MULTIVARIATE ANALYSES WITH fMRI DATA

Sudhir Shankar Raman

Translational Neuromodeling Unit (TNU)
Institute for Biomedical Engineering
University of Zurich & ETH Zurich

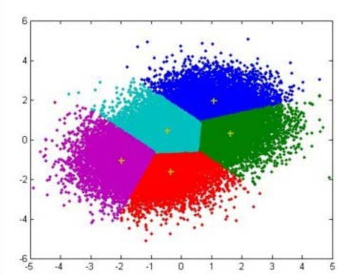
Motivation

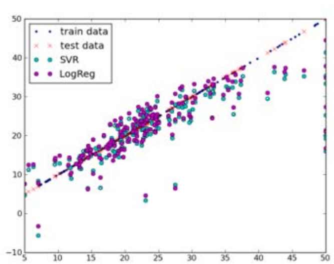
Modelling Concepts

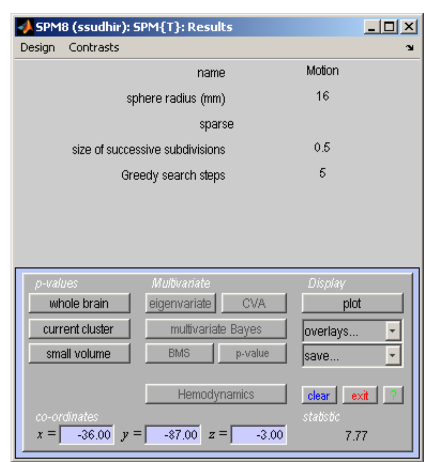
Learning From Data

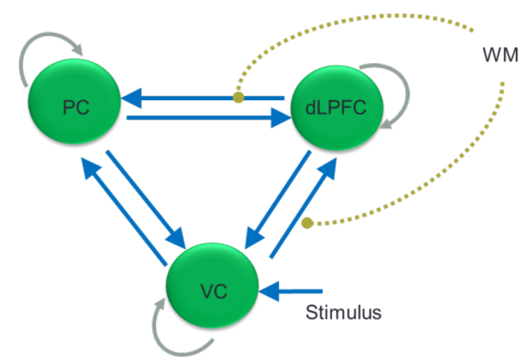
Multivariate Bayes in SPM

Generative Embedding



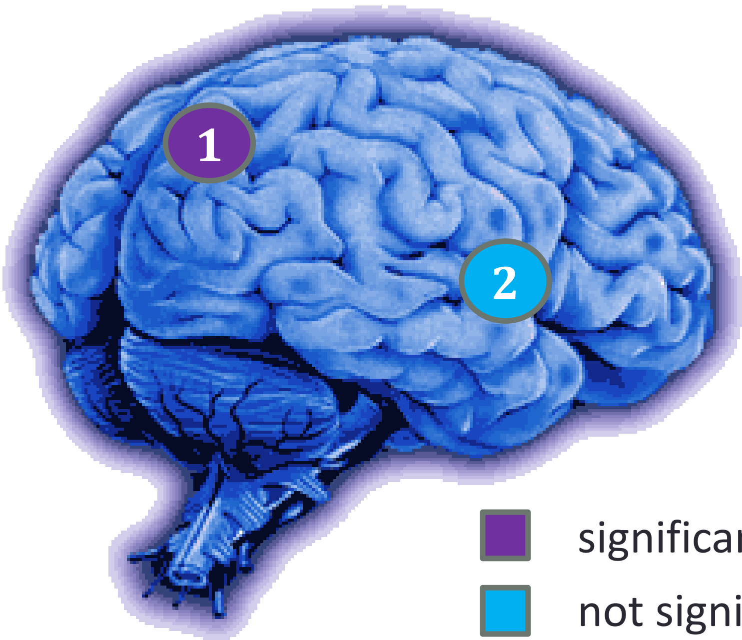


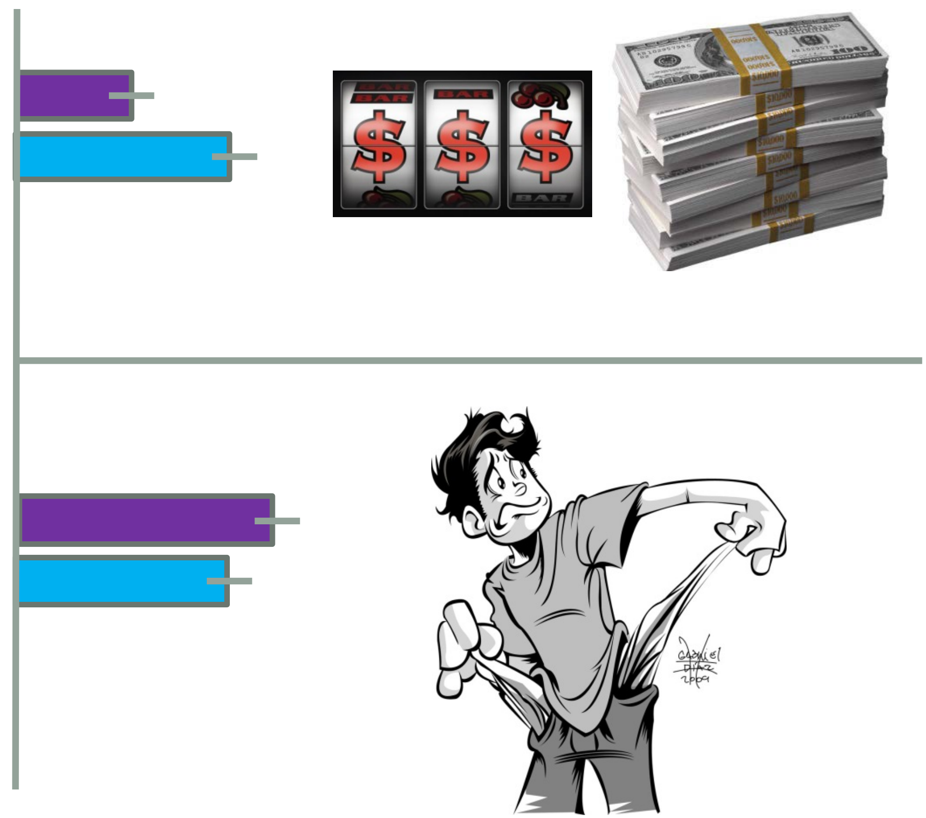




Motivation

- Local activations – Univariate approach

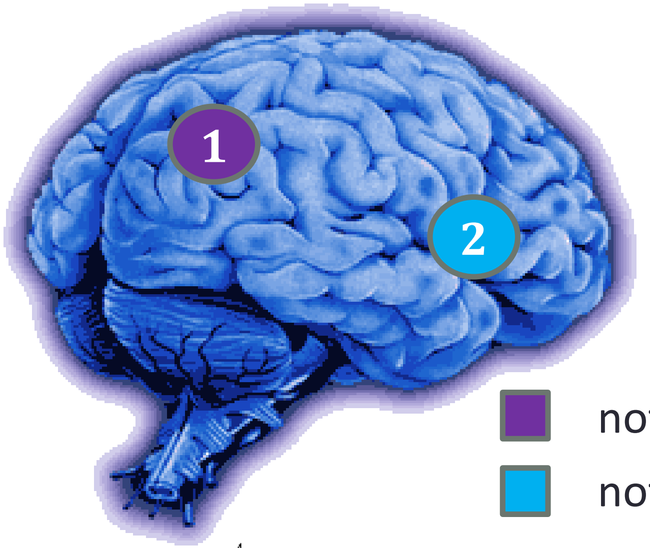




significant

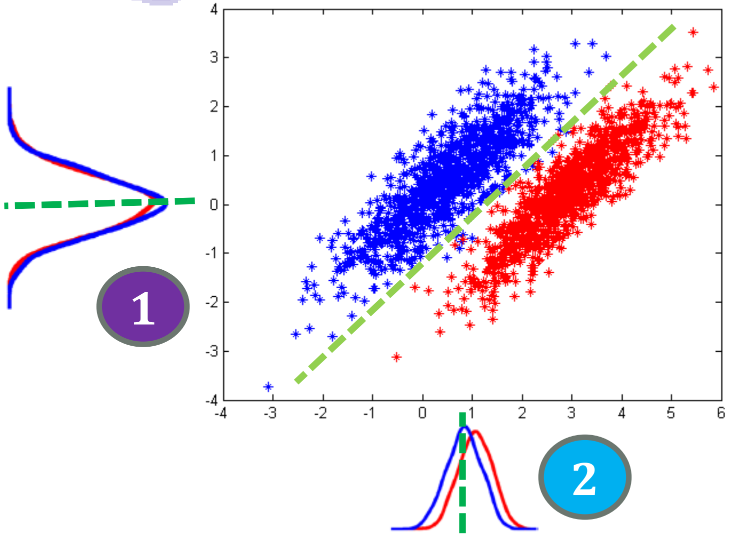
not significant

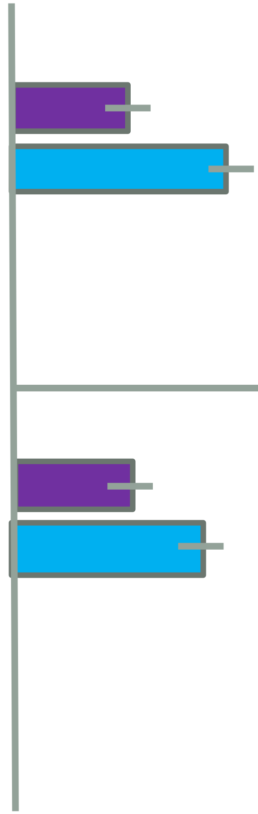
Univariate to Multivariate



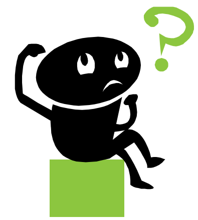
not significant

not significant

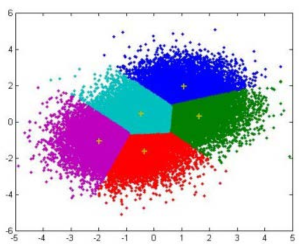


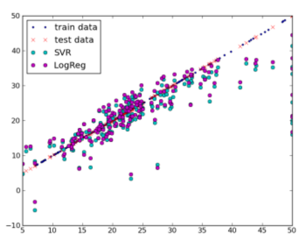


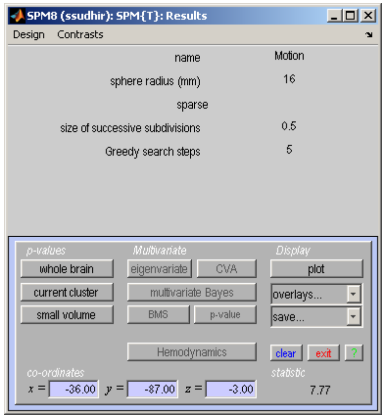
Motivation

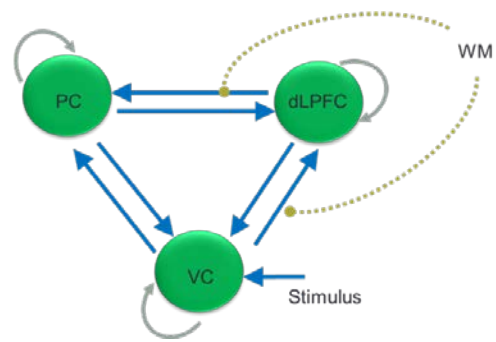


- Modelling Terminology
- Learning from data
- Multivariate Bayes in SPM
- Generative Embedding

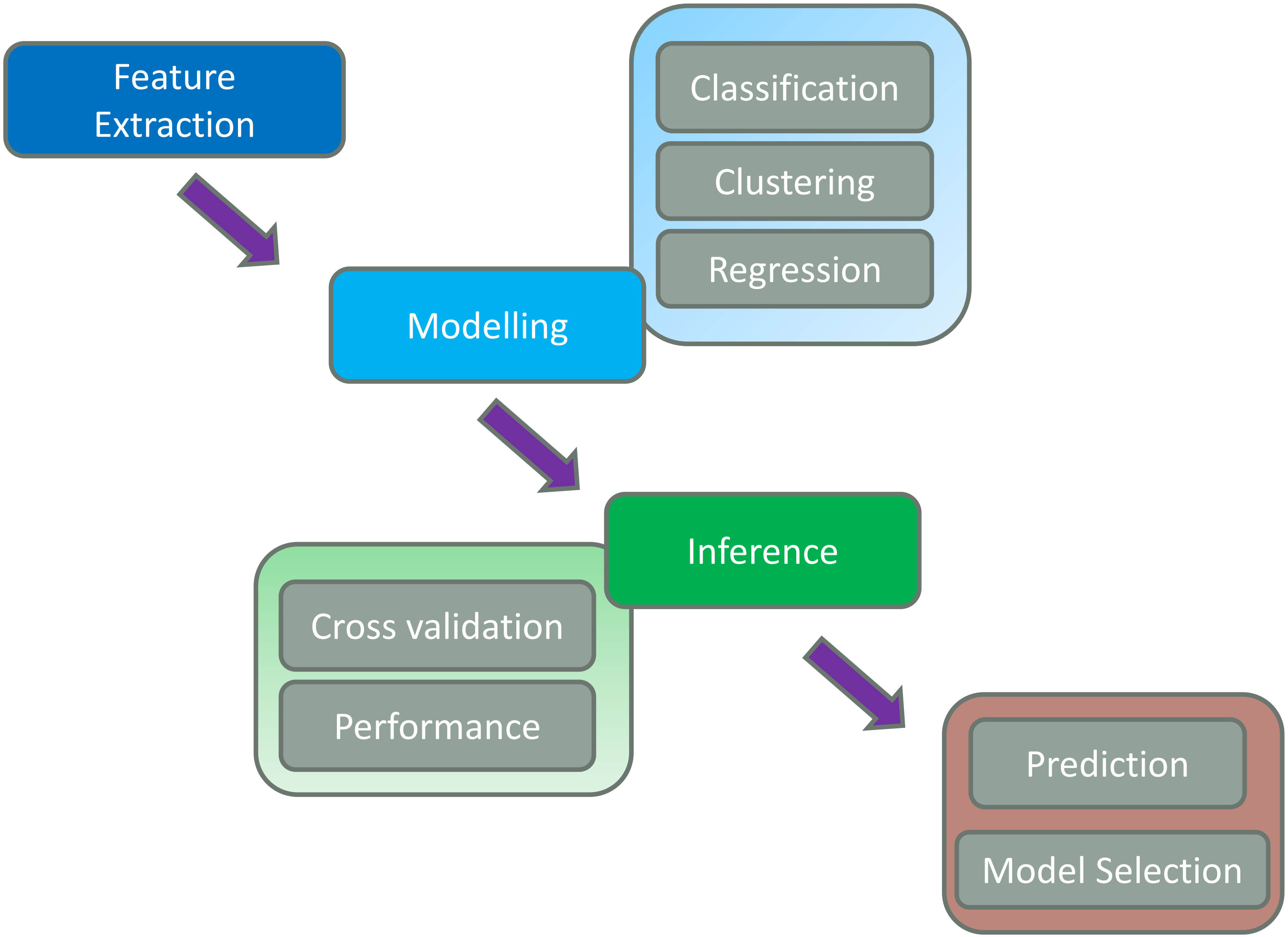








Steps for analysis



Feature Space

Features

Data Points

	F_1	F_2	.	.	.	F_P
S_1	1	0.5				
S_2	0	5.7				
.	1	4				
.	1	5.3				
S_N	1	6.6				

- Discrete
- Continuous

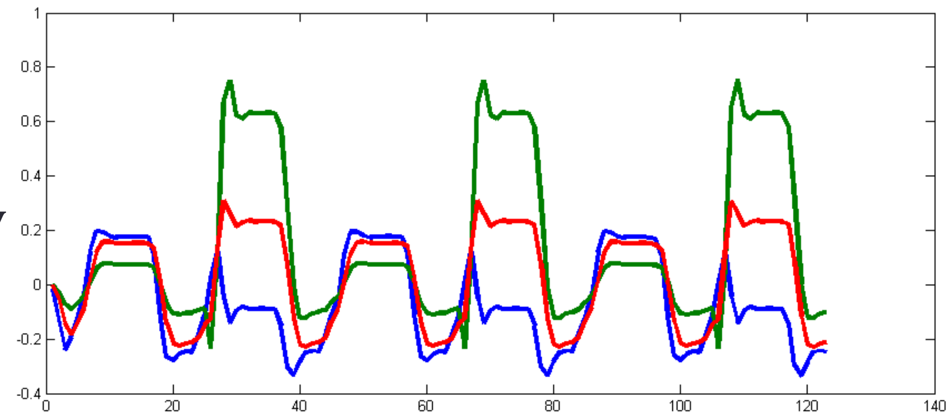
Feature Space Examples

	F_1	F_2	. . .	F_P
S_1				
S_2	Data Point or Feature Vector			
.				
.				
S_N				

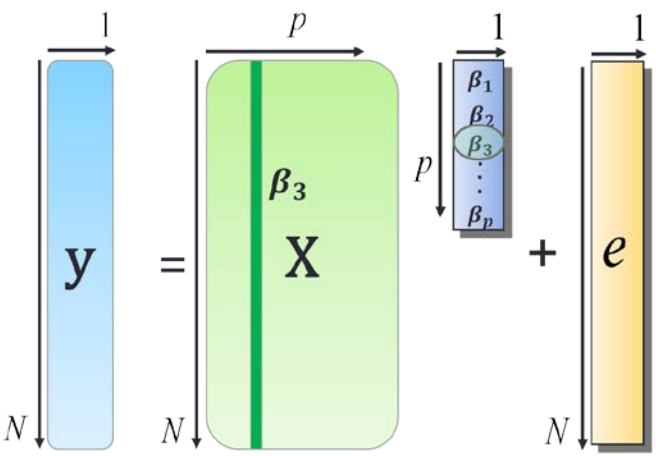
- High dimensionality
- Class/Cluster distributions
- Interpretation

Raw data

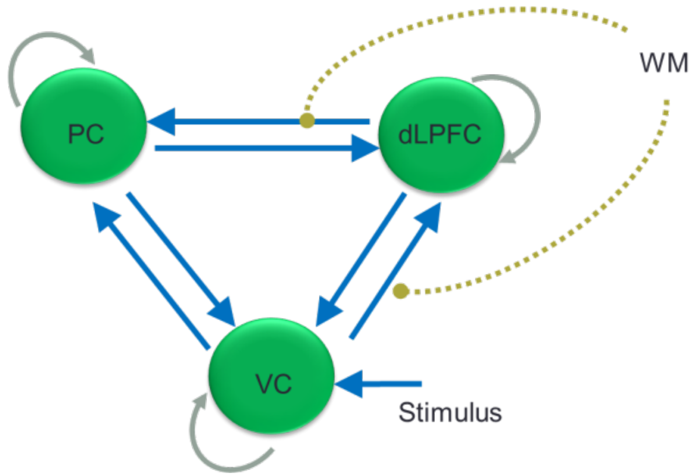
Voxel activity



Model parameters



GLM – regression coefficients

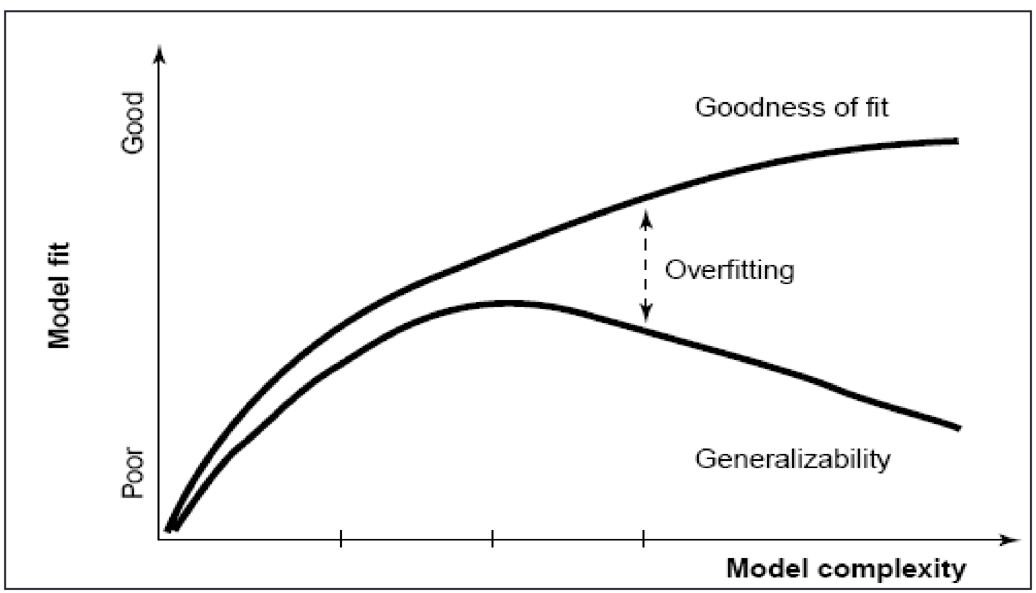


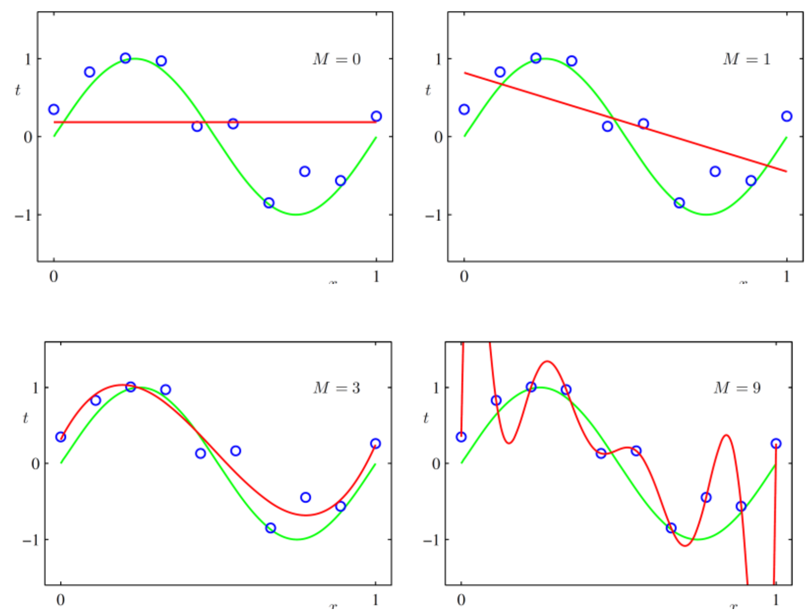
DCM – connection parameters

Model Selection - Generalizability

Model
Fit

Model
Complexity





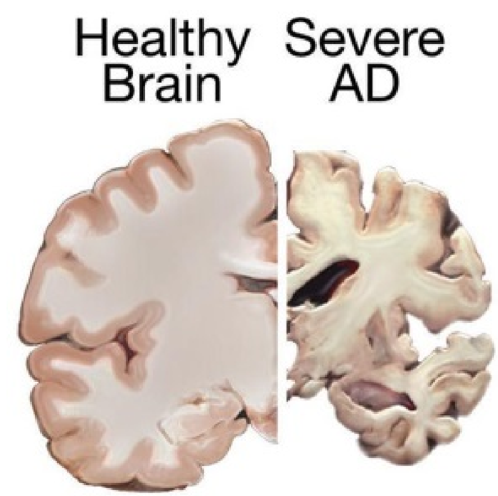
Bishop (2006), Pitt & Miyung (2002), *TICS*

Modelling goals

- Prediction

Predictive Density

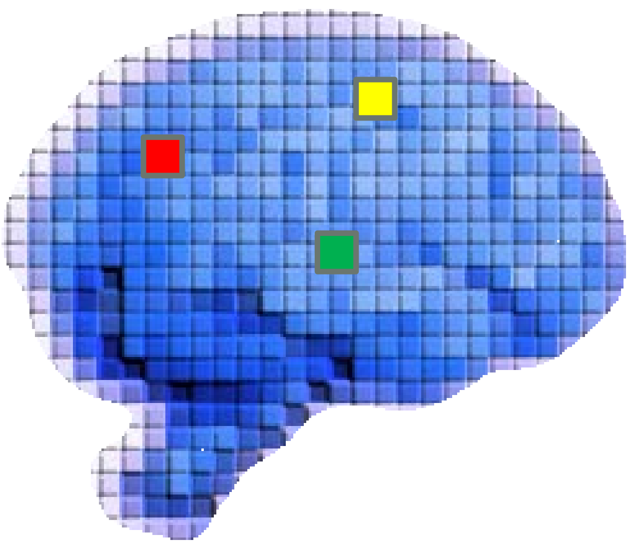
$$p(X_{\text{new}}|Y_{\text{new}}, X, Y) = \int p(X_{\text{new}}|\theta, Y_{\text{new}})q(\theta)d\theta$$



Modelling goals

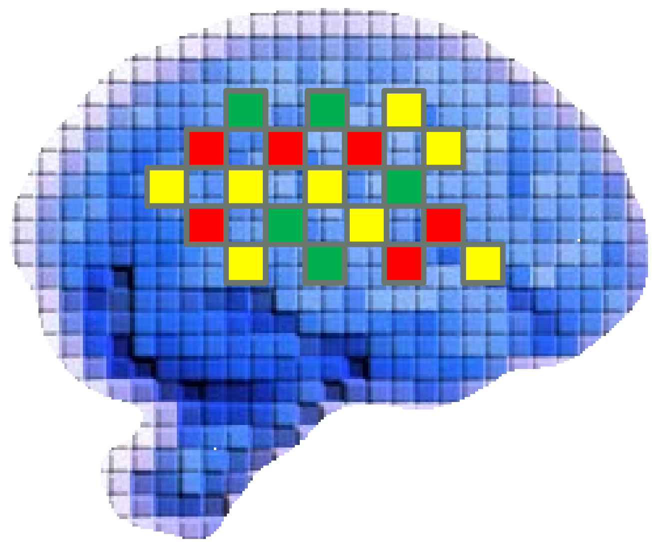
- Model Selection

Sparse Coding



VS

Distributed Coding

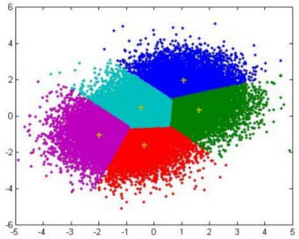


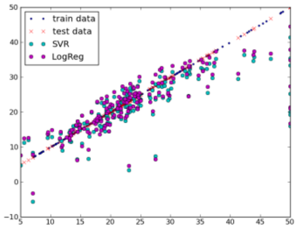
Model Evidence

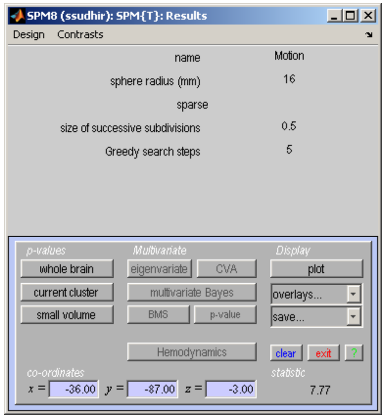
$$p(Y|X) = \int p(Y, \theta|X) d\theta$$

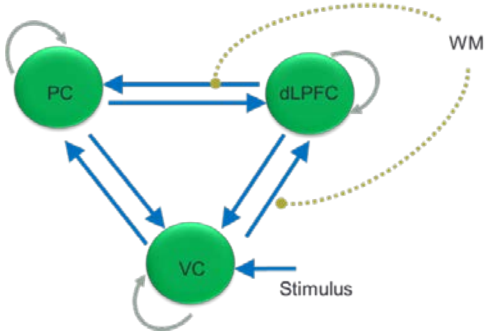
Motivation

- Modelling Concepts
- Learning From Data
- Multivariate Bayes in SPM
- Generative Embedding









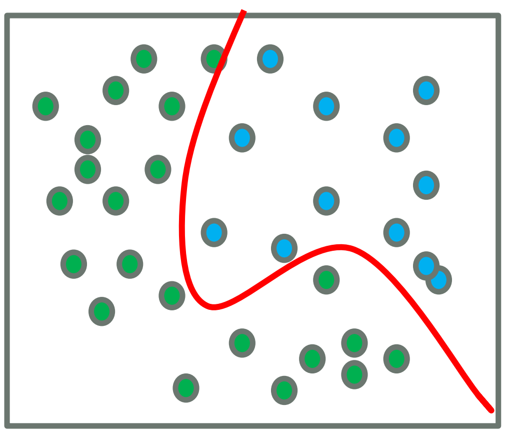
Learning from Data

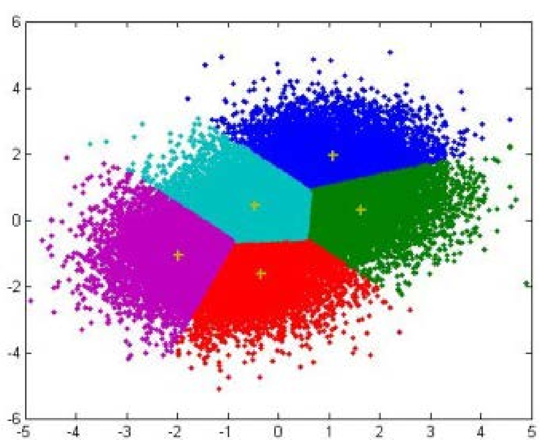
Supervised Learning

Unsupervised Learning

Reinforcement Learning

Semi-supervised Learning





Supervised Learning

Independent variables

X

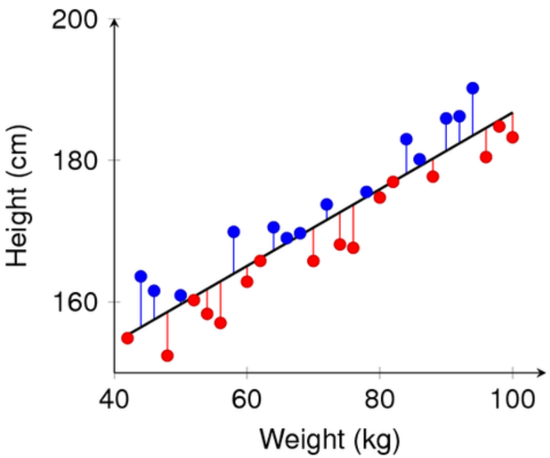
Function - f

dependent variable

Y

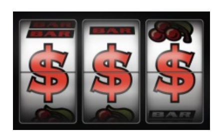
Regression

Continuous



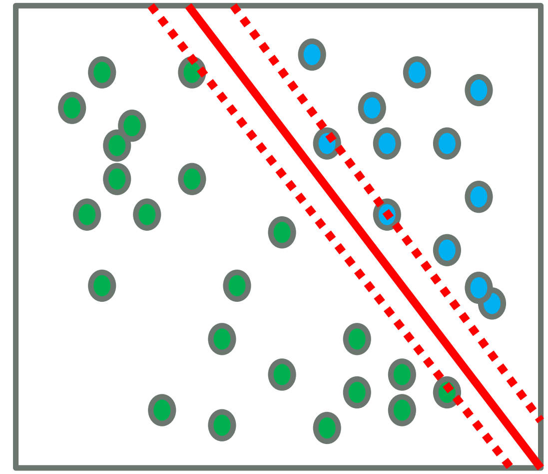
Classification

Categorical



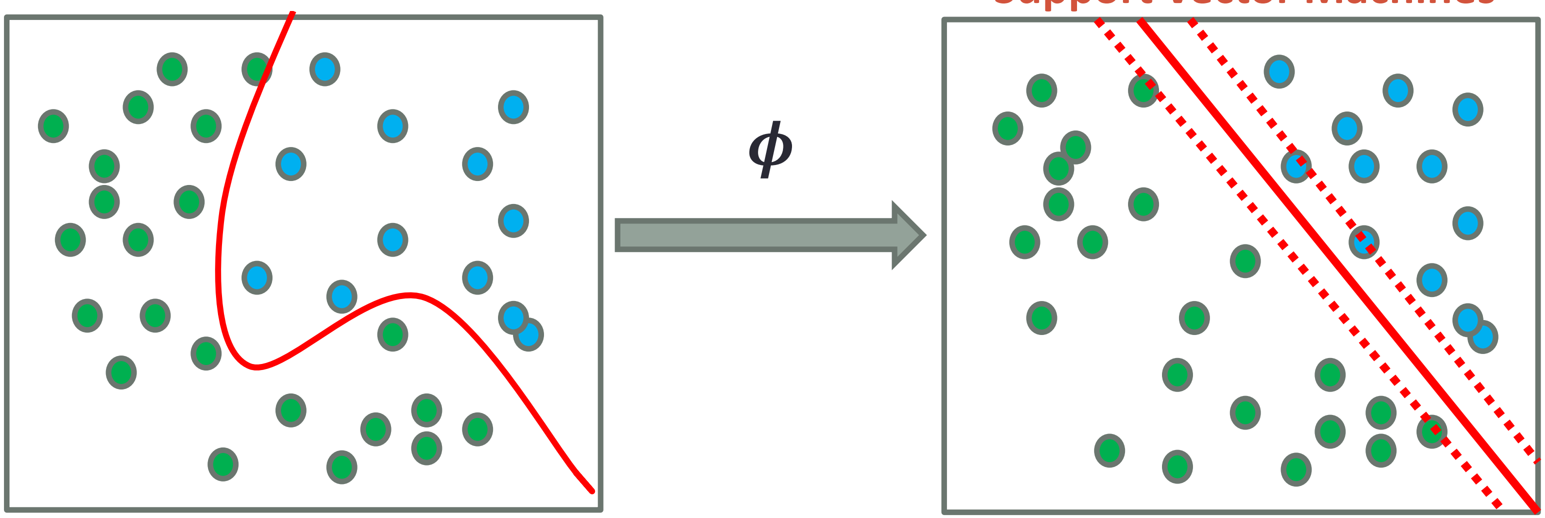
Classification

- Generative classifier
- Discriminative classifier



Kernel Methods

Support Vector Machines

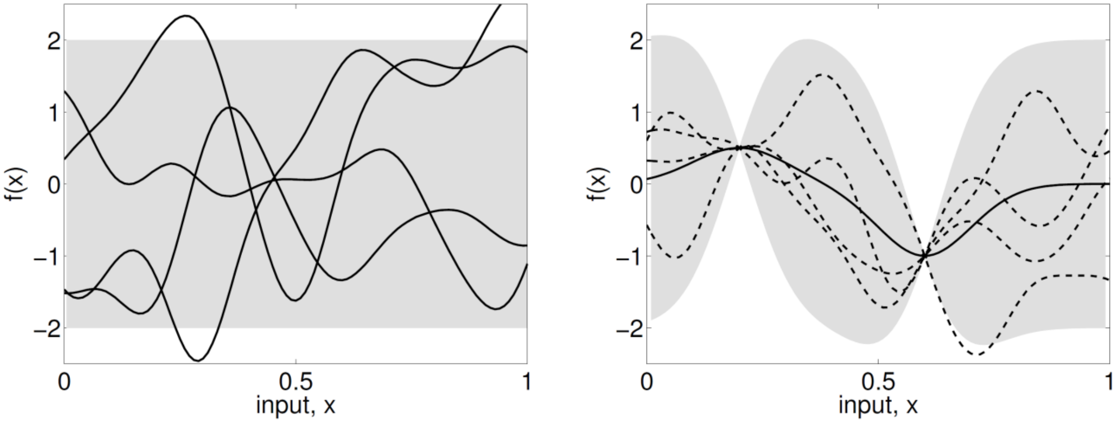


- Kernel Function – $K(x_i, x_j) = \phi(x_i).\phi(x_j)$

Kernel methods for pattern analysis, Taylor , Cristianini, 2004

Other popular classifiers

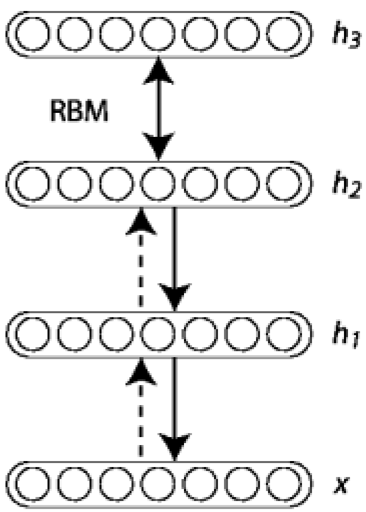
- Gaussian Processes



(a), prior (b), posterior

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006,

- Deep Belief networks



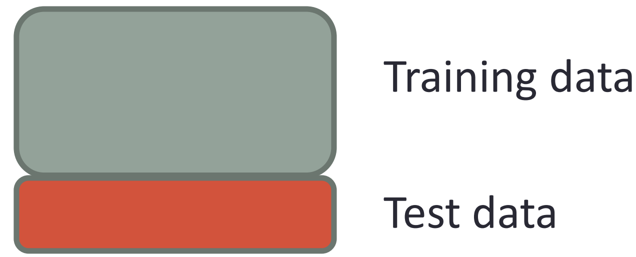
$$P(x, h^1, \ldots, h^\ell) = \left(\prod_{k=0}^{\ell-2} P(h^k | h^{k+1}) \right) P(h^{\ell-1}, h^\ell)$$

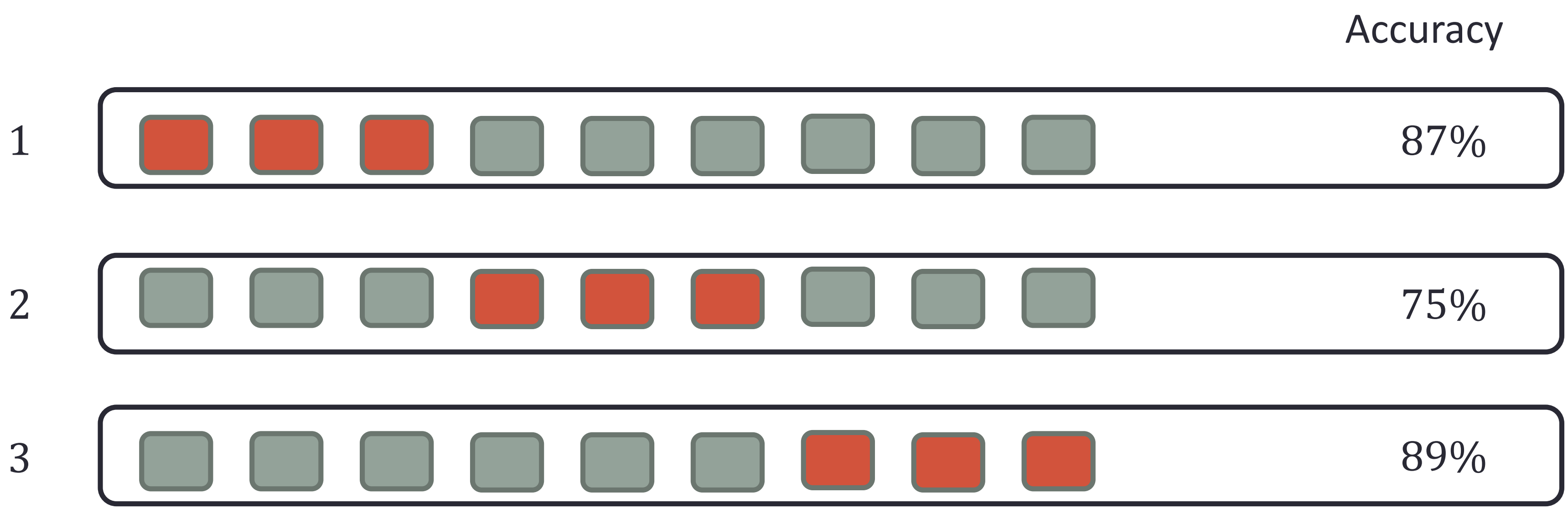
http://deeplearning.net/tutorial/DBN.html

G.E. Hinton, S. Osindero, and Y. Teh, “A fast learning algorithm for deep belief nets”, Neural Computation, vol 18, 2006

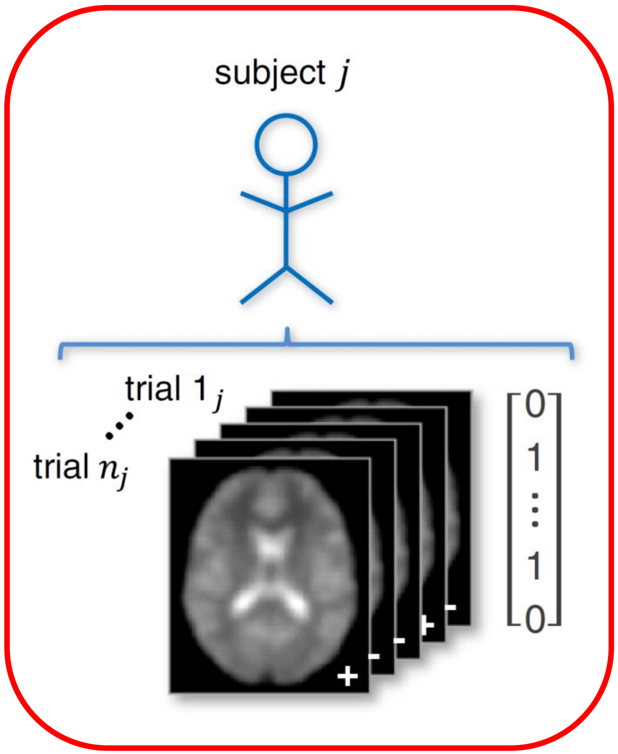
K-fold Cross Validation

- Model Selection
- Performance evaluation
  - Balanced Accuracy
  - F1 Score





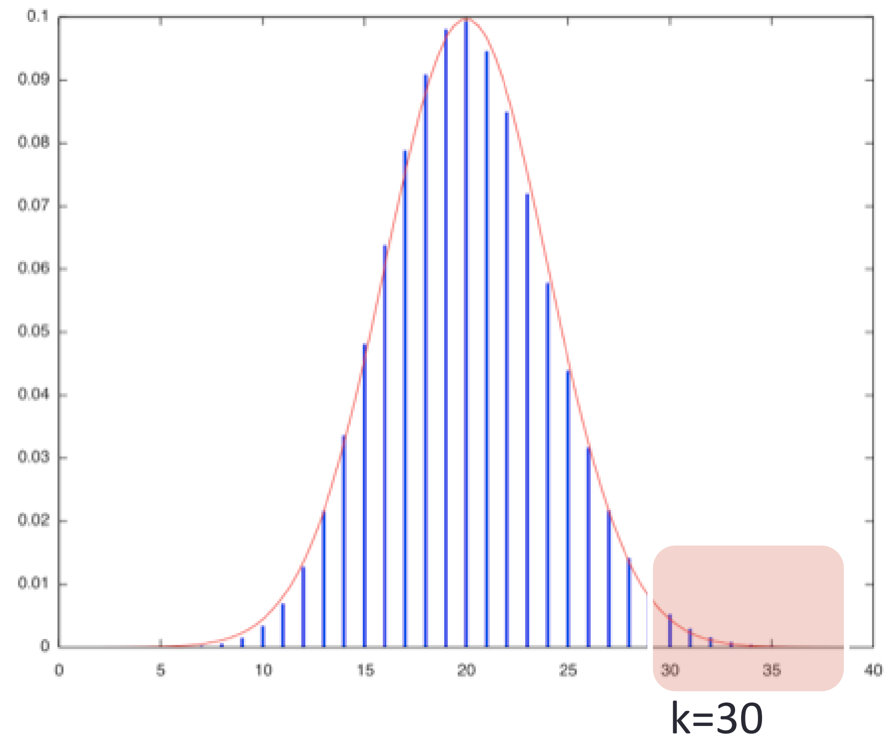
Performance – Single Subject



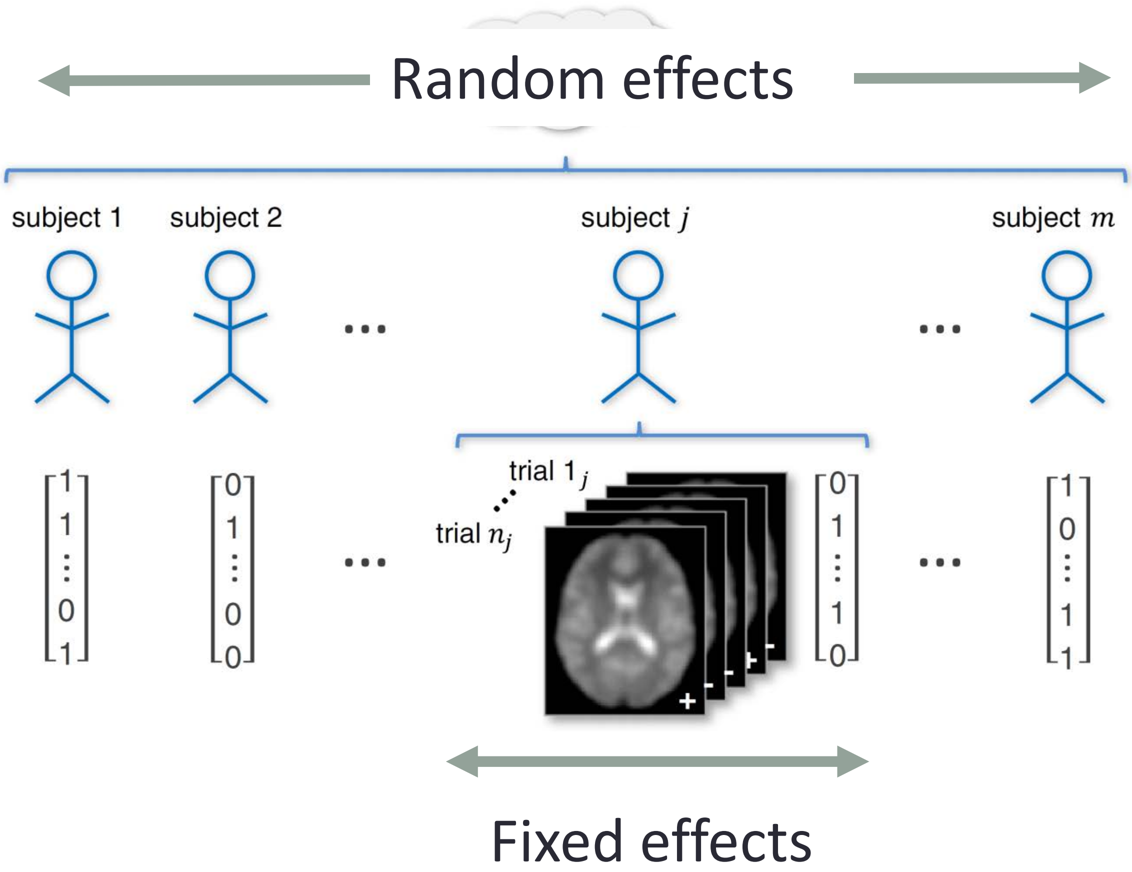
Brodersen et al. 2013, *NeuroImage*

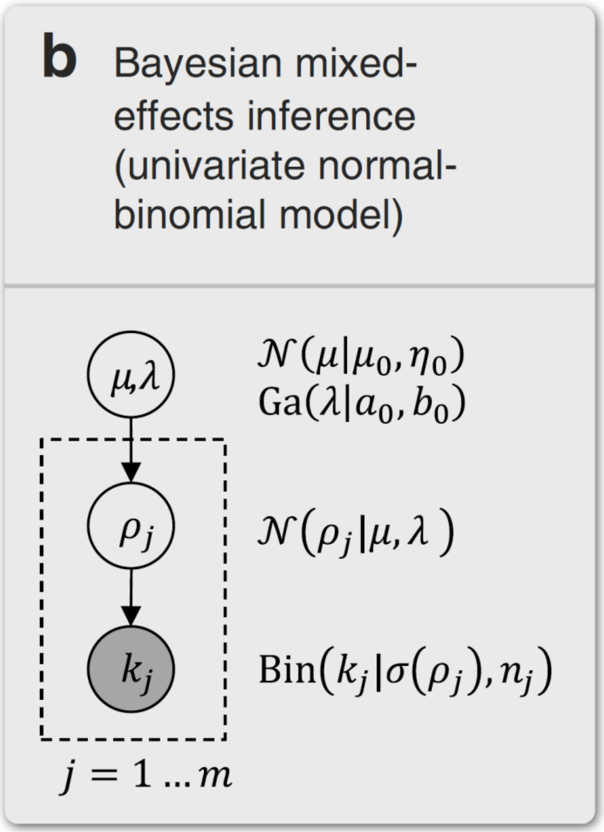
Binomial Test

$$p = P(X \geq k|H_0) = 1 - B(k|n, \pi_0)$$



Performance – Multiple Subjects



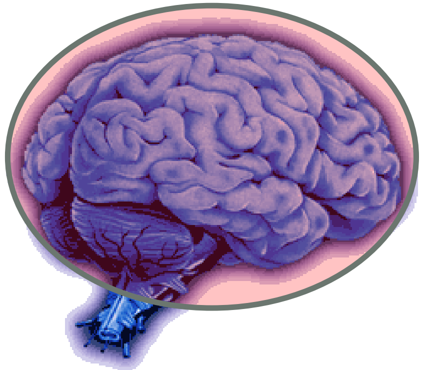


http://www.translationalneuromodeling.org/tapas/

Brodersen et al. 2013, *NeuroImage*

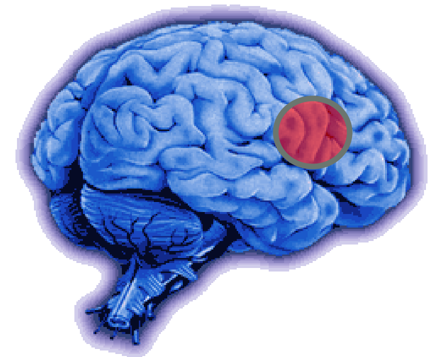
Using classification for fMRI data

Whole brain



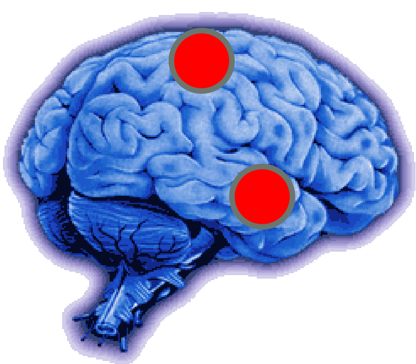
Mourao-Miranda et al. (2005) *NeuroImage*,
Marquand et al. *(2010) NeuroImage*

Searchlight classifier



Kriegeskorte et al. (2006) *PNAS*

Pattern localization

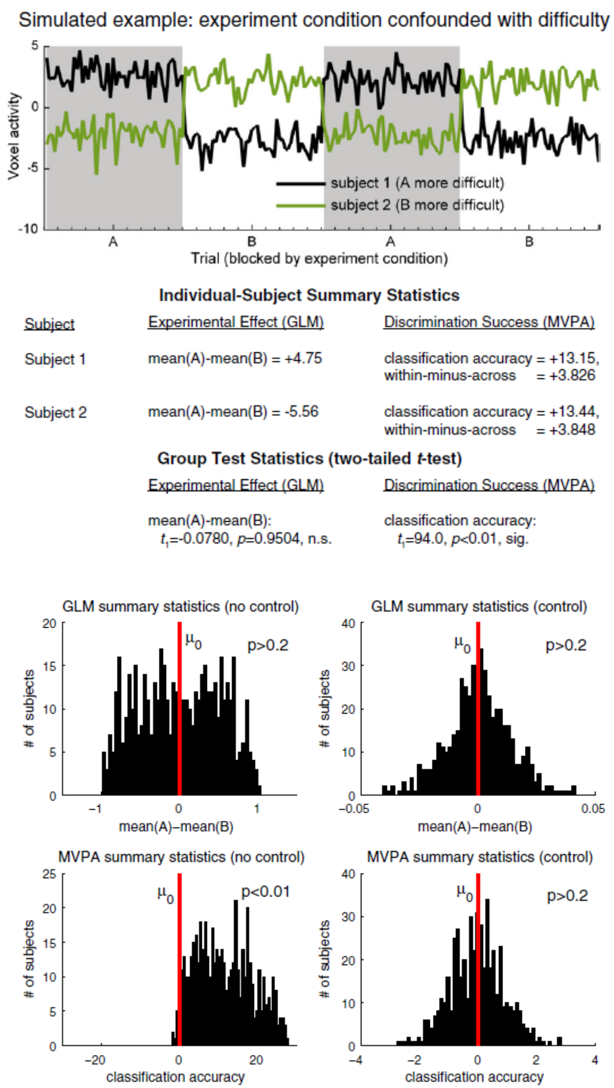


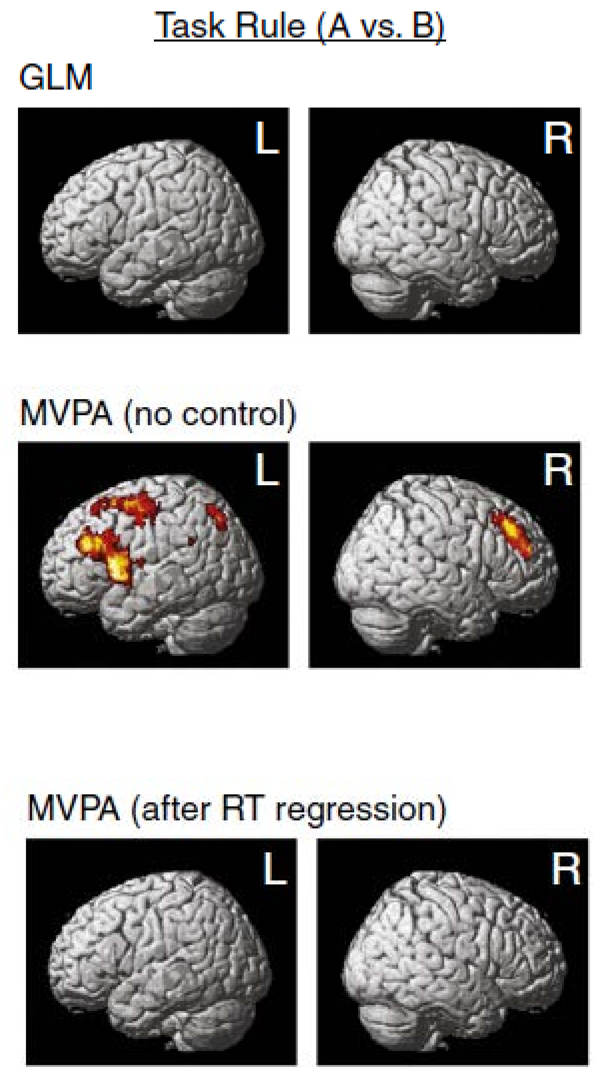
Word Category

Food
Building
People

Pereira et al. (2009) *NeuroImage*, Mitchell et al. (2004) *Machine Learning*

Confounds – GLM vs MVPA

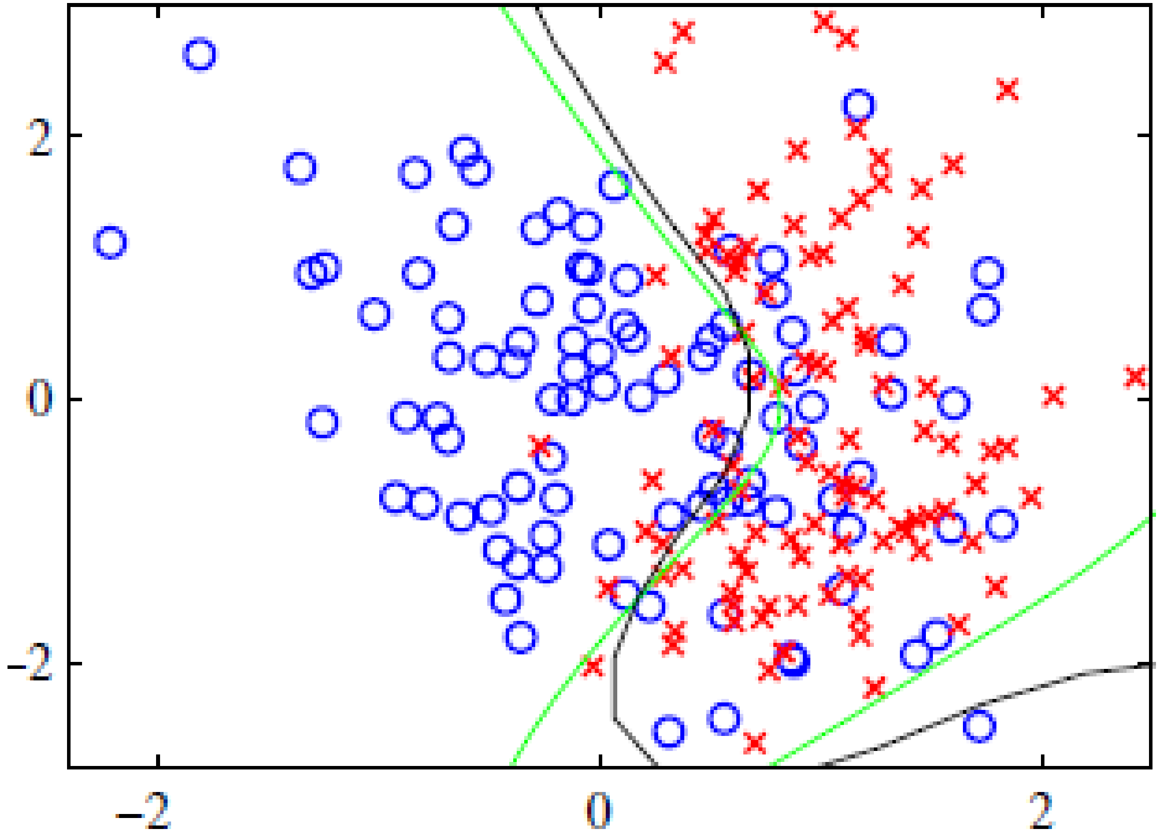




Todd et al. 2013, *NeuroImage*

Practicals

- Classification using SVM

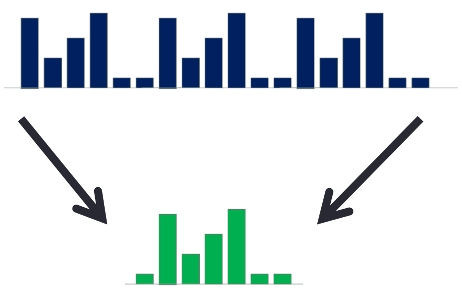


Bishop (2006) PRML

Unsupervised Learning

Building a representation of data

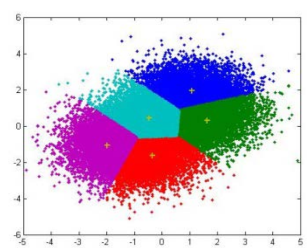
Dimensionality Reduction



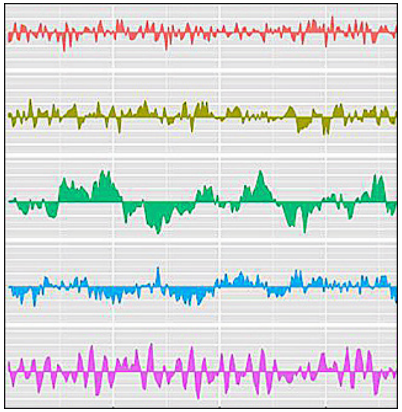
Clustering

K-means

Mixture models



Time series

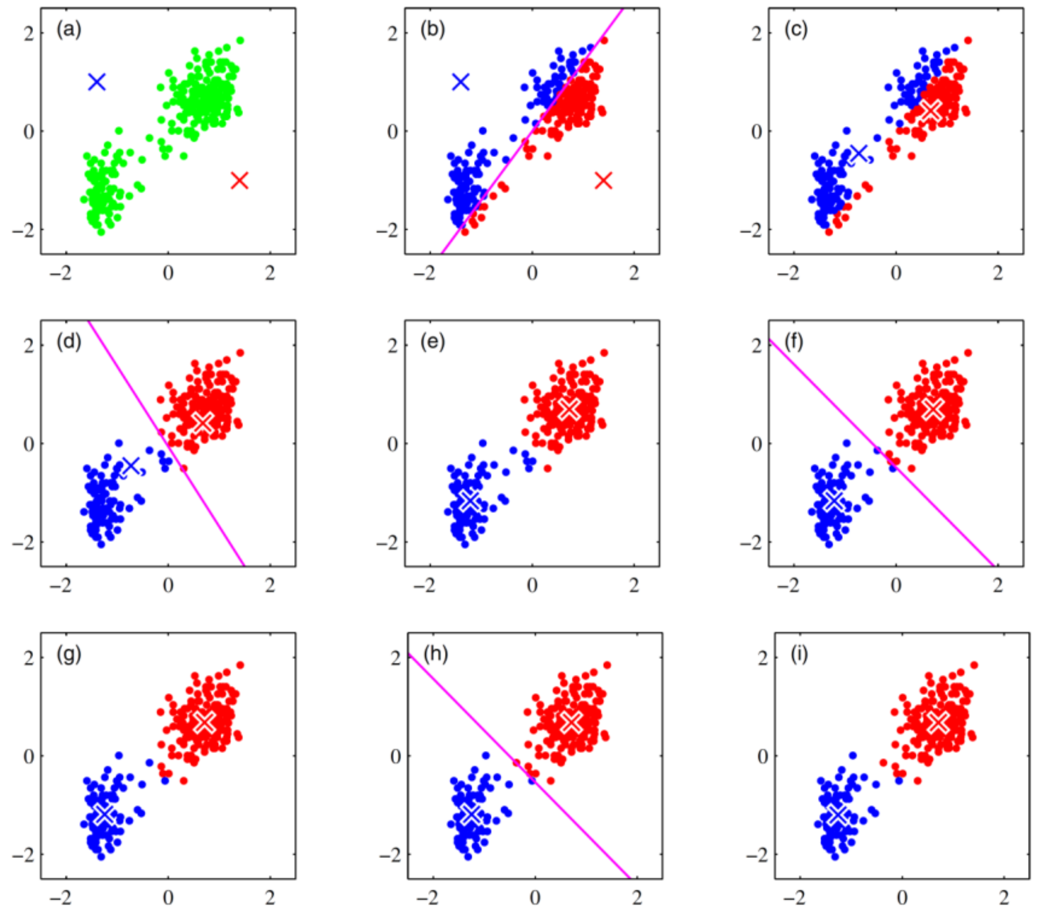


Clustering using K-means

- Cost function

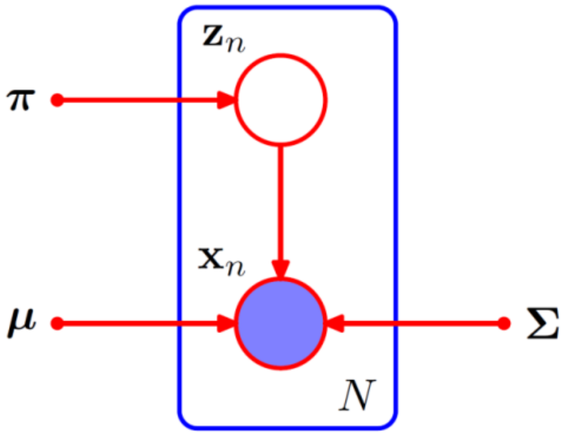
$$\widetilde{J} = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \mathcal{V}(\mathbf{x}_n, \boldsymbol{\mu}_k)$$

- Algorithm
    1. Initialize
    2. Estimate assignments
    3. Estimate cluster centroids
    4. Repeat 2,3 until convergence

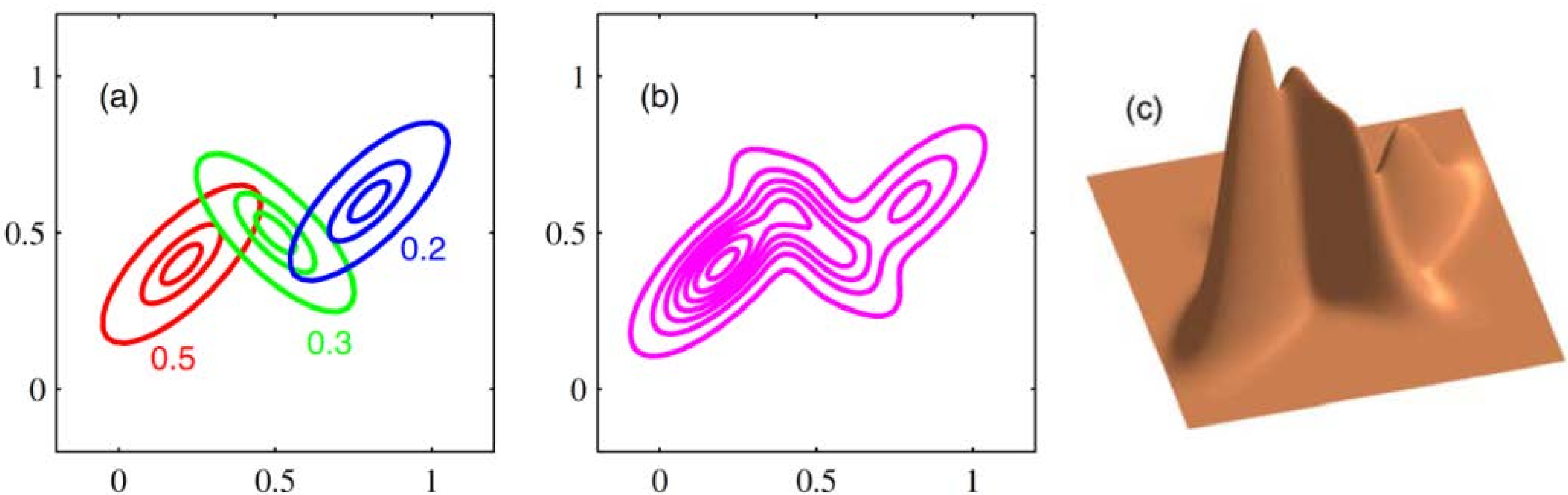


Bishop PRML (2006)

Clustering – Mixture of Gaussians



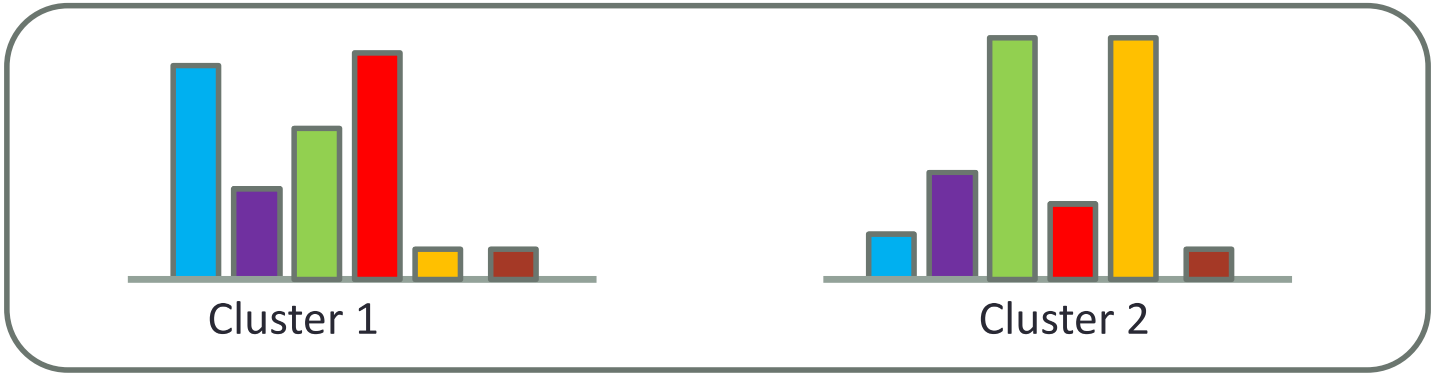
$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$



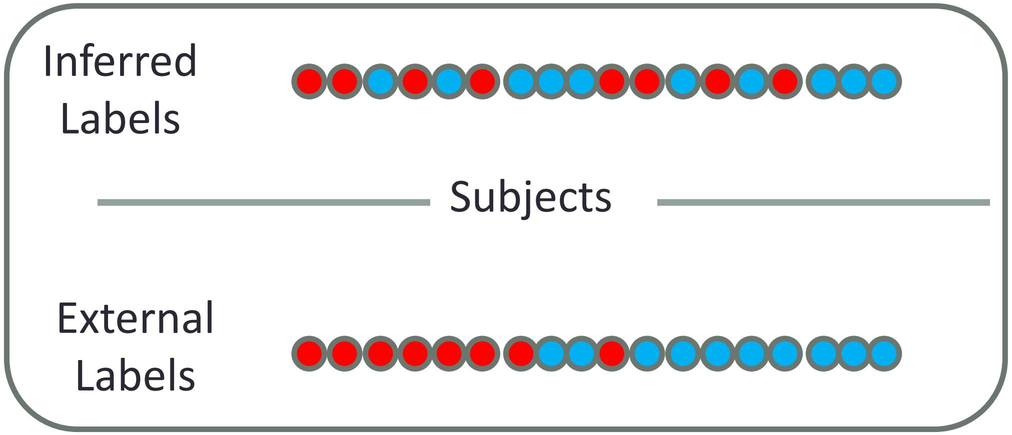
Bishop PRML (2006)

Interpretation

- Cluster parameters



- Internal Criterion – Model Evidence
- External Criterion - Purity



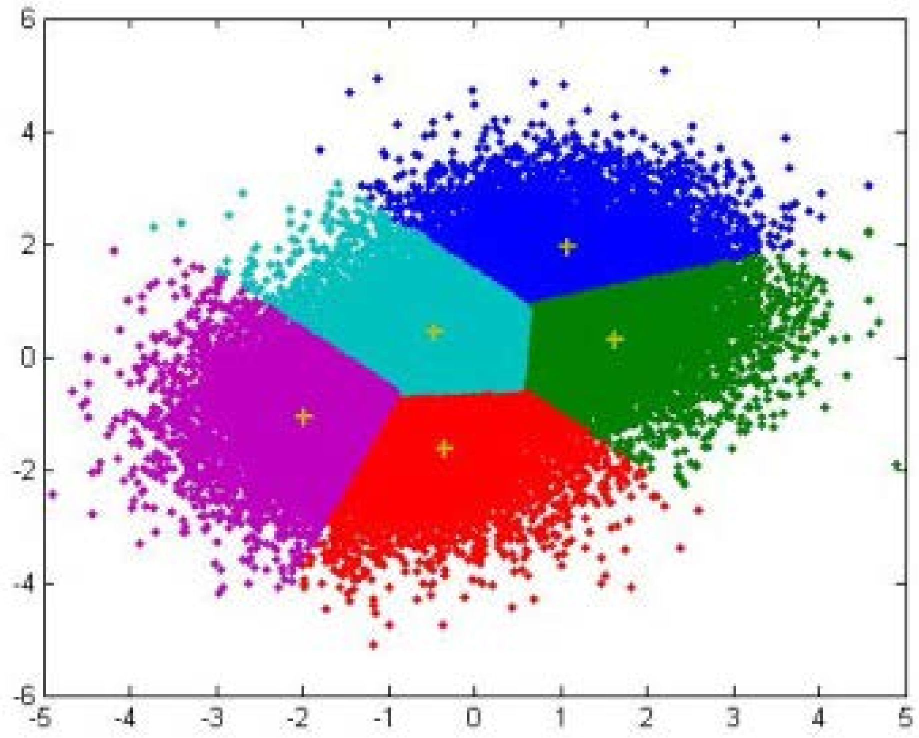
Normalized Mutual Information (NMI)

Balanced purity

Rand Index

Practicals

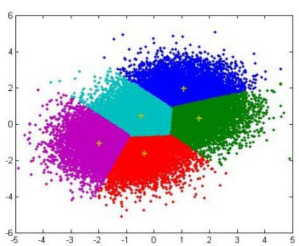
- Clustering
  - K-Means
  - Finite Gaussian mixture model

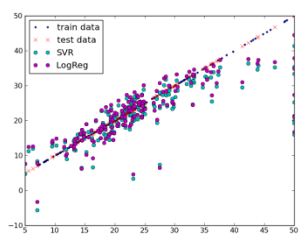


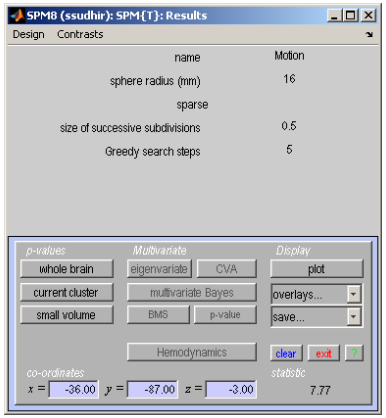
Bishop (2006) PRML

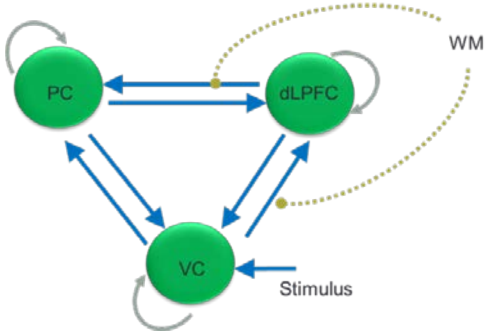
Motivation

- Modelling
- Learning from Data
- Multivariate Bayes in SPM
- Generative Embedding

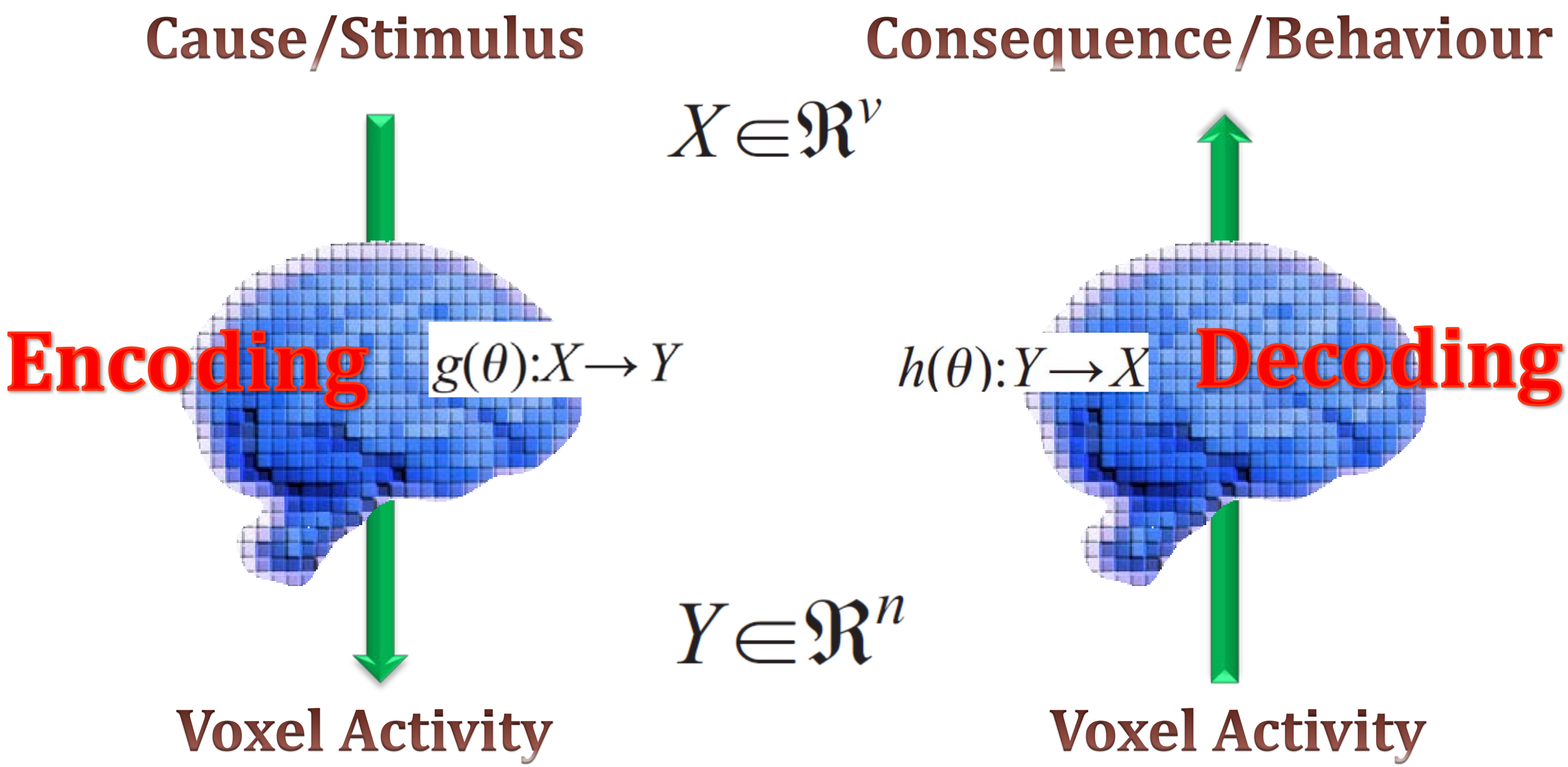




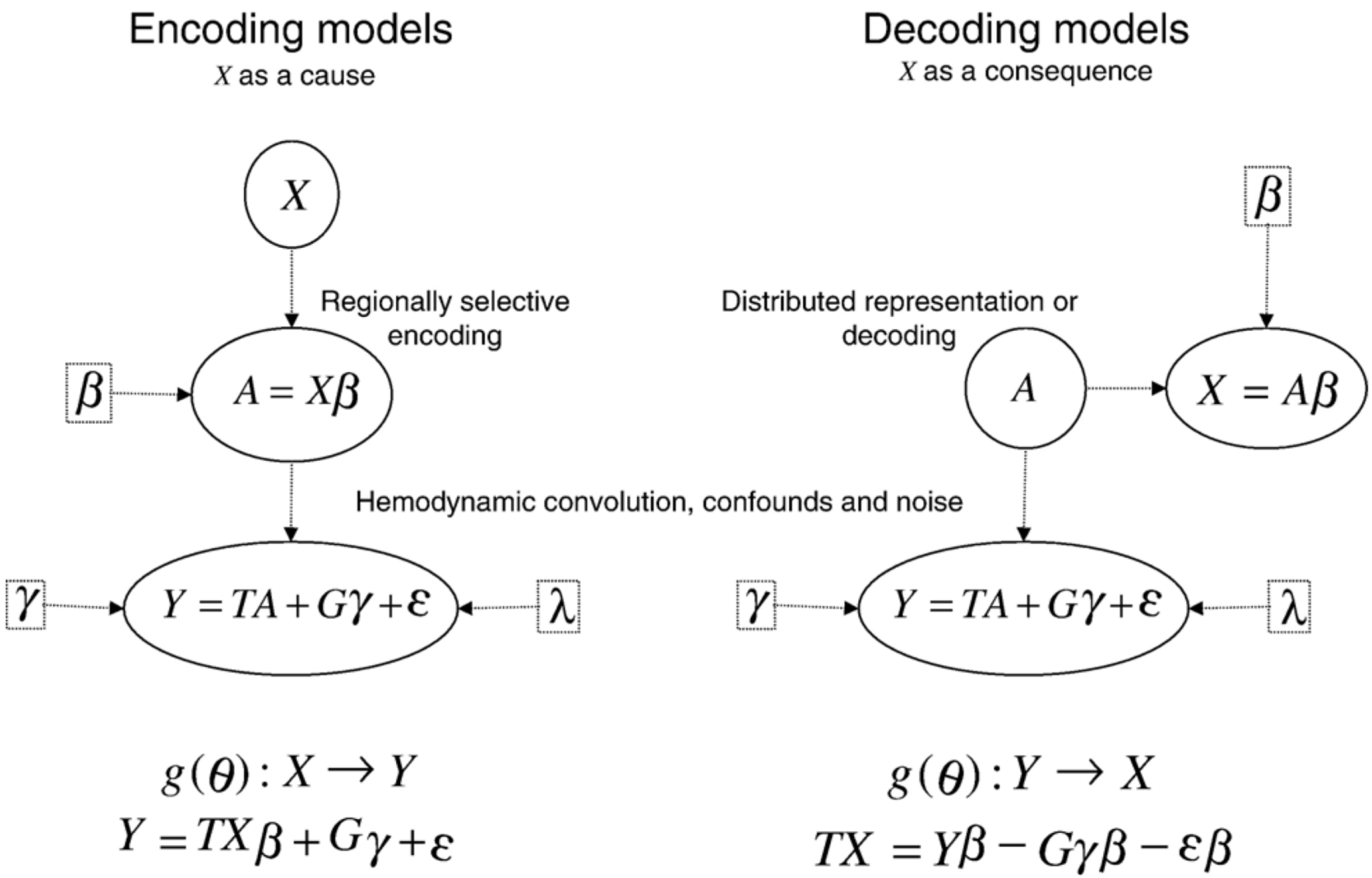




Encoding Vs Decoding Models



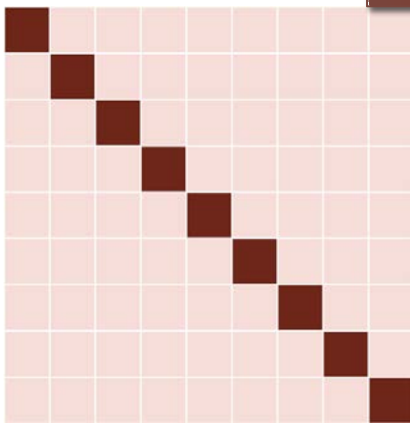
Encoding Vs Decoding



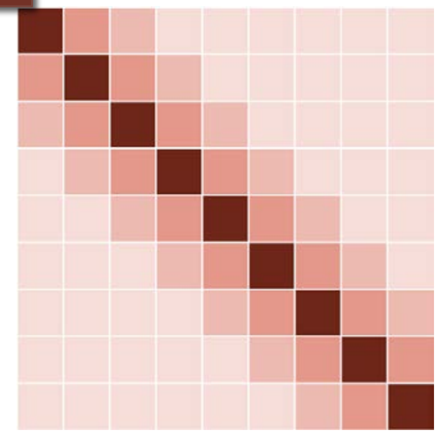
Friston et al. 2008 NeuroImage

Coding hypotheses

Sparse vectors



Spatial vectors

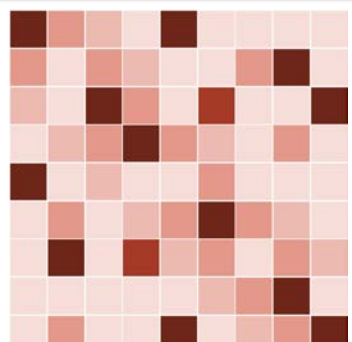


Smooth vectors

Distributed vectors

Singular vectors
of data

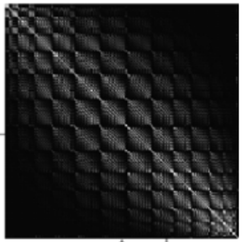
$$UDV^T = RY^T$$



Support vectors

$$U = RY^T$$

Coding hypotheses



Null:	$U = \varnothing$
Spatial vectors:	$U = I$
Smooth vectors:	$U(\vec{x}_i, \vec{x}_j) = \exp(-\tfrac{1}{2}(\vec{x}_i - \vec{x}_j)^2 \sigma^{-2})$
Singular vectors:	$UDV^T = RY^T$
Support vectors:	$U = RY^T$

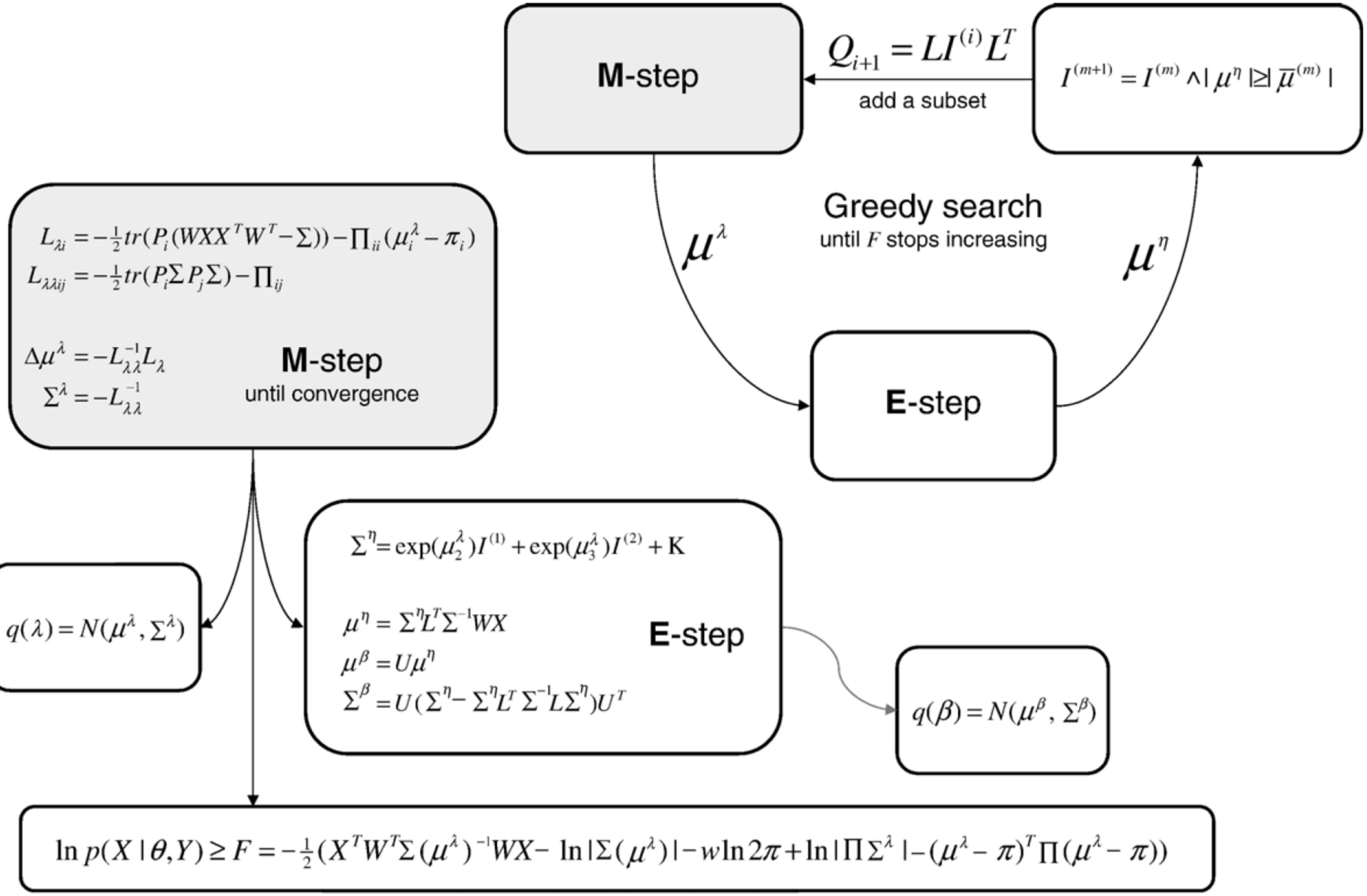
$$
\begin{aligned}
WX &= RY\beta + \varsigma \\
\beta &= U\eta \\
\mathrm{cov}(\varsigma) &= \Sigma^{\varsigma}(\lambda) = \exp(\lambda^{\varsigma})RVR^T \\
\mathrm{cov}(\eta) &= \Sigma^{\eta}(\lambda) = \exp(\lambda_1^{\eta})I^{(1)} + \ldots + \exp(\lambda_m^{\eta})I^{(m)}
\end{aligned}
$$

$$p(\beta) = N(0, U\Sigma^{\eta} U^T)$$

$$
\begin{aligned}
W &= RT \\
R &= orth(I - GG^{-})^T
\end{aligned}
$$

Friston et al. 2008 NeuroImage

VB Inference



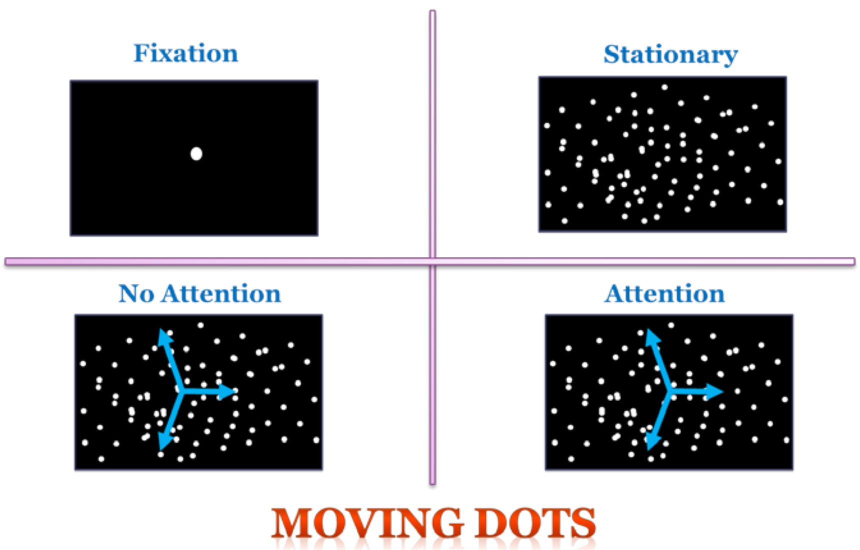
Friston et al. 2008 NeuroImage

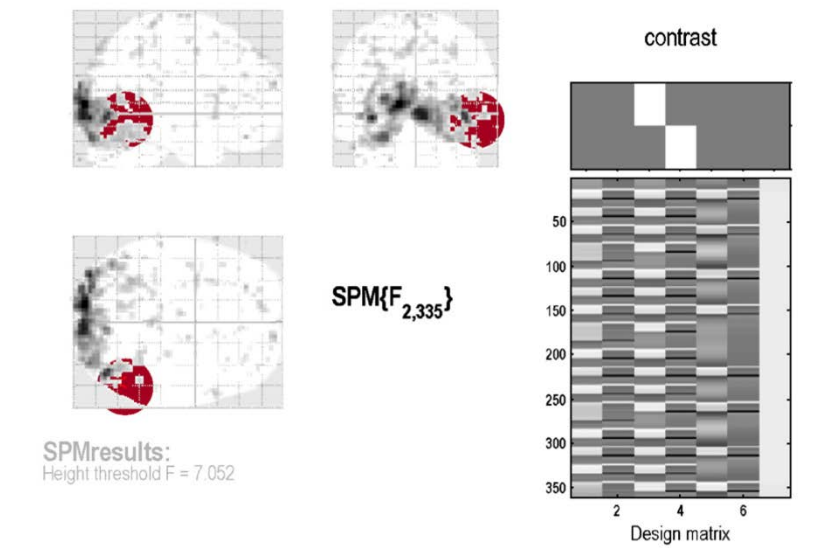
Bayesian decoding of motion

Attention to motion dataset - Büchel & Friston 1999 Cerebral Cortex

Experimental factors:

1. Photic
2. Motion
3. Attention





Statistics: ***search volume: 16.0mm sphere at [48,-63,0]***

set-level		cluster-level			voxel-level					mm mm mm		
p	c	$p_{corrected}$	k_E	$p_{uncorrected}$	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	F	$(Z_{\equiv})$	$p_{uncorrected}$			
0.006	4		39		0.000	0.000	40.66	Inf	0.000	39	-72	0
					0.000	0.000	40.09	Inf	0.000	42	-75	-3
			3		0.002	0.000	13.35	4.55	0.000	39	-63	-12
			1		0.124	0.003	8.67	3.52	0.000	51	-66	-3
			1		0.436	0.010	7.36	3.18	0.001	48	-63	0

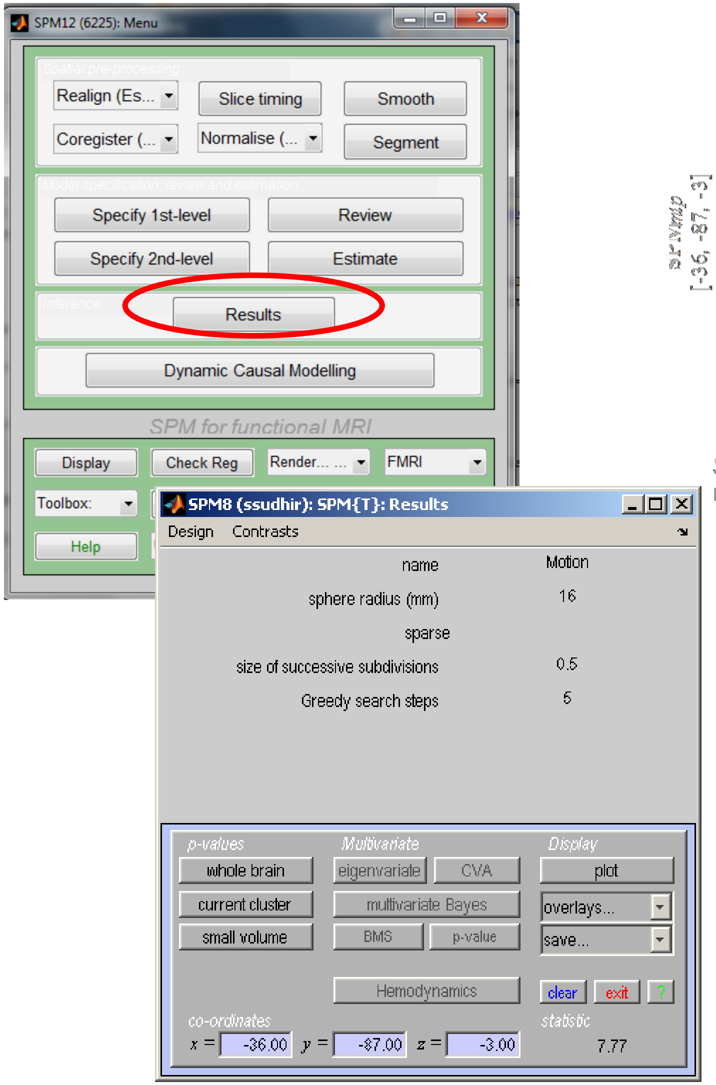
table shows 16 local maxima more than 4.0mm apart

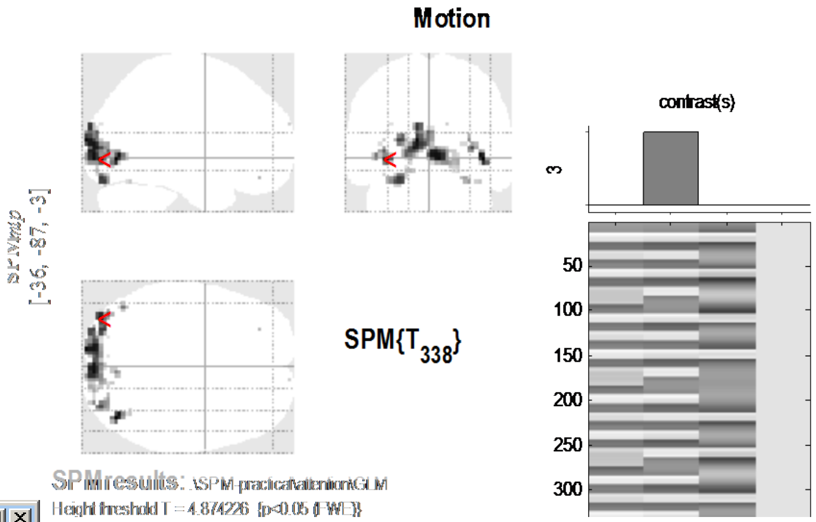
Height threshold: F = 7.05, p = 0.001 (0.513) {p<0.001 (unc.)}
Extent threshold: k = 0 voxels, p = 1.000 (0.513)
Expected voxels per cluster, <k> = 1.212
Expected number of clusters, <c> = 0.72
Expected false discovery rate, <= 0.01

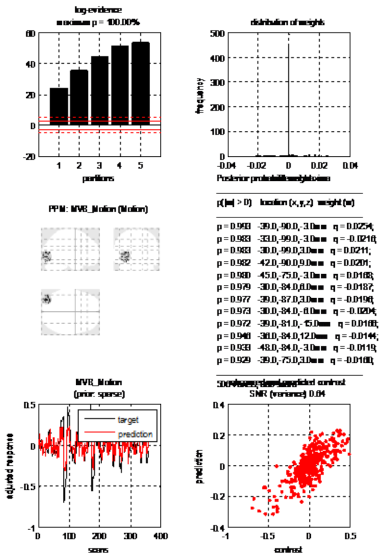
Degrees of freedom = [2.0, 335.0]
FWHM = 7.5 7.4 7.1 mm mm mm; 2.5 2.5 2.4 {voxels};
Volume: 15741; 583 voxels; 43.3 resels
Voxel size: 3.0 3.0 3.0 mm mm mm; (resel = 14.69 voxels)

Friston et al. 2008 NeuroImage

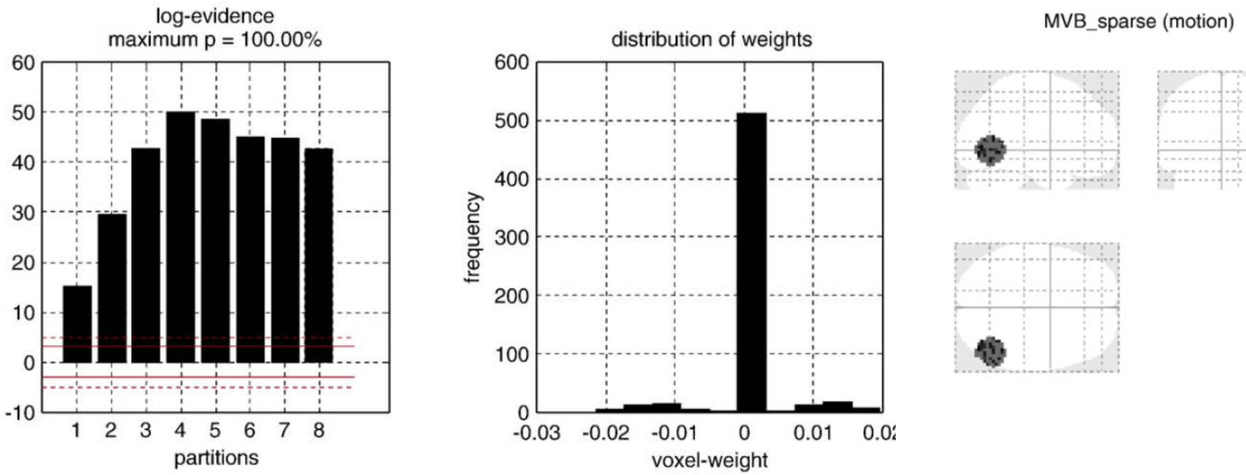
Multivariate Bayes in SPM







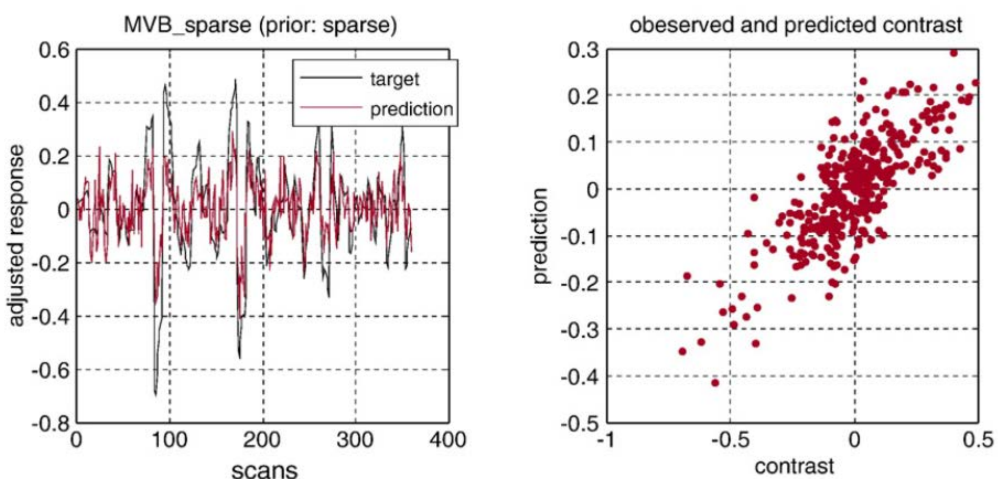
Results



Posterior probabilities at maxmia

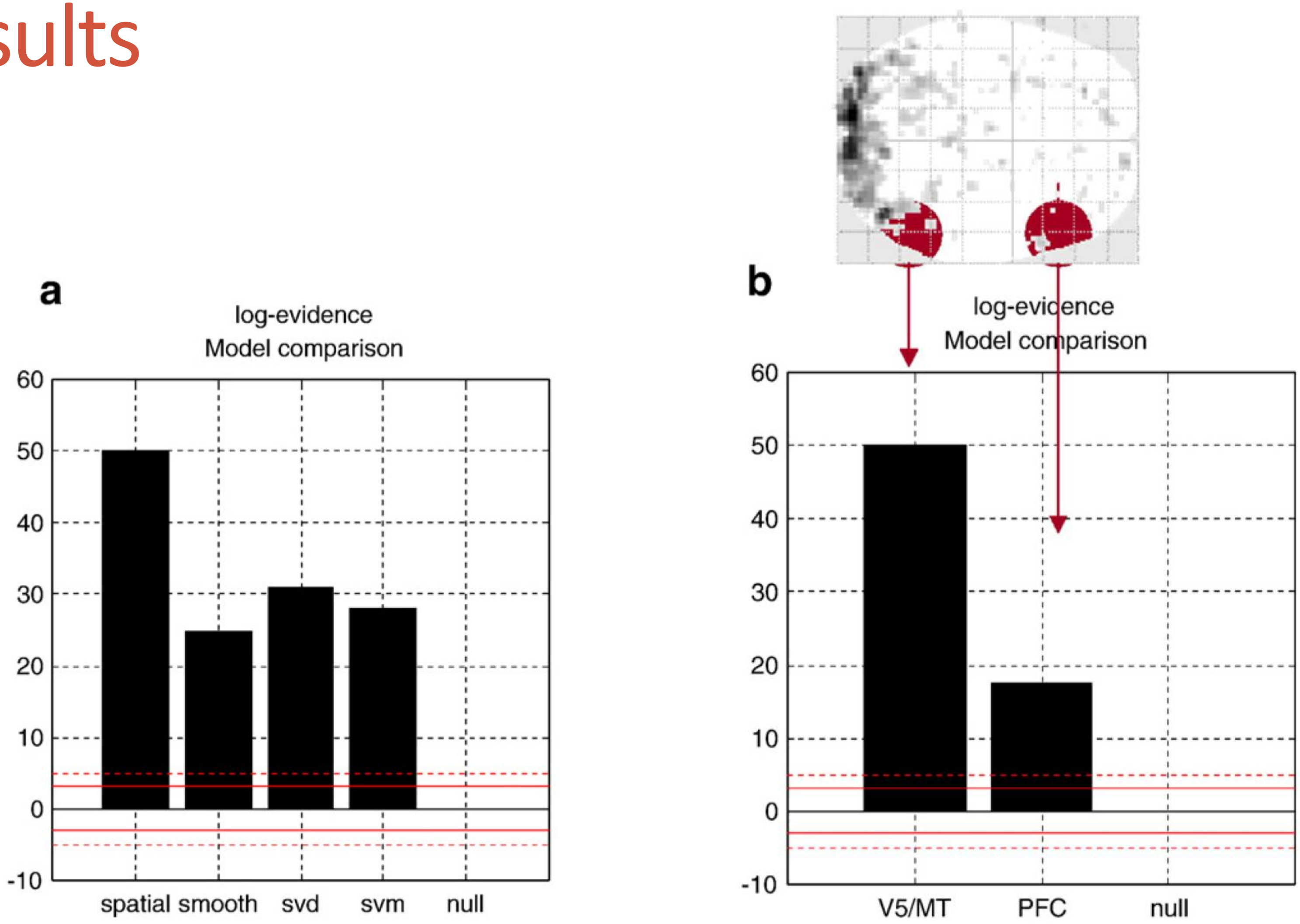
p(lwl > 0)	location (x,y,z)	weight (w)
p = 0.991	48,-78,0mm	q = -0.0208;
p = 0.977	48,-72,-3mm	q = -0.0215;
p = 0.973	36,-72,3mm	q = 0.0185;
p = 0.972	45,-51,9mm	q = 0.0188;
p = 0.968	39,-66,-6mm	q = -0.0180;
p = 0.966	42,-54,-3mm	q = -0.0168;
p = 0.963	45,-75,-6mm	q = 0.0196;
p = 0.954	54,-54,9mm	q = 0.0154;
p = 0.947	63,-60,3mm	q = -0.0161;
p = 0.945	42,-63,0mm	q = 0.0150;
p = 0.942	60,-60,-9mm	q = -0.0136;
p = 0.942	36,-57,6mm	q = -0.0167;

583 voxels; 360 scans



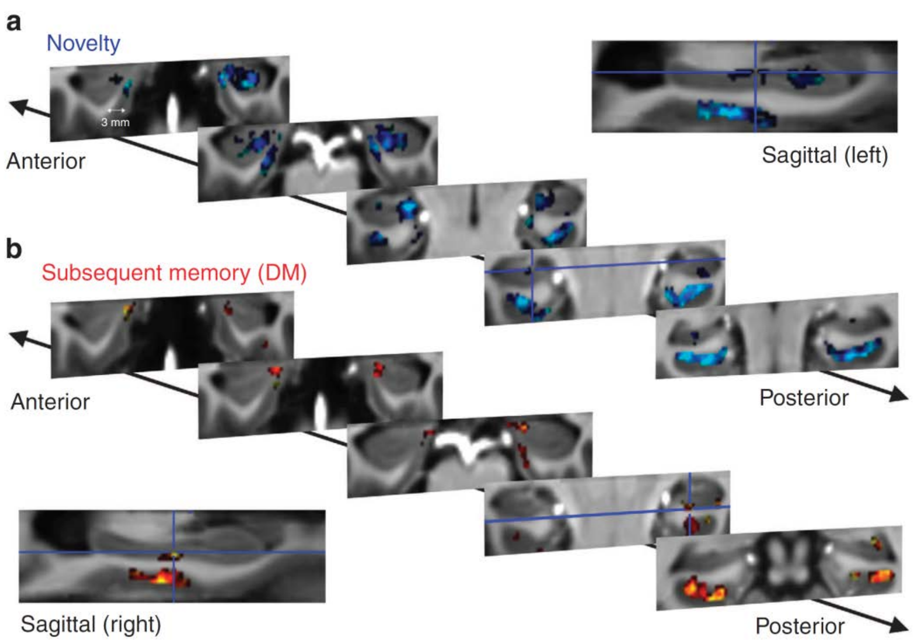
Friston et al. 2008 NeuroImage

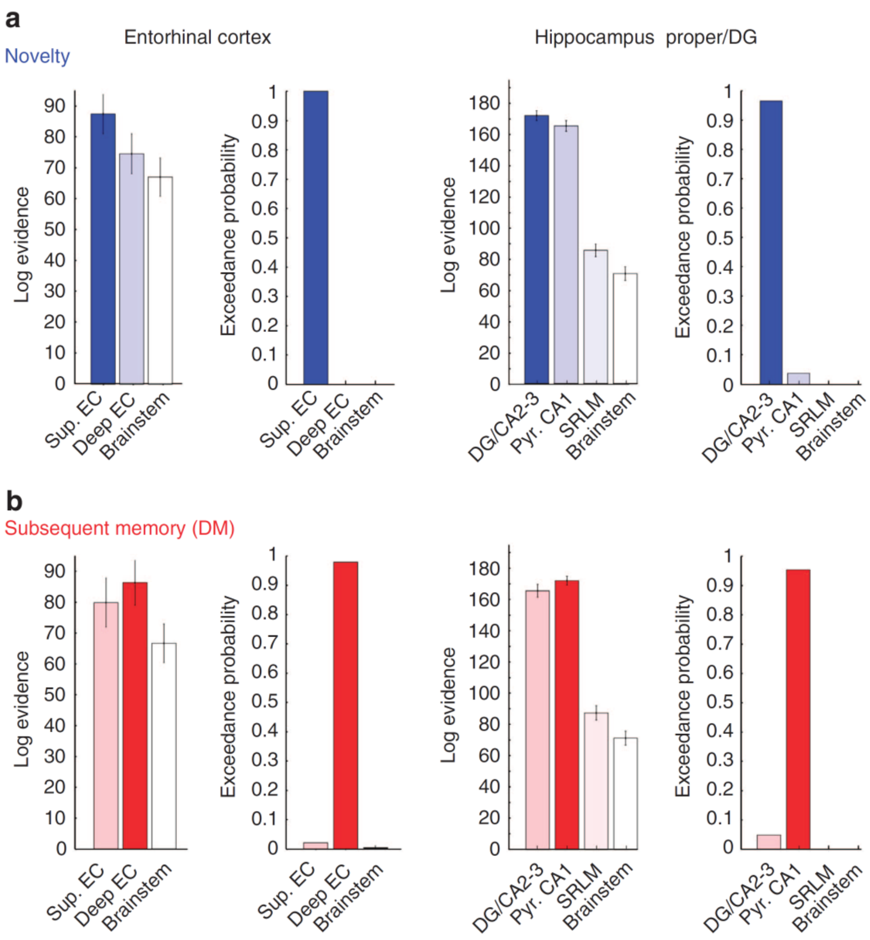
Results



a
log-evidence
Model comparison
60
50
40
30
20
10
0
-10
spatial smooth
svd
svm
null
b
log-evidence
Model comparison
60
50
40
30
20
10
0
-10
V5/MT
PFC
null

Laminar activity related to novelty and episodic encoding

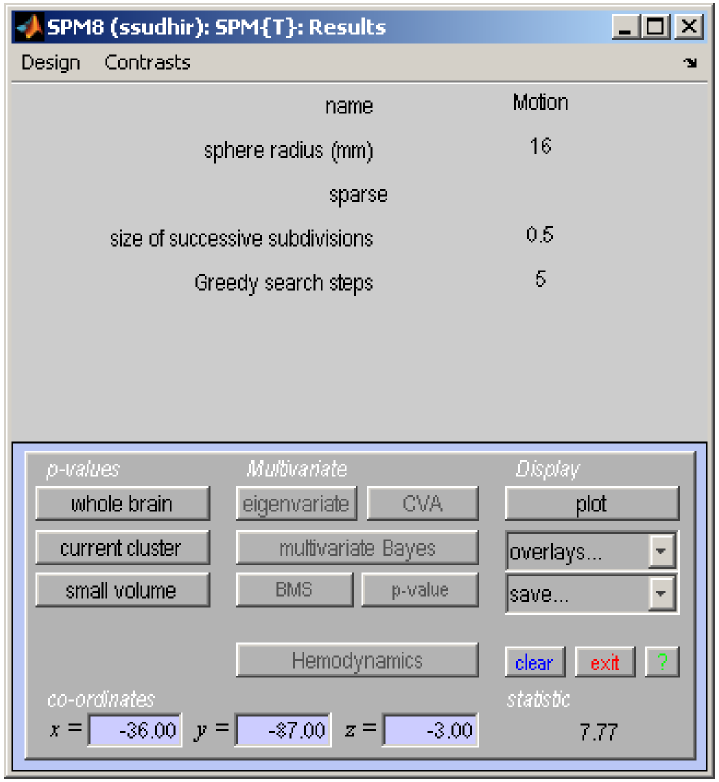




Maas et al. 2014 Nature Communications

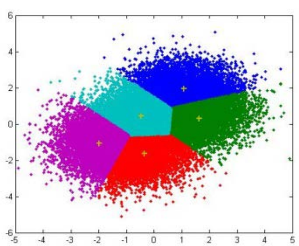
Demo

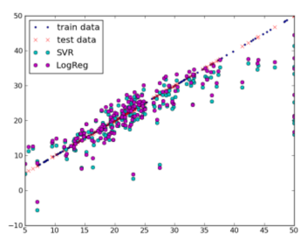
- Bayesian decoding of motion in SPM

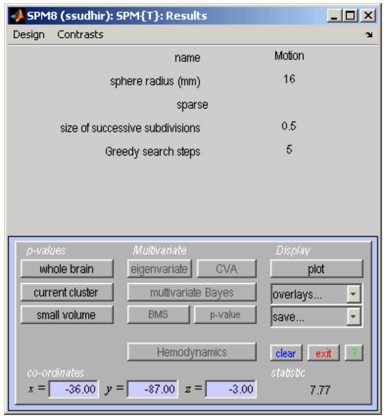


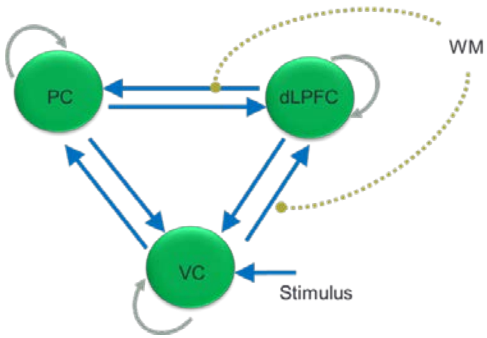
Motivation

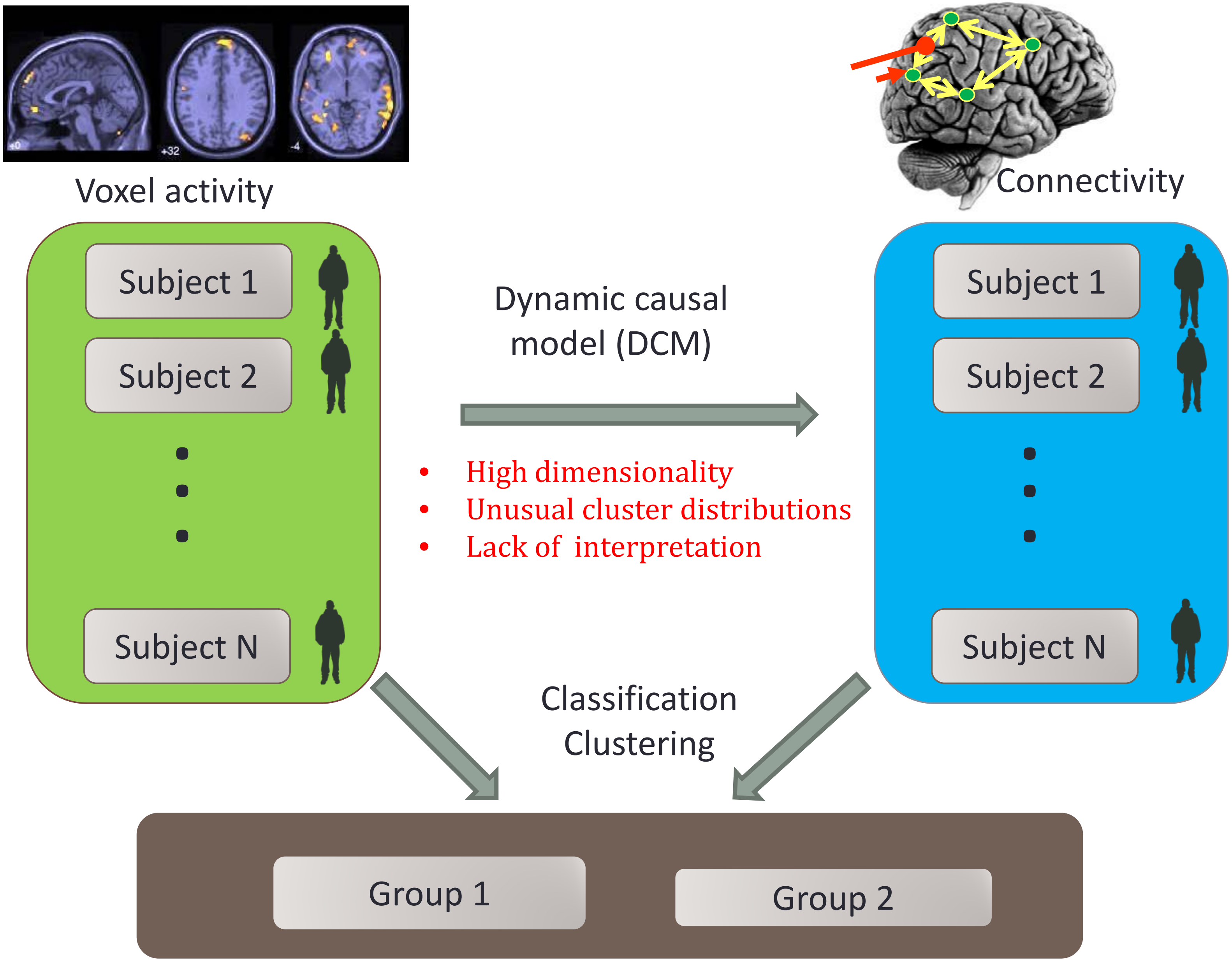
- Modelling Principles
- Learning from Data
- Multivariate Bayes in SPM
- Generative Embedding



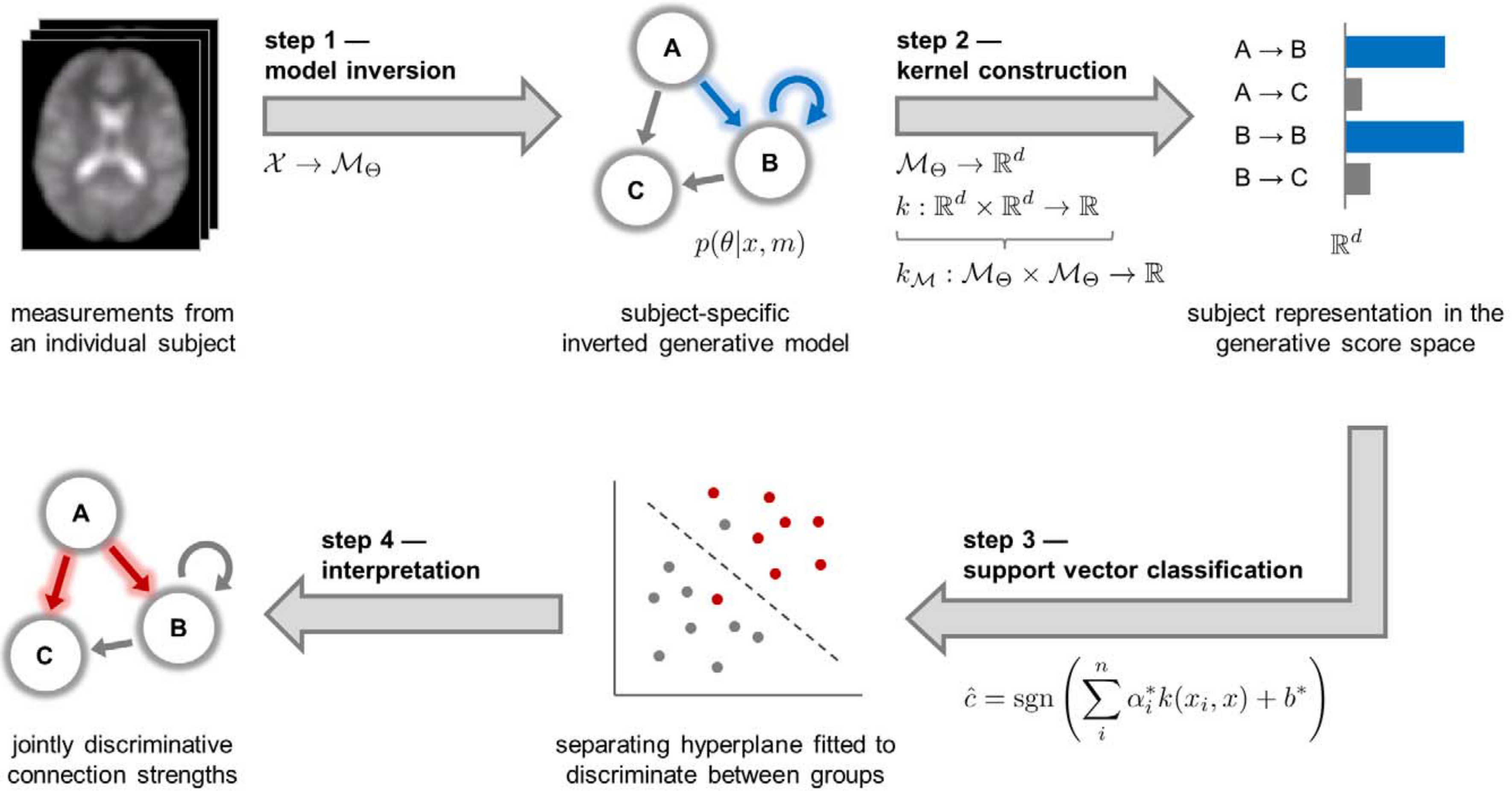






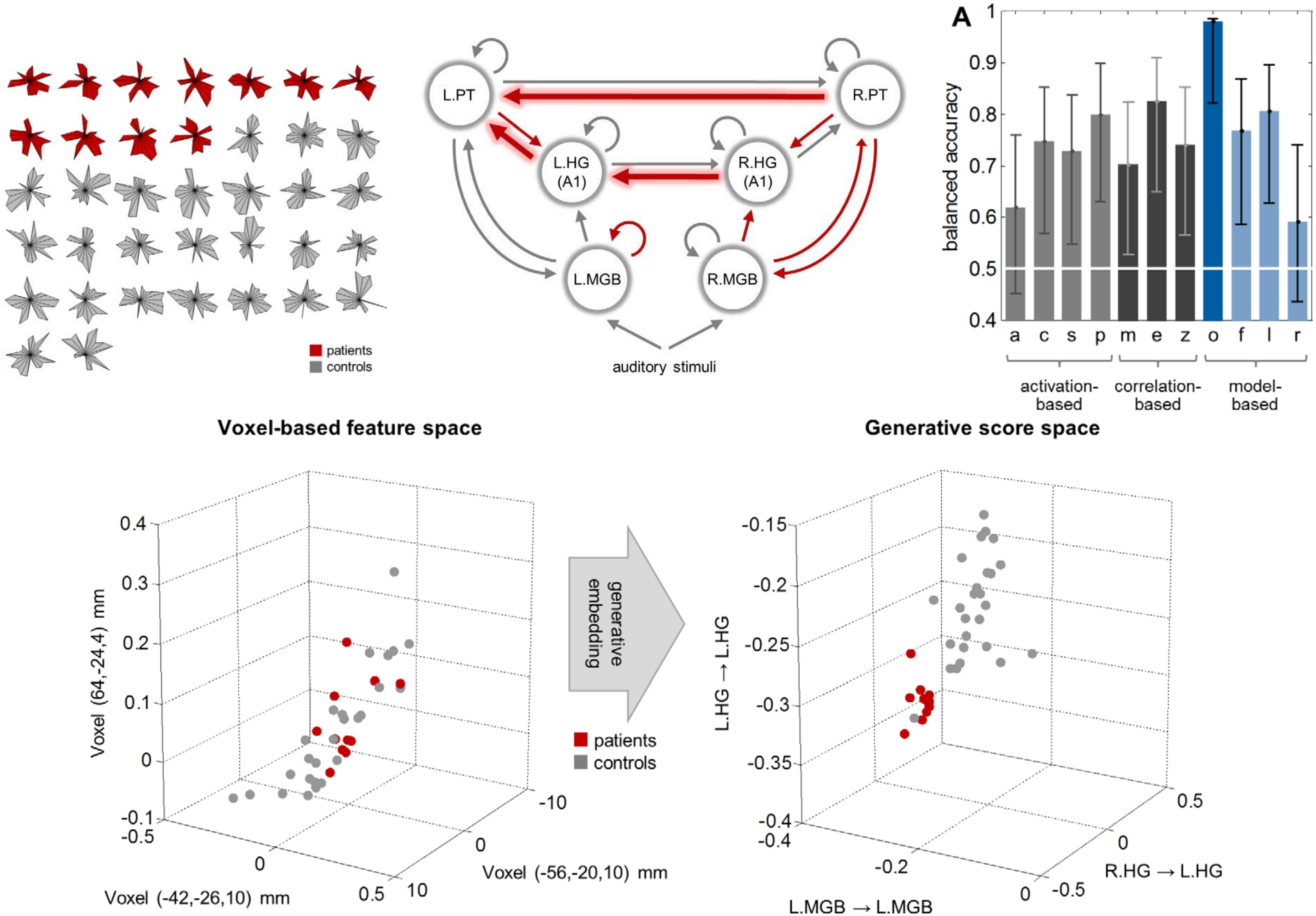

Voxel activity
Connectivity
Subject 1
Subject 2
Subject N
Dynamic causal
model (DCM)
High dimensionality
Unusual cluster distributions
Lack of interpretation
Subject 1
Subject 2
Subject N
Classification
Clustering
Group 1
Group 2

Generative Embedding - Classification



Brodersen et al. pLOS computation biology 2011.

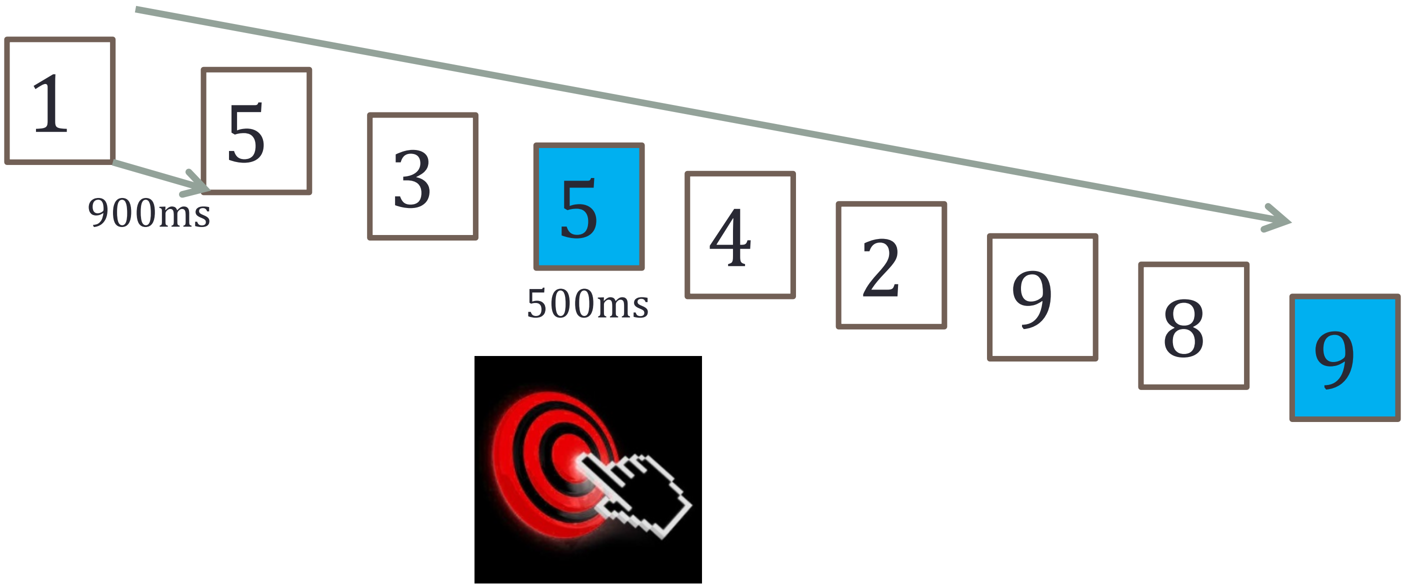
DCM - Speech processing



Brodersen et al. pLOS computation biology 2011.

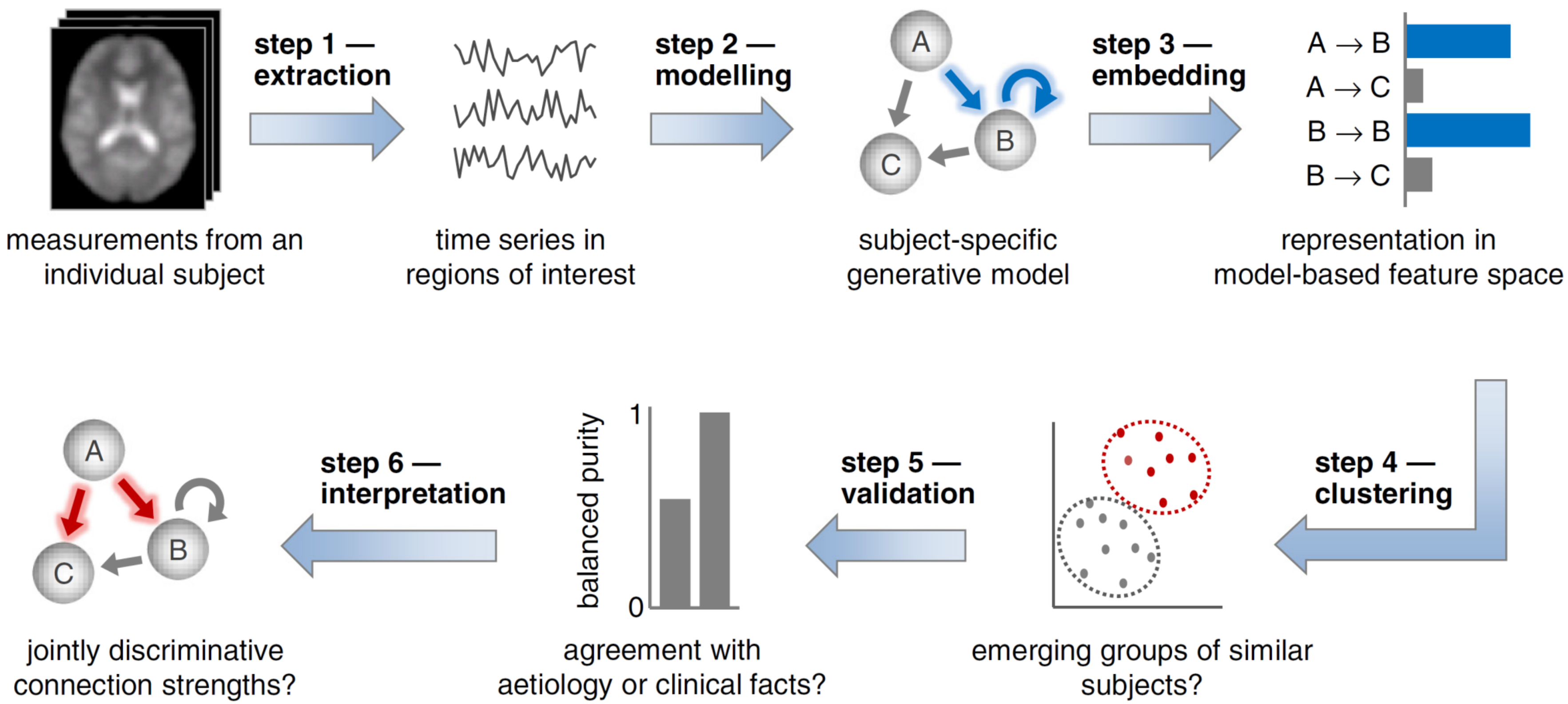
Working memory - fMRI

- 41 Schizophrenia patients (DSM IV,ICD 10), 42 controls
- Visual numeric n-back working memory task



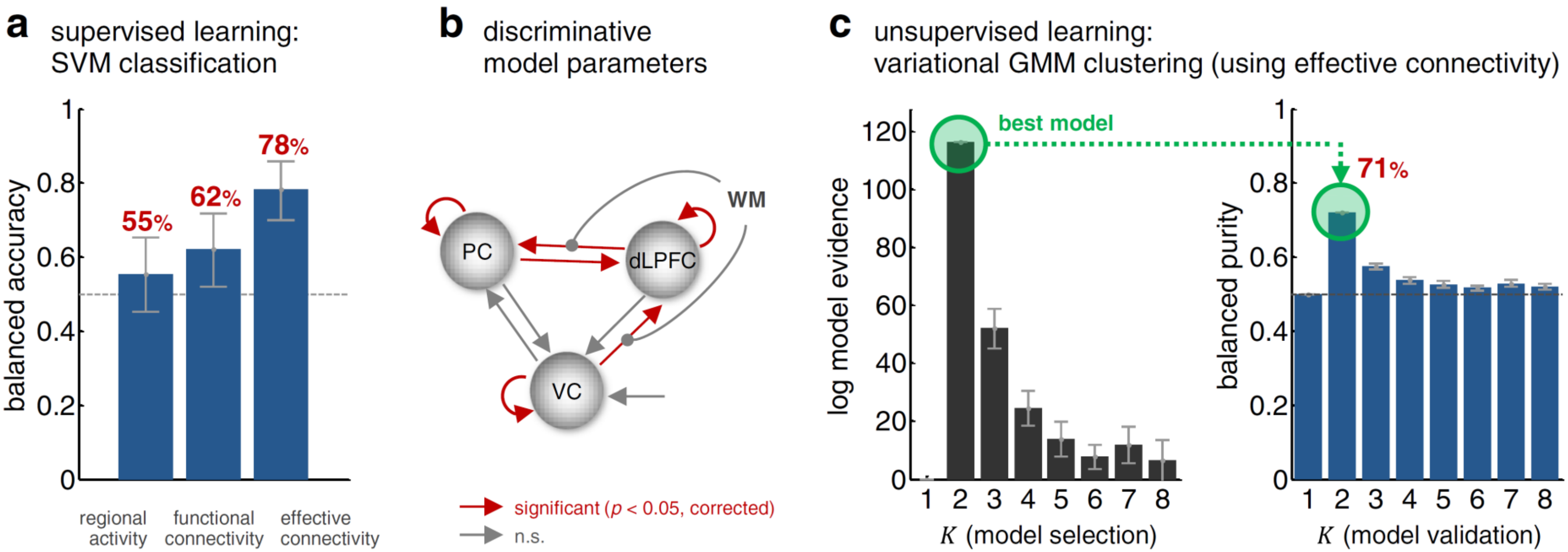
Deserno, Lorenz et al 2012. *The Journal of Neuroscience* 32 (1). Society for Neuroscience: 12–20.

Model based clustering



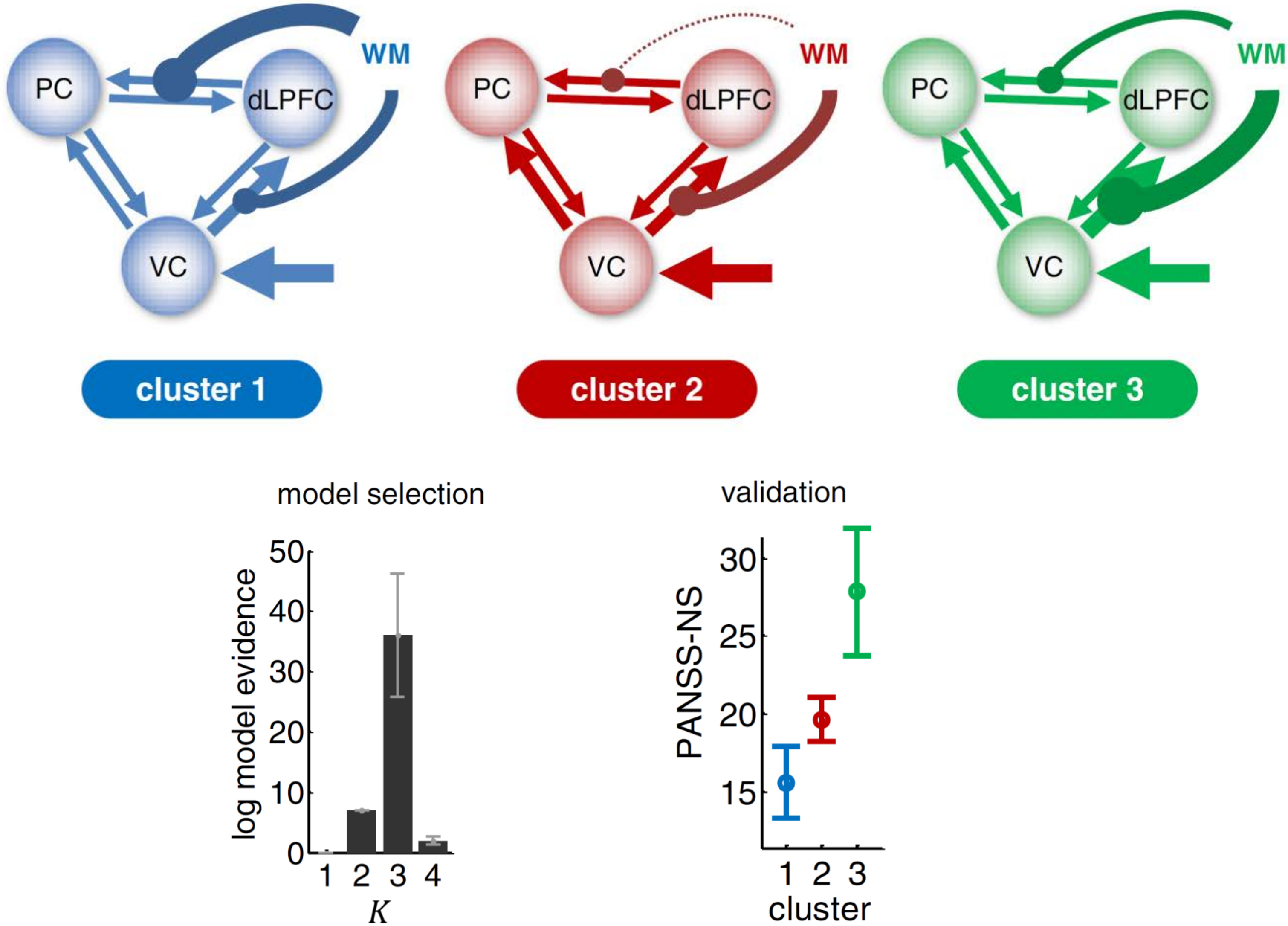
Brodersen et al 2014 Neuroimage

Results – Healthy vs Schizophrenic



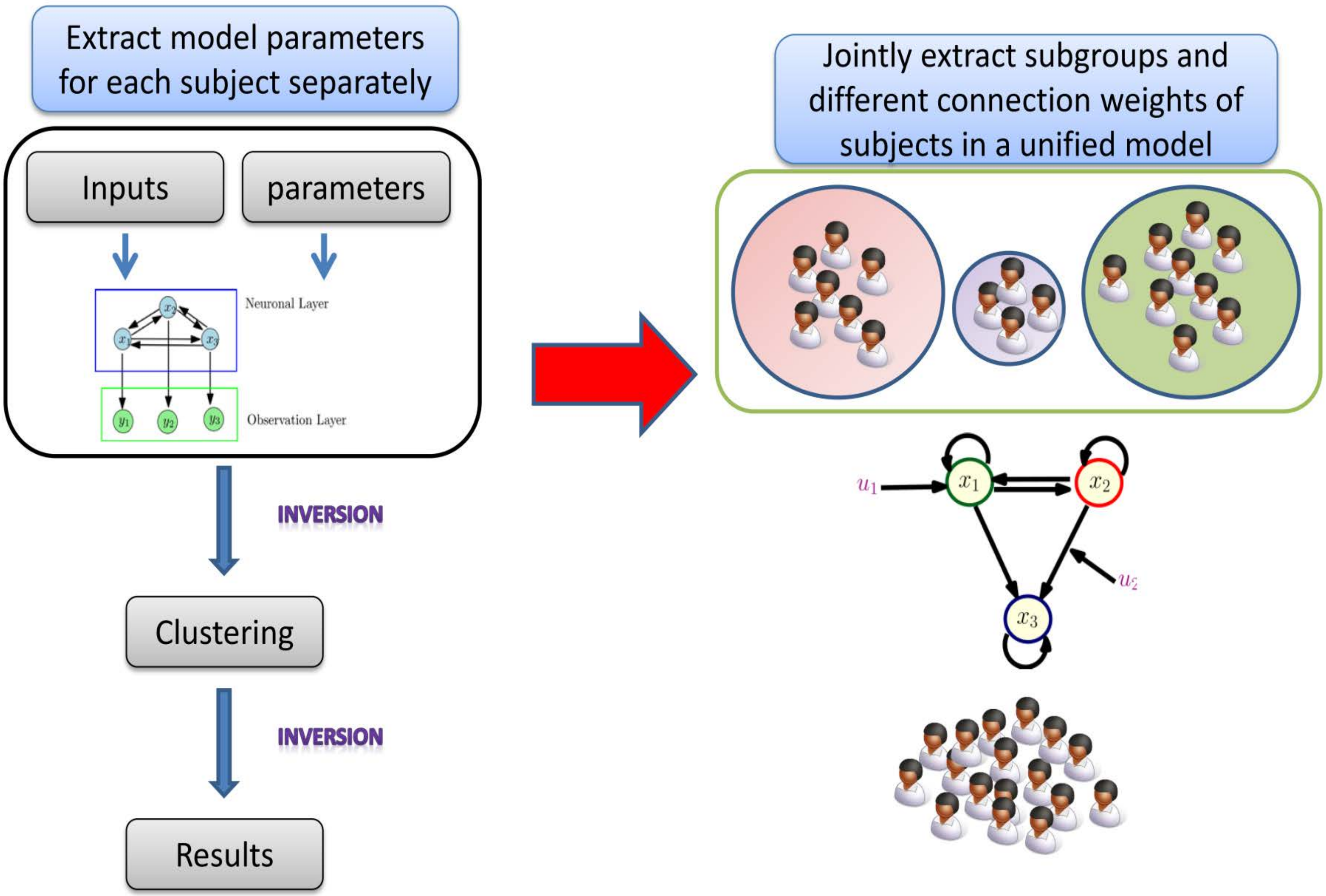
Brodersen et al 2014 Neuroimage

Results – Schizophrenia patients



Brodersen et al 2014 Neuroimage

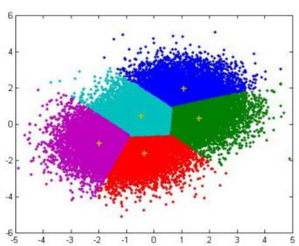
Unified model for identifying subgroups

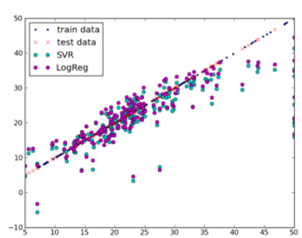


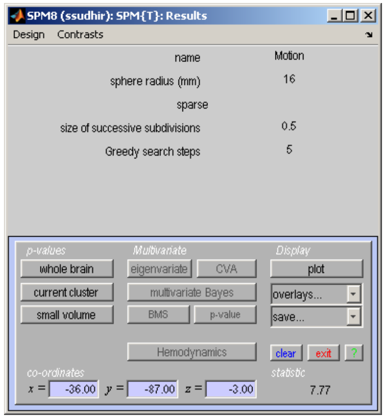
Raman et al A hierarchical model for integrating unsupervised generative embedding and empirical Bayes .Submitted

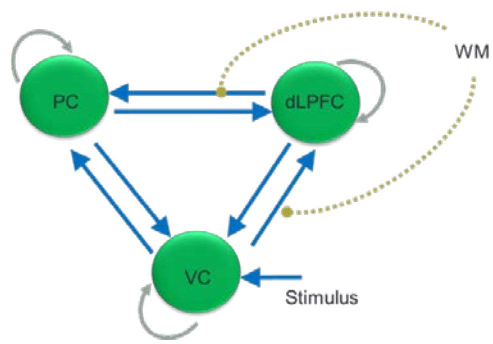
Summary

- Modelling Principles
- Learning from Data
- Multivariate Bayes in SPM
- Generative Embedding









Thank you