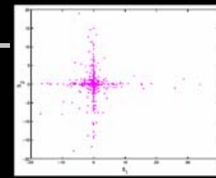
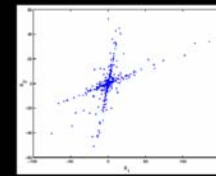

analysis
data
evidence
leads
Cognitive
present
level
unsupervised
well-aligned
abstract
social
low
speech
group
independent
activity
components
ensuing
huma
interactions
resulting
grouping
music
COCA
process
structure
component
text
defined





Outline

- Cognitive component analysis: A definition
- Cognitive component analysis: A motivation
- Machine learning tools (ICA, sparse representations)
- Examples
  - Phonemes as cognitive components
  - Communities in social networks
  - Example: Grouping of mixed media data
- Theory: Aspects of generalizability in unsupervised learning
- Conclusion and outlook

Cognitive Component Analysis

ENCYCLOPÆDIA BRITANNICA online

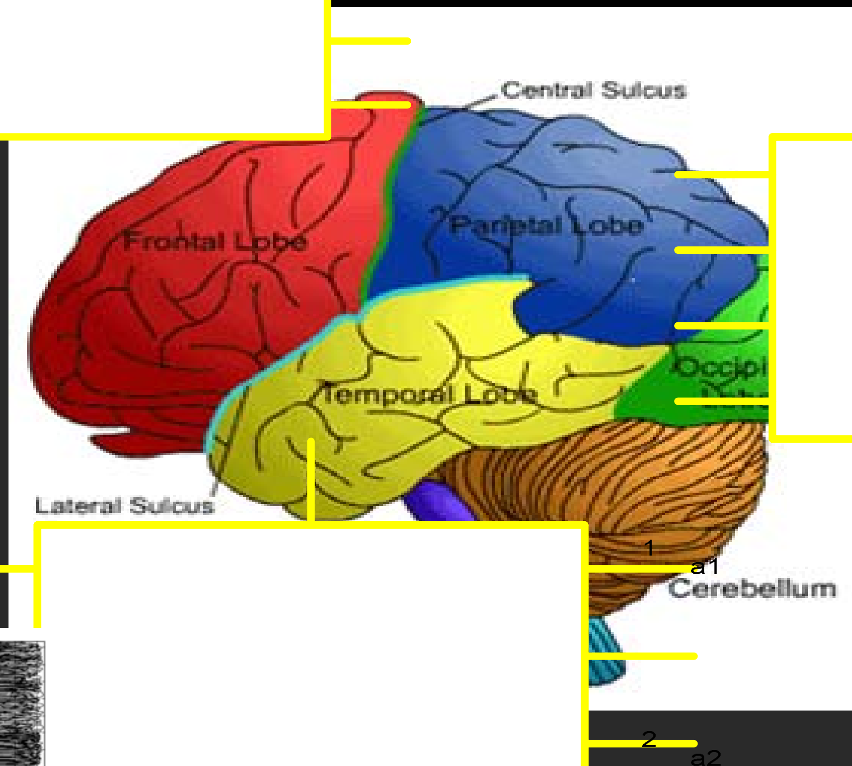
- A broad definition of cognition
  - "The act or process of knowing - Cognition includes every mental process that may be described as an experience of knowing (including perceiving, recognizing, conceiving and reasoning) as distinguished from an experience of feeling and willing."

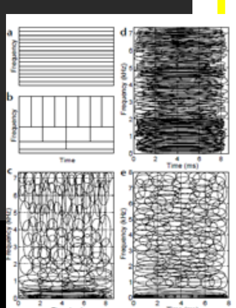
    -Brittanica Online (2005)

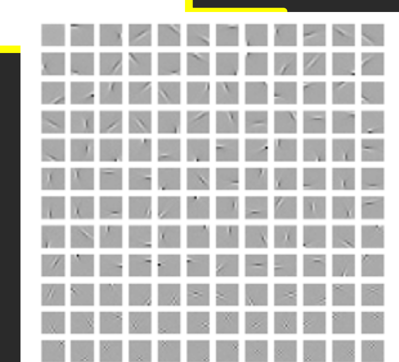
- Cognitive component analysis (COCA)
  - The process of unsupervised grouping of data so that the resulting group structure is well-aligned with grouping based on human cognitive activity:

Cognitive compatibility as "Micro Turing" tests....

Ecological modeling







Important for engineering proxies for human information processing…

Cf. efficient coding of ”context-to-action” mapping

Cognitive component analysis and the notion of an ”object”

- Cognitive component analysis (COCA)
  - The process of unsupervised grouping of data so that the ensuing group structure is well-aligned with that resulting from human cognitive activity

- The ”object” is a basic notion in cognitive psychology; E.g. ”number of objects in short time memory”.
  - Definition: ”…there is a strong focus on objects as perceptual signaling units. A pragmatic definition of an object is a "signal source that maintains a minimum of independent behaviors in a given environment".
  - Can cognitive component analysis be a step towards an general purpose definition of an “object”?

- Theoretical issues: we are interested in the relation between supervised and un-supervised learning. Related to the discussion of the utility of unlabeled examples in supervised learning.

Many generalizations are possible – which ones will make sense to a human?

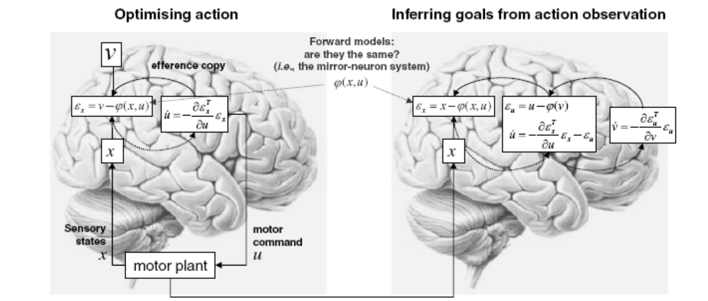
Cognive modeling: Mental models

- Human cognition is often to act on *weak signals*, i.e., lack of information or poor signal to noise conditions.

- Mental models can be more or less well-aligned with actual physics/ecology, c.f.

  Friston et al.'s *Predictive coding* model

  – J.M.Kilner, K.J.Friston C.D.Frith. Predictive coding: an account of the mirror neuron system. Cogn Process. 2007 Sep;8(3):159-66



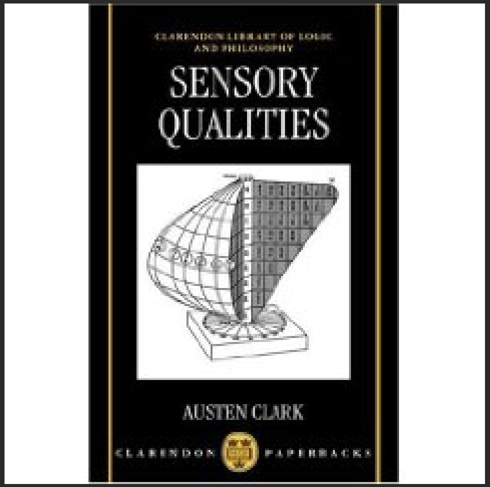
Simplified scheme for **motor control**: The motor plant receives commands (u) and changes the sensory input (x). These commands are constructed by a controller (inverse model) to minimise the difference between the desired trajectory of the states (v) and those predicted by the forward model. The forward (predictor) model is a function of the [efference] copy of the motor command. In this case, the goal is known and only u is optimised. The inverse model or controller is represented as a recognition function that minimises prediction error by gradient descent (the dot above a variable means rate of change).

Simplified scheme for **action-perception**: A hierarchical generative or forward model of sensory states is inverted to infer their [unknown] causes. These causes include the motor commands (u) of the observed agent that are inferred by minimising the difference between the observed and predicted states (using a forward model of the motor plant). The agent's goals are inferred by minimising the error between the inferred commands (u) and those predicted by their forward model, which is a function of goals.

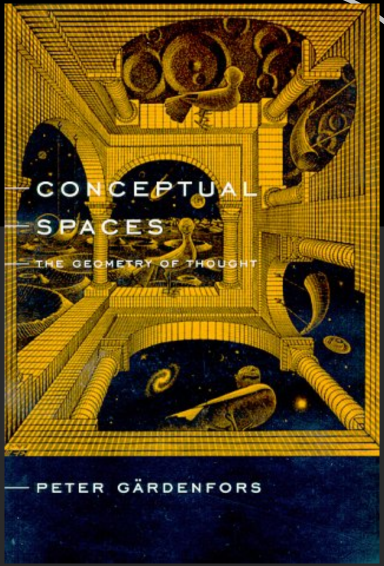
Qualia, sensory data

- Qualitative data often mapped with MDS multidimensional scaling: low-dimensional, neighbor preserving Euclidean representation

- Austen Clark in Sensory Qualities (1993):
  - "*The number of dimensions of the MDS space corresponds to the number of independent ways in which stimuli in that modality can be sensed to resemble or differ, but the dimensions per se have no meaning*"



Gärdenfors' "Conceptual Spaces"

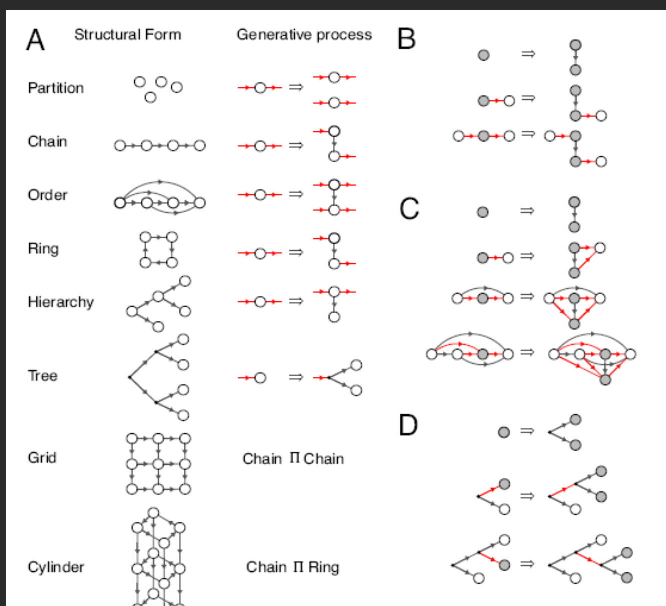


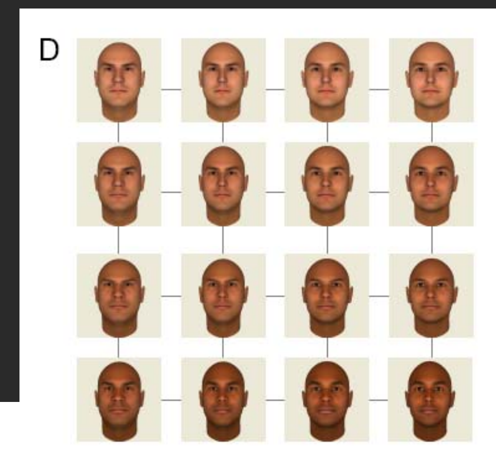
- Cognitive models:
  - Symbolic, associative/connectionist, geometrical
- Human cognition ~ similarity judgments ~ Gestalt theory ~ geometrical proximity
- How to identify conceptual spaces, i.e., geometrical representations?

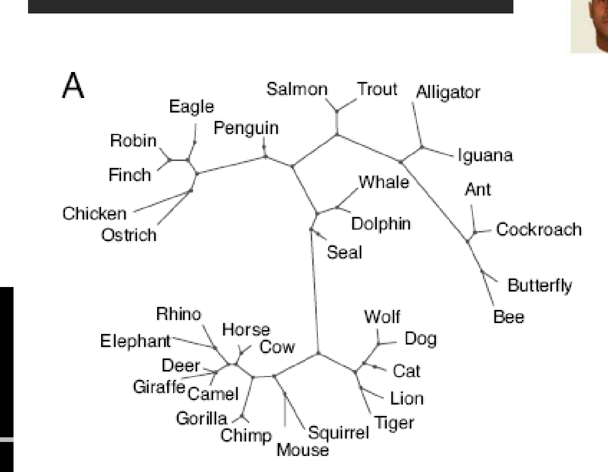
(Gärdenfors, 2000)

Kemp-Tenenbaum – Discovery of structural form
(PNAS, 2008)

Human mind has only access to relatively low complexity modeling tools.







Cognitive Component Analysis: Why independence?

- Cognitive component analysis (COCA)
  - The process of unsupervised grouping of cognitive events so that the resulting group structure is well-aligned with manual grouping

- The **object** is a basic notion in cognitive psychology;
  - E.g. "modeling number of objects in short time memory".
  - A pragmatic definition of **an object** is "**a signal source that maintains a minimum of independent behavior in a given environment**".
  - Thus, independent component analysis could attain a key role in understanding cognition? (Hypothesis presented at AKRR, Helsinki 2005)

- Theoretical issues: we are interested in the relation between supervised and un-supervised learning. **How compatible are the hidden representations of supervised and unsupervised models?** Related to the discussion of the utility of unlabeled examples in supervised learning.

Cognitive compatibility

Unsupervised Learning

Hidden variable

$$p(\mathbf{s} \mid \mathbf{x}, \mathbf{w}_u) \propto p(\mathbf{x} \mid \mathbf{s}, \mathbf{w}_u) p(\mathbf{s} \mid \mathbf{w}_u)$$

”Cognitive event”:
Data, sound, image, behavior

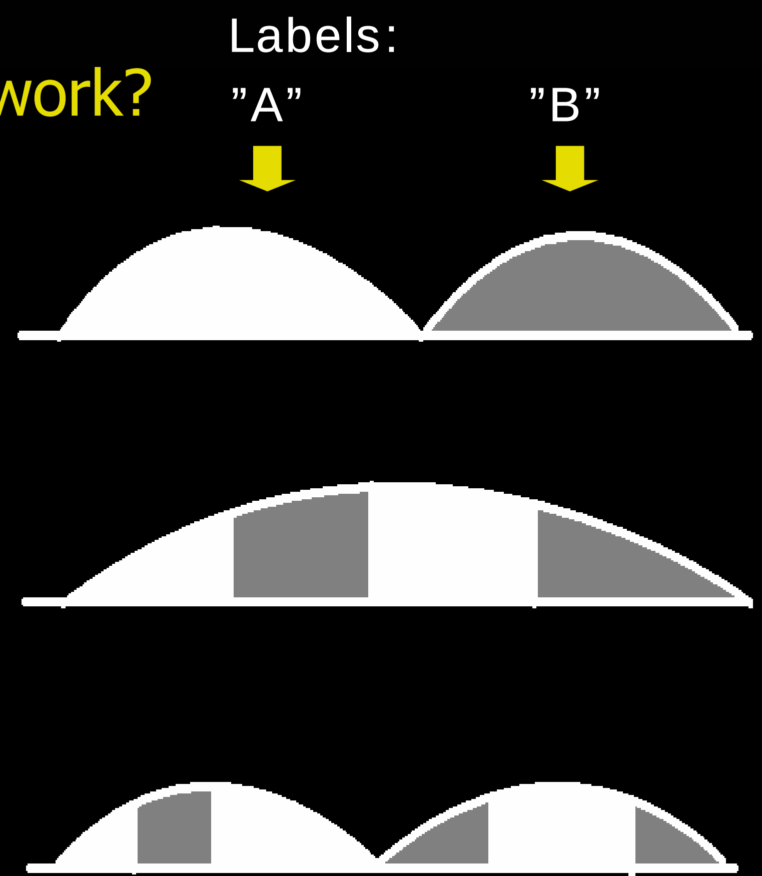
Supervised learning

$$p(\mathbf{y} \mid \mathbf{x}, \mathbf{w}_s)$$

”Cognitive” label, i.e. provided by a human observer

How well do these learned representations match: s = y ?

When can COCA be expected to work?

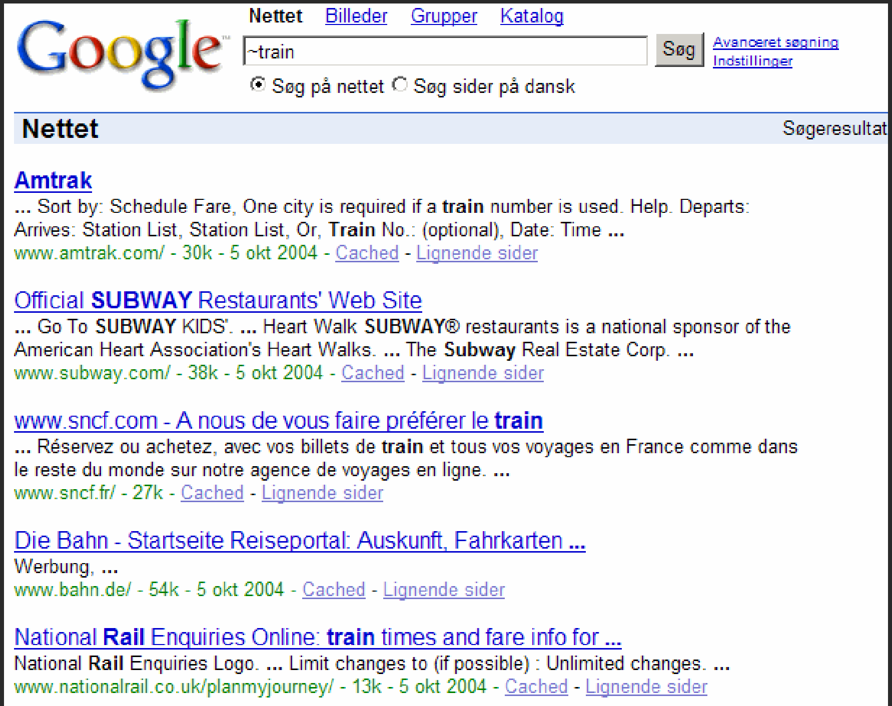


- If "statistical structure" in the relevant feature space is well aligned with the label structure we expect high **cognitive compatibility**
- Unsupervised-then-supervised learning can explain "learning from single example"
- Benign case,
  malign case,
  worst case....

How will COCA help computers understand media content?

- Understand = simulate cognitive processing in humans
- Metadata estimation from media (sound/images/video/ deep web objects)

- Basic signal processing tools are known (perceptual models…)
- ISO/MPEG standardization



Vector space representation

- Abstract representation - can be used for all digital media
- A “cognitive event” is represented as a point in a high-dimensional ”feature space” – document similarity ~ spatial proximity in a given metric

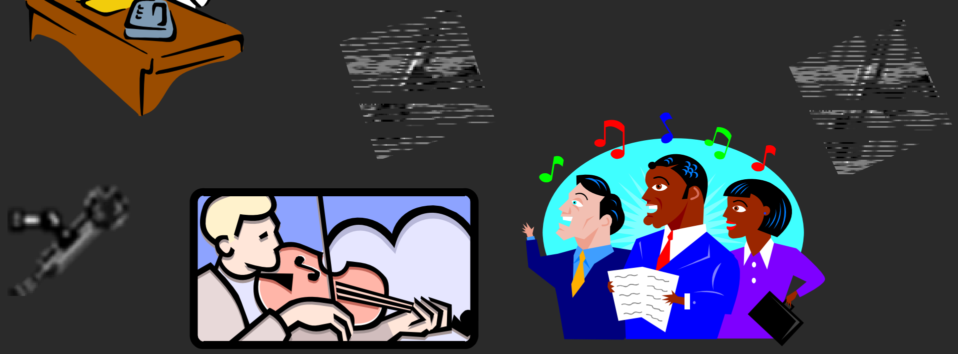
- Text: Term/keyword histogram, N-grams
- Image: Color histogram, texture measures
- Video: Object coordinates (tracking), active appearance models
- Sound: Spectral coefficients, mel cepstral coefficients, gamma tone filters

Contexts can be identified by their feature associations (= Latent semantics)

The independent context hypothesis (ICA)

- Challenge: Presence of multiple agents/contexts
- Need to ”blindly” separate source signals = learn contexts
- Machine learning come to rescue!

$$x(feature, time) = \sum_k A(feature, k)\ s(k, time)$$



Linear mixing generative model ICA - "Synthesis"
simplistic model incorporating sparsity and independence

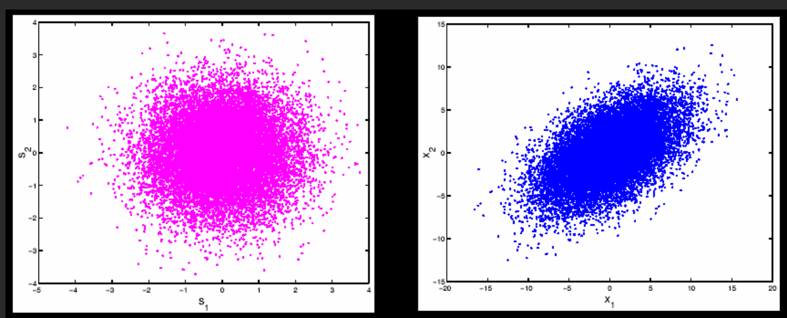
Space-time matrix

Component's "where"

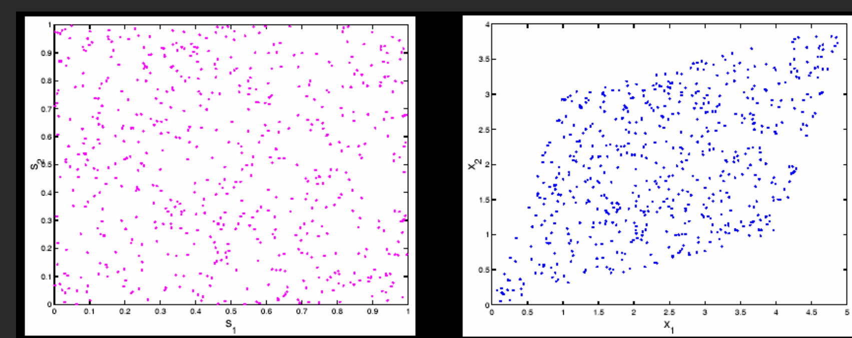
Vector of "what"

$$x(loc,time) = \sum_k A(loc,k)\ s(k,time)$$

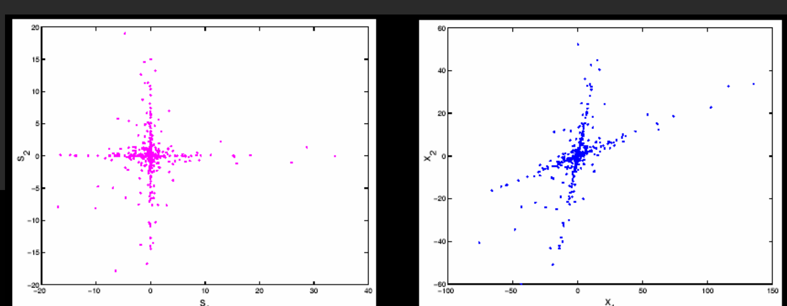
Normal sources



Dense sources



Sparse sources



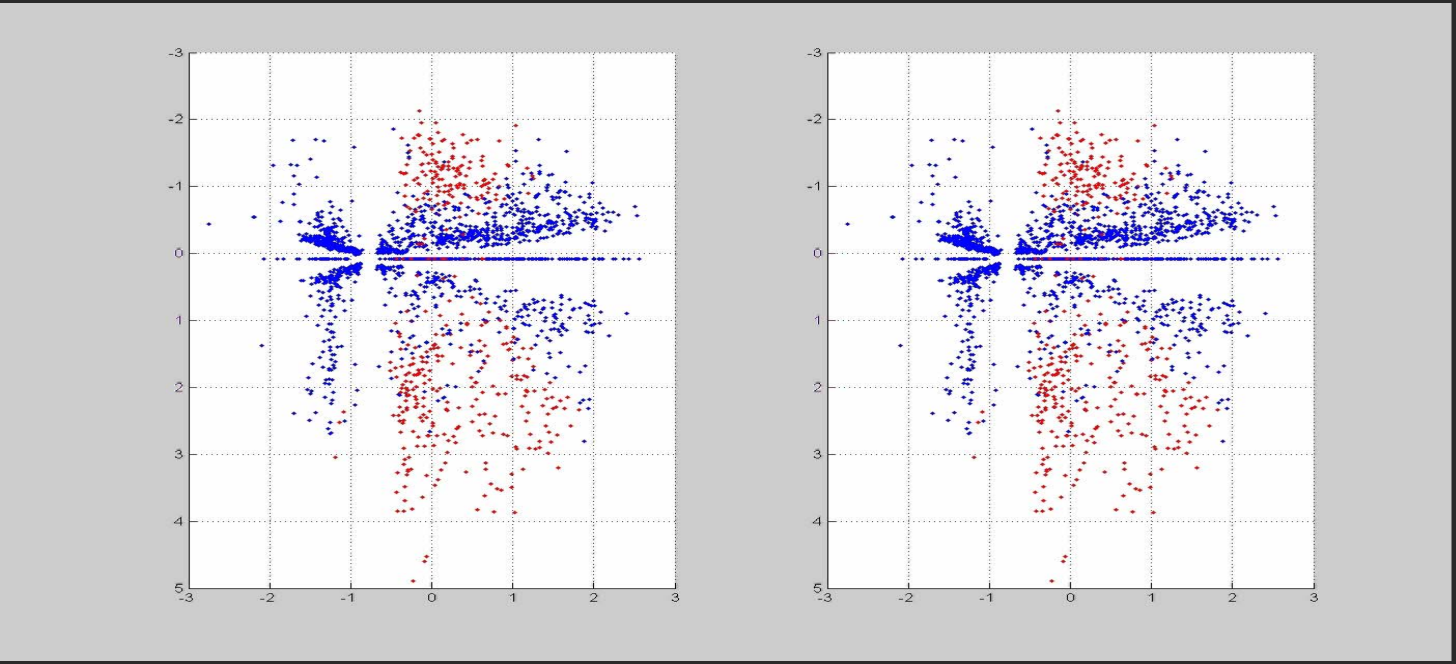
Protocol for comparing supervised and unsupervised learning

- Use the “unsupervised-then-supervised” scheme to implement a classifier:
  - Train the unsupervised scheme, eg., ICA
  - Freeze the ICA representation (A matrix)
  - Train a simple (e.g. Naïve Bayes) classifier using the features obtained in unsupervised learning Use
- Compare with supervised classifier
  - Error rates of the two systems
  - Compare posterior probabilities

Research question: Can this simple linear model based on independence account for the pattern of human errors?

Phoneme classification

Nasal vs oral: ”Esprit project ROARS” (Alinat et al., 1993)



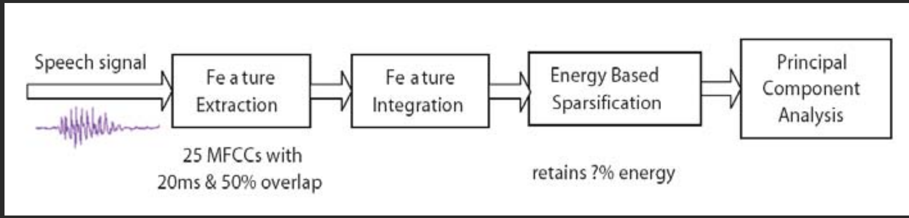
Binary classification

Error rates: **0.23** (sup.), **0.22** (unsup.)

Bitrates: **0.48** (sup.), **0.39** (unsup.)

Cognitive components of speech

- Basic representation: Mel weigthed cepstral coefficients (MFCCs)
- Modeling at different time scales 20 msec – 1000 msec



- Phonemes
- Gender
- Speaker identity

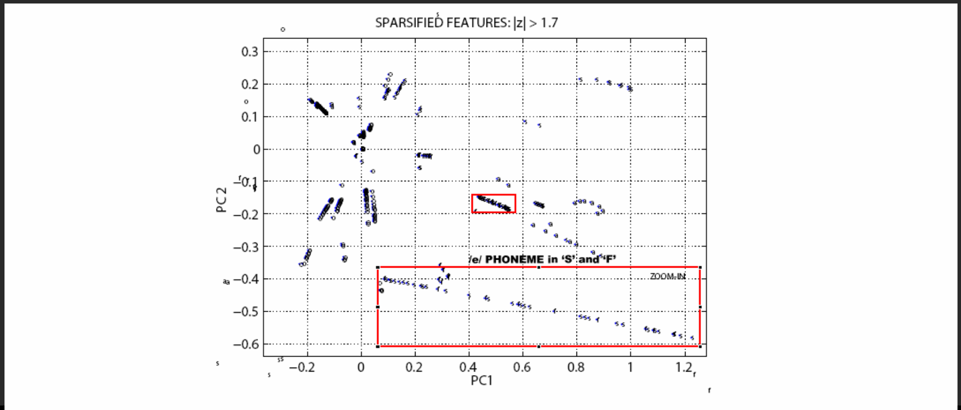
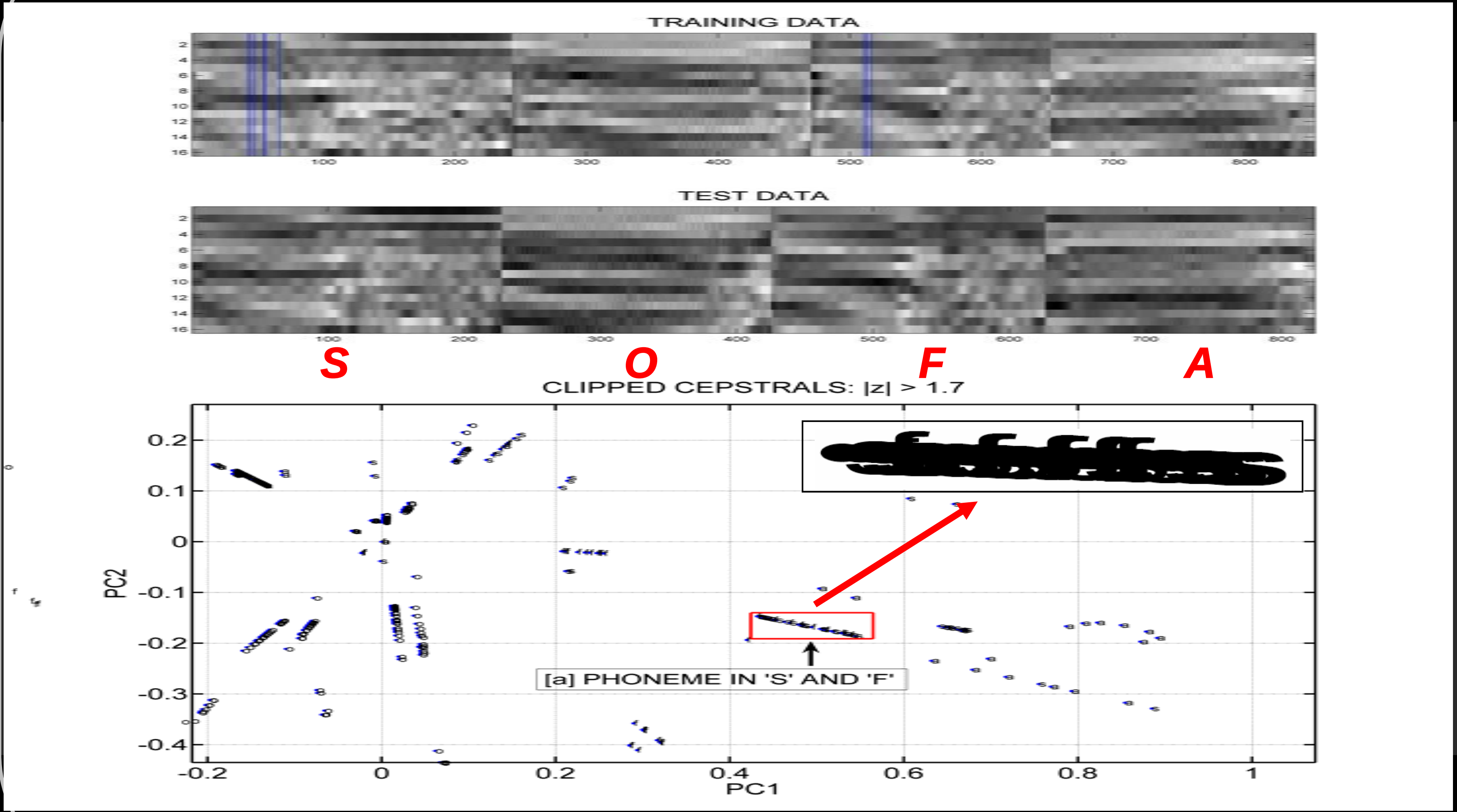
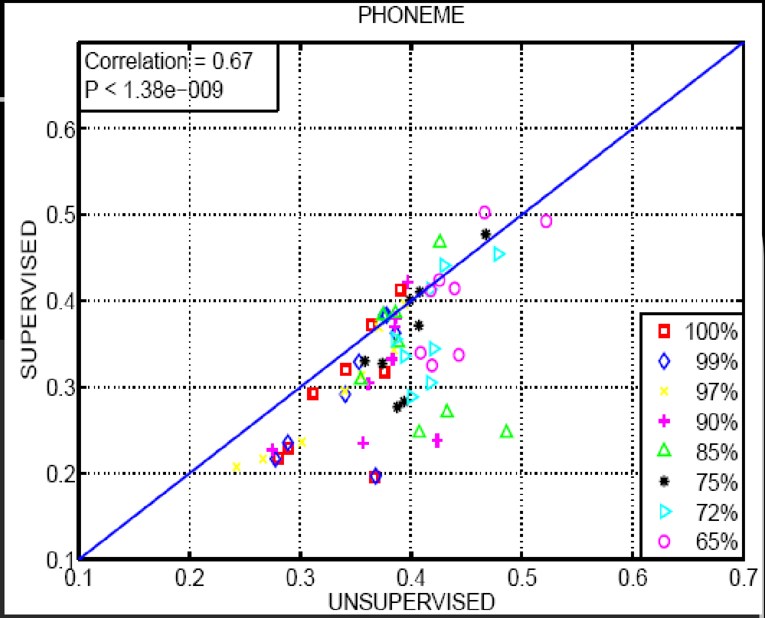


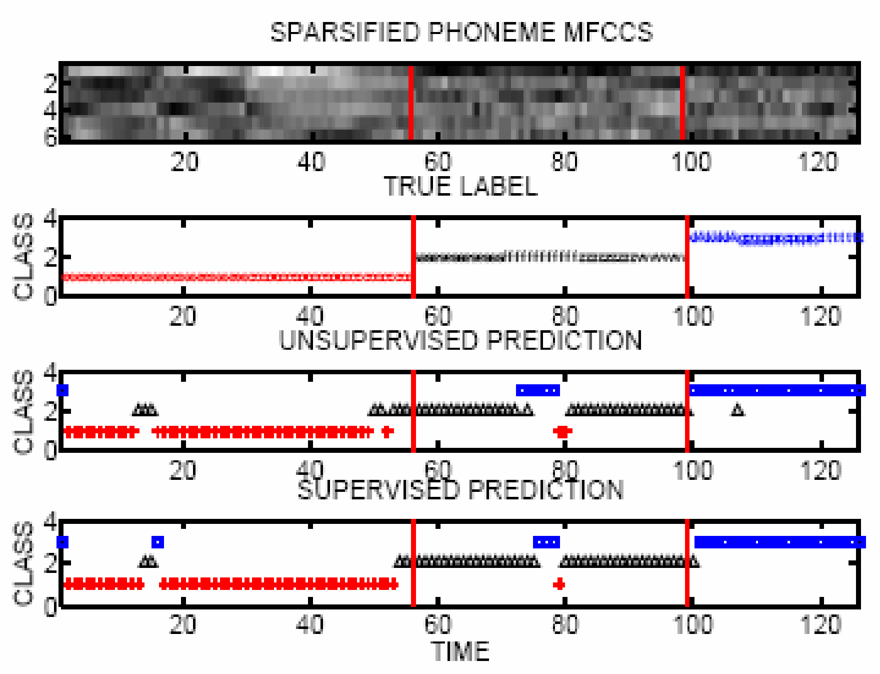
Figure 3: The latent space is formed by the two first principal components of data consisting of four separate utterances representing the sounds 's', 'o', 'f', 'a'. The structure clearly shows the sparse component mixture, with 'rays' emanating from the origin (0,0). The ray embraced in a rectangle contains a mixture of 's' and 'f' features, a cognitive component associated with the vowel /e/ sound.


TRAINING DATA
TEST DATA
S
O
F
A
CLIPPED CEPSTRALS: |z| > 1.7
PC2
PC1
[a] PHONEME IN 'S' AND 'F'



Error rate comparison

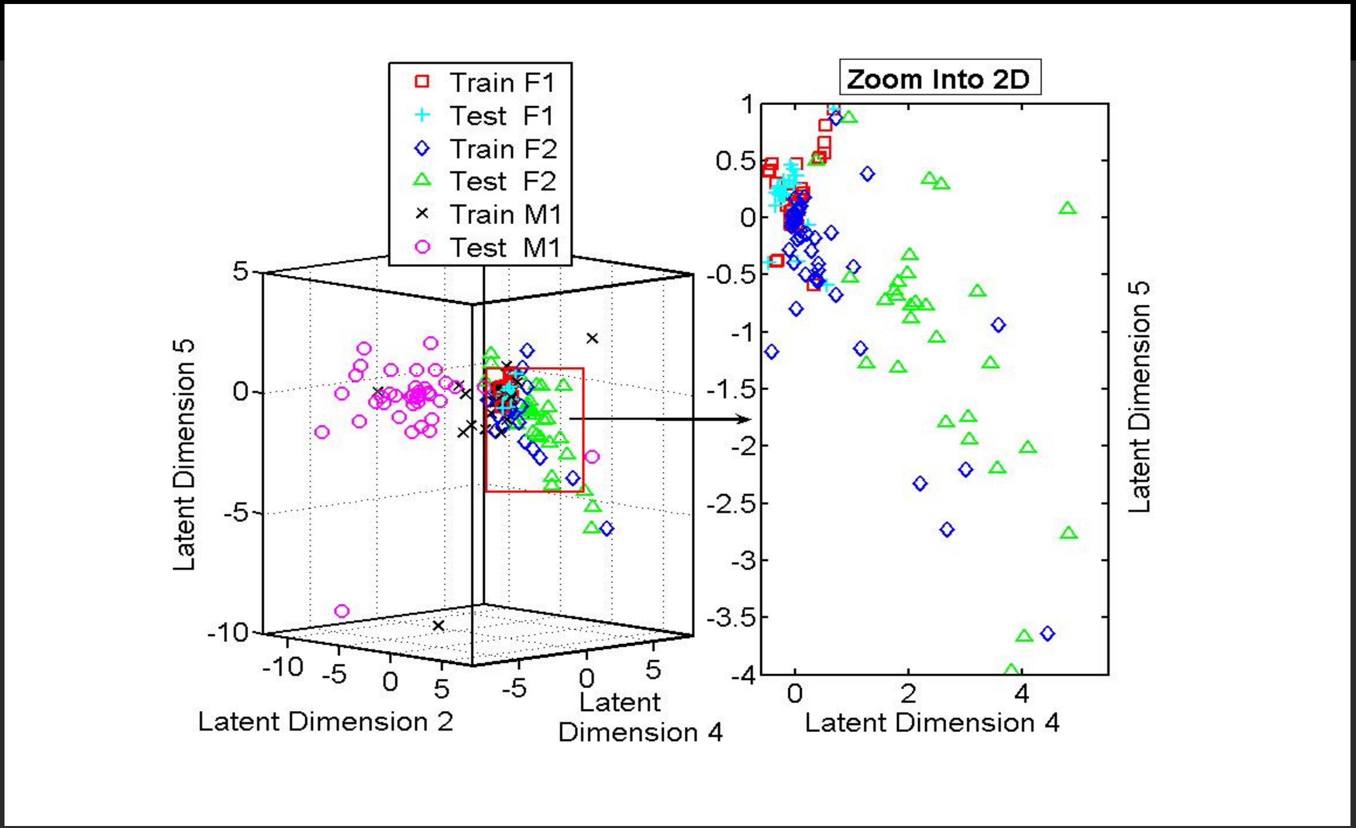
For the given time scales and thresholds, data locate around $y = x$, and the correlation coefficient $\rho = 0.67$, $p < 1.38e{-}09$.



Sample-to-sample correlation

- Three groups: vowels eh, ow; fricatives s, z, f, v; and stops k, g, p, t.
- 25-d MFCCs; EBS to keep 99% energy; PCA reduces dimension to 6.
- Two models had a similar pattern of making correct predictions and mistakes, and the percentage of matching between supervised and unsupervised learning was 91%.

Longer time scales



Time integrated (1000ms) MFCC's: text independent speaker recognition....

Feng & Hansen (CIMCA, 2005)

Error rate correlations super/unsupervised learning for different cognitive time scalesevents (phoneme, gender, height, speaker identity)

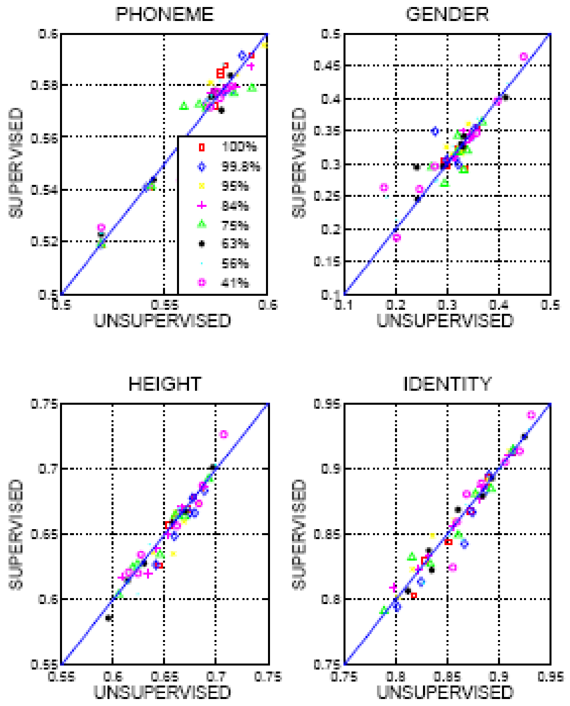


Fig. 4. Figure shows test error rates of both supervised and unsupervised learning on four topics: phonemes, gender, height and identity. Solid lines indicate $y = x$ in the coordinate systems. All data located along this line, meaning high correlation between supervised and unsupervised learning.

Digital media: Vector space representation

- Abstract representation - can be used for all digital media
- Document is represented as a point in a high-dimensional ”feature space” – document similarity ~ spatial proximity in a given metric

- Text: Term/keyword histogram, N-grams
- Image: Color histogram, texture measures
- Video: Object coordinates (tracking), active appearance models
- Sound: Spectral coefficients, cepstral coefficients, gamma tone filters

Contexts can be identified by their feature associations (= Latent semantics)

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R:
Indexing by latent semantic analysis.
Journal of the American Society for Information Science, 41(6), 391-407, (1990)

Latent semantics

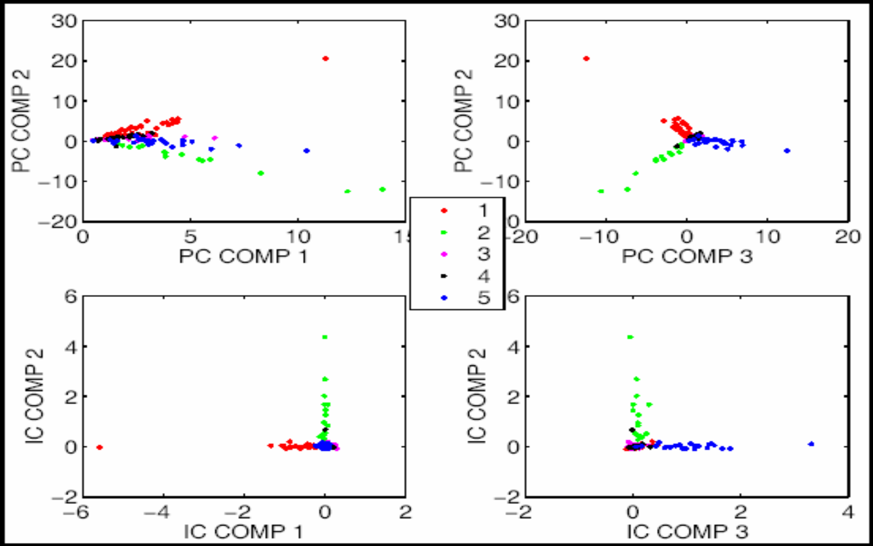
- Document features are correlated, the pattern of correlation reflects ”associations”.
- Associations are context specific
- Word sets are activated in concert in a given context

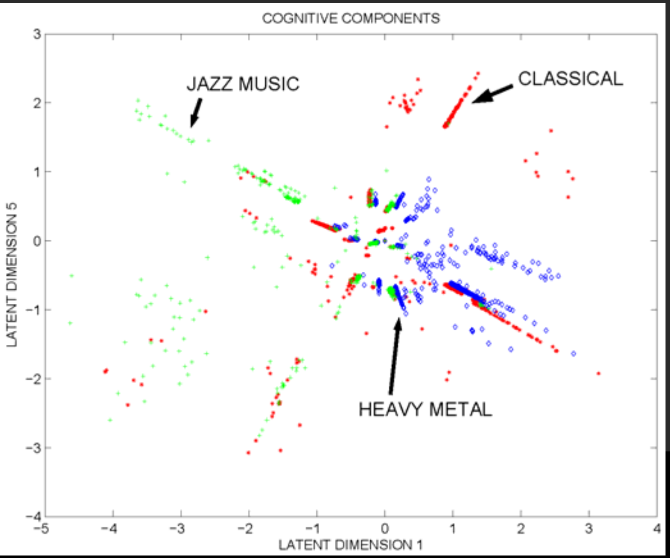
  ape ~ zoo, zoo ~ elephant => ape ~ elephant

Latent semantic analysis: Contexts can be identified by term co-variance patterns (PCA)

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R:
Indexing by latent semantic analysis.
Journal of the American Society for Information Science, 41(6), 391-407, (1990)

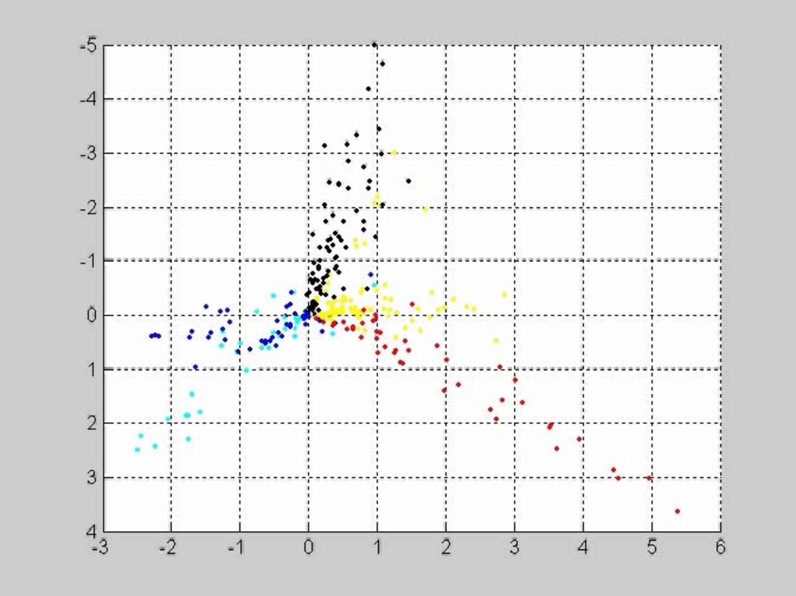
Linear mixture of independent agents in term-document scatterplots





Linear mixture of independent contexts observed in short time features (mel-ceptrum) in a music database.

Social networks: Linear mixtures of independent communities?

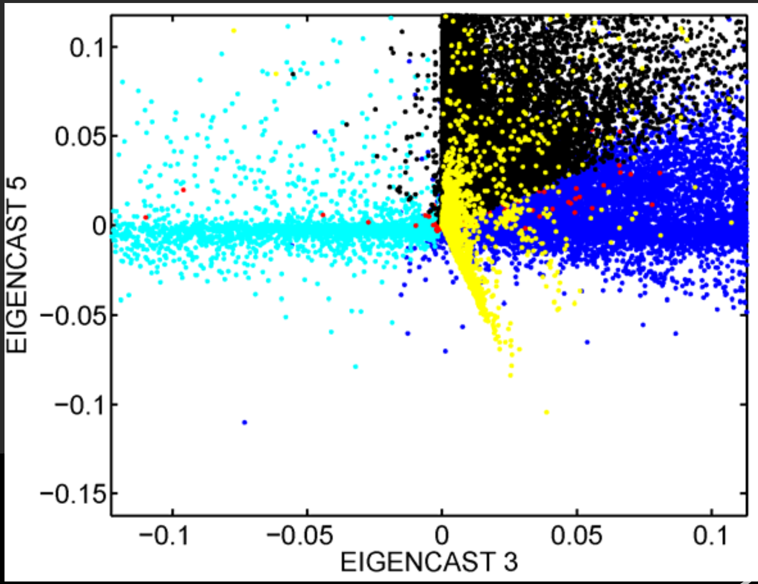


Genre patterns in expert's opinion on similar music artists

(AMG400, Courtesy D. Ellis)

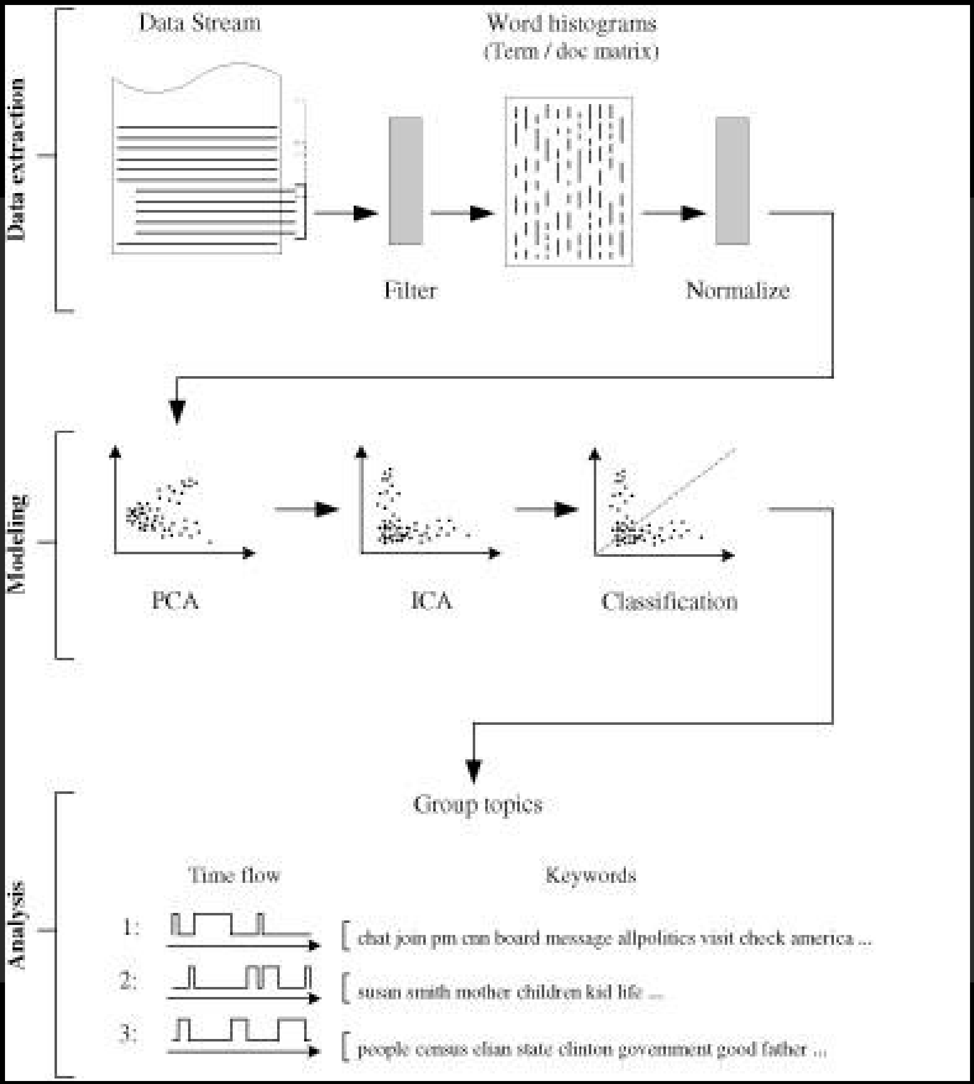
"Movie actor network"
- A collaborative small world network

128.000 movies
380.000 actors



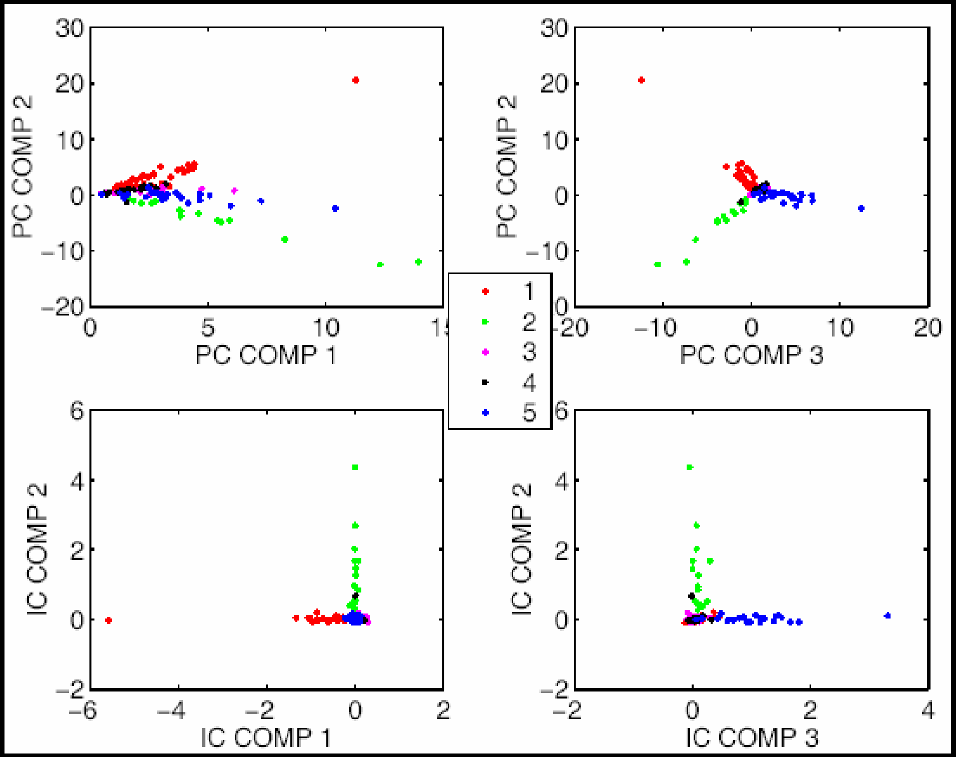
Independent contexts in document data bases

- x(j,t) is the occurence of the j'th word in the t'th document.
- s(k,t) quantifies how much the k'th context is expressed in t'th document.
- A(j,k) quantifies the typical importance of the j'th word in the k'th context



Isbell and Viola: "Restructuring sparse high dimensional data for effective retrieval" NIPS*11, 361-362 (1999)

PCA vs ICA document scatterplots



Independent contexts in dynamic text:

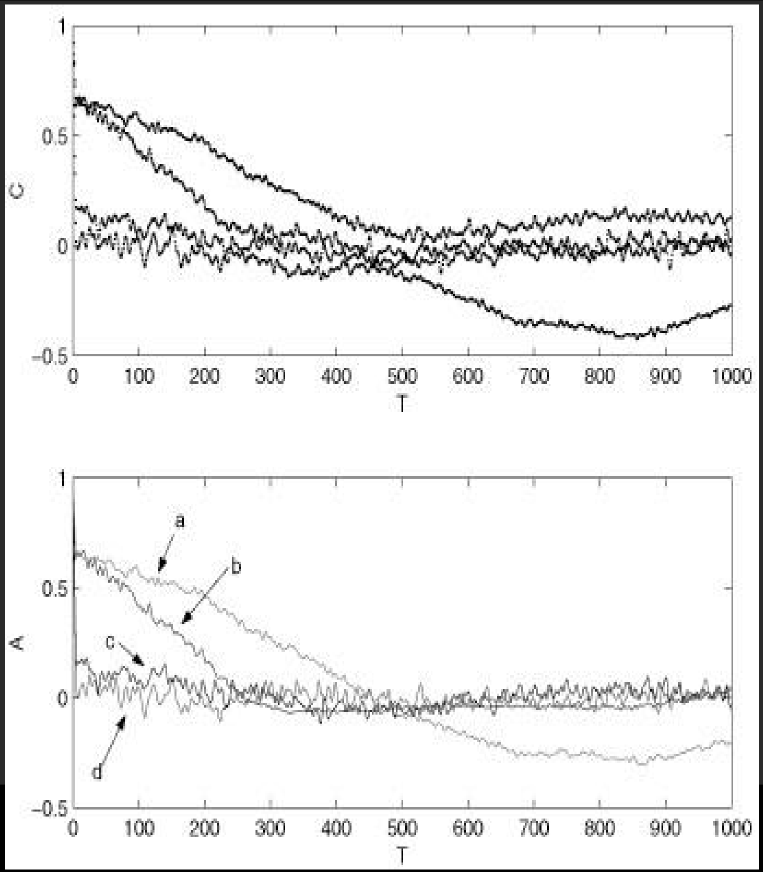
Chat room analysis

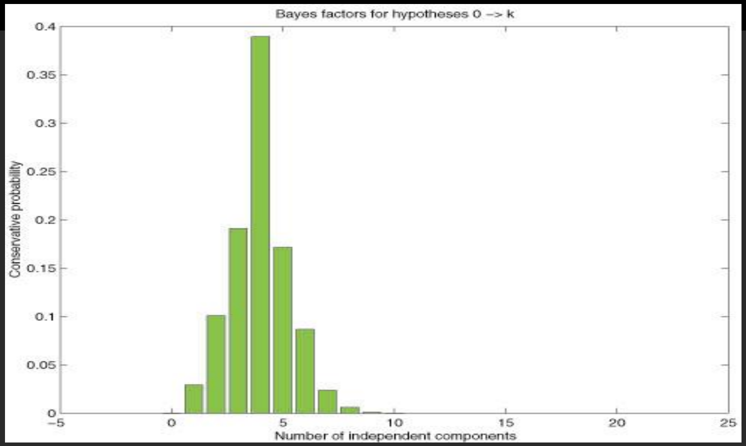
We logged a days
chat in a CNN ”news cafe”.

The database involves
120 users chatting
during an 8 hour period

... just statements like that - over the past
few weeks.
<Miez> heyy seagate
<Recycle> denise: he deserved it for stealing os code in his early days
<Zeno> ok Sharonelle
<denise> LOL @ Recycle
<HaleyCNN> Join Book chat at 10am ET in #auditorium. Chat with Robert Ballard author of "Eternal Darkness: A Personal History of Deep-Sea Exploration," after his appearance on CNN Morning News at 9:30am ET.
<heartattackagain> Ed Shore....lol....We might have an operating system that doesn't crash every thirty minits....lololol.....
<EdShore> Shooby, I don't believe you. I've been doing this sine PET, TRS-80, and PIRATES! Don't tell me you've been CHATTING! PROVE IT!
<Zeno> Recycle LOL ethical and criminal laws are different for the business world
<_Seagate_> Recycle, thats what the technology business is all about.
<tribe> I heard a local radio talk show host saying last night that he has noticed everytime this Elian issue slows down, something happens to either the family in Miami or in Cuba to put it right back in the headlines. He mentioned the cousin's hospitalization as just the latest saga
<Diogenes> If Bill Gates was in Silicon Valley never a word would you have ever heard.
<Zeno> EdC you may have been doing sine but i have been doing cosine.
<shooby> EdShore: Compuserve since, heck, 76?
<Zeno> i mean EdShore
<Recycle> rumor has it that he was even dumpster diving ...

ICA by dynamic decorrelation (Molgedey-Schuster, Kolenda et al. 2001)





The Bayes factor - P(M|D) - of each model is estimated in the BIC approximation

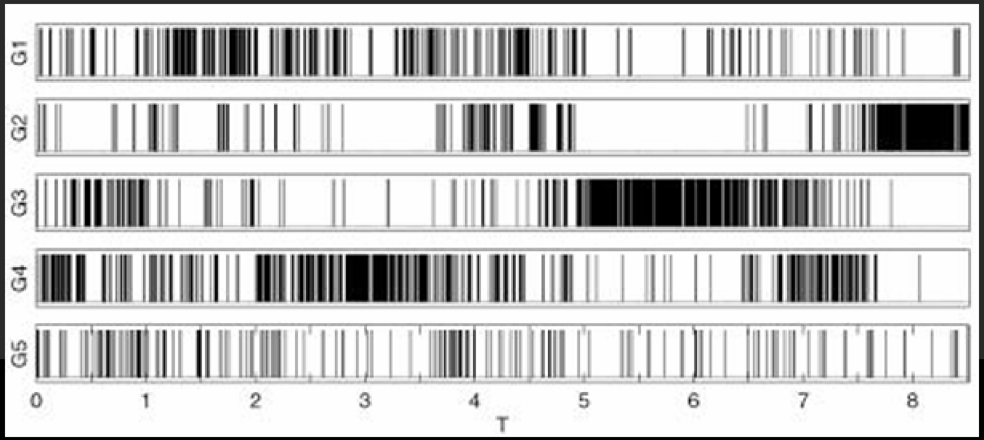
Source autocorrelations

Chat room analysis

Keywords from prototype histograms (A-columns)

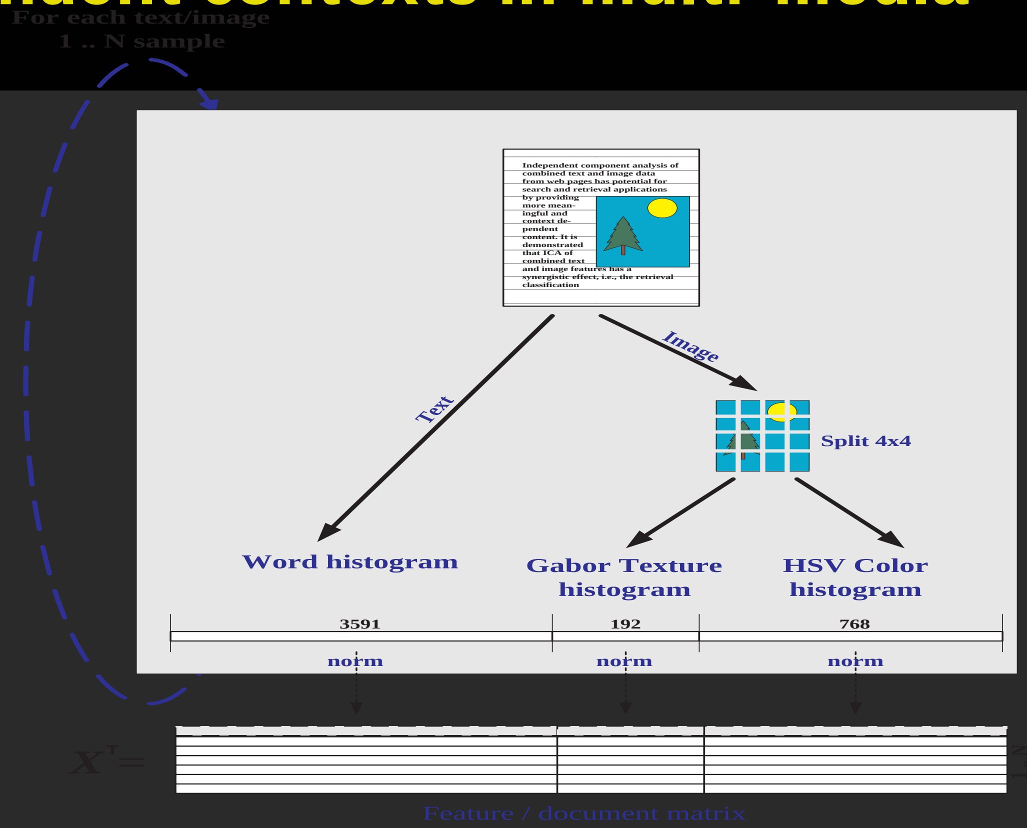
Topic 1: Chat, join, pm, cnn, board,...
Topic 2: Gun show
Topic 3: Susan, Smith, mother, children, kid, life
Topic 4: People, census, elian, state, clinton,...

"Topic 5" is a rest group



Example: Independent contexts in multi-media

- Organizing webpages in categories
- Labels obtained from
  Yahoo's directory
- Features: Text, color, and texture subsets of MPEG image features



L.K. Hansen, J. Larsen and T. Kolenda "On Independent Component Analysis for Multimedia Signals".
In L. Guan et al.: *Multimedia Image and Video Processing*, CRC Press, Ch. 7, pp. 175-199, 2000.

Performance of the system trained by associating unsupervised independent components with labels – generalization based on Yahoo cathegories

Modality	Classification Error
Color	23.0%
Texture	18.0%
Texture/Color	11.5%
Text	5.7%
Combined (texture/color/text)	2.8%

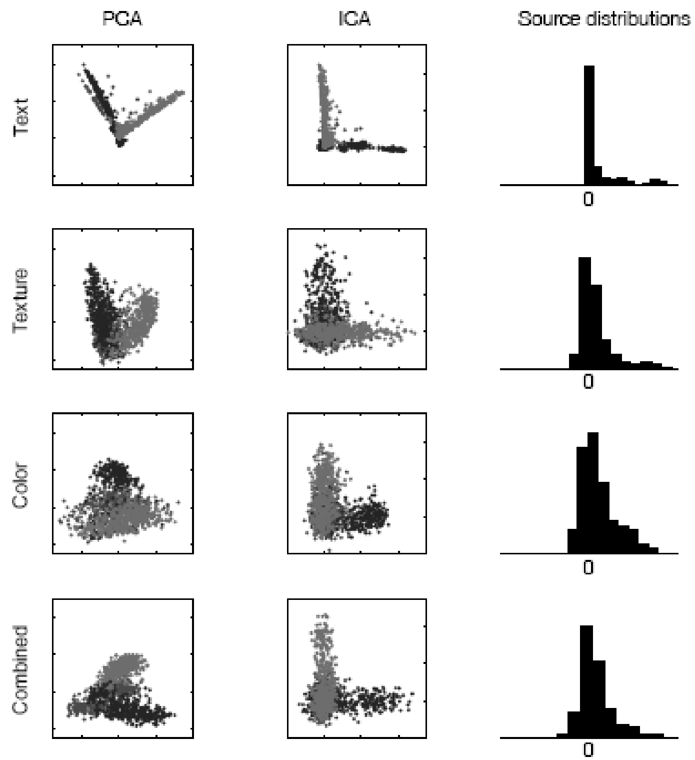


Fig. 3. Scatterplots of the text and image multimedia data, projected to a two-dimensional subspace found by PCA. Grey value of points corresponds to the three classes considered, see Fig. 4. The ray like structure strongly suggest an ICA interpretation, however, the relevance of this representation can only be determined by a subsequent inspection of the recovered source signals. As we will see in section 4.6, it turns out that there is an interesting alignment of the source signals and a manual labeling of the multimedia documents.

Performance

Texture (K=13)

69.75	7.75	6.5
11.5	88.5	5.75
18.75	3.75	87.75

Color (K=16)

70.75	3.75	10
12	81.5	11.25
17.25	14.75	78.75

Text (K=45)

93	2	2.25
0.5	94.75	2.5
6.5	3.25	95.25

Combined errorrate: 2.8%
Single best errorrate: 5.7%

Texture Color (K=26)

82	1.75	4.5
9	93.75	5.75
9	4.5	89.75

Texture Color Text (K=26)

98.25	0.25	0.75
0.75	98	3.75
1	1.75	95.5

I_1 I_2 I_3

Image	Label	Keywords
I_1	Sports	position college weight born lbs height guard
I_2	Aviation	na air convair wing
I_3	Paintball	check darkside force gog strike odt

”Brede” -- tools for a neuroimaging search engine

- Exponential growth in publications/images/data
- Distributed (www) heterogeneous databases
- Multimedia facilities needed
- Tools:
  - Novelty detection (Nielsen&Hansen, HBM 2002)
  - Finding similar volumes (Nielsen&Hansen, AIM 2003)
  - Brede toolbox

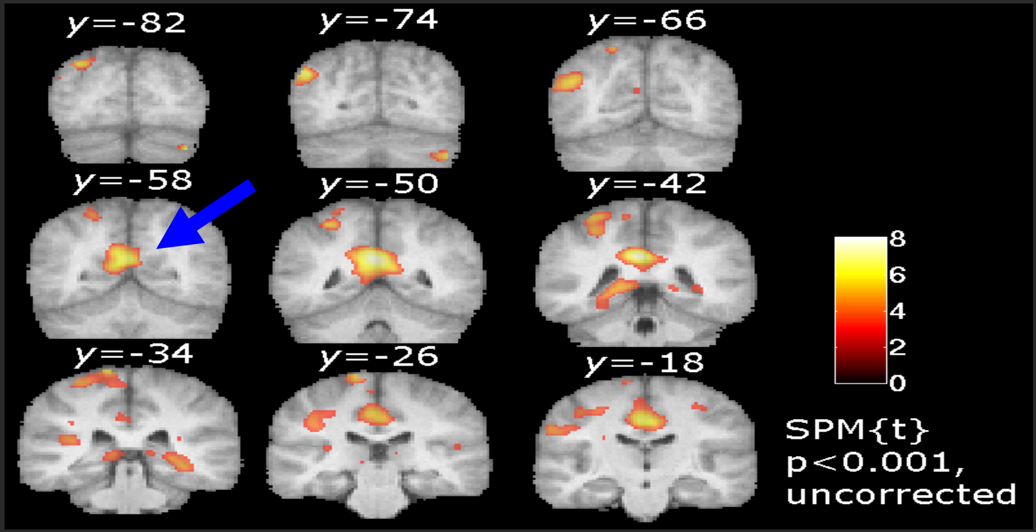
Primary co-worker: Finn Årup Nielsen

...consult his homepage for a neuroinformatics informed search engine

Analytic search example: “The posterior cingulate cortex”

- cyto-architecturally well defined brain region (Vogt et al, 2001)

- no “consensus” about its function: Several functions are reported including involvement in emotion, pain, memory etc

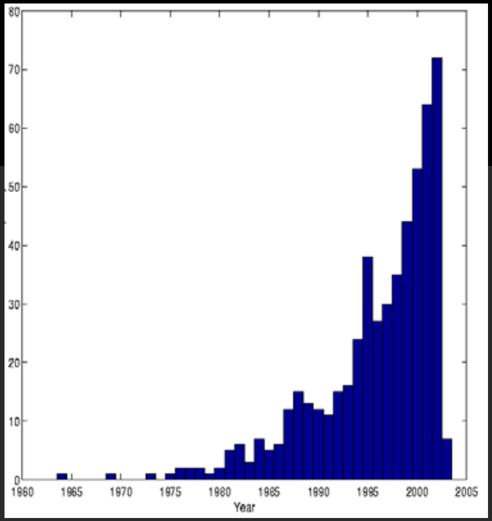


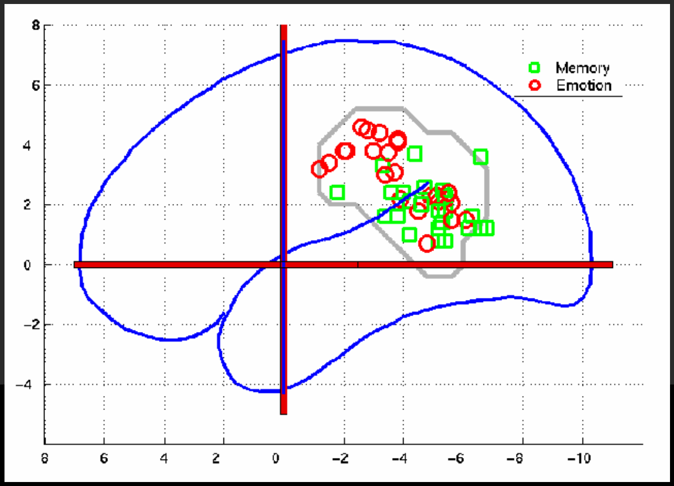
Nielsen et al., Neuroinformatics 2004

Materials

- 245 abstracts of functional imaging studies that responded to PubMed query (march 13th, 2003) :

("posterior cingulate" OR "posterior cingulum" OR "retrosplenial" OR "retrosplenium")
AND
("magnetic resonance imaging" OR "positron emission tomography")





We find independent components of abstracts and locate them using the Brede database: Two components are rich in terms related to ”memory” and ”emotion”. Localization may actually be slightly different pointing to a regional specialization.

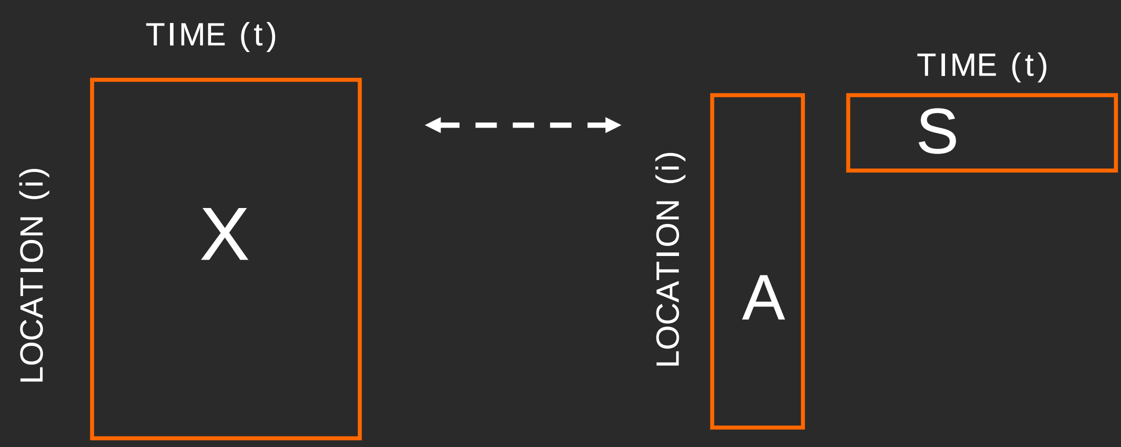
Linear independent components (Factor) models seem to be relevant to many cognitive ecologies

Let us take a look under the hood...

- Can we understand swift human learning?
- ...a closer look at generalizability of unsupervised learning in factor models

Factor models

- Represent a datamatrix by a low-dimensional approximation



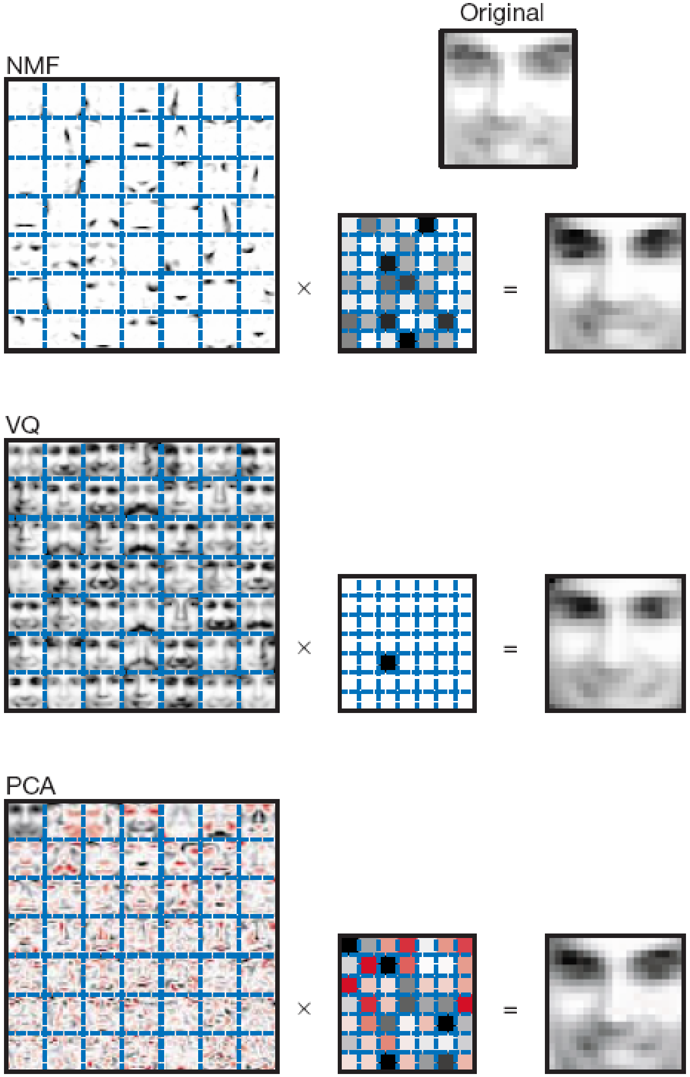


Figure 1 Non-negative matrix factorization (NMF) learns a parts-based representation of faces, whereas vector quantization (VQ) and principal components analysis (PCA) learn holistic representations. The three learning methods were applied to a database of $m = 2{,}429$ facial images, each consisting 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

Generative model for hidden variables

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

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

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

Source distribution:
PCA: … normal
ICA: … other
IFA: … Gauss. Mixt.

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

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

S known:	GLM
$(1-A^{-1})$ sparse:	SEM
S,A positive:	NMF

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

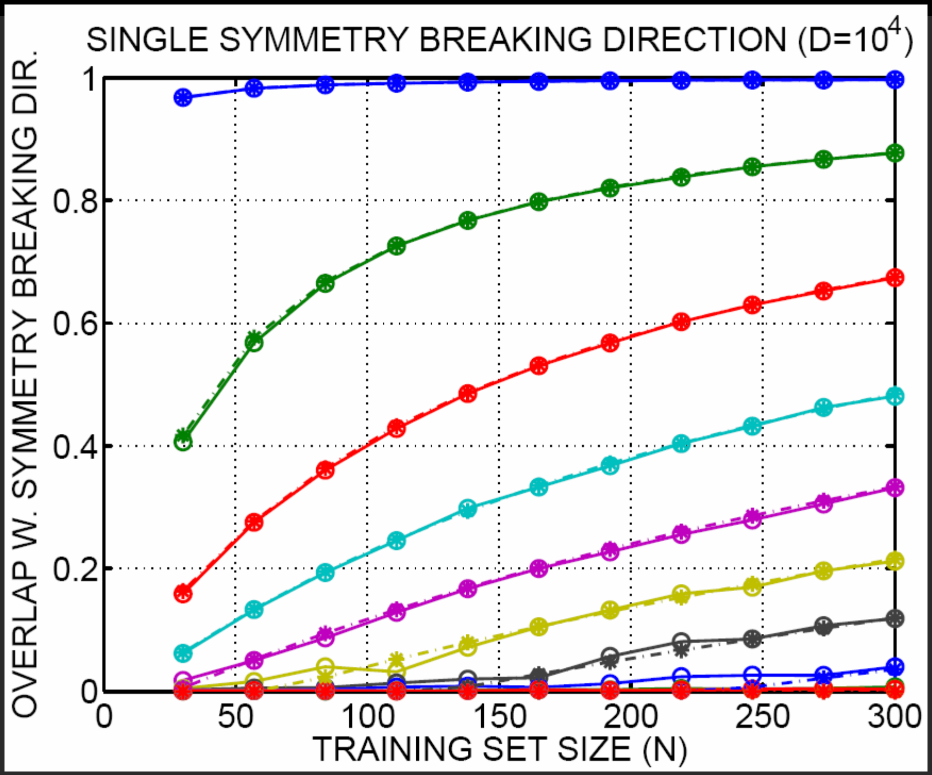
Modeling the generalizability of FA

- 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 $R^2 = \langle (u_1^T * u_0)^2 \rangle$ 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)



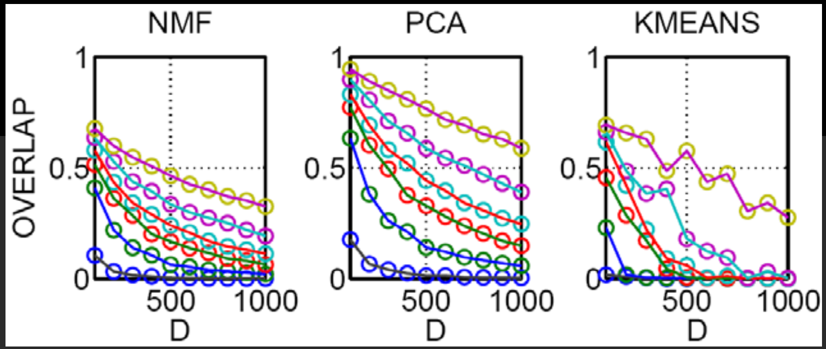
N_c = (0.0001, 0.2, 2, 9, 27, 64, 128, 234, 400, 625)

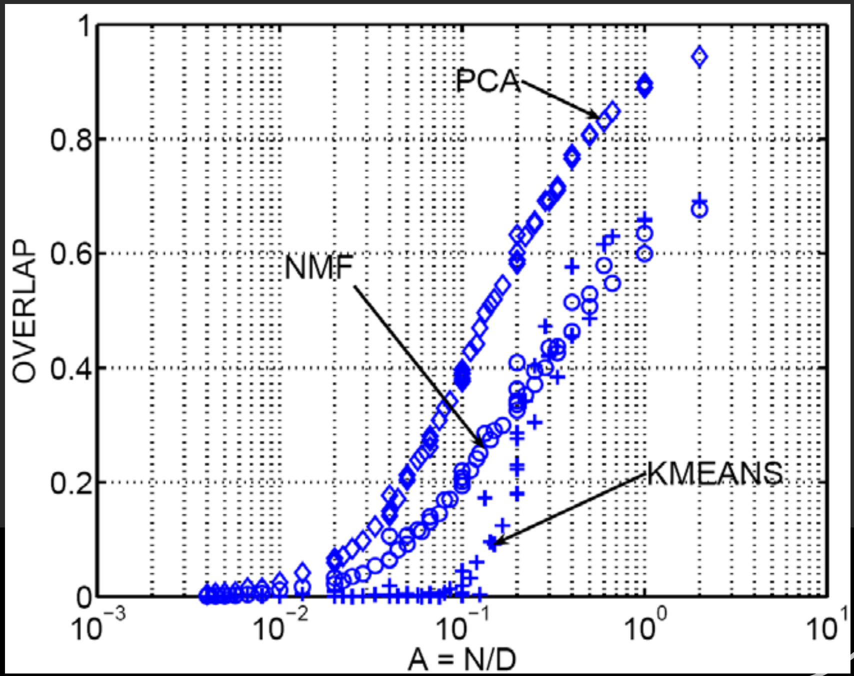
σ = (0.01, 0.06, 0.12, 0.17, 0.23, 0.28, 0.34, 0.39, 0.45, 0.5)

Generalization in unsupervised learning

- 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.

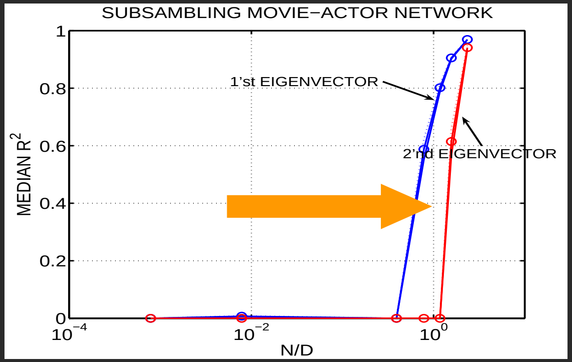
Experiment
Variable: N, D
Fixed: SNR

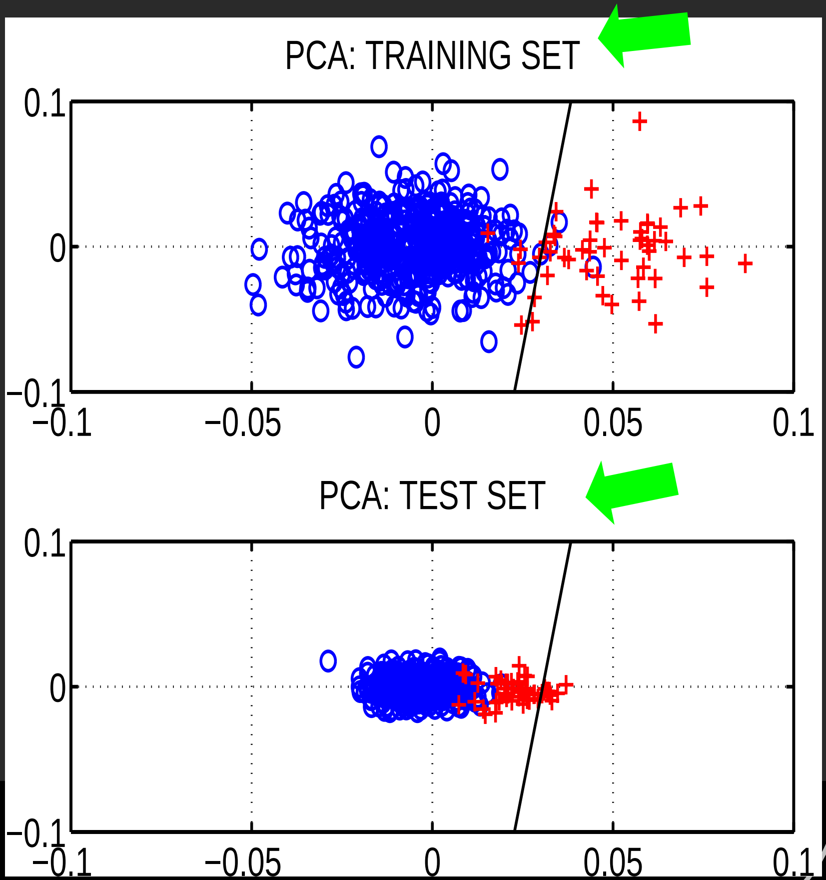




Restoring the generalizability of FA

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

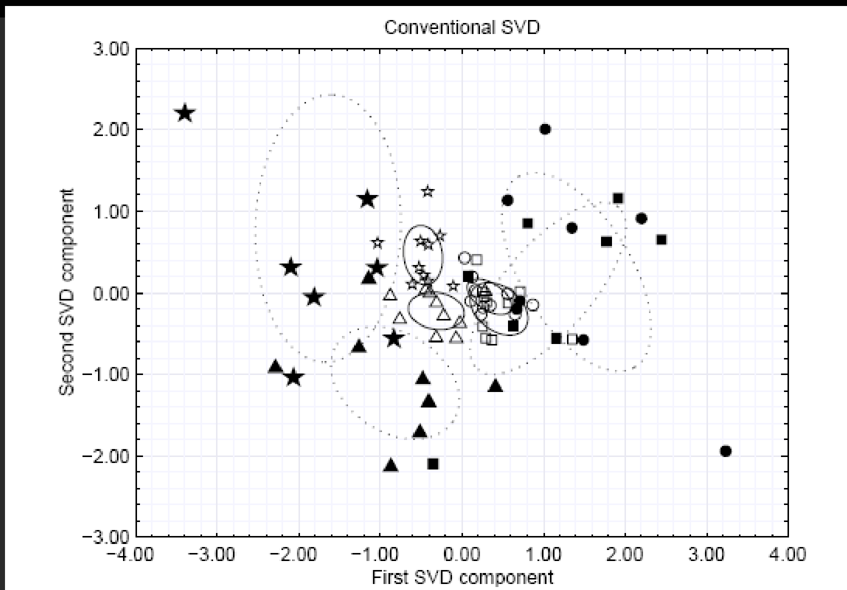


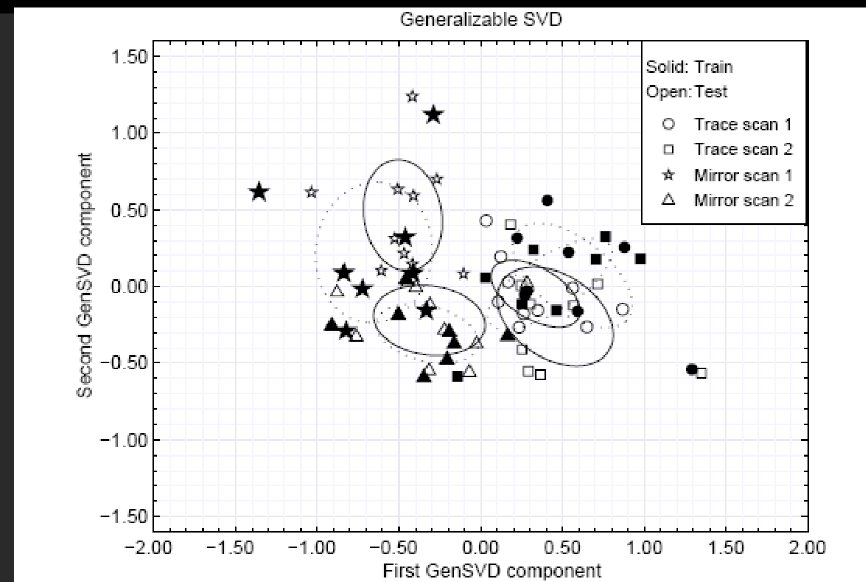


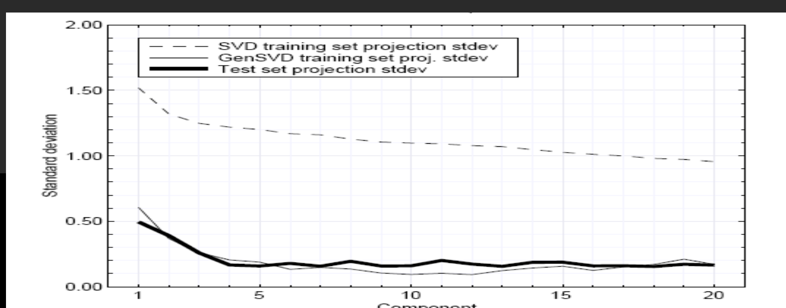
- The estimated projection is offset, hence, future projections will be too small!

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

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





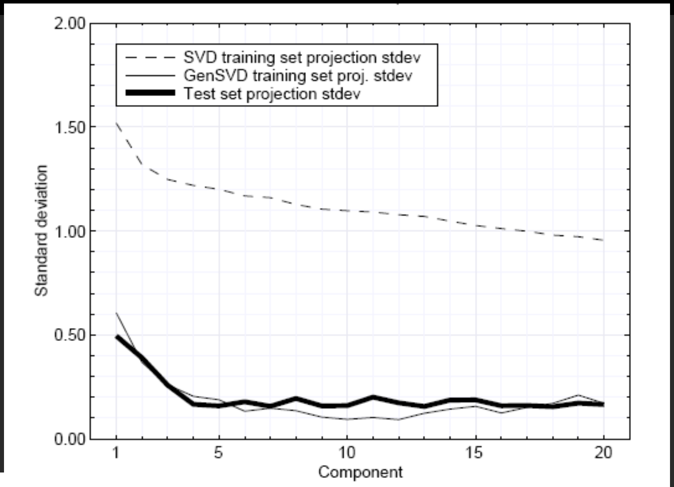


N=72, D=2.5 10^4

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



$$
\begin{aligned}
&\text{Compute } \boldsymbol{U}_0 \boldsymbol{\Lambda}_0 \boldsymbol{V}_0^\top = \mathrm{svd}(X) \text{ and } \boldsymbol{Q}_0 = [\,\boldsymbol{q}_j\,] = \boldsymbol{\Lambda}_0 \boldsymbol{V}_0^\top \\
&\text{foreach } j = 1 \ldots N \\
&\qquad \bar{\boldsymbol{q}}_{-j} = \tfrac{1}{N-1} \textstyle\sum_{j' \neq j} \boldsymbol{q}_{j'} \\
&\qquad \text{Compute } \boldsymbol{B}_{-j} \boldsymbol{\Lambda}_{-j} \boldsymbol{V}_{-j}^\top = \mathrm{svd}(\boldsymbol{Q}_{-j} - \bar{\boldsymbol{Q}}_{-j}) \\
&\qquad \boldsymbol{z}_j = \boldsymbol{B}_{-j} \boldsymbol{B}_{-j}^\top (\boldsymbol{q}_j - \bar{\boldsymbol{q}}_{-j}) \\
&\hat{\lambda}_i^2 = \tfrac{1}{N-1} \textstyle\sum_j z_{ij}^2
\end{aligned}
$$

Kjems, Hansen, Strother: NIPS (2001)

What about other factorizations?

- NMF can be adjusted similarly using histogram equlization

...

- Note the re-scaling problem is almost trivial if the classifier is Naïve Bayes-like (threshold adaption) taking the independent components as features
- More complex classifiers may need more coordination

Conclusions

- Evidence that phonemes, gender, identity are independent components ’objects’ in the (time stacked) MFCC representation
- Evidence that human categorization is based on sparse independent components in social networks, text, digital media
- Conjecture: that objects in digital media can be identified as independent components: The brain uses old tricks from perception to solve complex ”modern” problems.

Outlook

- Compatibility in social networks
  - Are links and features compatible?
- The independent context hypotesis:
  - Are the Gestalt laws simply results of ICA representations?
  - Next step to understand dynamics

Acknowledgments

- Danish Research Councils
- EU Commission
- NIH Human Brain Project grant (P20 MH57180)

References

A. J. Bell and T. J. Sejnowski, “The ‘independent components’ of natural scenes are edge filters,” *Vision Research*, vol. 37, pp.3327–3338, 1997.

P. Hoyer and A. Hyvrinen, “Independent component analysis applied to feature extraction from colour and stereo images,” *Network: Comput. Neural Syst.*, vol. 11, pp. 191–210, 2000.

M. S. Lewicki, “Efficient coding of natural sounds,” *Nature Neuroscience*, vol. 5, pp. 356–363, 2002.

E. Doi and T. Inui and T. W. Lee and T. Wachtler and T. J. Sejnowski , “ Spatiochromatic Receptive Field Properties Derived from Information-Theoretic Analyses of Cone Mosaic Responses to Natural Scenes, ” Neural Comput., vol. 15(2), pp. 397-417, 2003.

J. H. van Hateren and D. L. Ruderman, “Independent Component Analysis of Natural Image Sequences Yields Spatio-Temporal Filters Similar to Simple Cells in Primary Visual Cortex,” Proc. Biological Sciences, vol. 265(1412), pp. 2315-2320, 1998.

H.B. Barlow, “Unsupervised learning,” Neural Computation, vol. 1, pp. 295–311, 1989.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R: *Indexing by latent semantic analysis*. Journal of the American Society for Information Science, 41(6), 391-407, (1990)

J. Larsen, L. K. Hansen, T. Kolenda, F. Å. Nielsen: *Independent Component Analysis in Multimedia Modeling*, Proc. of ICA2003, Nara Japan, 687-696, (2003)

L.K. Hansen, P Ahrendt, J Larsen. Towards cognitive component analysis. In Proc. AKRR’05 Conf on Adaptive Knowledge Representation and Reasoning, Helsinki 2005.

L. Feng and L.K. Hansen. On Low-level Cognitive Components of Speech. International Conference on Computational Intelligence for Modelling, vol.2, pp 852-857, 2005.

L. Feng and L.K. Hansen. Phonemes as Short Time Cognitive Components. The 31st International Conference on Acoustics, Speech, and Signal Processing, vol.5, pp869-872, 2006.

L.K. Hansen and L. Feng. Cogito Componentiter Ergo Sum. The 6th International Conference on Independent Component Analysis and Blind Source Separation, pp 446-453, 2006.

L. Feng and L.K. Hansen. Cognitive Components of Speech at Different Time Scales. The 29th annual meeting of the Cognitive Science Society, pp 983-988, 2007.

L. Feng and L.K. Hansen. On Phonemes as Cognitive Components of Speech. The 1st IAPR Workshop on Cognitive Information Processing, pp 205-210, 2008.

L. Feng and L.K. Hansen. Is Cognitive Activity of Speech Based on Statistical Independence? The 30th annual meeting of the Cognitive Science Society, pp 1197-1202, 2008.