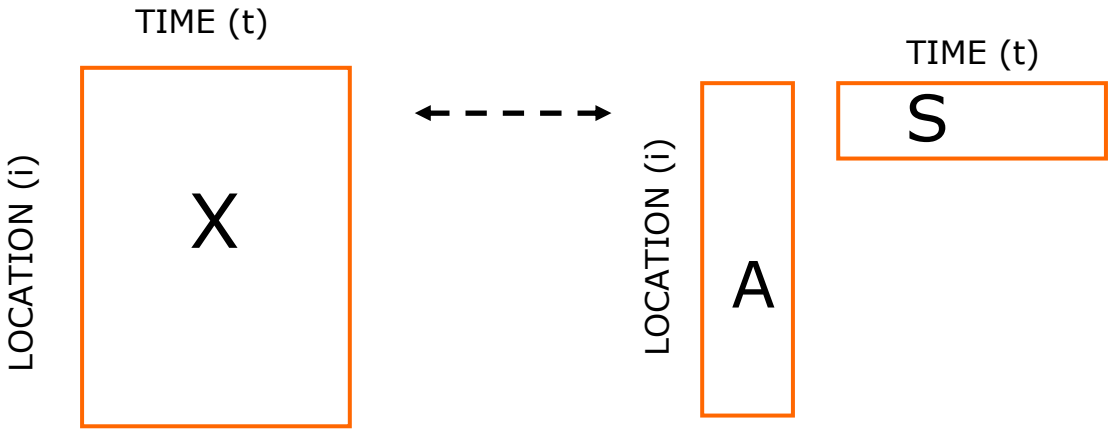
FA: $\boldsymbol{\Sigma} = \mathbf{D}$

S known: GLM
$(1\text{-}A)^{-1}$ sparse: SEM
S,A positive: NMF
S binary, simplex k-Means

Højen-Sørensen, Winther, Hansen,
Neural Compution (2002), Neurocomputing (2002)

Factor models

- Represent a datamatrix by a low-dimensional approximation
- fMRI: Identify spatio-temporal networks of activation



$$X(i,t) \approx \sum_{k=1}^{K} A(i,k)S(k,t)$$

Matrix factorization: SVD / PCA, NMF, Clustering

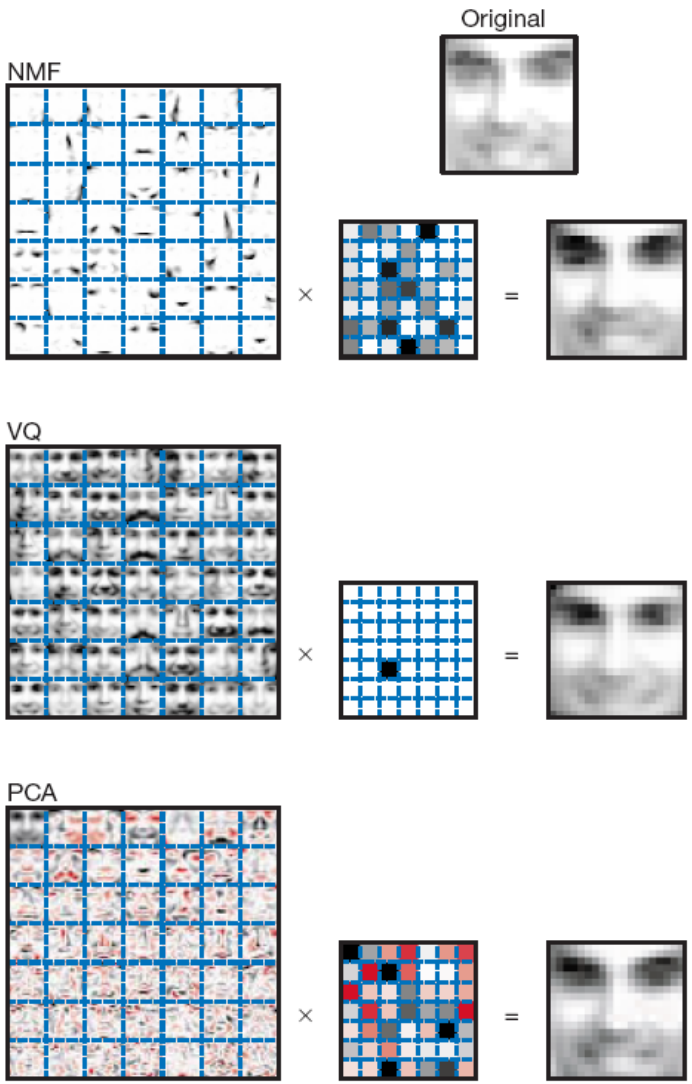


Figure 1 Non-negative matrix factorization (NMF) learns a parts-based representation of faces, whereas vector quantization (VQ) and principal components analysis (PCA) learn holistic representations. The three learning methods were applied to a database of $m = 2{,}429$ facial images, each consisting of $n = 19 \times 19$ pixels, and constituting an $n \times m$ matrix V. All three find approximate factorizations of the form $V \approx WH$, but with three different types of constraints on W and H, as described more fully in the main text and methods. As shown in the 7×7 montages, each method has learned a set of $r = 49$ basis images. Positive values are illustrated with black pixels and negative values with red pixels. A particular instance of a face, shown at top right, is approximately represented by a linear superposition of basis images. The coefficients of the linear superposition are shown next to each montage, in a 7×7 grid, and the resulting superpositions are shown on the other side of the equality sign. Unlike VQ and PCA, NMF learns to represent faces with a set of basis images resembling parts of faces.

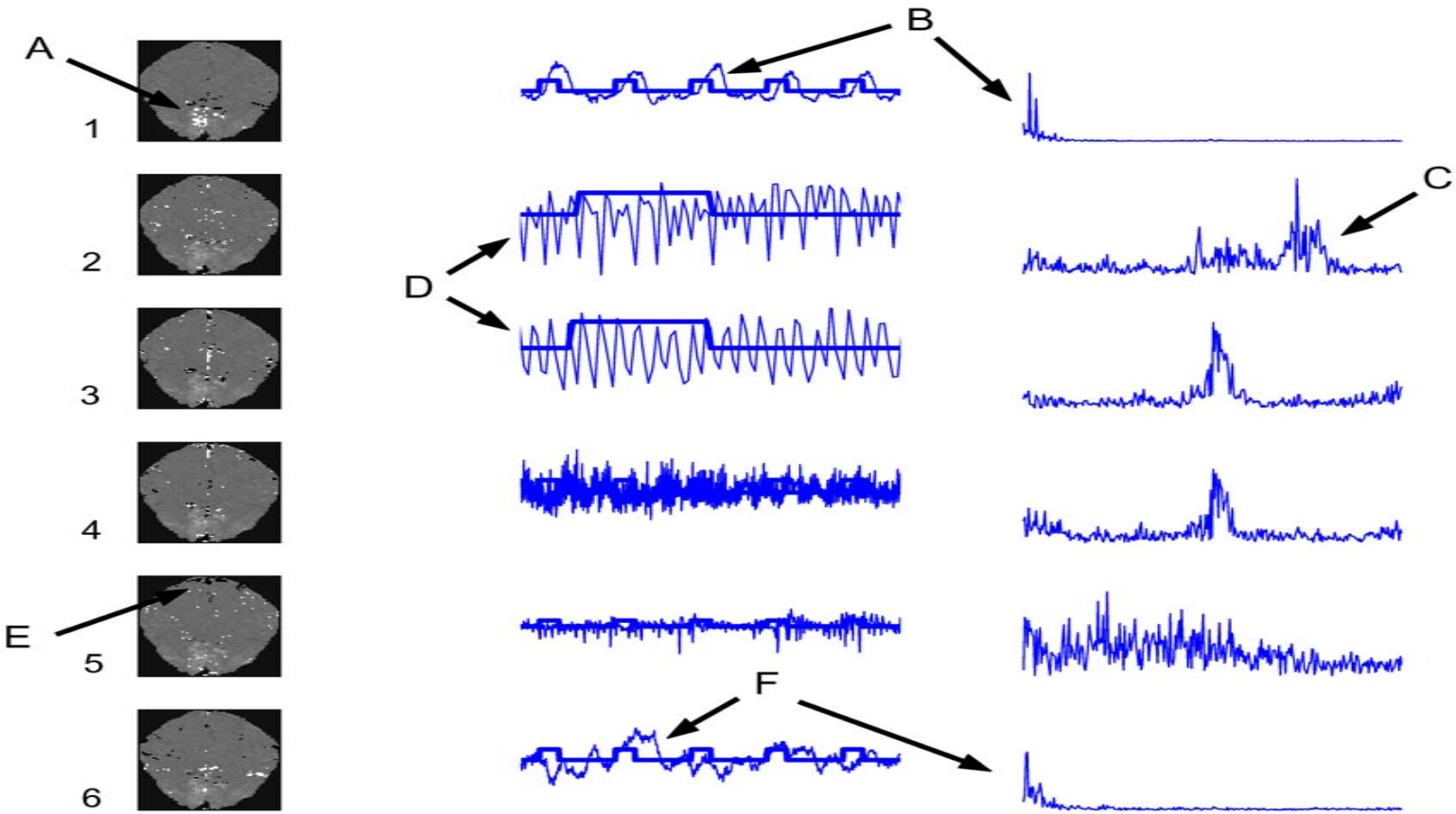
Learning the parts of objects by non-negative matrix factorization

Daniel D. Lee* **& H. Sebastian Seung*†**

** Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA*
† Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

NATURE | VOL 401 | 21 OCTOBER 1999 | www.nature.com

ICA: Assume S(k,t), S(k',t) statistically independent



(McKeown, Hansen, Sejnowski, Curr. Op. in Neurobiology (2003)

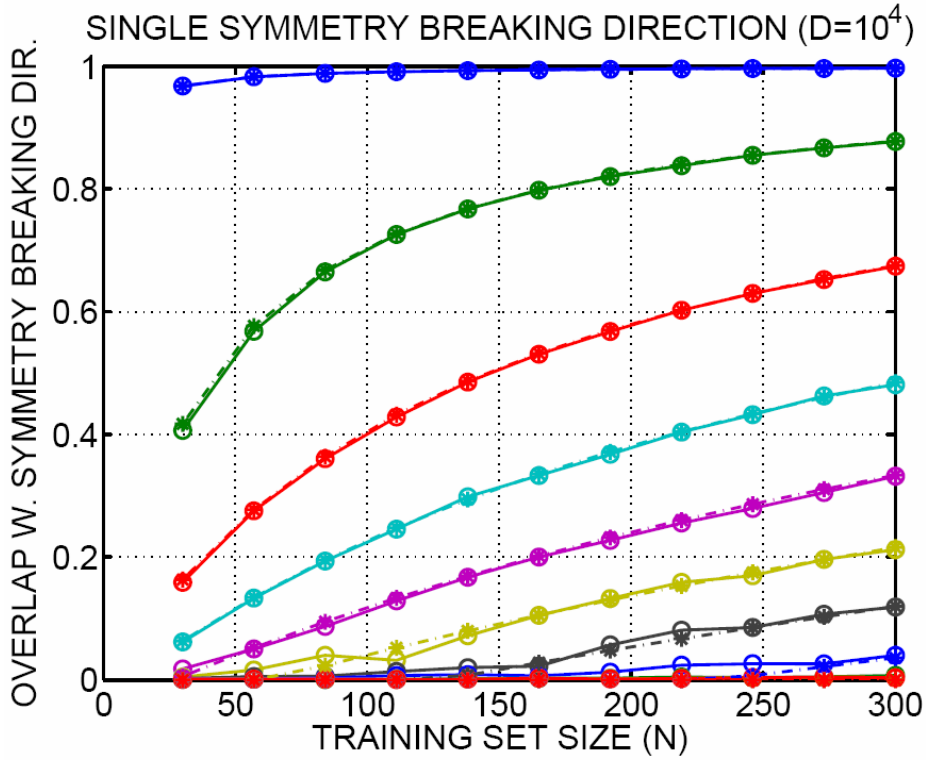
Modeling the generalizability of SVD manifold learning

- Rich physics literature on ”retarded” learning
- **Universality**
  - Generalization for a ”single symmetry breaking direction” is a function of ratio of N/D and signal to noise S
  - For subspace models-- a bit more complicated -- depends on the component SNR’s and eigenvalue separation
  - For a single direction, the mean squared overlap $R^2 = \langle (u^T_1 * u_0)^2 \rangle$ is computed for N,D -> ∞

$$R^2 = \begin{cases} (\alpha S^2 - 1)/S(1+\alpha S) & \alpha > 1/S^2 \\ 0 & \alpha \le 1/S^2 \end{cases}$$

$$\alpha = N/D \qquad S = 1/\sigma^2 \qquad N_c = D/S^2$$

Hoyle, Rattray: Phys Rev E **75** 016101 (2007)



N_c = (0.0001, 0.2, 2, 9, 27, 64, 128, 234, 400, 625)

σ = (0.01, 0.06, 0.12, 0.17, 0.23, 0.28, 0.34, 0.39, 0.45, 0.5)

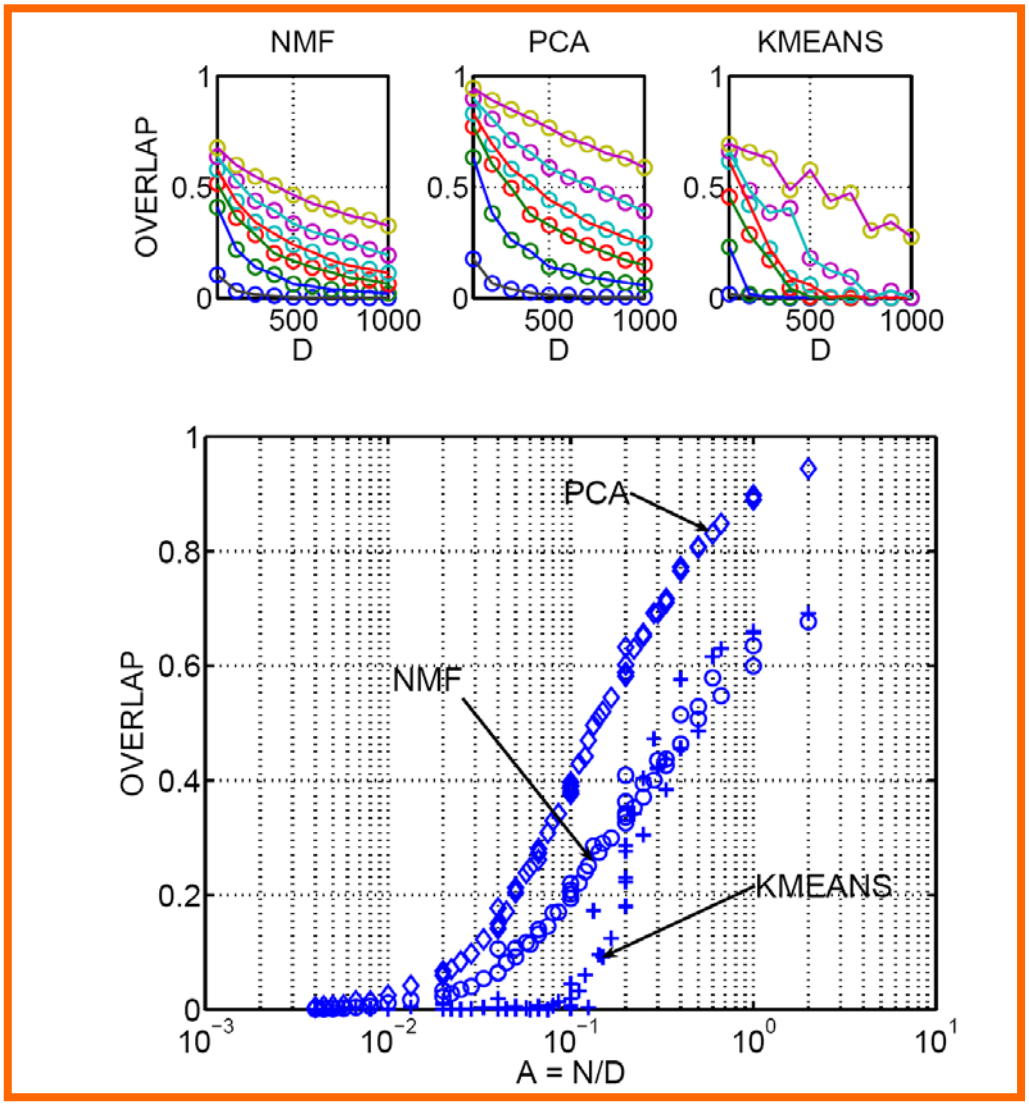
Universality in PCA, NMF, Kmeans

- Looking for universality by simulation
  - learning two clusters in white noise.
- Train K=2 component factor models.
- Measure overlap between line of sigth and plane spanned by the two factors.

Experiment

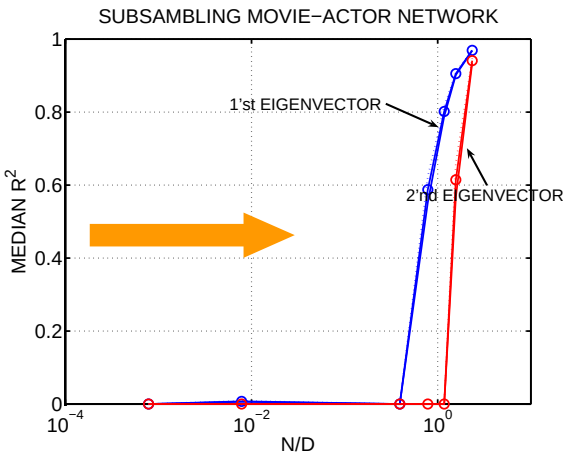
Variable: N, D

Fixed: SNR



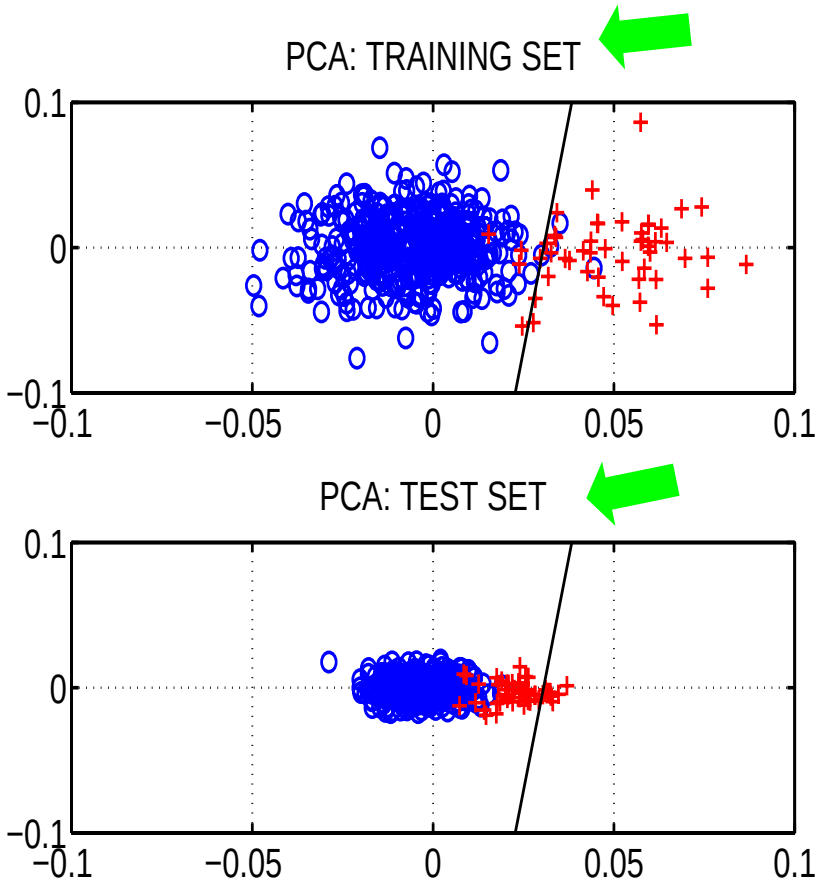
Restoring the generalizability of SVD

- Now what happens if you are on the slope of generalization, i.e., N/D is just beyond the transition to retarded learning ?

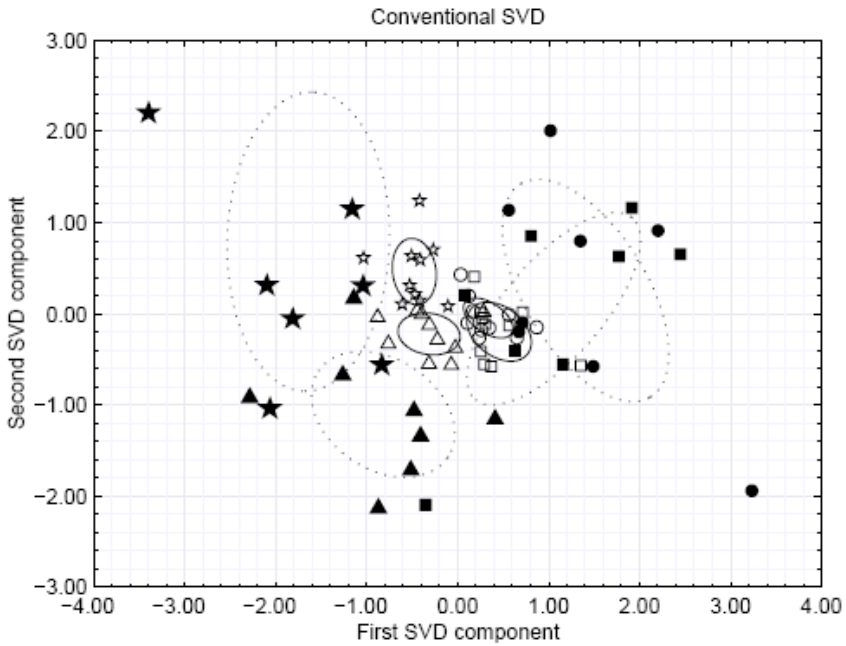


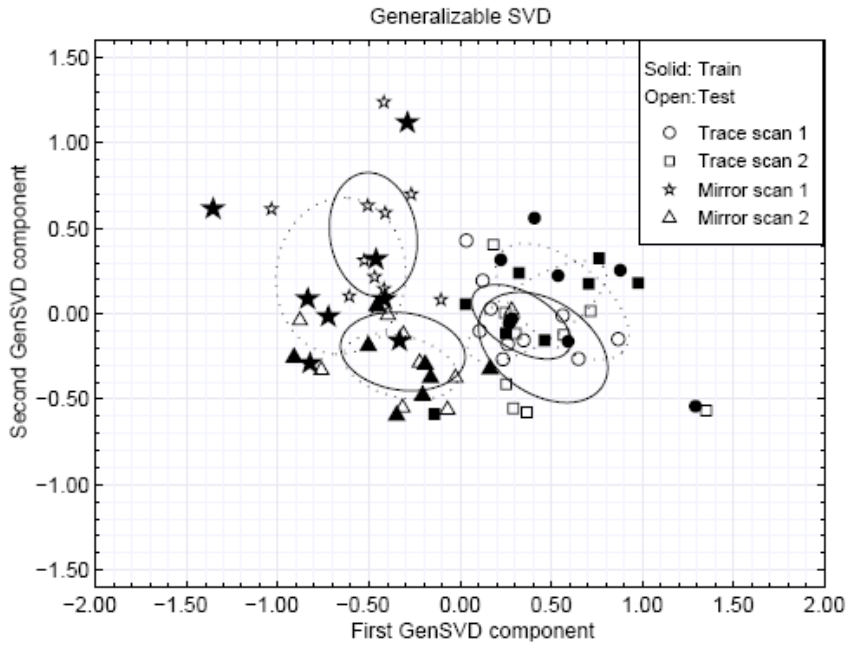
- The estimated projection is offset, hence, future projections will be too small!

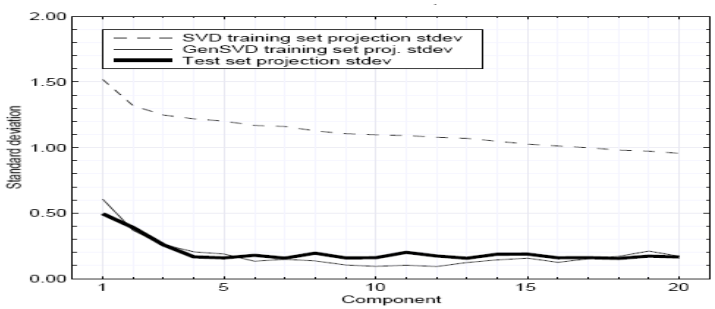
- ...problem if discriminant is optimized for unbalanced classes in the training data!



Heuristic: Leave-one-out re-scaling of SVD test projections





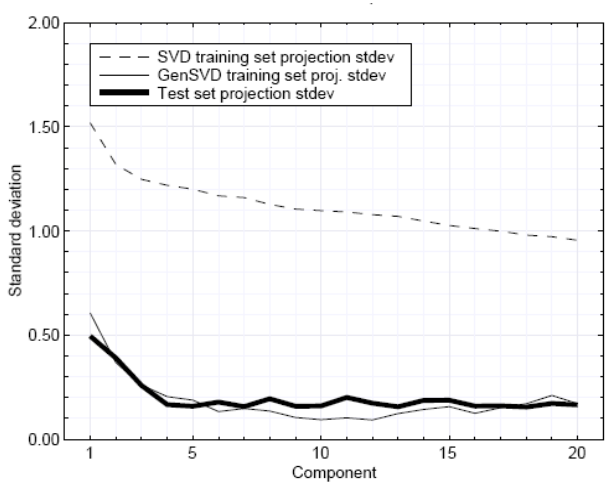


N=72, D=2.5 10^4

Kjems, Hansen, Strother: "Generalizable SVD for Ill-posed data sets" NIPS (2001)

Re-scaling the component variances

- Possible to compute the new scales by leave-one-out doing N SVD's of size N << D



$$\text{Compute } \boldsymbol{U}_0\boldsymbol{\Lambda}_0\boldsymbol{V}_0^{\mathsf{T}} = \text{svd}(X) \text{ and } \boldsymbol{Q}_0 = \left[\boldsymbol{q}_j\right] = \boldsymbol{\Lambda}_0\boldsymbol{V}_0^{\mathsf{T}}$$
$$\text{foreach } j = 1 \ldots N$$
$$\bar{\boldsymbol{q}}_{-j} = \tfrac{1}{N-1}\textstyle\sum_{j' \neq j} \boldsymbol{q}_{j'}$$
$$\text{Compute } \boldsymbol{B}_{-j}\boldsymbol{\Lambda}_{-j}\boldsymbol{V}_{-j}^{\mathsf{T}} = \text{svd}(\boldsymbol{Q}_{-j} - \bar{\boldsymbol{Q}}_{-j})$$
$$\boldsymbol{z}_j = \boldsymbol{B}_{-j}\boldsymbol{B}_{-j}^{\mathsf{T}}(\boldsymbol{q}_j - \bar{\boldsymbol{q}}_{-j})$$
$$\hat{\lambda}_i^2 = \tfrac{1}{N-1}\textstyle\sum_j z_{ij}^2$$

Kjems, Hansen, Strother: NIPS (2001)

Joint work with Trine Abrahamsen (2009)

Non-linear manifolds: kPCA

- Kernel PCA is based on non-linear mapping of data to

$$\mathbf{x}_n \rightarrow \varphi(\mathbf{x}_n) \equiv \varphi_n, \qquad n = 1, \ldots, N$$

- Aim is to locate maximum variance directions in the feature space, i.e.

$$\mathbf{l}_1 \equiv \arg\max_{\|\mathbf{l}\|=1} \left\langle \left(\mathbf{l}^T \cdot \varphi\right)^2 \right\rangle, \qquad \varphi(\mathbf{x}_n) = \sum_k \mathbf{l}_k s_{k,n}$$

- The principal direction is in the span of data: $\mathbf{l}_1 = \sum_{n=1}^{N} a_{1,n} \varphi_n$

$$\mathbf{a}_1 = \arg\max_{\|\mathbf{a}\|=1} \left\langle \mathbf{a}^T \cdot \mathbf{K} \cdot \mathbf{a} \right\rangle, \qquad \mathbf{K}_{n,n'} = \varphi_n^T \cdot \varphi_{n'} = \exp\left(-\frac{\|x_n - x_{n'}\|^2}{2c} \right)$$

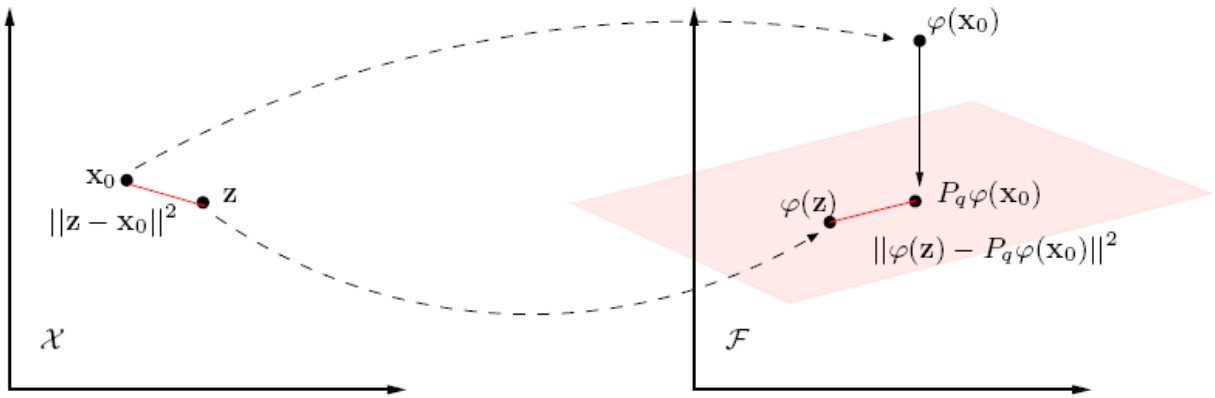
TJ Abrahamsen and LK Hansen. "Input Space Regularization Stabilizes Pre-image for Kernel PCA De-noising". To appear in Proc. of Int. Workshop on Machine Learning for Signal Processing, Grenoble, France (2009).

- Now, assume that we have a point of interest in feature space, e.g. a certain projection on to a principal direction "Φ", can we find its position "**z**" in measurement space?

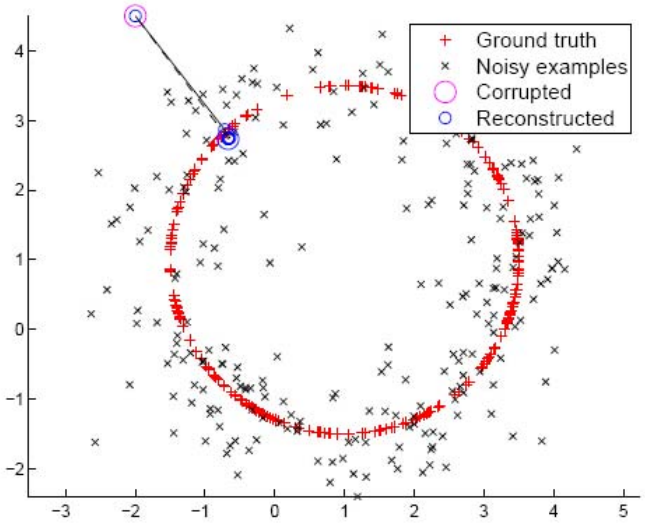
$$\mathbf{z} = \varphi^{-1}(\phi)$$

- Problems: (1) Such a point need not exist, and (2) if it does there is no reason that it should be unique!

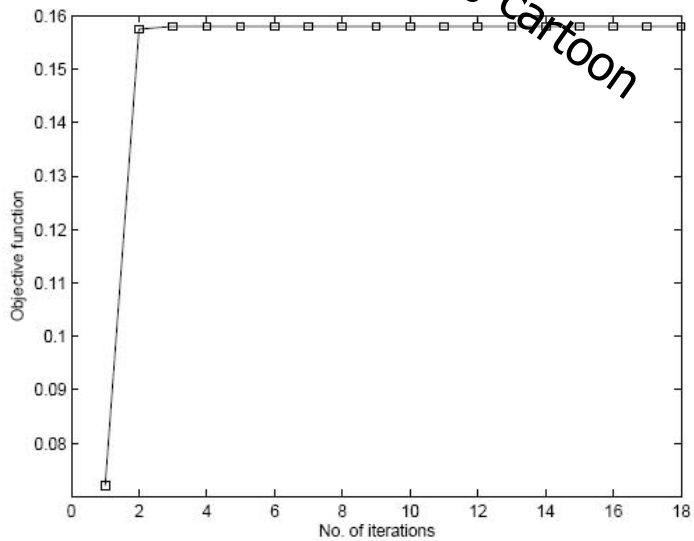
- Mika et al. (1999): Find the closest match.



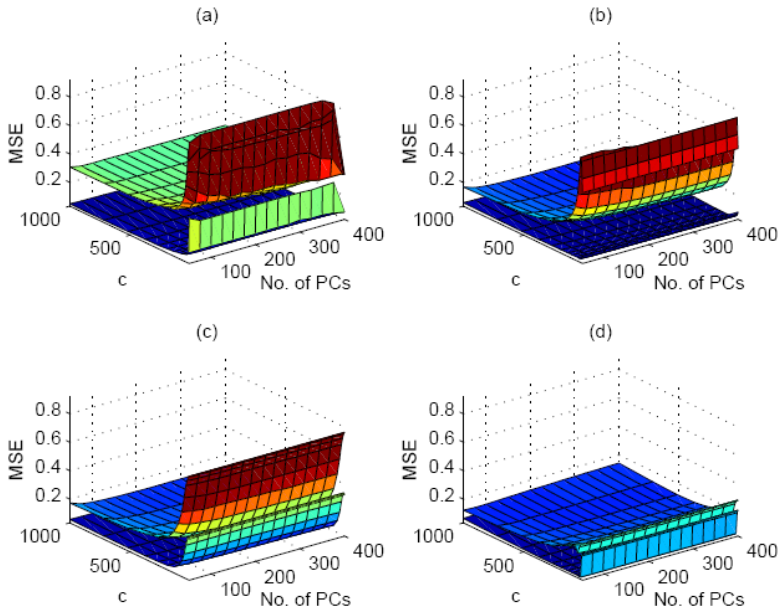
De-noising based on manifold learning

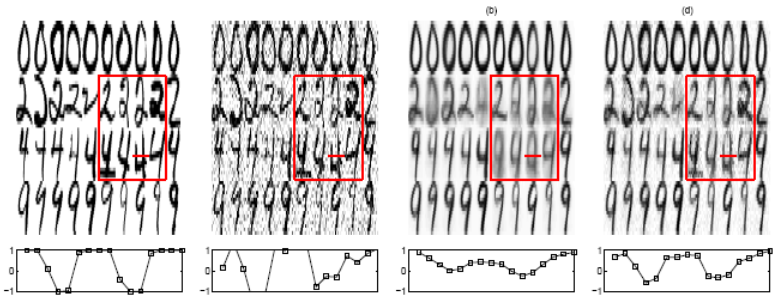


Manifold denosing cartoon



De-noising hand written digits





Comparing MSE confidence intervals for

(a) Mika et al. (1999)

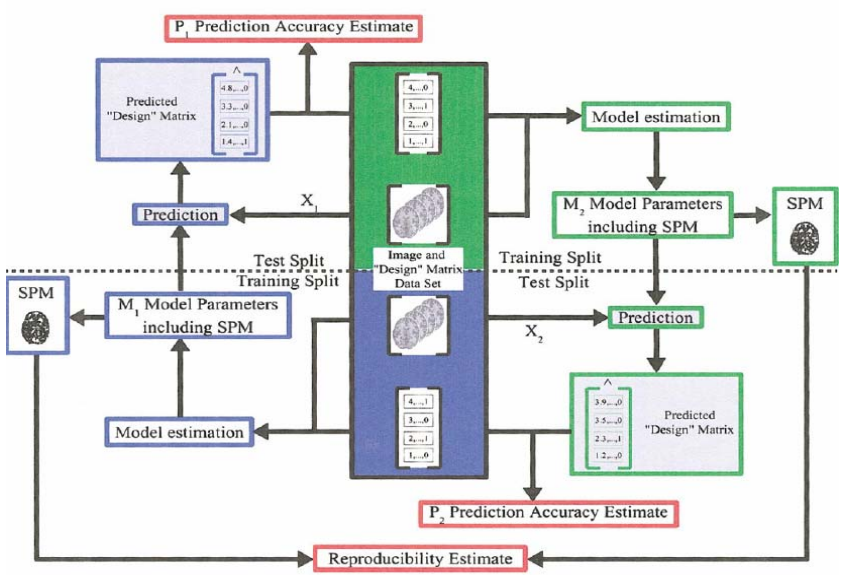
(b) Kwok et al.(2004)

(c) Dambreville et al. (2006)

(d) distance regularized pre-image estimation

Generalizable supervised models

- Non-linear sparse kernel machines SVM, kernel regression, Gaussian processes
- Parameters to be optimized wrt generalizability and reproducibility
  - Non-linearity, sparsity



Dimensional reduction using Laplacian eigenmaps

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 102 (2007) 186–191

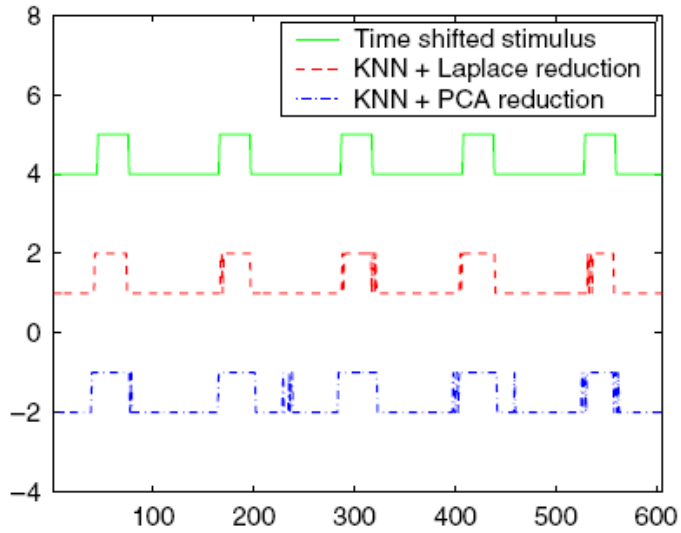
Brain and Language
www.elsevier.com/locate/b&l

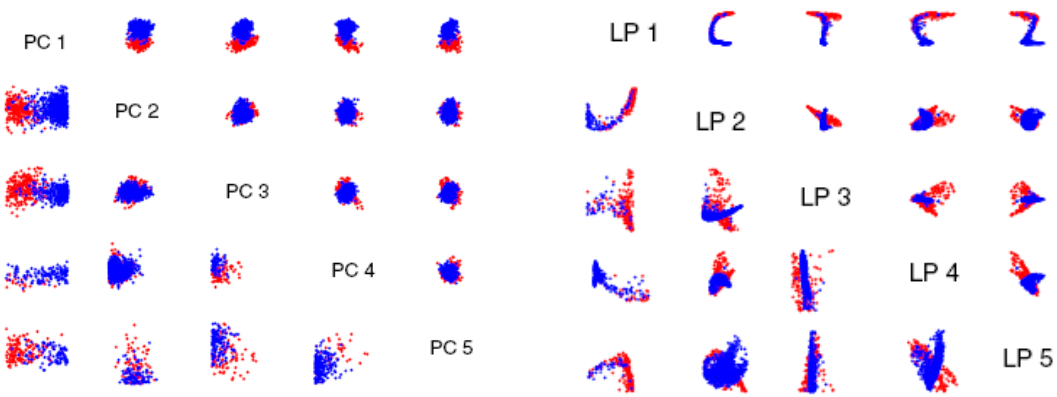
Multivariate strategies in functional magnetic resonance imaging

Lars Kai Hansen

Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Accepted 7 December 2006
Available online 16 January 2007





Belkin, Niyogi (2002)
Thirion, Faugeras (2004)
Hansen (2007)

Generalizable supervised models

- Non-linear kernel machines, SVM

$$s(n') \approx \sum_{n=1}^{N} \alpha(n) K(x_n, x_{n'})$$

$$K(x_n, x_{n'}) = \exp\left\{ -\frac{\|x_n - x_{n'}\|^2}{2c} \right\}$$

Visualization of SVM learning from fMRI

- Visualization of kernel machines
  - How to create an SPM for a kernel machine?
  - The sensitivity map for kernels

$$s(n') \approx \sum_{n=1}^{N} \alpha(n) K(x_n, x_{n'})$$

$$K(x_n, x_{n'}) = \exp\left\{ -\frac{(x_n - x_{n'})^2}{2c} \right\}$$

Visualization of kernel machine internal representations

- Existing visualization methods
  - Pre-image (Mika et al., NIPS 1998, Schölkopf et al., 1999)
    Basically an ill-defined objective, useful for denoising
  - Multi-dimensional scaling (Kwok & Tsang, ICML 2003)
    Interpolates nearest neighbors, suffers in high dimensions, has a non-intuitive weighting of neighbors

Problem: Existing methods provide local visualization, which point should be visualized? –non linearity makes this issue non-trivial.

The sensitivity map

NeuroImage 15, 772–786 (2002)
doi:10.1006/nimg.2001.1033, available online at http://www.idealibrary.com on IDEAL®

The Quantitative Evaluation of Functional Neuroimaging Experiments: Mutual Information Learning Curves

U. Kjems,*,1 L. K. Hansen,* J. Anderson,†,‡ S. Frutiger,‡,§ S. Muley,§ J. Sidtis,§ D. Rottenberg,†,‡,§ and S. C. Strother†,‡,§,¶

*Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark; †Radiology Department, §Neurology Department, and ¶Biomedical Engineering, University of Minnesota, Minneapolis, Minnesota 55455; and ‡PET Imaging Center, VA Medical Center, Minneapolis, Minnesota 55417

$$m_j = \left\langle \left(\frac{\partial \log p(s|x)}{\partial x_j} \right)^2 \right\rangle$$

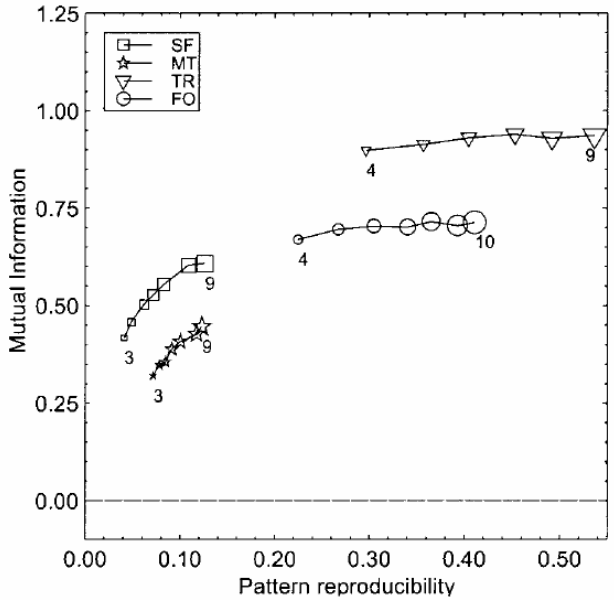


FIG. 3. Plot of scan/label mutual information versus reproducibility signal/noise for the four data sets, for varying numbers of subjects in the training set. There were 2 labels/4 scans per subject (balanced data set; Setup 1, Table 1) corresponding to the dashed solid line in Fig. 4. We see that both measures indicate improved performance of the model as the number of subjects increases.

- The sensitivity map measures the impact of a specific feature/location on the predictive distribution
- Kernel machine is manipulated to produce posterior probabilities.

Sensitivity maps for non-linear kernel regression

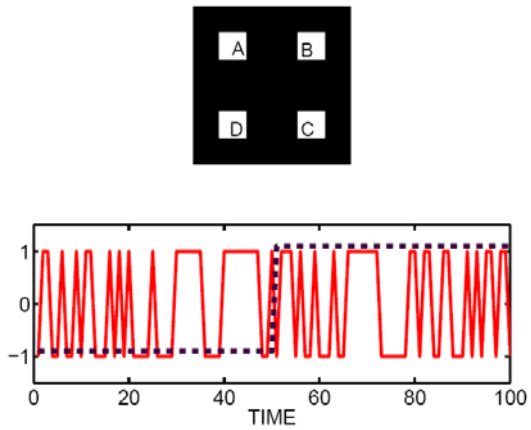


Fig. 1. XOR-image set define by four activated regions (A,B,C,D). Initially we let regions (A,B,D) be activated by random sequence taking values ± 1, as shown in example in the bottom panel (full curve). The target signal, also taking values $t_n = pm1$, and is also indicated in the bottom panel (dashed line). The region (C) is activated with an XOR-sequence relative to (A) and t_n, so that $C_n = A_n * t_n$, hence, in the active state the two regions (A,C) are randomly, but identically activated, while in the resting condition, they are random, but opposite

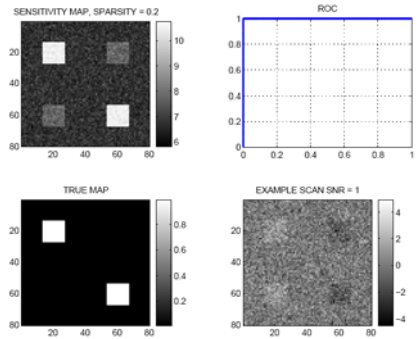


Fig. 2. XOR-image set define by four activated regions. The results of analyzing a image set with $N = 400$ examples. The image signal-to-noise ratio is $SNR = 1$, i.e., the additive noise is unit variance. The target function has in addition been contaminated by 10% random label noise. The four subplots show: The sensitivity map (upper left), the near-perfect receiver operating curve (ROC, upper right), the true activation map (lower left), and a random example of the simulated brain images. We modeled the data set using the kernel regression method. The linear model was estimated using the so-called least angle elastic net method (LARSEN) with a degree of sparsity of 0.2, i.e., using $N = 0.2 \times 400 = 80$ support vectors.

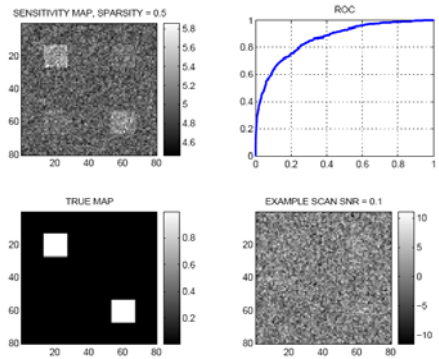


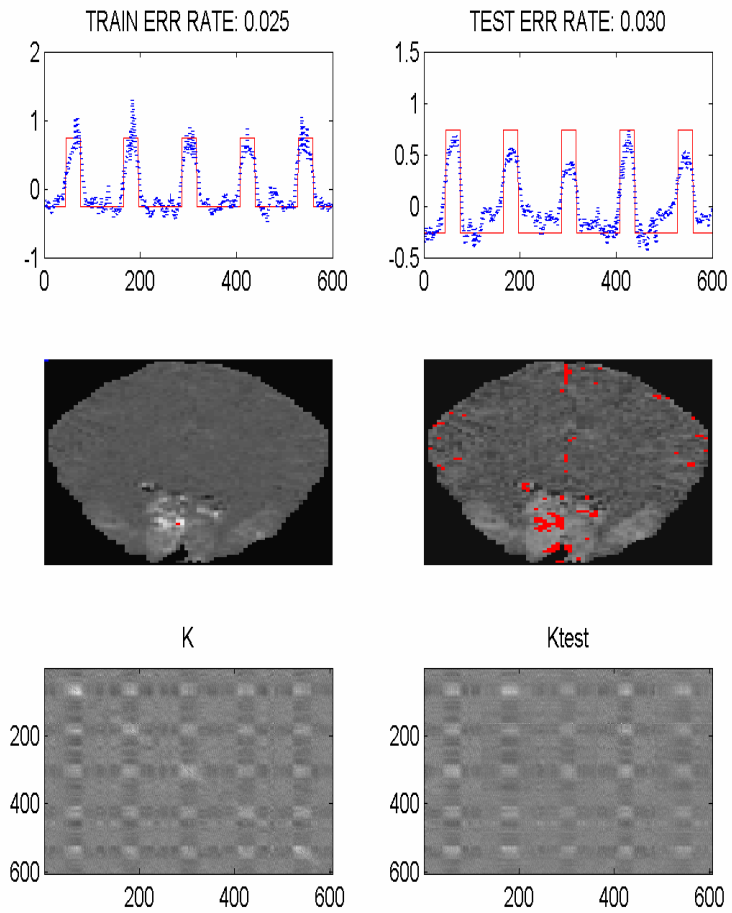
Fig. 4. XOR-image set define by four activated regions. Similar to figure 2, however the image signal-to-noise ratio is $SNR = 0.1$.

Initial dip data: Visual stimulus (TR 0.33s)

- Gaussian kernel, sparse kernel regression
- Sensitivity map computed for whole slice
- Error rates about 0.03
- How to set
  - Kernel width?
  - Sparsity?

$$s(n') \approx \sum_{n=1}^{N} \alpha(n) K(x_n, x_{n'})$$

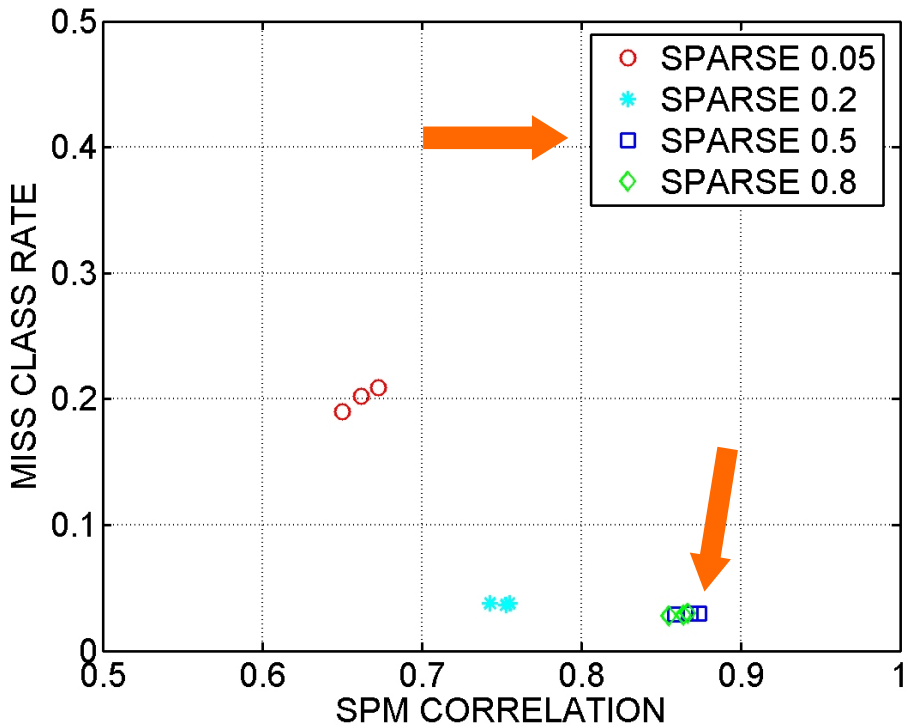
$$K(x_n, x_{n'}) = \exp\left\{-\frac{(x_n - x_{n'})^2}{2c}\right\}$$



Lars Kai Hansen
IMM, Technical University of Denmark

Initial dip data: Visual stimulus (TR 0.33s)

- Select hyperparameters of kernel machine using NPAIRS resampling
  - Degree of sparsity
  - Kernel width, localization of map



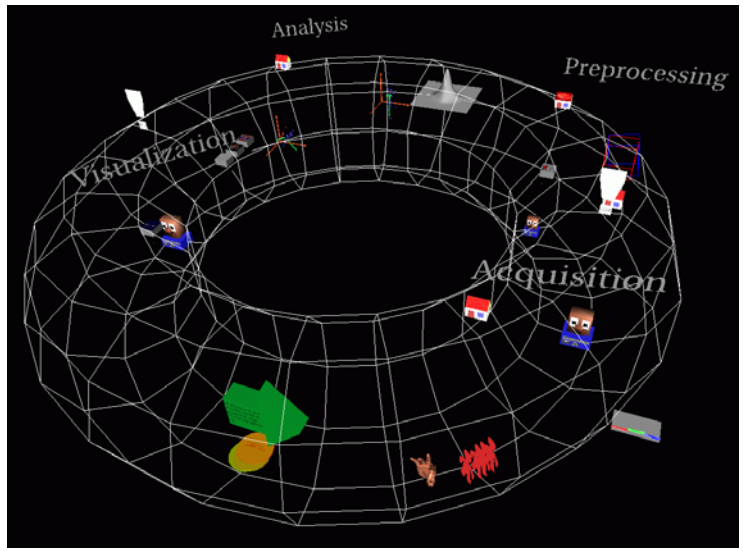
Conclusion

- Machine learning in statistical modeling has two equally important objectives
  - Generalizability
  - SPM reproducibility
- Can visualize general brain state decoders maps with perturbation based methods (saliency maps, sensitivity maps etc)
- NPAIRS split-half based framework for optimization of generalizability and robust visualizations

Acknowledgments

Lundbeck Foundation (www.cimbi.org)
NIH Human Brain Project (grant P20 MH57180)
PERCEPT / EU Commission
Danish Research Councils

www.imm.dtu.dk/cisp
hendrix.imm.dtu.dk



References

- Belkin, M., & Niyogi, P. (2002). Laplacian eigenmaps and spectral techniques for embedding and clustering. In T. G. Dietterich, S.Becker, & Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems 14 (pp. 585–591). Cambridge, MA: MIT Press.
- Dambreville S, Y. Rathi, and A. Tannenbaum, "Statistical shape analysis using kernel pca," in IS&T/SPIE Symposium on Electronic Imaging, 2006.
- Hansen L.K., F.AA. Nielsen, S.C. Strother, N. Lange Consensus Inference in Neuroimaging, NeuroImage 13 1212-1218, (2001).
- Hansen, L. K., Paulson, O. B., Larsen, J., Nielsen, F. A., Strother, S. C., Rostrup, E., et al. (1999). Generalizable patterns in neuroimaging: how many principal components? NeuroImage, 9, 534–544.
- Kjems, U., Hansen, L., Anderson, J., Frutiger, S., Muley, S., Sidtis, J., et al. (2002). The quantitative evaluation of functional neuroimaging experiments: Mutual information learning curves. NeuroImage, 15(4), 772–786.
- Kjems, U., Hansen, L. K., & Strother, S. C. (2000). Generalizable singular value decomposition for ill-posed datasets. In NIPS (pp. 549–555).
- Kwok JT and I. WH Tsang, "The preimage problem in kernel methods," IEEE transactions on neural networks, vol. 15, no. 6, pp. 1517–1525, 2004.
- McKeown, M., Hansen, L. K., & Sejnowski, T. J. (2003). Independent component analysis for fMRI: What is signal and what is noise? Current Opinion in Neurobiology, 13(5), 620–629.
- Mika S, B. Sch¨olkopf, A. Smola, K. R. Muller, M. Scholz, and G. R¨atsch, "Kernel pca and de-noising in feature spaces," in Advances in Neural Information Processing Systems 11. 1999, pp. 536–542, MIT Press.
- Mørch, N., Hansen, L., Strother, S., Svarer, C., Rottenberg, D., Lautrup, B., (1997). Nonlinear vs. linear models in functional neuroimaging: learning curves and generalization crossover. In Proc. IPMI 1997. Vol. Springer Lecture Notes in Computer Science 1230, pp. 259–270.
- Mørch, N., Kjems, U., Hansen, L., Svarer, C., Law, I., Lautrup, B., (1995). Visualization of neural networks using saliency maps. In:Proceedings of the 1995 IEEE Int. Conf. on Neural Networks. Vol. 4 (pp. 2085–2090).
- Scholkopf B, S.Mika, C. J. C. Burges, P. Knirsch, K. R. Muller,et al. Input space versus feature space in kernel-based methods," IEEE Transactions On Neural Networks, vol. 10, no. 5, pp. 1000–1017, 1999.
- Strother, S., Anderson, J., Hansen, L., Kjems, U., et al. (2002). The quantitative evaluation of functional neuroimaging experiments: The NPAIRS data analysis framework. NeuroImage,15(4), 747–771.
- Thirion, B., & Faugeras, O. (2004). Nonlinear dimension reduction of fMRIdata: The Laplacian embedding approach. In Proceedings of the 2nd Proceedings of IEEE ISBI (pp. 372–375). Arlington, VA.