Multivariate models of inter-subject anatomical variability

John Ashburner

Wellcome Trust Centre for Neuroimaging,
UCL Institute of Neurology,
12 Queen Square,
London WC1N 3BG,
UK.

"The only relevant test of the validity of a hypothesis is comparison of prediction with experience."

Milton Friedman

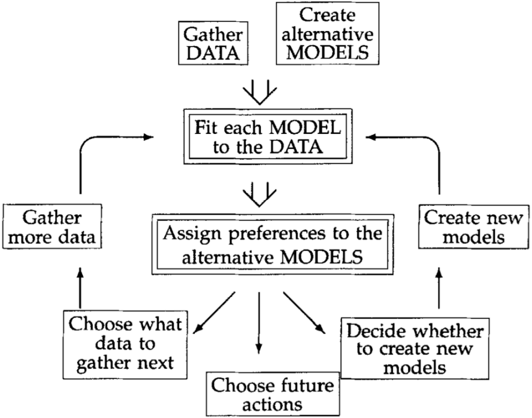
Choosing models/hypotheses/theories

David J. C. MacKay

Professor

David John Cameron MacKay, FRS FInstP FICE, is the Regius Professor of Engineering in the Department of Engineering at the University of Cambridge and chief scientific adviser to the UK Department of Energy and Climate Change. Wikipedia

MacKay, DJC. "Bayesian interpolation." Neural computation 4, no. 3 (1992): 415-447.



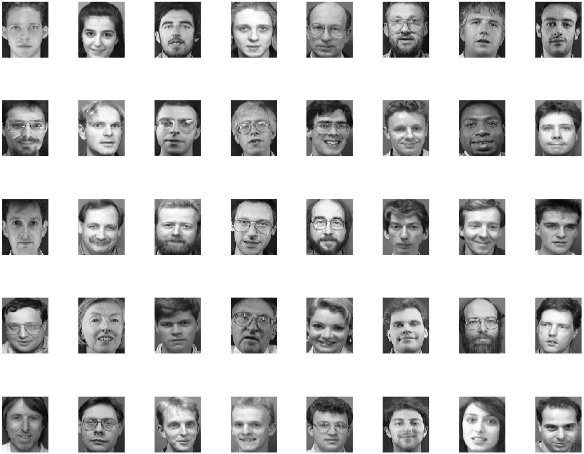
Evidence-based Science

...also just known as "science".

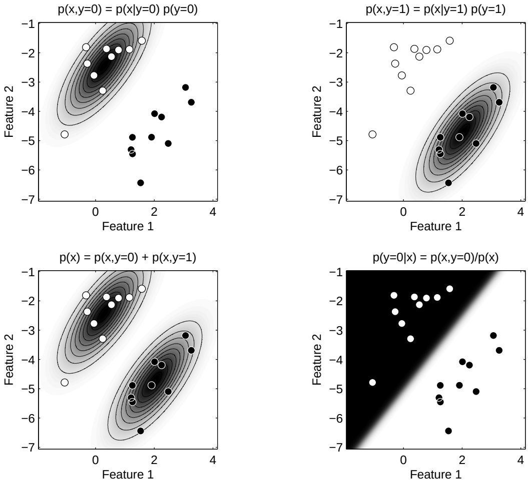
- Researchers claim to find differences between groups. Do those findings actually discriminate?
- How can we most accurately diagnose a disorder from image data?
- Pharma wants biomarkers. How do we most effectively identify them?
- There are lots of potential imaging biomarkers. Which are most (cost) effective?

Pattern recognition provides a framework to compare data (or preprocessing strategy) to determine the most accurate approach.

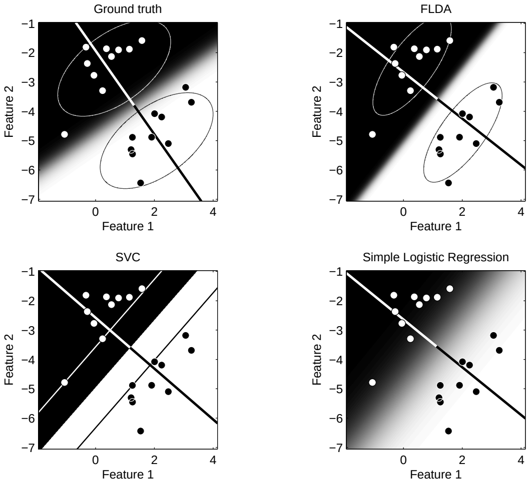
Biological variability is multivariate



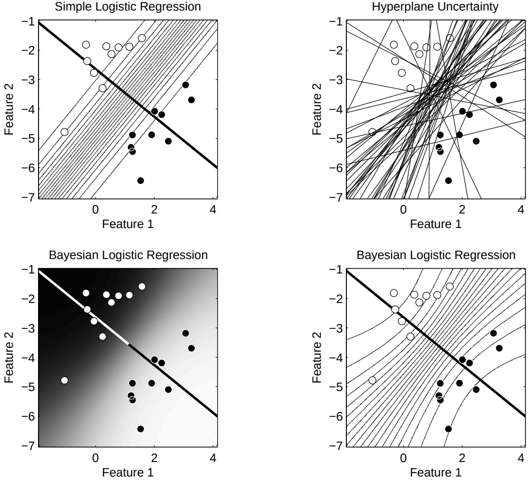
A generative classification approach



Discriminative classification approaches



Bayesian classification



Why Bayesian?

- To deal with different priors.
  - Consider a method with 90% sensitivity and specificity.
  - Consider using this to screen for a disease afflicting 1% of the population.
  - On average, out of 100 people there would be 10 wrongly assigned to the disease group.
  - A positive diagnosis suggests only about a 10% chance of having the disease.

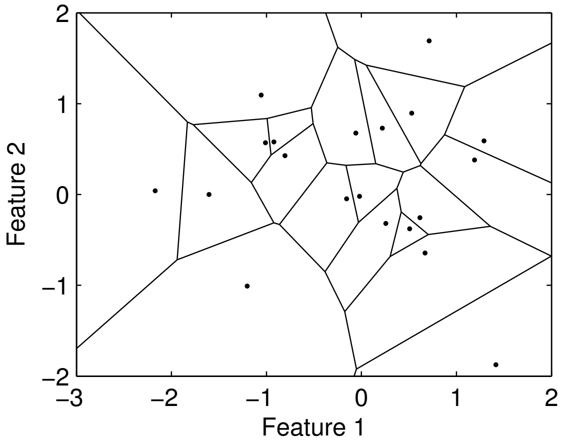
$$
\begin{aligned}
P(\text{Disease}|\text{Pred}+) &= \frac{P(\text{Pred}+|\text{Disease})P(\text{Disease})}{P(\text{Pred}+|\text{Disease})P(\text{Disease})+P(\text{Pred}+|\text{Healthy})P(\text{Healthy})} \\
&= \frac{\text{Sensitivity}\times P(\text{Disease})}{\text{Sensitivity}\times P(\text{Disease})+(1-\textit{Specificity})\times P(\text{Healthy})}
\end{aligned}
$$

- Better decision-making by accounting for utility functions.

Curse of dimensionality

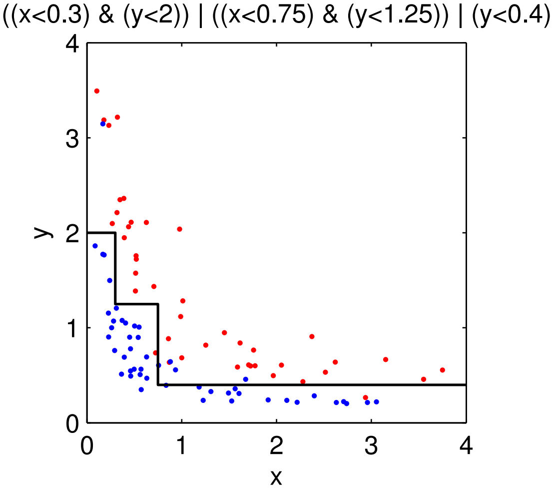
Large p, small n.

Nearest-neighbour classification



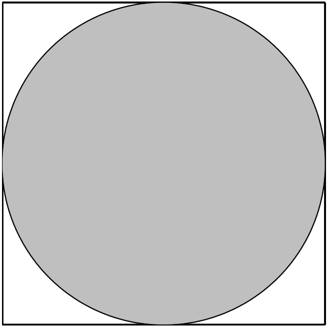
- Not nice smooth separations.
- Lots of sharp corners.
- May be improved with *K-nearest neighbours*.

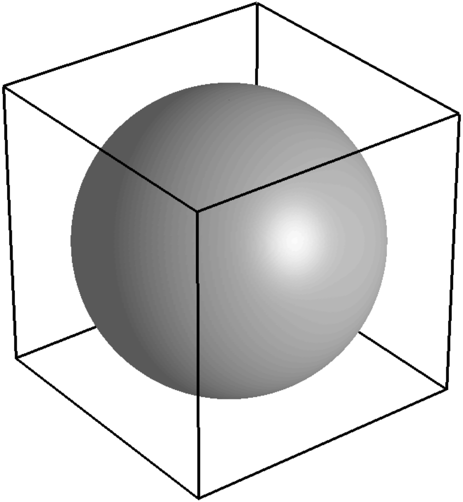
Rule-based approaches



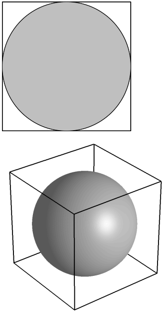
- Not nice smooth separations.
- Lots of sharp corners.

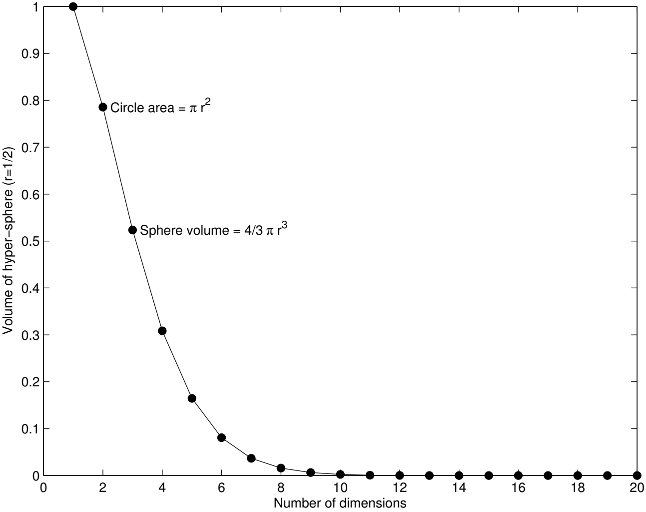
Corners matter in high-dimensions





Corners matter in high-dimensions





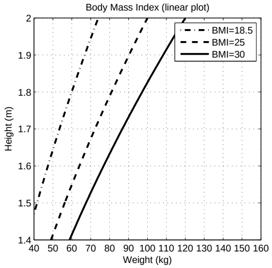
Dimensionality $\neq$ number of voxels

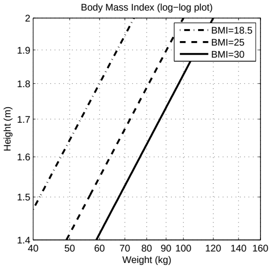
- Little evidence to suggest that most voxel-based feature selection methods help.
  - Little or no increase in predictive accuracy.
  - Commonly perceived as being more "interpretable".
- Prior knowledge derived from independent data is the most reliable way to improve accuracy.
  - e.g. search the literature for clues about which regions to weight more heavily.

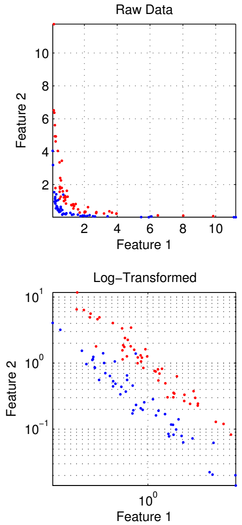
Cuingnet, Rémi, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, and Olivier Colliot. "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database." Neuroimage 56, no. 2 (2011): 766-781.
Chu, Carlton, Ai-Ling Hsu, Kun-Hsien Chou, Peter Bandettini, and ChingPo Lin. "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images." Neuroimage 60, no. 1 (2012): 59-70.
See winning strategies in `http://www.ebc.pitt.edu/PBAIC.html`

Linear versus Nonlinear methods

- Linear methods are more interpretable.
- Nonlinear methods usually increase dimensionality.
- Better to preprocess to obtain features that behave more linearly.

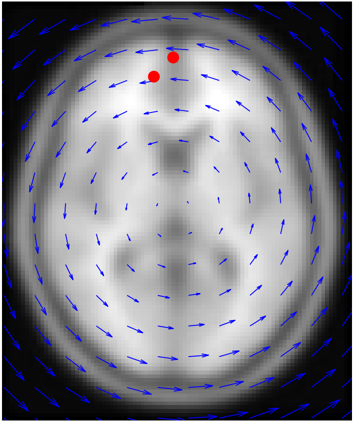


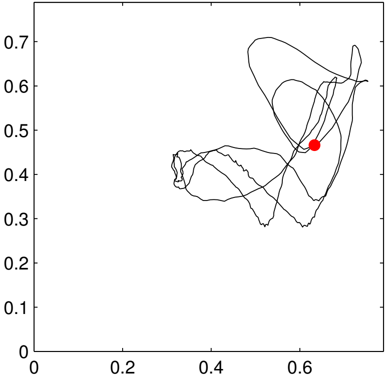




Transformed images fall on manifolds

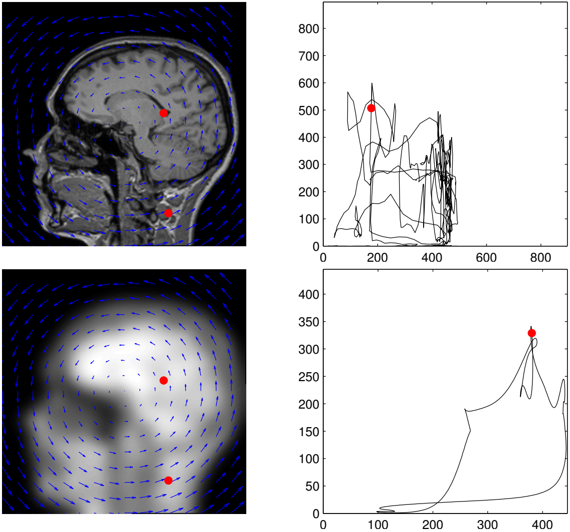
Rotating an image leads to points on a 1D manifold.





Rigid-body motion leads to a 6-dimensional manifold (not shown).

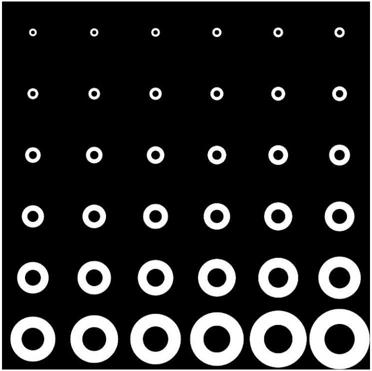
Local linearisation through smoothing



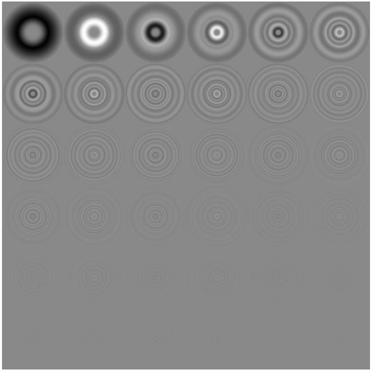
Spatial smoothing can make the manifolds more linear with respect to small misregistrations. Some information is inevitably lost.

One mode of geometric variability

Simulated images



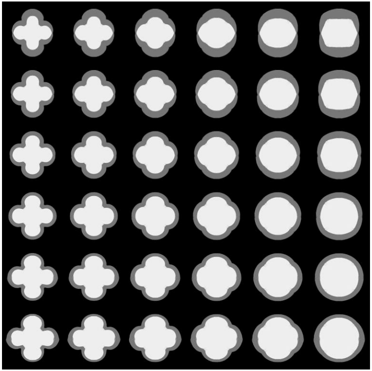
Principal components



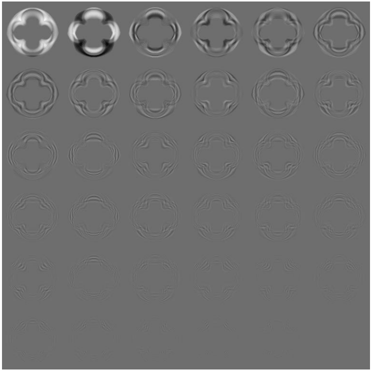
A suitable model would reduce these data to a single dimension.

Two modes of geometric variability

Simulated images



Principal components



A suitable model would reduce these data to two dimensions.

Similarity Measures

- Many methods are based on similarity measures.
- A common similarity measure is the dot product.

$$\text{Similarity: } k(\mathbf{x},\mathbf{y}) = \sum_k x_k y_k$$

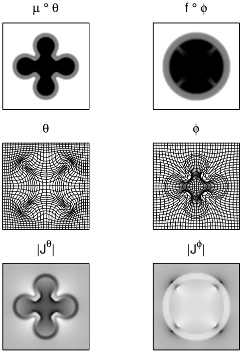
- Nonlinear methods are often based on distances.

$$\begin{aligned} \text{Distance:} \quad d(\mathbf{x},\mathbf{y}) &= \sqrt{\sum_k (x_k - y_k)^2} \\ \text{Similarity:} \quad k(\mathbf{x},\mathbf{y}) &= exp(-\lambda d(\mathbf{x},\mathbf{y})^2) \end{aligned}$$

- How do we best measure distances between brain images?

Image Registration

- Image registration measures distances between images.
- Often involves minimising the sum of two terms:
  - Distance between the image intensities.
  - Distance of the deformation from zero.
- The sum of these terms gives the distance.



Different ways of measuring distances

David Mumford

Mathematician

David Bryant Mumford is an American mathematician known for distinguished work in algebraic geometry, and then for research into vision and pattern theory. He won the Fields Medal and was a MacArthur Fellow. Wikipedia

Born: June 11, 1937 (age 76), Worth village, West Sussex, Crawley

Children: Steve Mumford

Education: Phillips Exeter Academy, Harvard University

Awards: Fields Medal, Wolf Prize in Mathematics, MacArthur Fellowship, The Shaw Prize in Mathematical Sciences, National Medal of Science for Mathematics and Computer Science

Empirical Statistics and Stochastic Models for Visual Signals

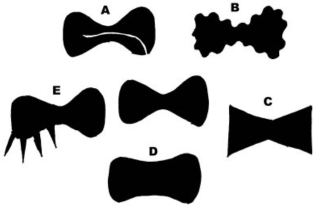
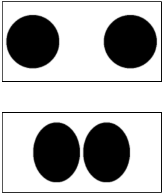
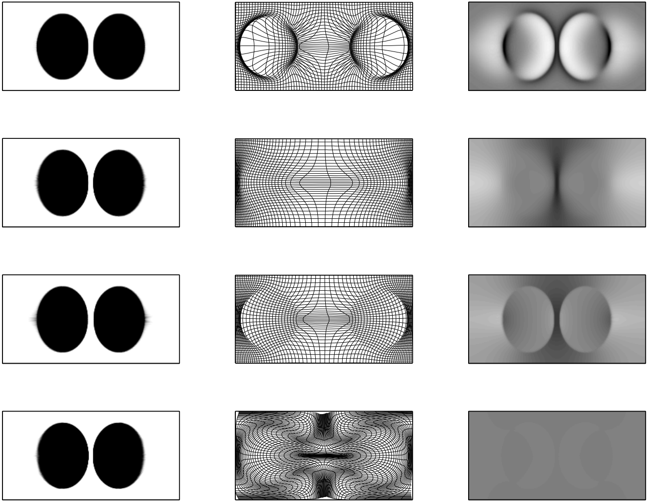


Figure 1.11 Each of the shapes A,B,C,D and E is similar to the central shape, but *in different ways*. Different metrics on the space of shape bring out these distinctions.

Different ways of measuring distances

Two simulated images





Metrics

Distances need to satisfy the properties of a *metric*:

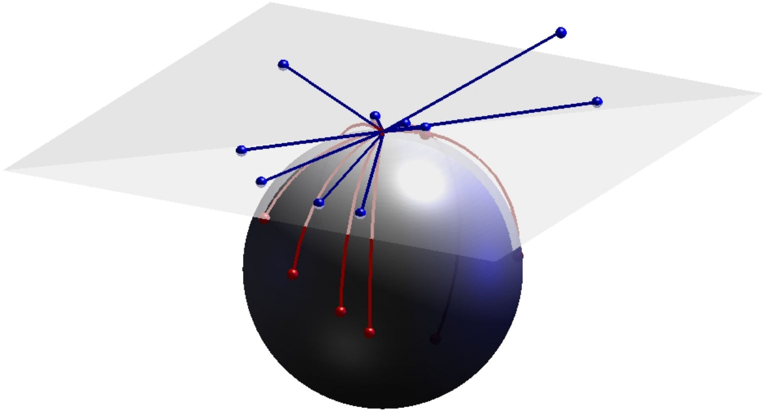
1. $d(\mathbf{x},\mathbf{y}) \geq 0$ (non-negativity)
2. $d(\mathbf{x},\mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$ (identity of indiscernibles)
3. $d(\mathbf{x},\mathbf{y}) = d(\mathbf{y},\mathbf{x})$ (symmetry)
4. $d(\mathbf{x},\mathbf{z}) \leq d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})$ (triangle inequality).

Satisfying (3) requires inverse-consistent image registration.
Satisfying (4) requires a specific family of image registration algorithm.

Non-Euclidean geometry

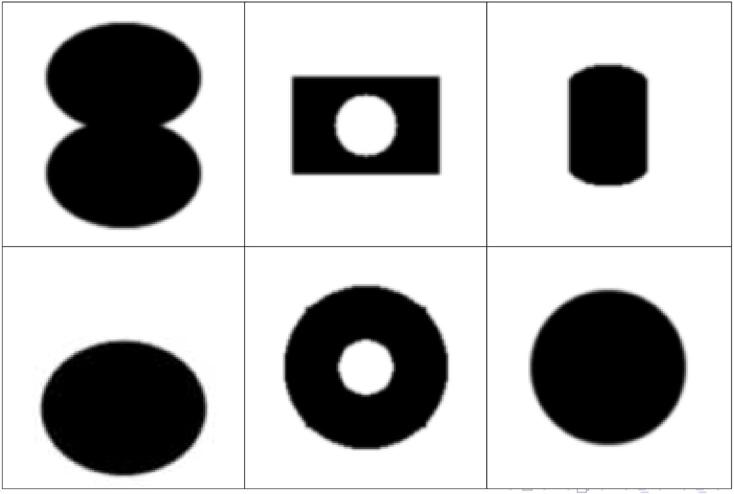
- Distances are not always measured along a straight line.
- *“Shapes are the ultimate non-linear sort of thing”*

Linear approximations to nonlinear problems



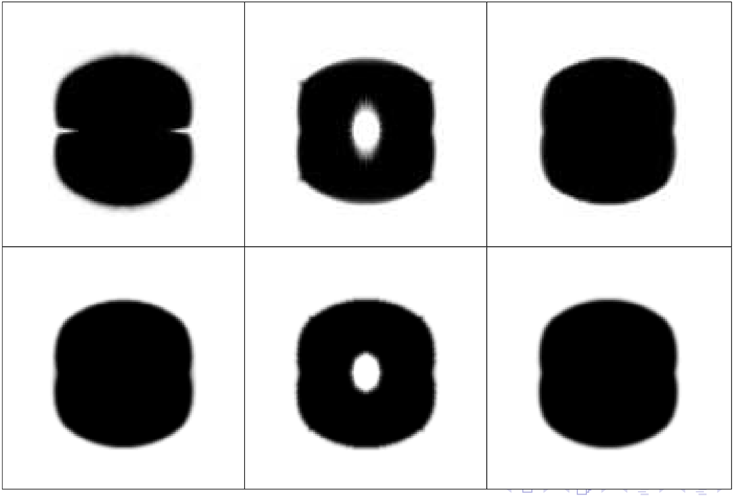
Example Images

Some example (non-brain) images.



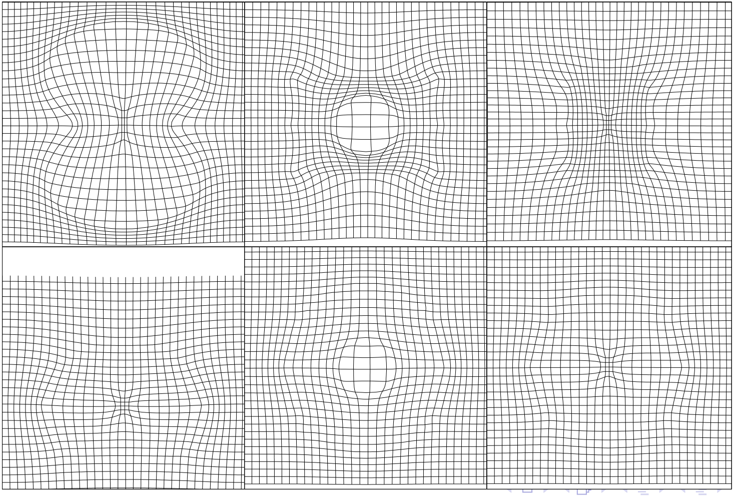
Registered Images

We could register the images to their average shape...



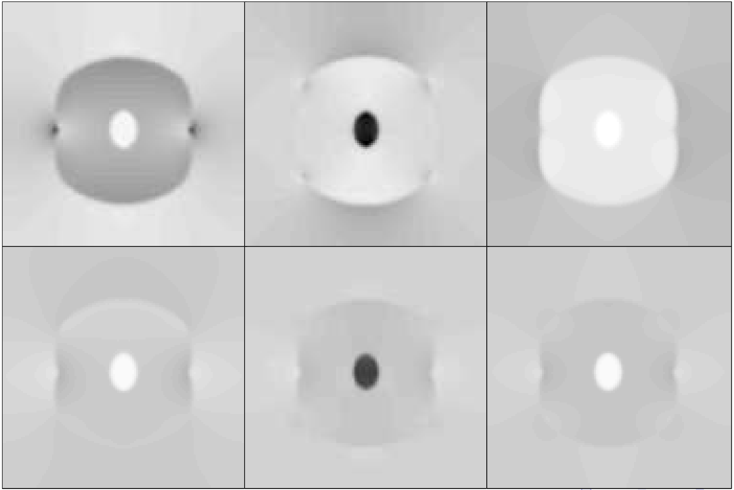
Deformations

...and study the deformations...



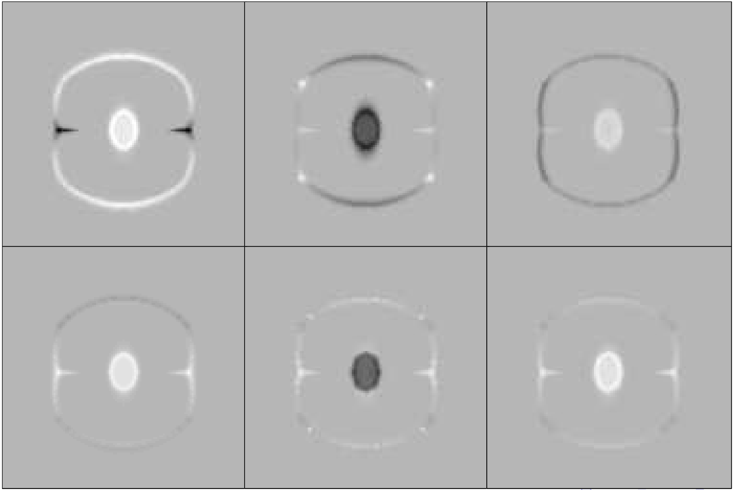
Jacobian Determinants

...or the relative volumes...



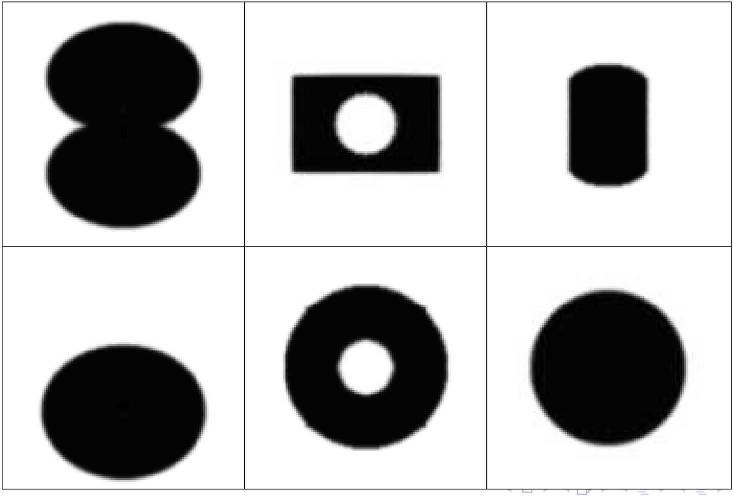
Scalar Momentum

... or "scalar momentum"



Reconstructed Images

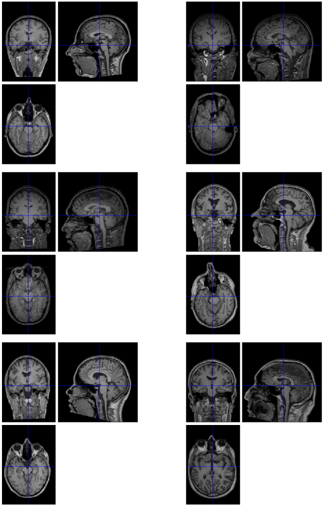
Reconstructions from template and scalar momenta.



Real data

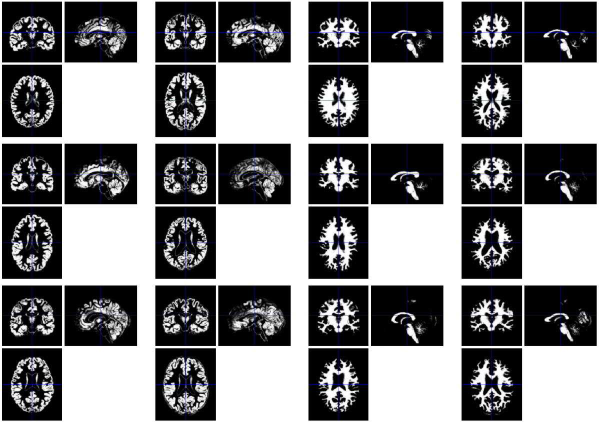
Used 550 T1w brain MRI from IXI (**I**nformation e**X**traction from **I**mages) dataset.
`http://www.brain-development.org/`
Data from three different hospitals in London:

- Hammersmith Hospital using a Philips 3T system
- Guy's Hospital using a Philips 1.5T system
- Institute of Psychiatry using a GE 1.5T system



Grey and White Matter

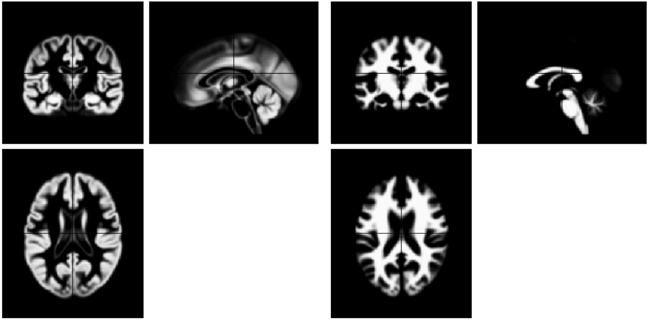
Segmented into GM and WM. Approximately aligned via rigid-body.



Ashburner, J & Friston, KJ. *Unified segmentation*. NeuroImage 26(3):839–851 (2005).

Diffeomorphic Alignment

All GM and WM were diffeomorphically aligned to their common average-shaped template.



Ashburner, J & Friston, KJ. *Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation*. NeuroImage 55(3):954–967 (2011).
Ashburner, J & Friston, KJ. *Computing average shaped tissue probability templates*. NeuroImage 45(2):333–341 (2009).

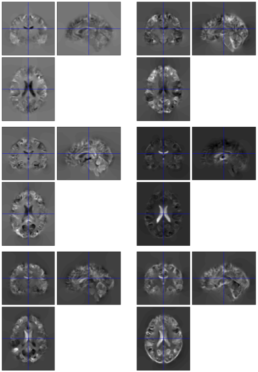
Volumetric Features

A number of features were used for pattern recognition.

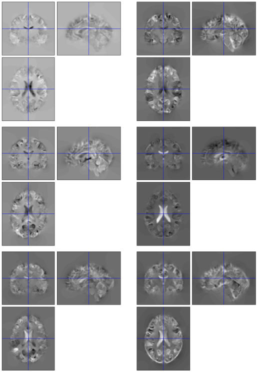
Firstly, two features relating to relative volumes.

Initial velocity divergence is similar to logarithms of Jacobian determinants.

Jacobian Determinants

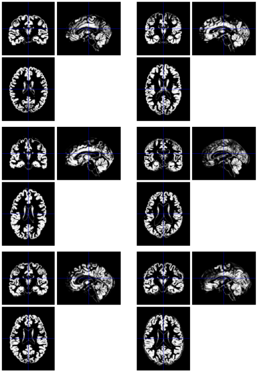


Initial Velocity Divergence

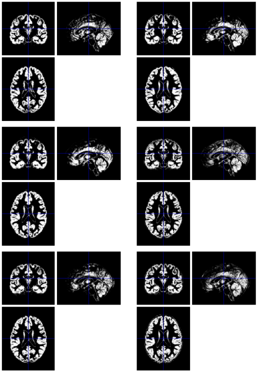


Grey Matter Features

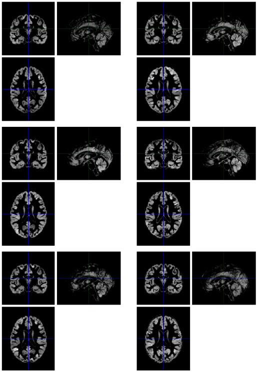
Rigidly Registered GM



Nonlinearly Registered GM



Registered and Jacobian Scaled GM

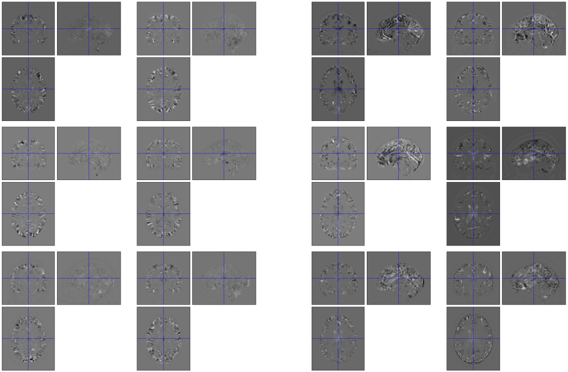


"Scalar Momentum" Features

"Scalar momentum" actually has two components because GM was matched with GM and WM was matched with WM.

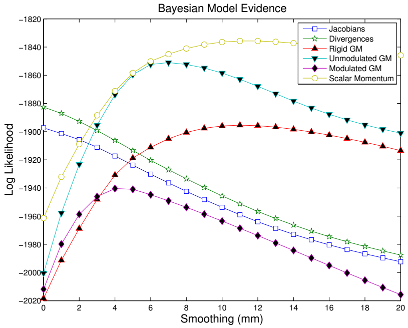
First Momentum Component

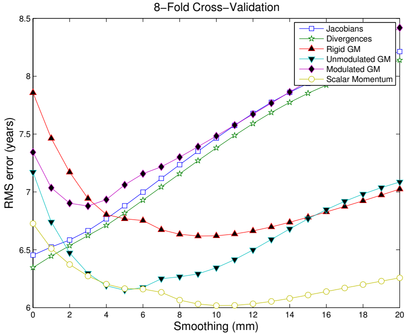
Second Momentum Component



Age Regression

Linear Gaussian Process Regression to predict subject ages.

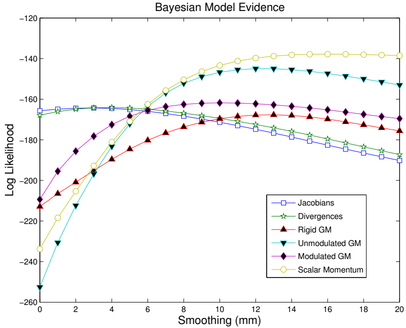


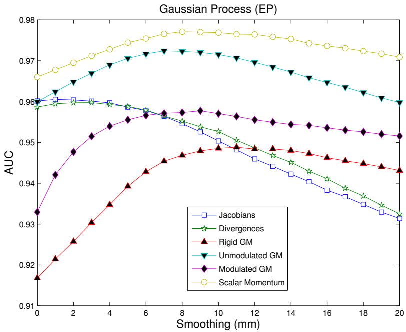


Rasmussen, CE & Williams, CKI. *Gaussian processes for machine learning*. Springer (2006).

Sex Classification

Linear Gaussian Process Classification (EP) to predict sexes.

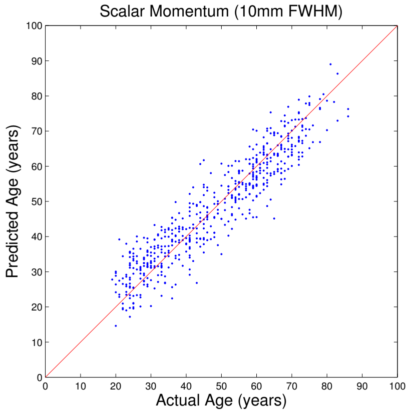




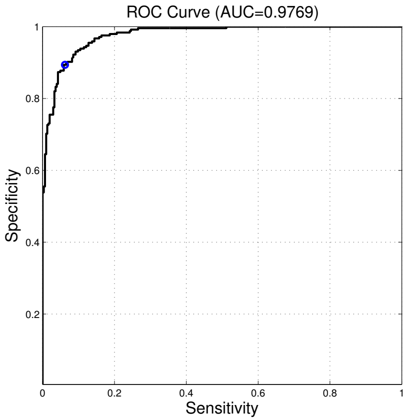
Rasmussen, CE & Williams, CKI. *Gaussian processes for machine learning*. Springer (2006).

Predictive Accuracies

Age



Sex



Conclusions

- Scalar momentum (with about 10mm smoothing) appears to be a useful feature set.
- Jacobian-scaled warped GM is surprisingly poor.
- Amount of spatial smoothing makes a big difference.
- Further dependencies on the details of the registration still need exploring.