Multivariate Analysis

John Ashburner.
Wellcome Trust Centre for Neuroimaging,
UCL Institute of Neurology,
London, UK.
john@fil.ion.ucl.ac.uk

Overview

- **Multivariate Approaches**
- Scalar Momentum
- Some Evaluations

Mass-univariate analyses

- Tests hypotheses at each voxel.
  - Not a test of a single hypothesis.
- A hybrid between:
  - Exploratory analysis
  - Hypothesis testing
- No clear separation between exploratory data analysis and hypothesis testing.

Multivariate approaches

- A single hypothesis test for the entire brain.
- Multivariate approaches include:
  - Hotelling's T^2 tests.
    - Essentially just an F test.
  - MANCOVA/ANCOVA tests.
    - Wilk's lambda statistic
  - Pattern recognition.

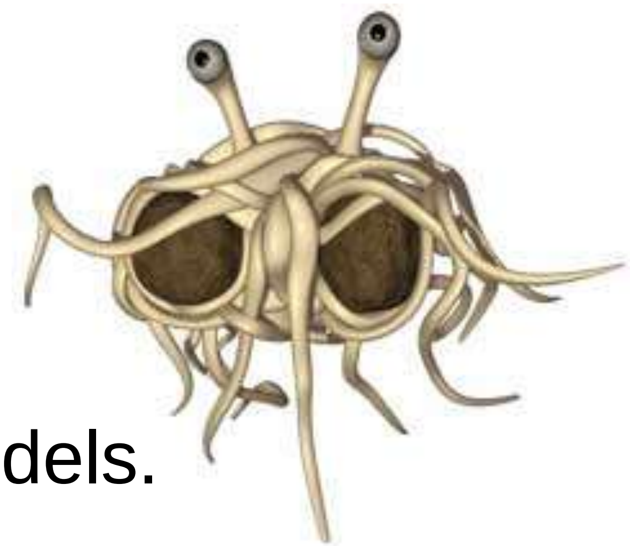
Pattern Recognition

- Training: analogous to exploratory analysis.
- Testing: analogous to hypothesis testing.

- Clear separation between exploratory analysis and hypothesis testing.

Hypothesis Testing

- Roughly, hypothesis testing involves comparing two models, to determine which models the probability of the data more accurately.
  – $p(\mathbf{Y}|M_0)$ or $p(\mathbf{Y}|M_1)$



- Why only compare two models?
- An infinite number of possible models.

Model Selection

- Search for the best of a number of models:

  $p(\mathbf{Y}|M_0)$, $p(\mathbf{Y}|M_1)$, $p(\mathbf{Y}|M_2)$, $p(\mathbf{Y}|M_3)$, $p(\mathbf{Y}|M_4)$…

- Cross-validation is essentially hypothesis testing.
  - Learn a hypothesis/model from the training data.
  - Test it on the data that was left out.

- Other model selection strategies are also possible – eg Bayesian Model Selection.

- The complexity of the best model depends on how much data is available.
  - Brains are a bit complicated

Multivariate models of form

- In theory, assumptions about structural covariance among brain regions are more biologically plausible.
  Form determined (in part) by spatio-temporal modes of gene expression.
- Empirical evidence in (eg)
  Mechelli, Friston, Frackowiak & Price. *Structural covariance in the human cortex*. Journal of Neuroscience 25(36):8303-8310 (2005).

- We should work with the most accurate modelling assumptions available.
  - If a model is accurate, it will make accurate predictions.

Generative Model for Discrimination

- Generative:

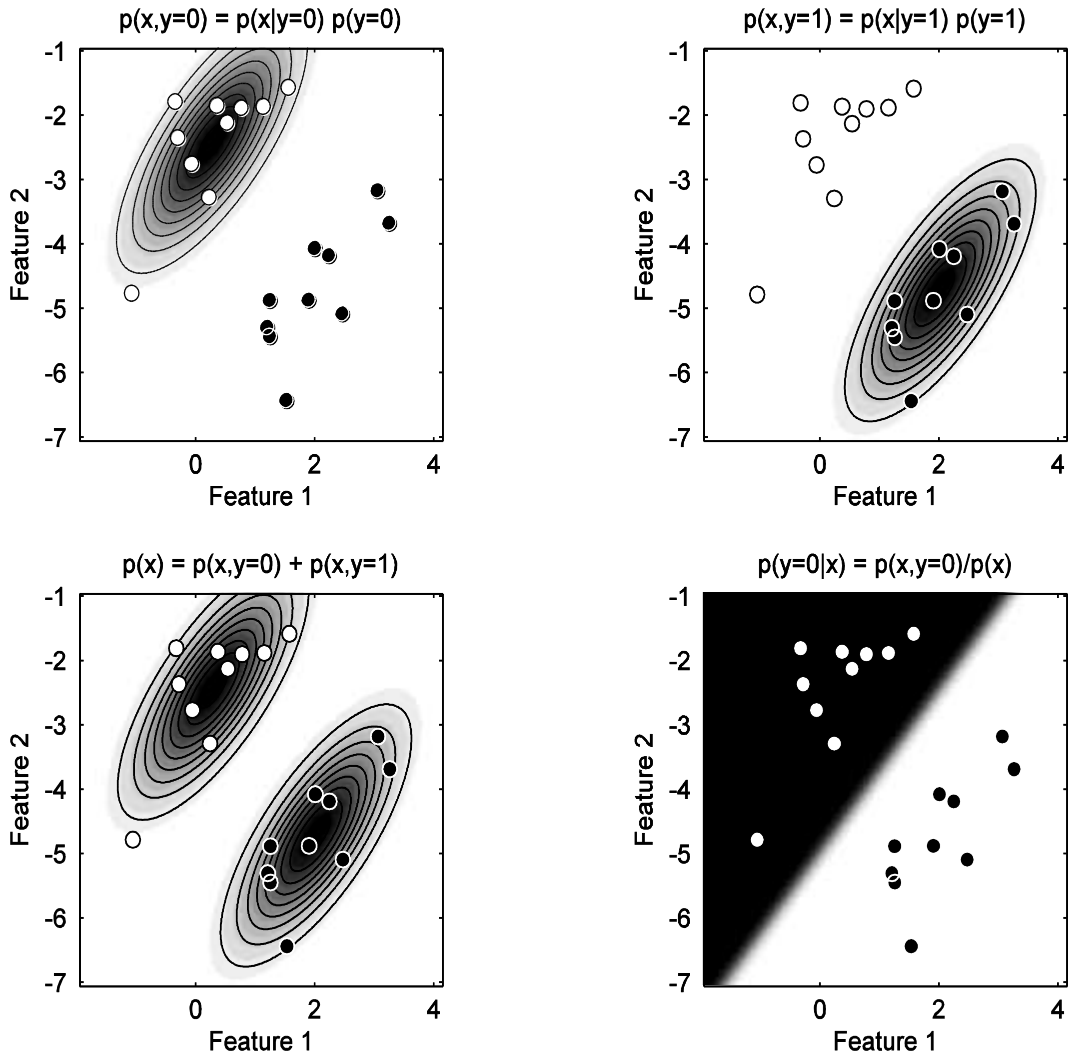
$$P(t=1|\mathbf{x}) = \frac{p(\mathbf{x}|t=1)P(t=1)}{p(\mathbf{x}|t=0)P(t=0) + p(\mathbf{x}|t=1)P(t=1)}$$

Where	$\mathbf{x}$	feature data
	t	prediction

- Discriminative:
  - Directly learns to give $P(t=1|\mathbf{x})$
  - We are not normally interested in all the variables needed to represent within-group variability.
  - Only after a discriminative direction.

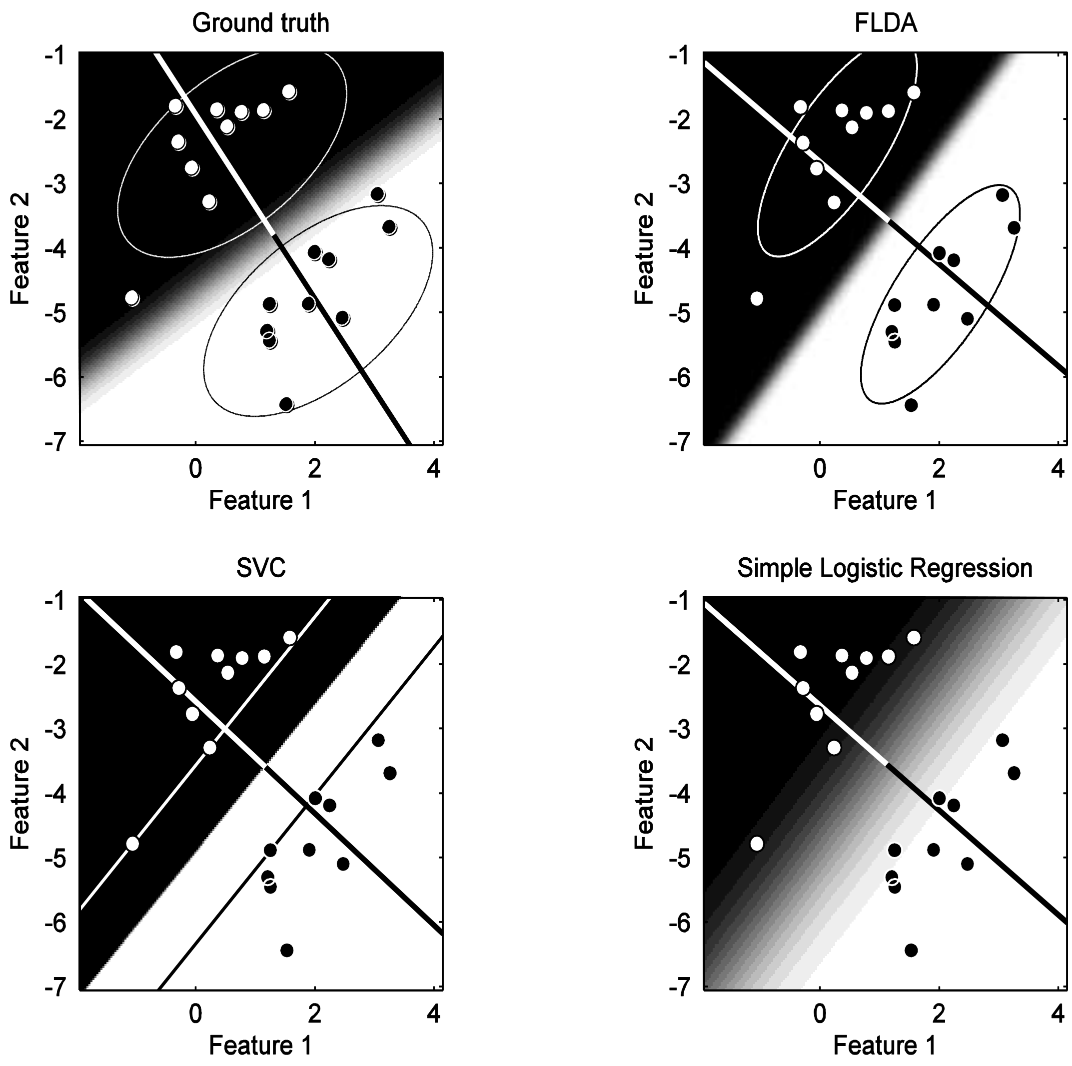
Fisher's Linear Discriminant Analysis

- A multivariate model.
- Special case of canonical variates analysis.
- A **generative model**.

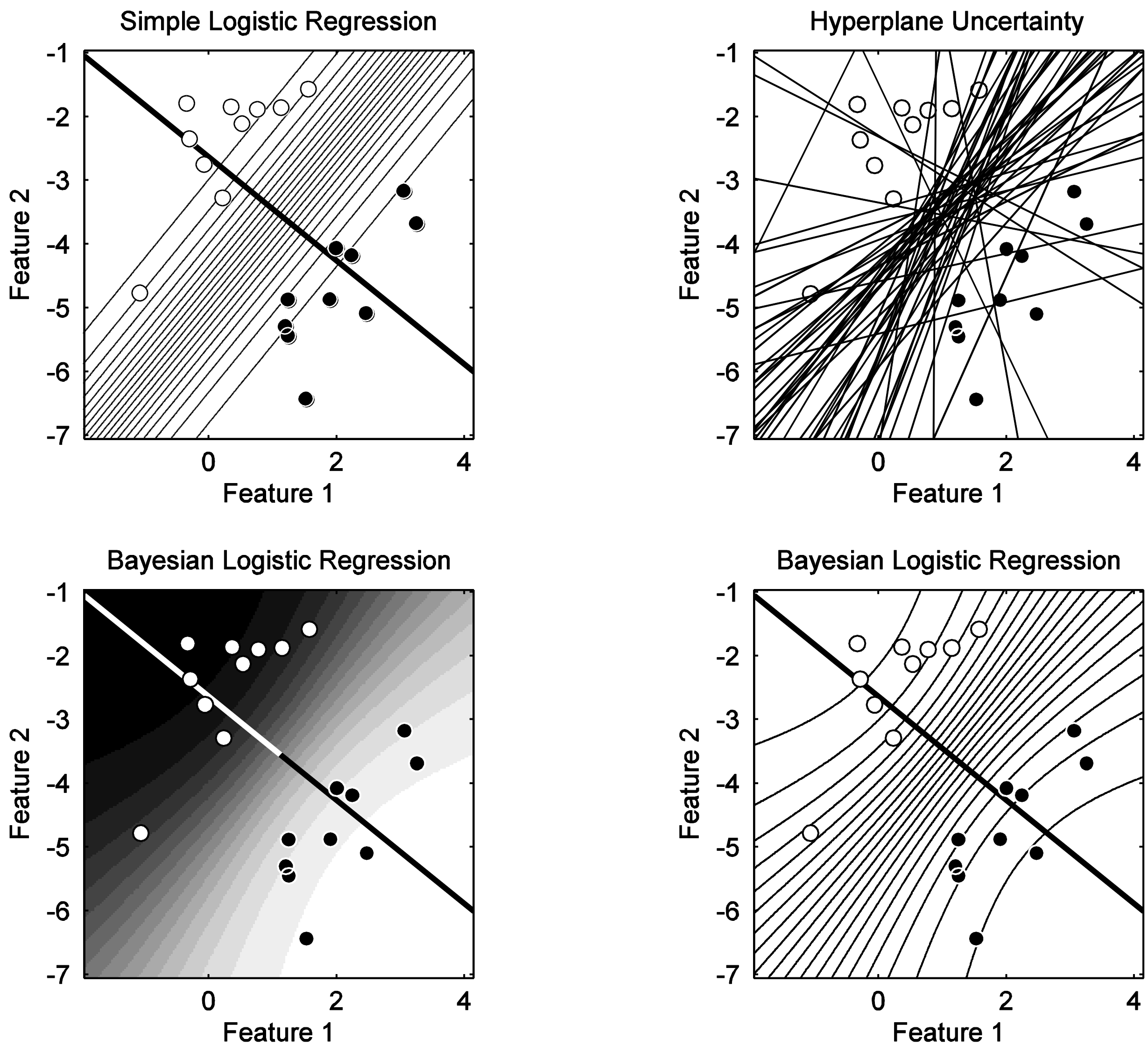


Other linear discrimination approaches

- Can also use **discriminative models**.
- Anatomical differences are encoded by the vector orthogonal to the separating hyper-plane.
- The most accurate model of difference is the one that best

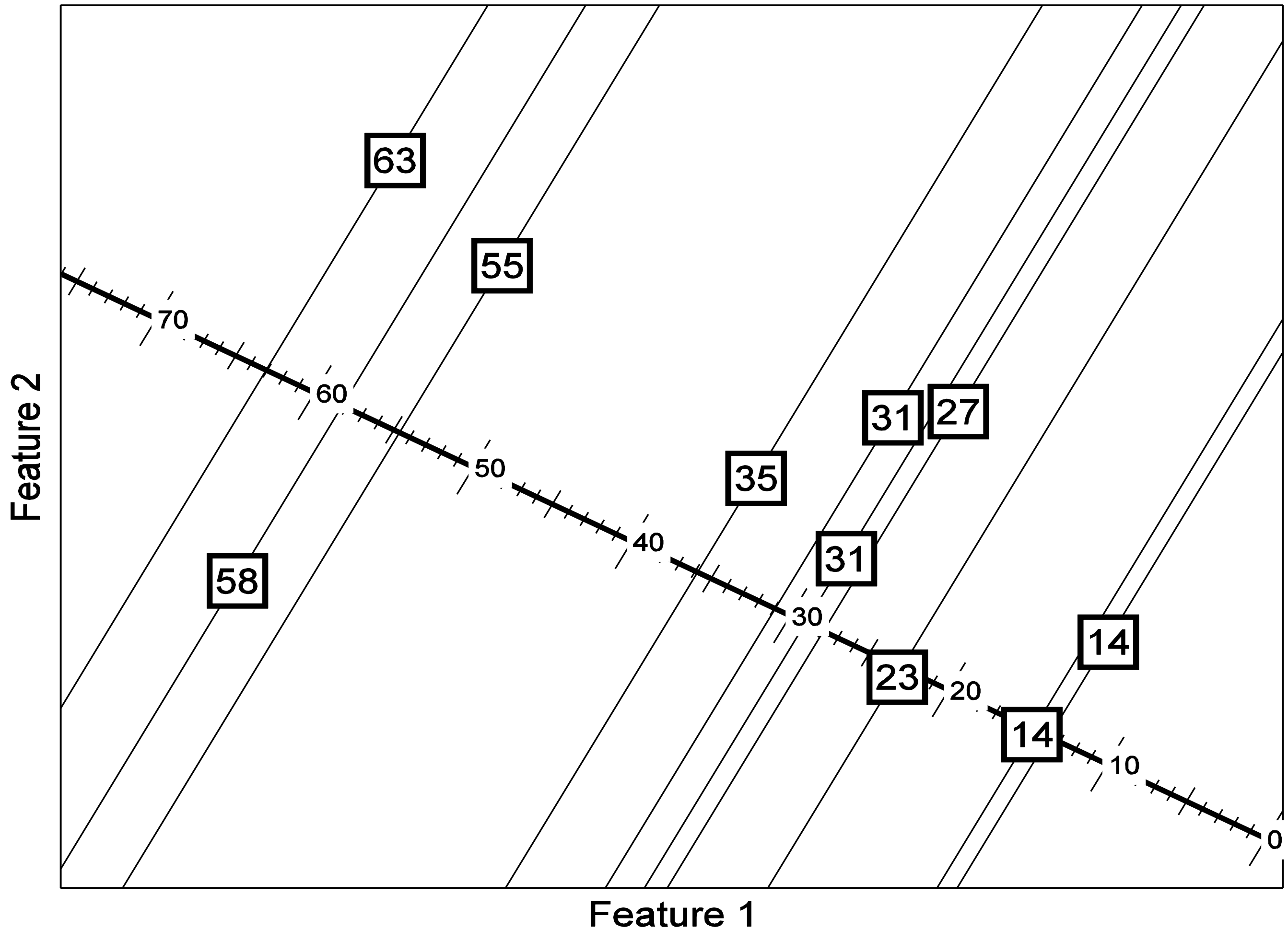


Probabilistic Approaches

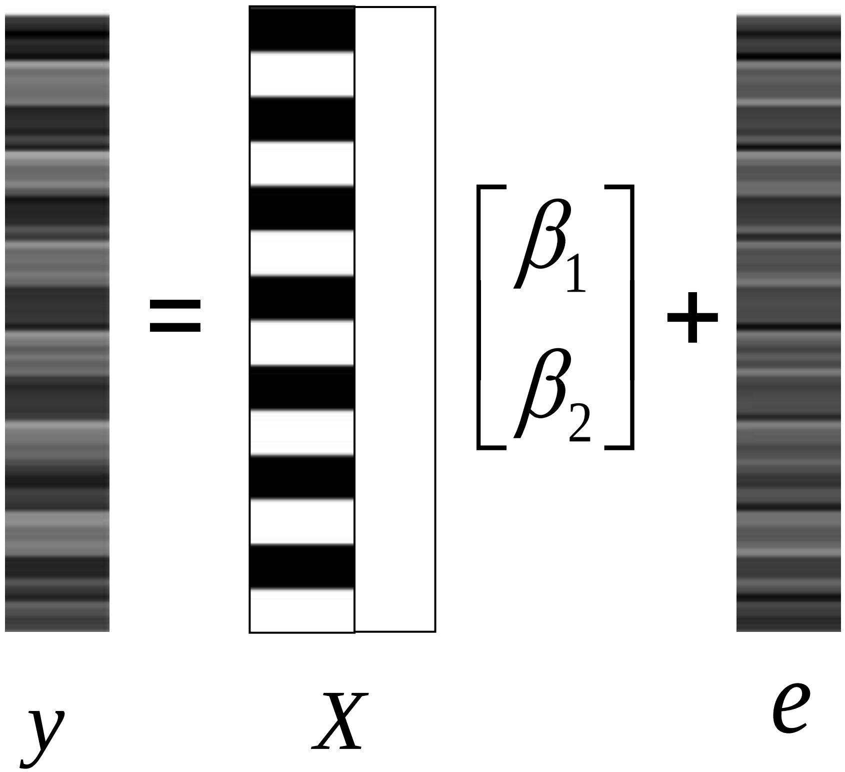


Regression

- For predicting a continuous variable



Regression



$$y = X\beta + e$$

Objective: estimate parameters to minimize

$$\sum_{t=1}^{N} e_t^2$$

Ordinary least squares estimation (OLS) (assuming i.i.d. error):

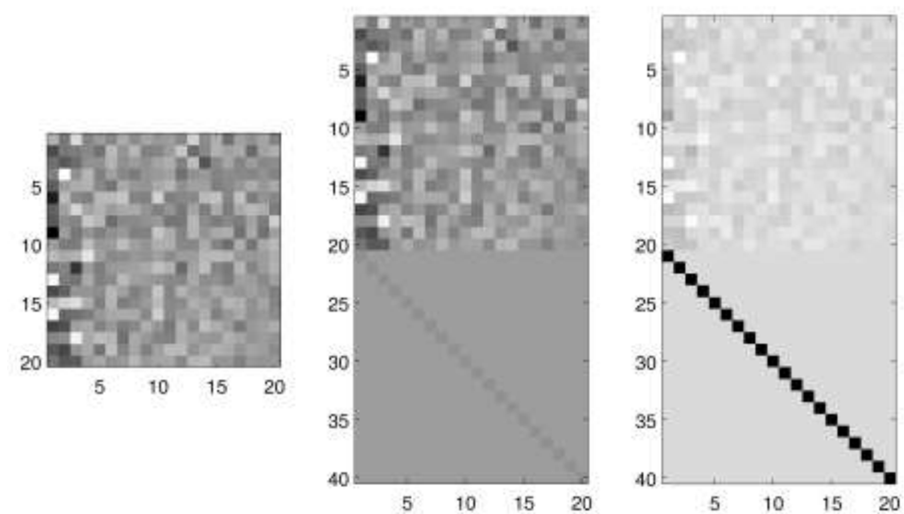
$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Curse of Dimensionality

- More voxels in an image than images in a study

  $\mathbf{t} = \mathbf{X}\mathbf{w} + \mathbf{e}$
- System is under-determined, so need to regularise
- Could use PCA, but more principled to use ridge-regression.
- Learn the regularisation parameters with REML.

- Turns out to be same as Gaussian Process model.



Gaussian Process Regression

- Estimate a covariance matrix by maximising.

  $\log p(\mathbf{t}|\boldsymbol{\theta}) = -\tfrac{1}{2}\log |\mathbf{C}(\boldsymbol{\theta})| - \tfrac{1}{2}\mathbf{t}^T\mathbf{C}(\boldsymbol{\theta})^{-1}\mathbf{t}$

  Where e.g. $\mathbf{C}(\boldsymbol{\theta}) = \theta_1\mathbf{I} + \theta_2 + \theta_3\, \mathbf{X_1}^T\mathbf{X_1} + \theta_4\, \mathbf{X_2}^T\mathbf{X_2}$

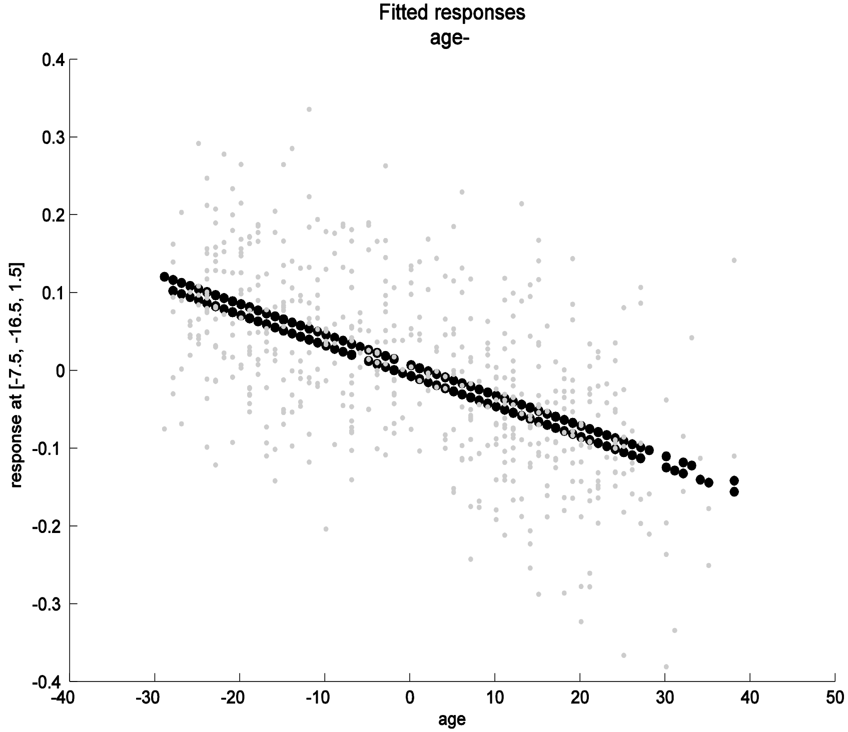
- Augment covariance matrix with data for testing:

$$\mathbf{C}_{full} = \begin{pmatrix} \mathbf{C} & \mathbf{k} \\ \mathbf{k}^T & \mathbf{c} \end{pmatrix}$$

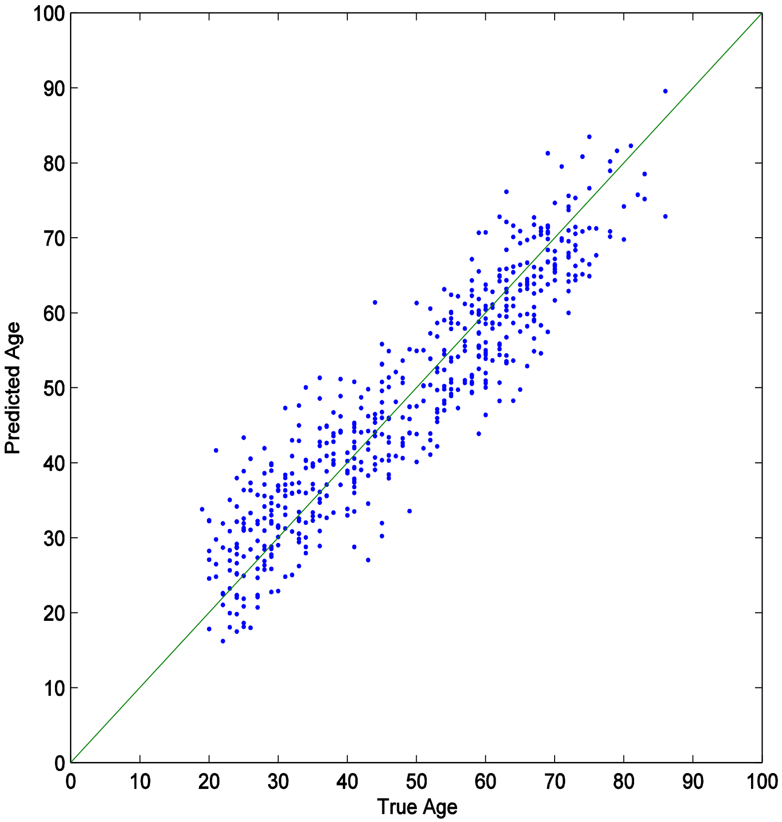
- Make inference from
  - $p(\mathbf{t}_{new}|\mathbf{t},\boldsymbol{\theta}) = N(\mathbf{k}^T\mathbf{C}^{-1}\mathbf{t},\ c - \mathbf{k}^T\mathbf{C}^{-1}\mathbf{k})$
- Classification is a bit more complicated

Predicting Age – univariate v multivariate

Single Voxel



Combining All Voxels



Weight Map

For linear classifiers, predictions are made by:

$$y = a_1 \times x_1 + a_2 \times x_2 + a_3 \times x_3 + \ldots + b$$

where: y is the prediction

x_1, x_2, x_3 etc are voxels in the image to classify

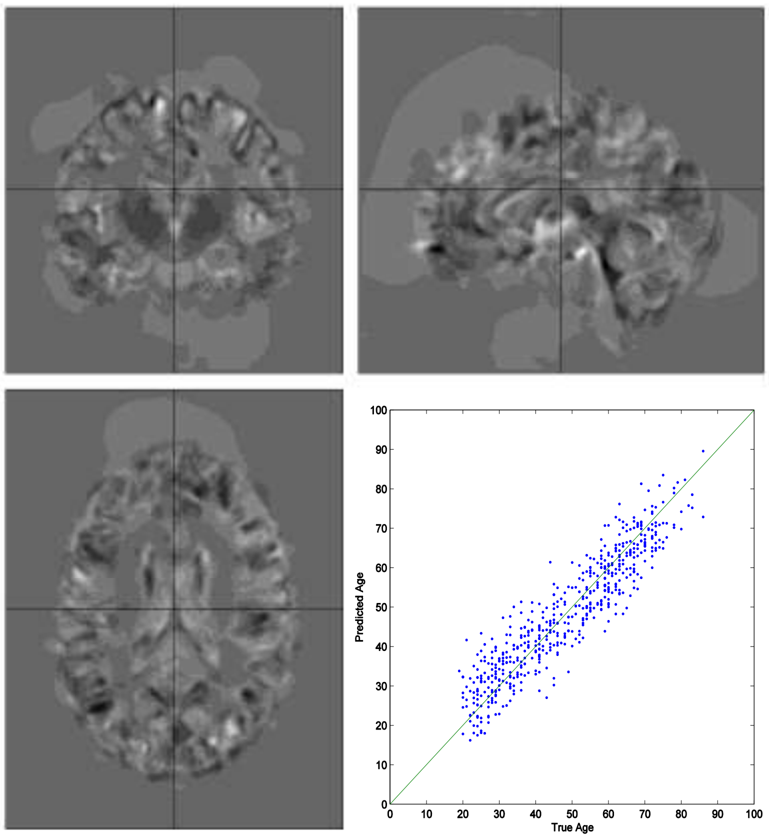
a_1, a_2, a_3 etc are voxels in a weight map

b is a constant offset.

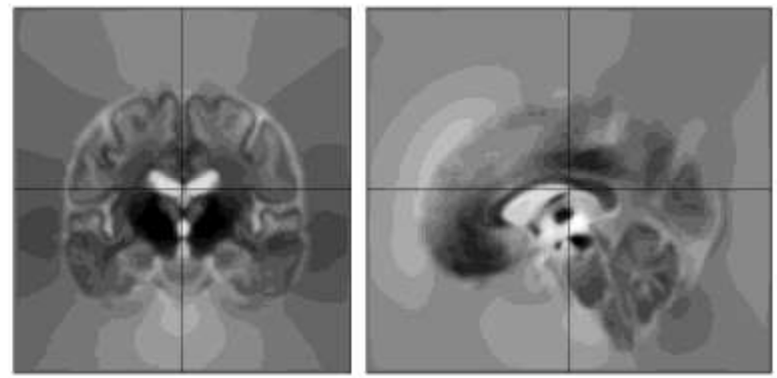
The weight map can be visualised

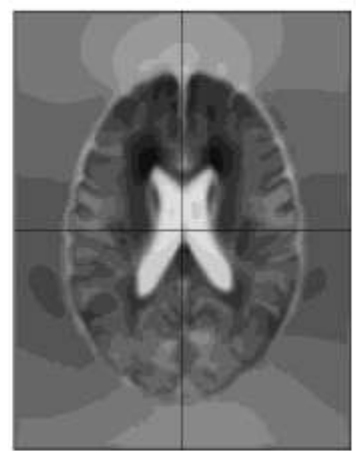
Maps

Multivariate weight map



Simple T statistic image

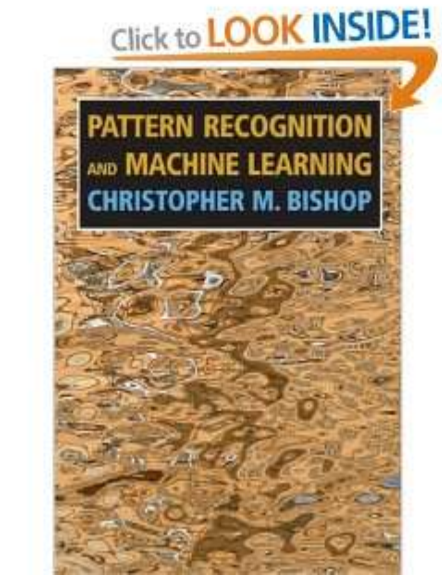


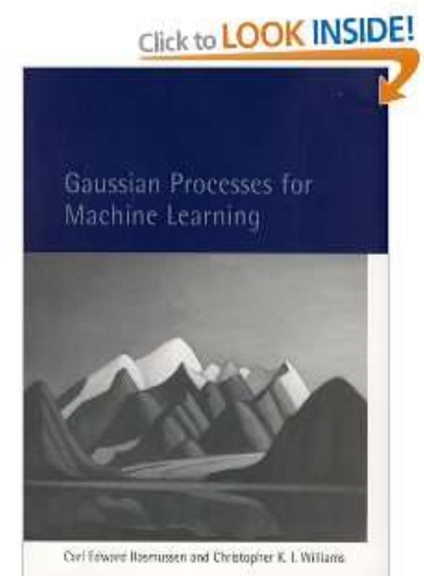


Prettier – but does not accurately characterise the effects of age.

Some References

- Bishop. *Pattern Recognition and Machine Learning*. 2006.
- Rasmussen & Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. ISBN-10 0-262-18253-X, ISBN-13 978-0-262-18253-9. http://www.gaussianprocess.org/gpml/
- Ashburner & Klöppel. "*Multivariate models of inter-subject anatomical variability*". NeuroImage 56(2):422-439, 2011.





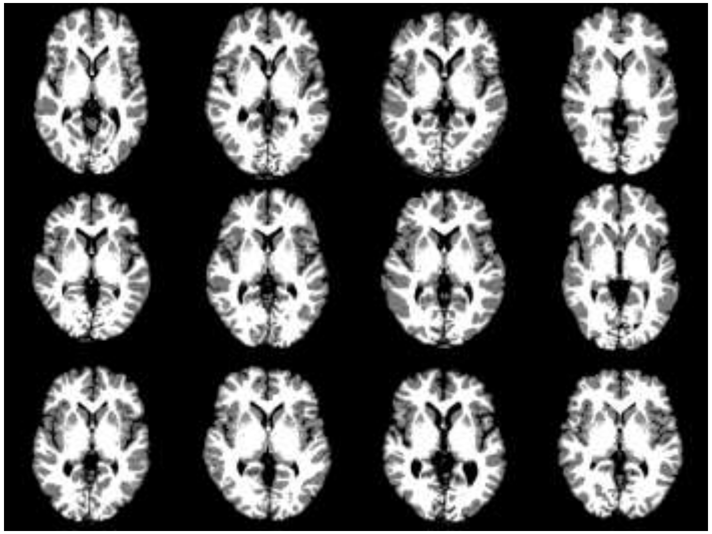
Overview

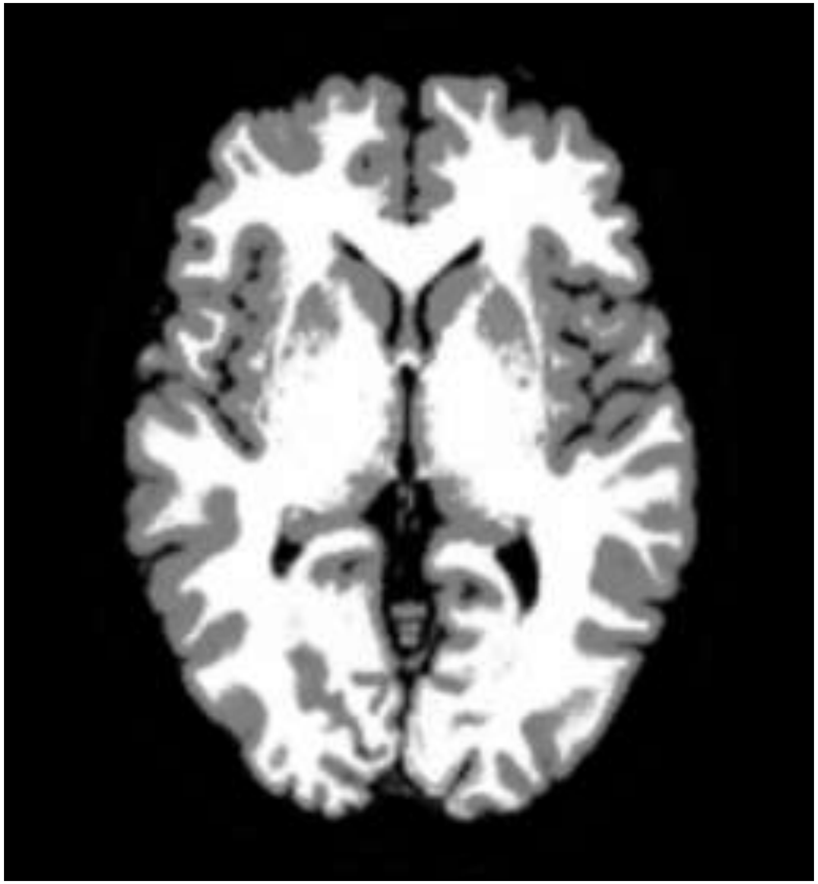
- Multivariate Approaches
- **Scalar Momentum**
- Some Evaluations

Distances

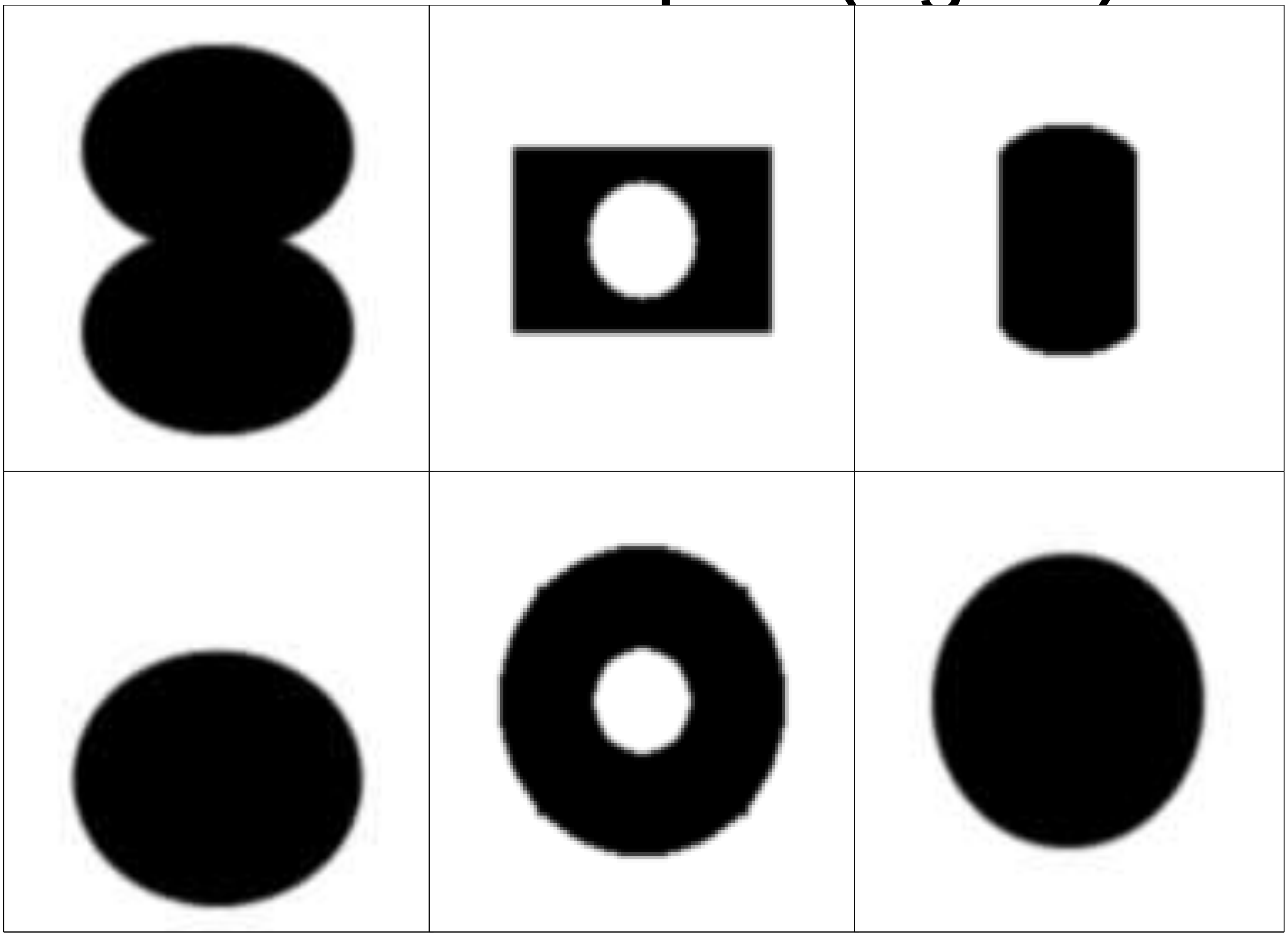
Biologically plausible measures of anatomical similarity.

Nonlinear distance measures

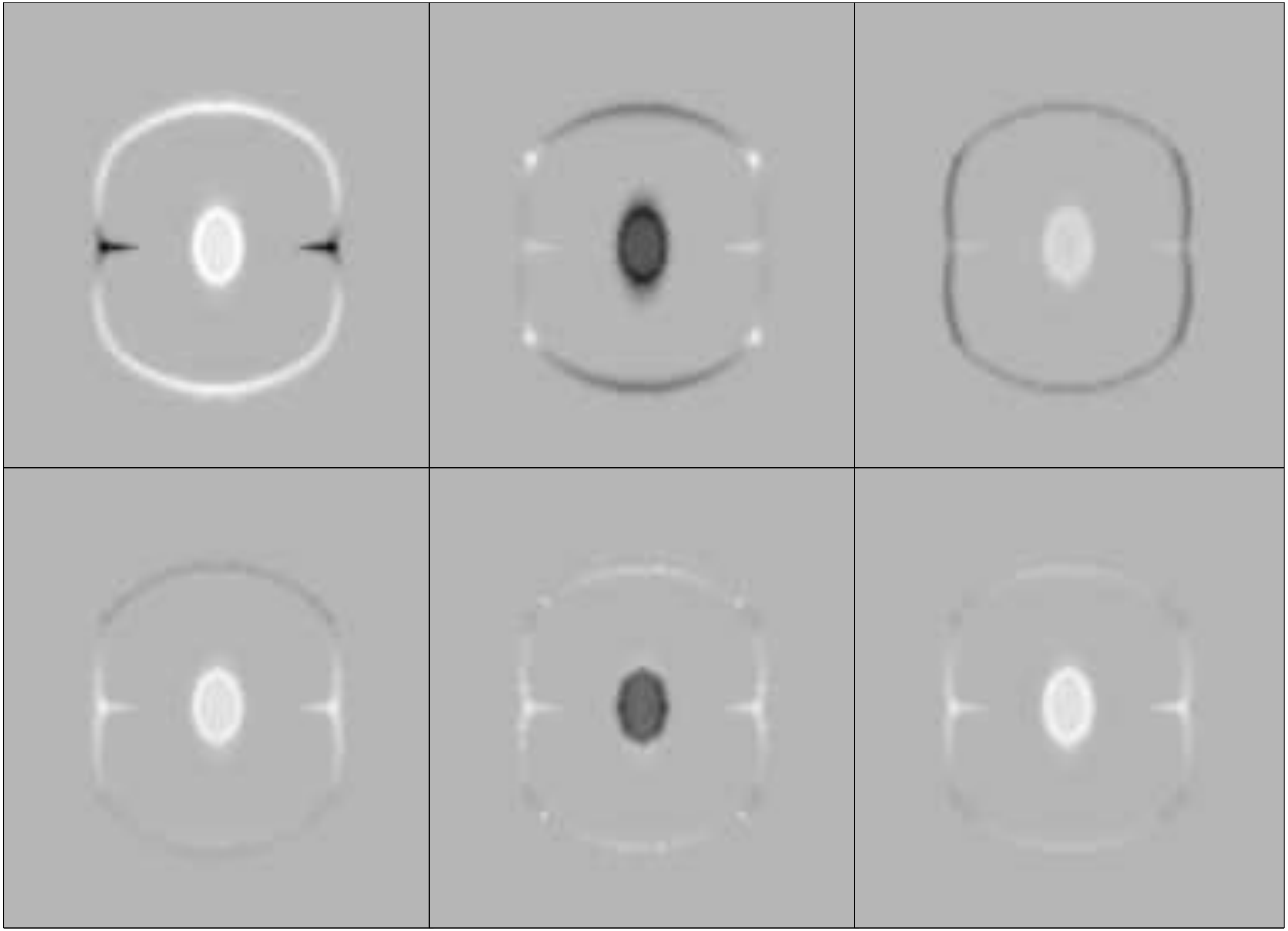




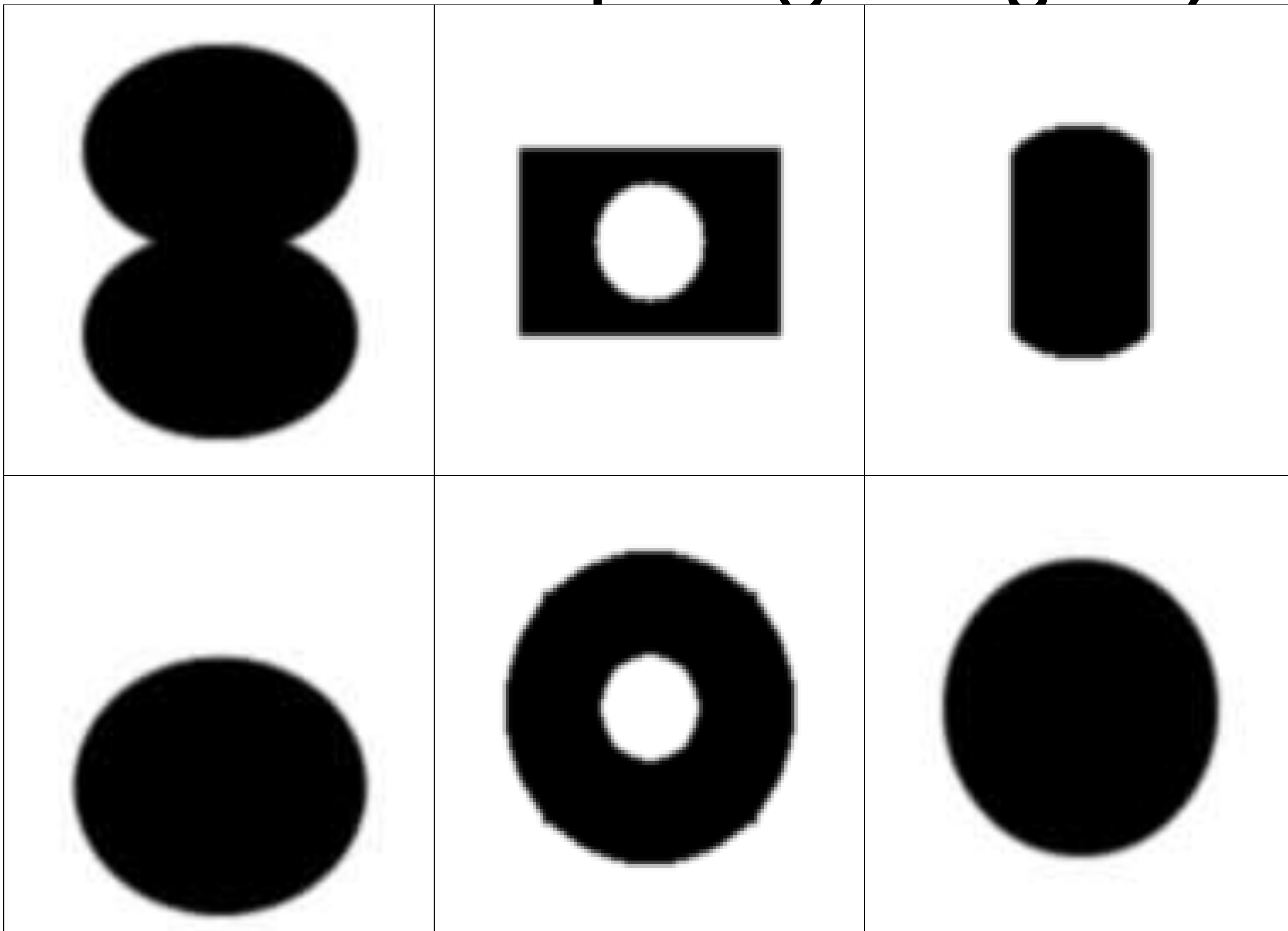
The 2D shapes (again)



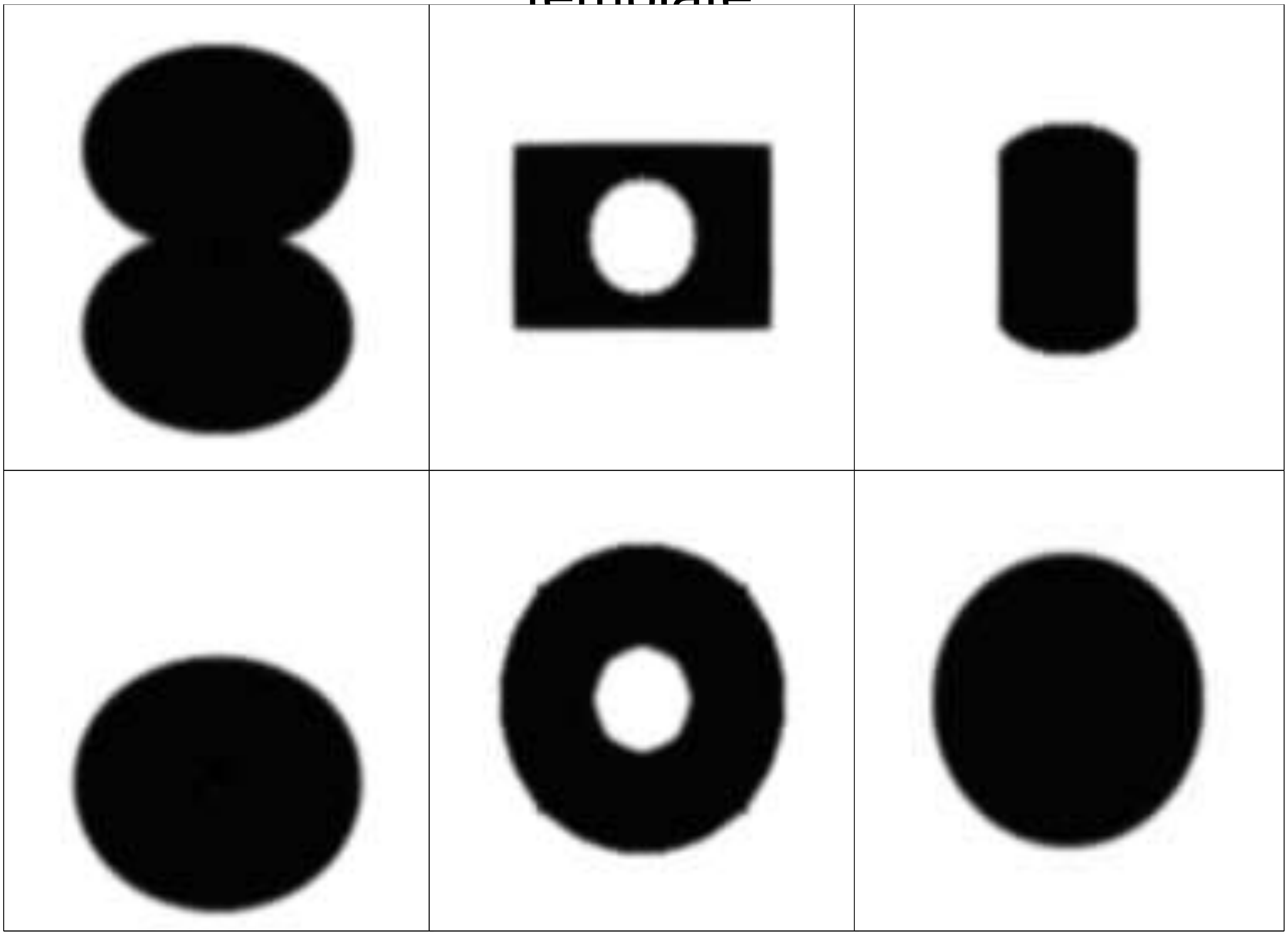
“Scalar momentum” – encodes the original shapes



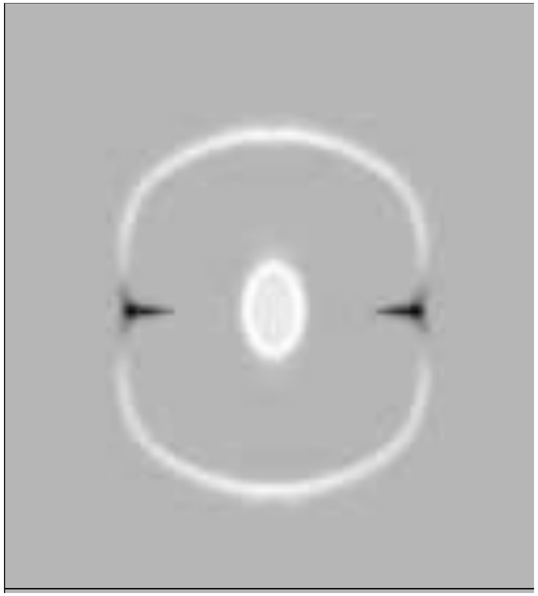
The 2D shapes (yet again)



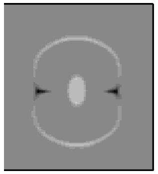
Reconstructed from scalar momentum and template



“Scalar momentum” – encodes the original shapes



Residuals

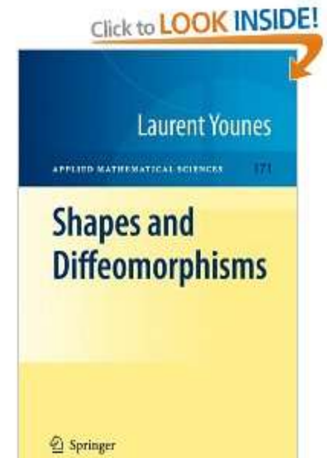


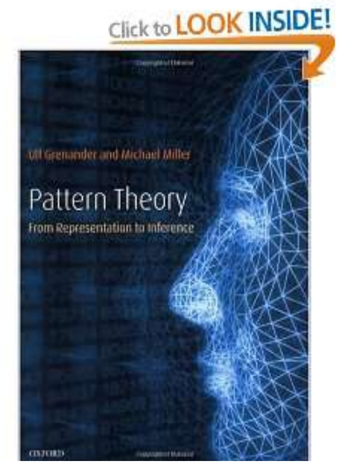
Some References

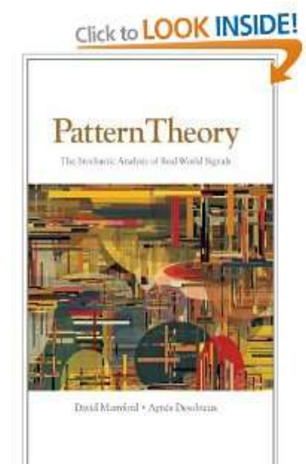
- Younes, Arrate & Miller. "*Evolutions equations in computational anatomy*". NeuroImage 45(1):S40-S50, 2009.

- Singh, Fletcher, Preston, Ha, King, Marron, Wiener & Joshi (2010). *Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures*. T. Jiang et al. (Eds.): MICCAI 2010, Part III, LNCS 6363, pp. 529–537, 2010.

- Various textbooks







Overview

- Multivariate Approaches
- Scalar Momentum
- **Some Evaluations**

Ugly Duckling Theorem

- An argument asserting that classification is impossible without some sort of bias.

Watanabe, Satosi (1969). Knowing and Guessing: A Quantitative Study of Inference and Information. New York: Wiley. pp. 376–377.

7.6. THEOREM OF THE UGLY DUCKLING

The purposes of this section is to show that from the formal point of view there exists no such thing as a class of similar objects in the world, insofar as all predicates (of the same dimension) have the same importance. Conversely, if we acknowledge the empirical existence of classes of similar objects, it means that we are attaching nonuniform importance to various predicates, and that this weighting has an extralogical origin.

David Mumford’s version

Empirical Statistics and Stochastic Models for Visual Signals

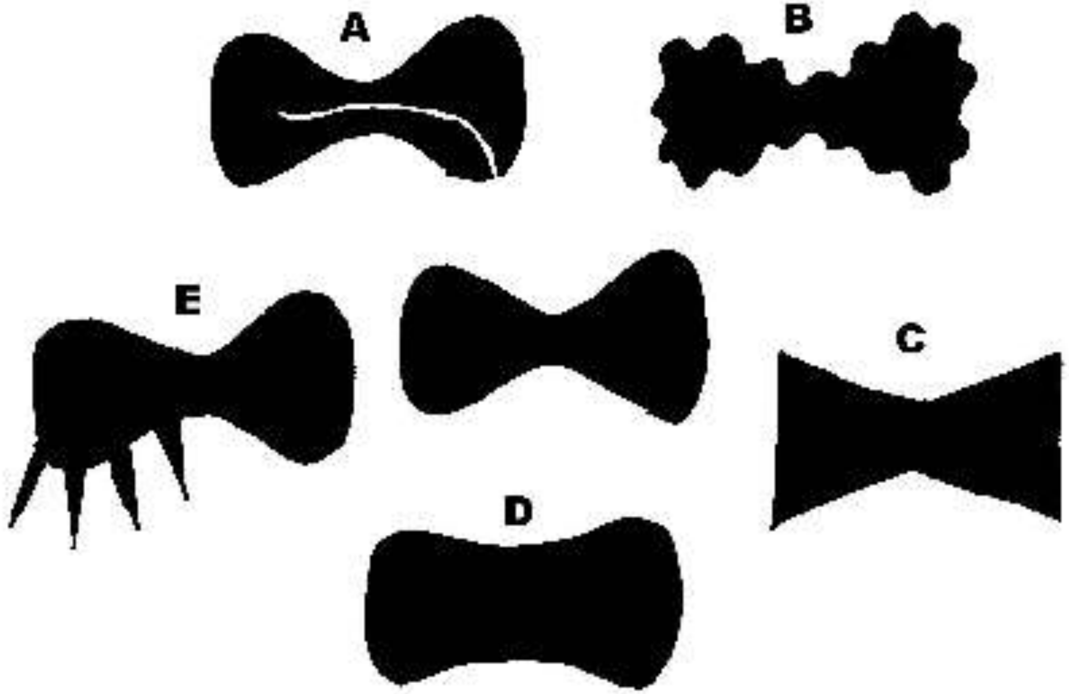
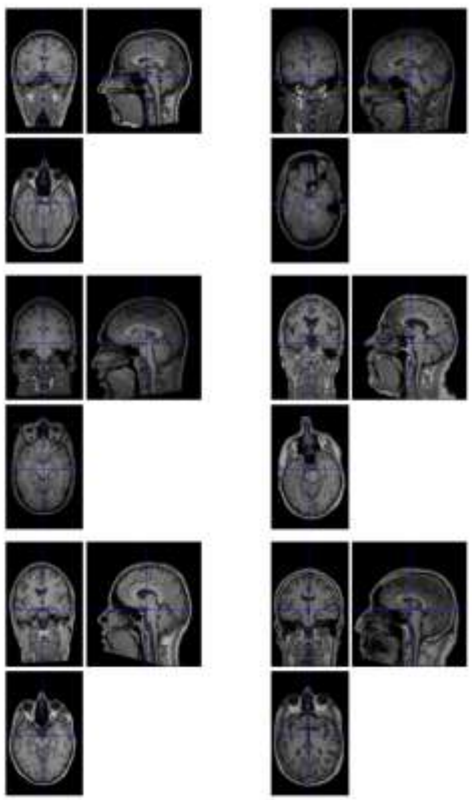


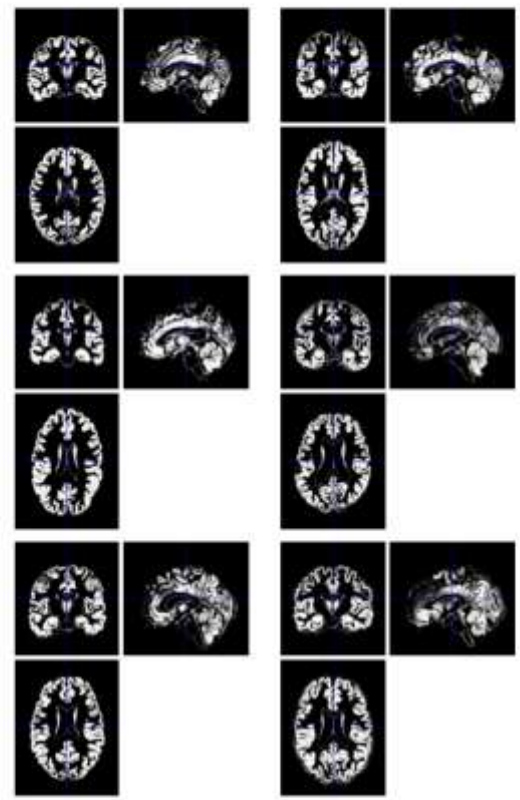
Figure 1.11 Each of the shapes A,B,C,D and E is similar to the central shape, but *in different ways*. Different metrics on the space of shape bring out these distinctions.

IXI Data

Original Images

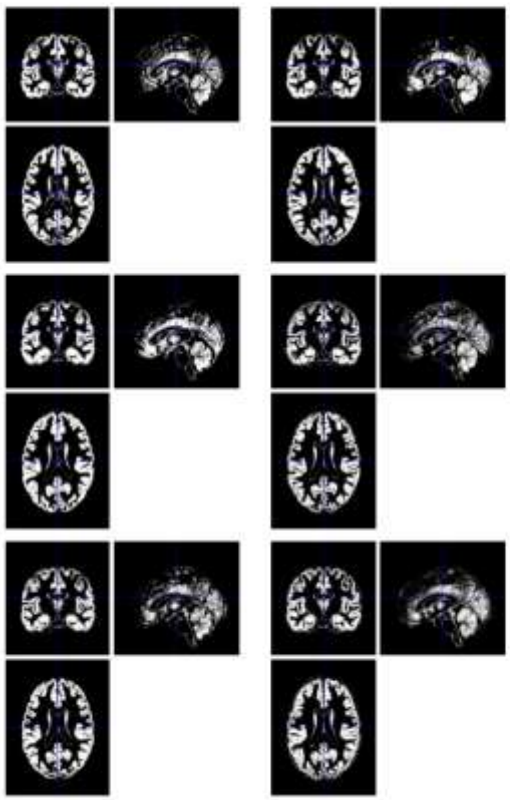


Rigidly Aligned Grey Matter

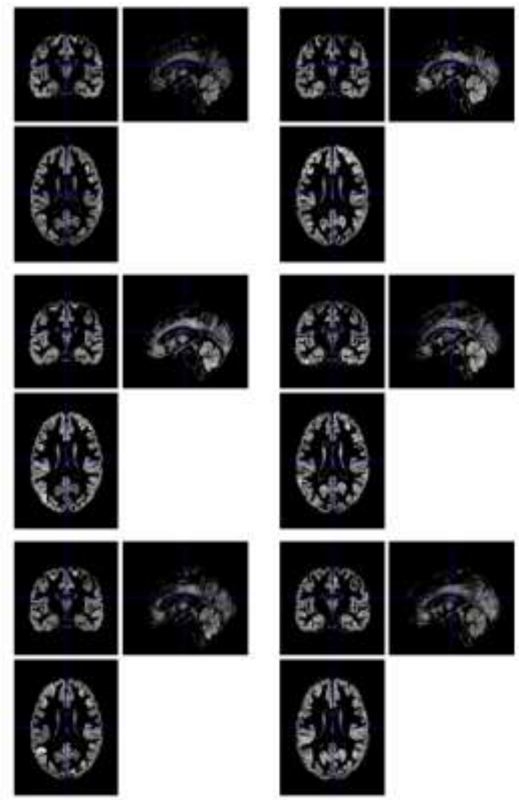


VBM-type Features

Warped Grey Matter

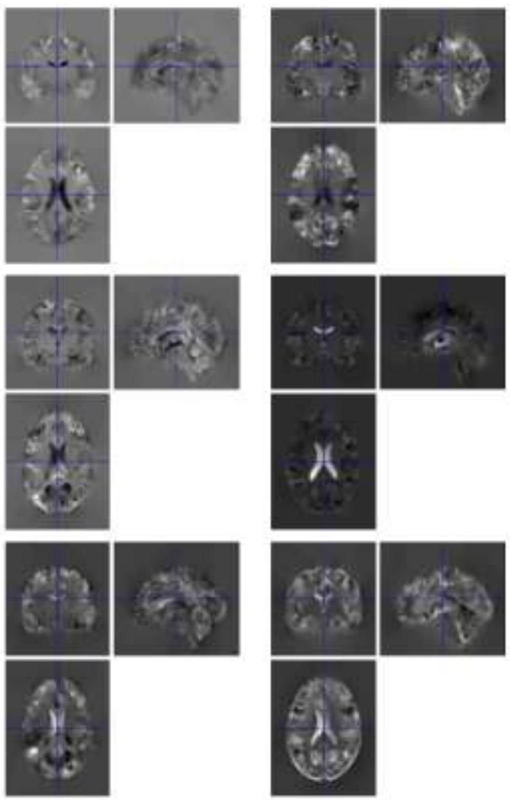


“Modulated” Warped GM

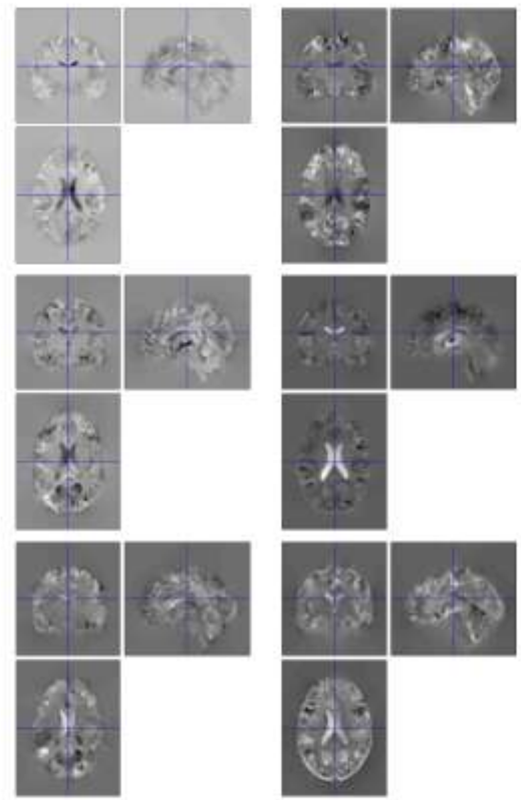


Volumetric Measures from Deformation Fields

Jacobian determinants

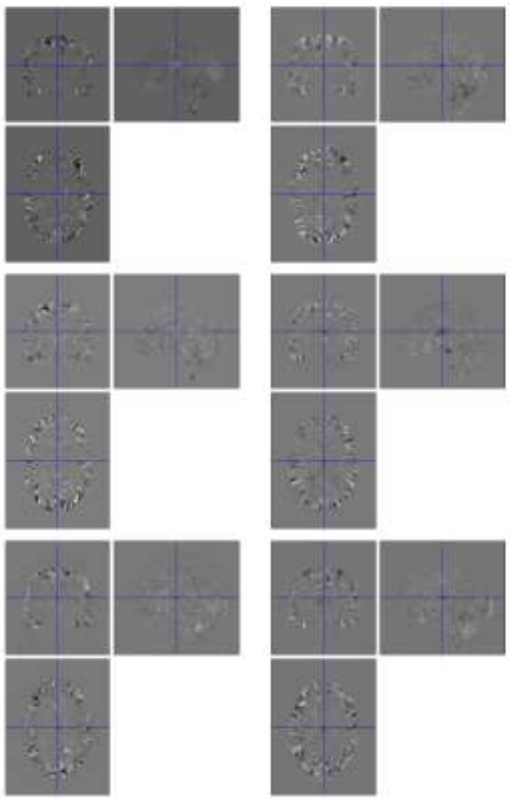


Initial Velocity Divergence

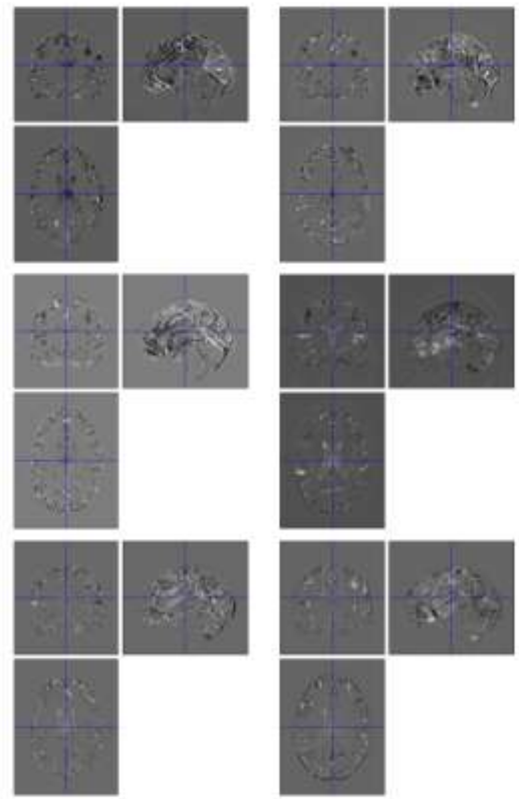


Scalar Momentum

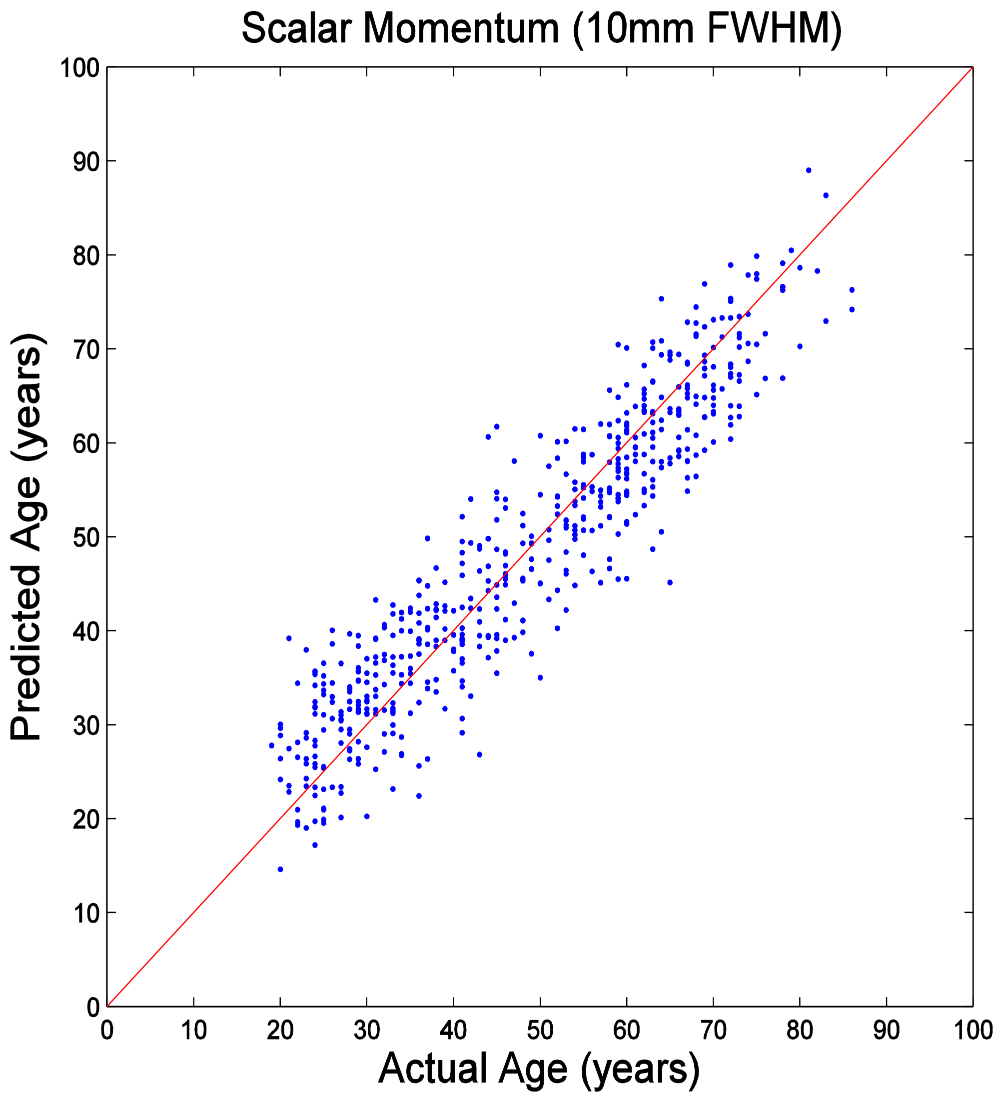
1st Component



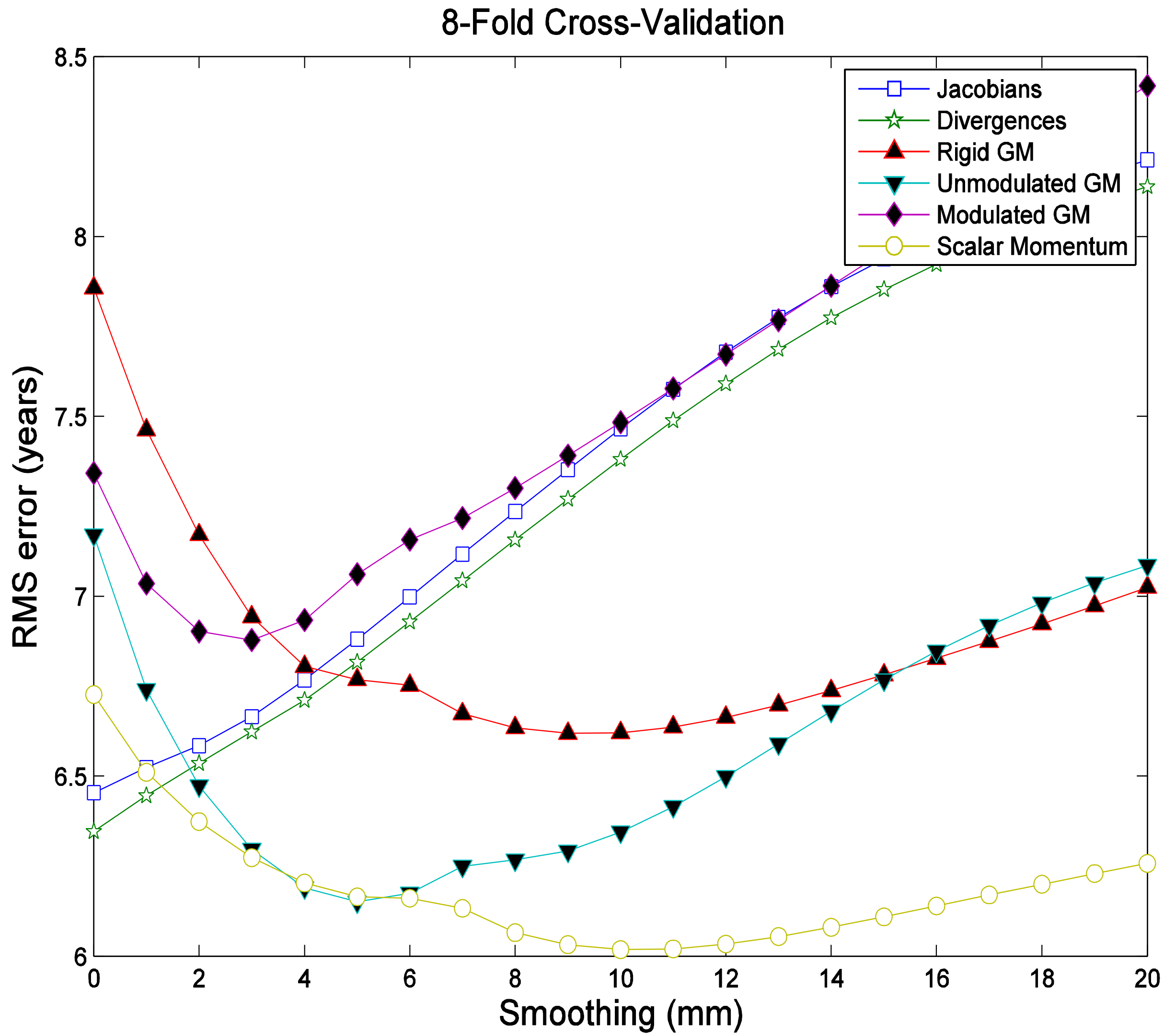
2nd Component



Age Prediction - Best Result

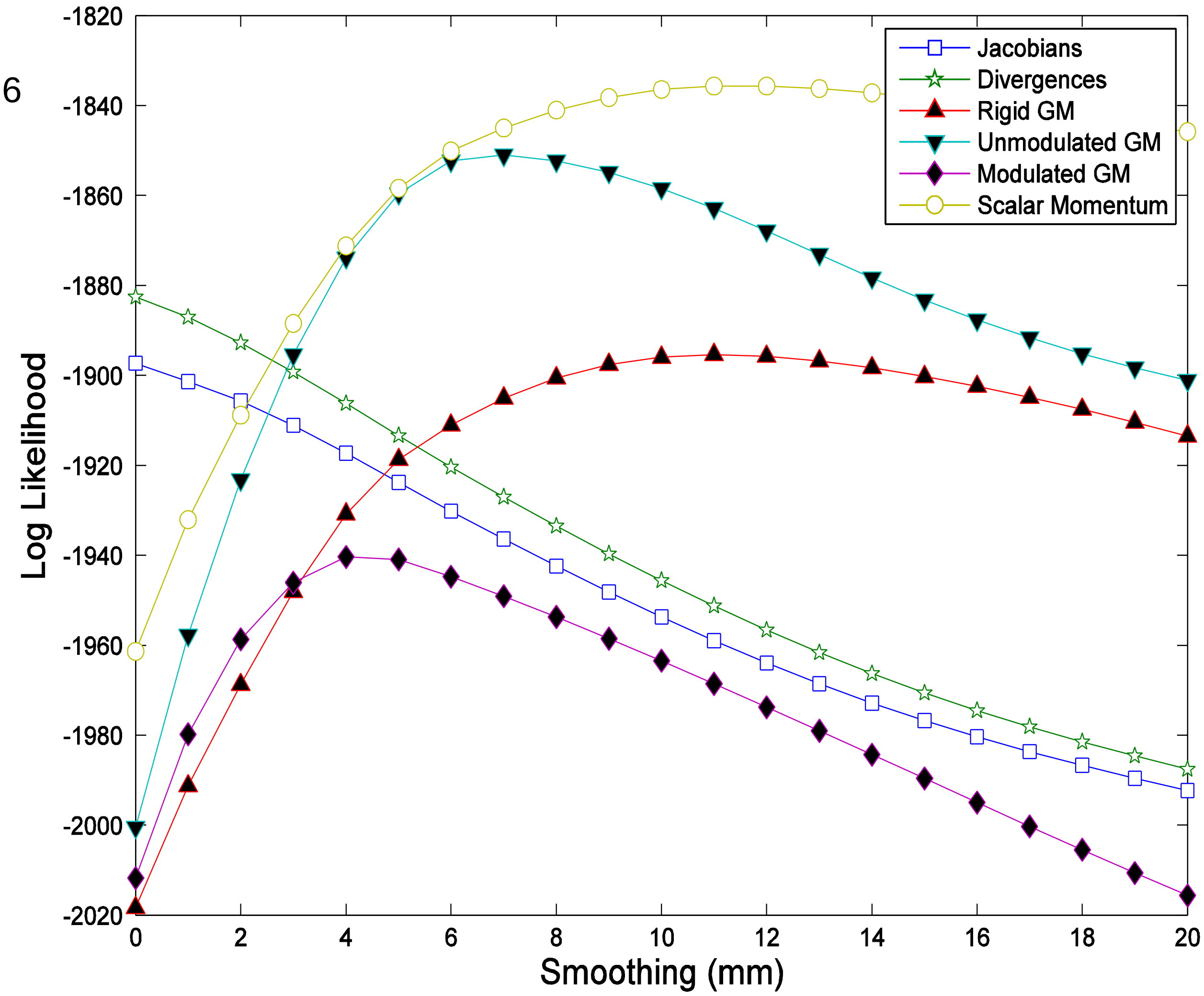


Age Prediction – Comparison Among Features

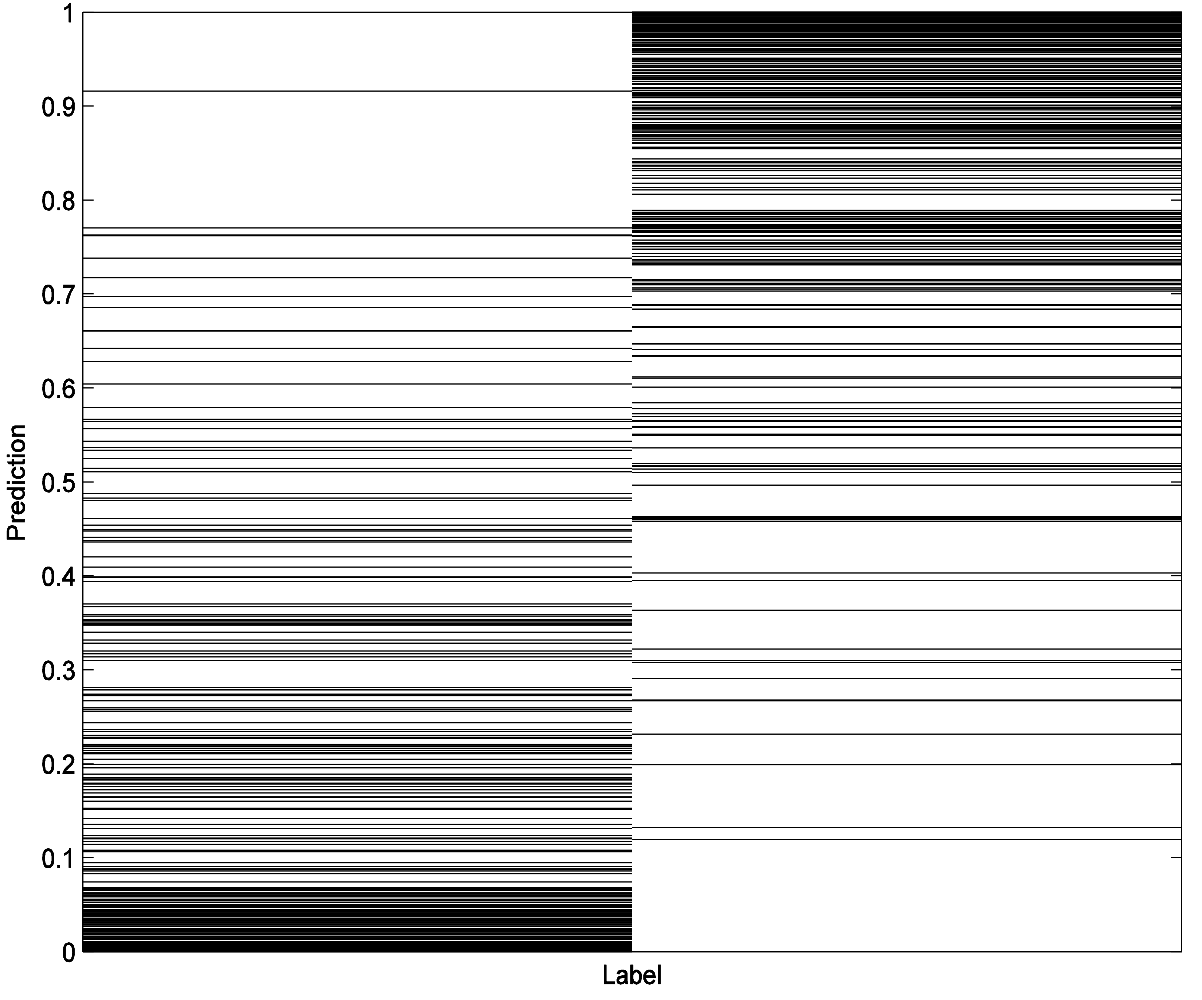


Age Prediction – Model Log Likelihoods

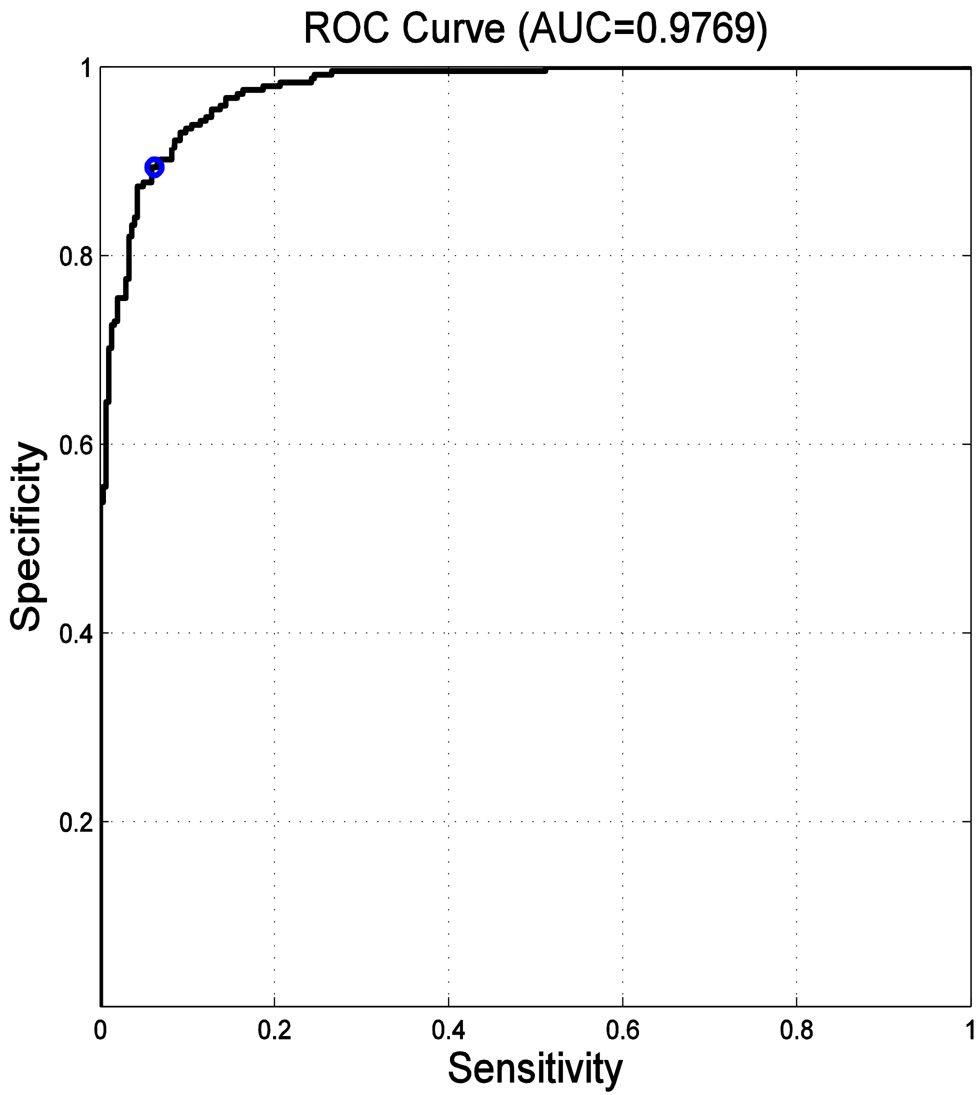
Differences > 4.6 indicate “decisive” evidence in favour of one approach over another.



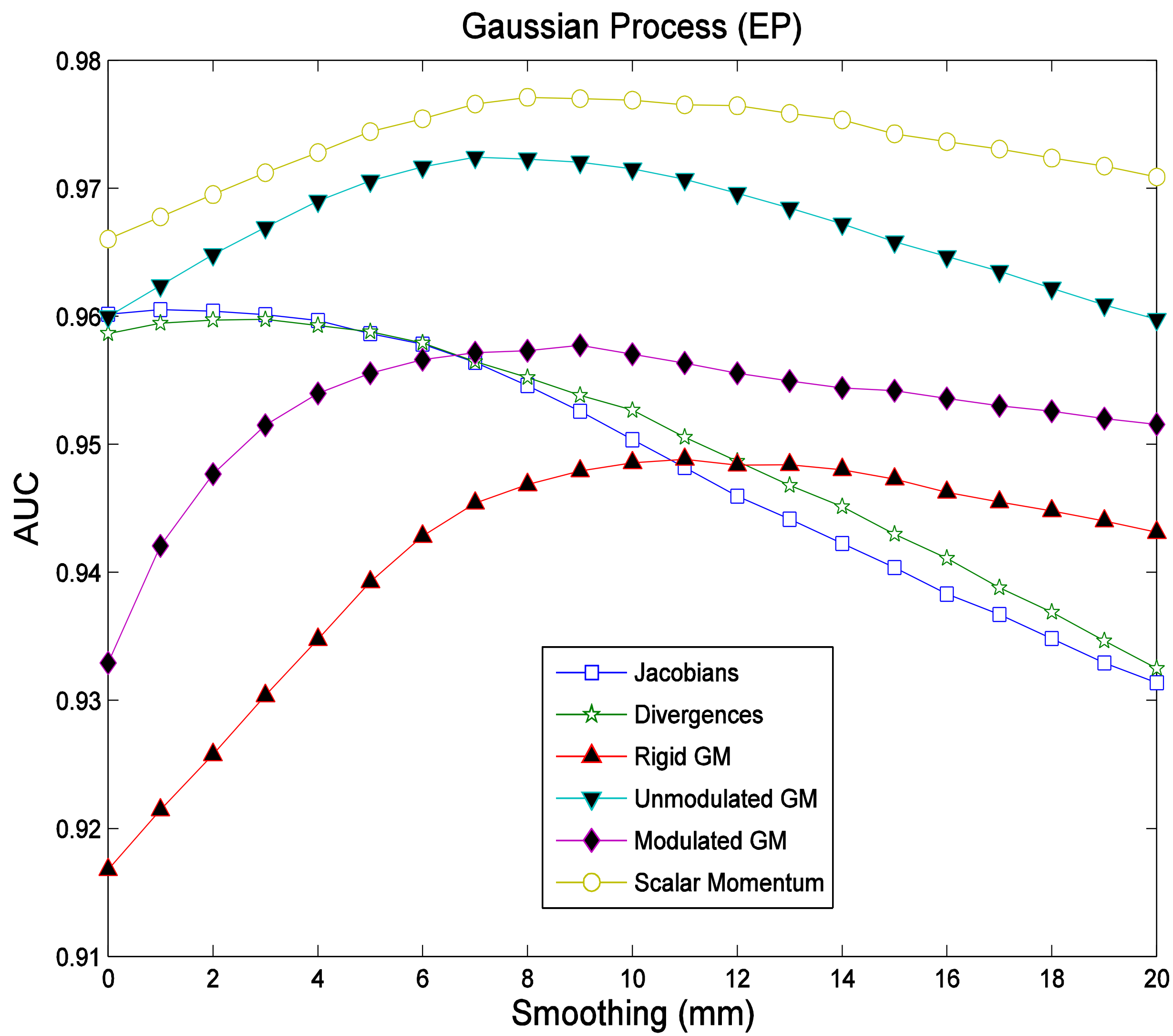
Sex Prediction – Best Result



Sex Prediction – Best Result

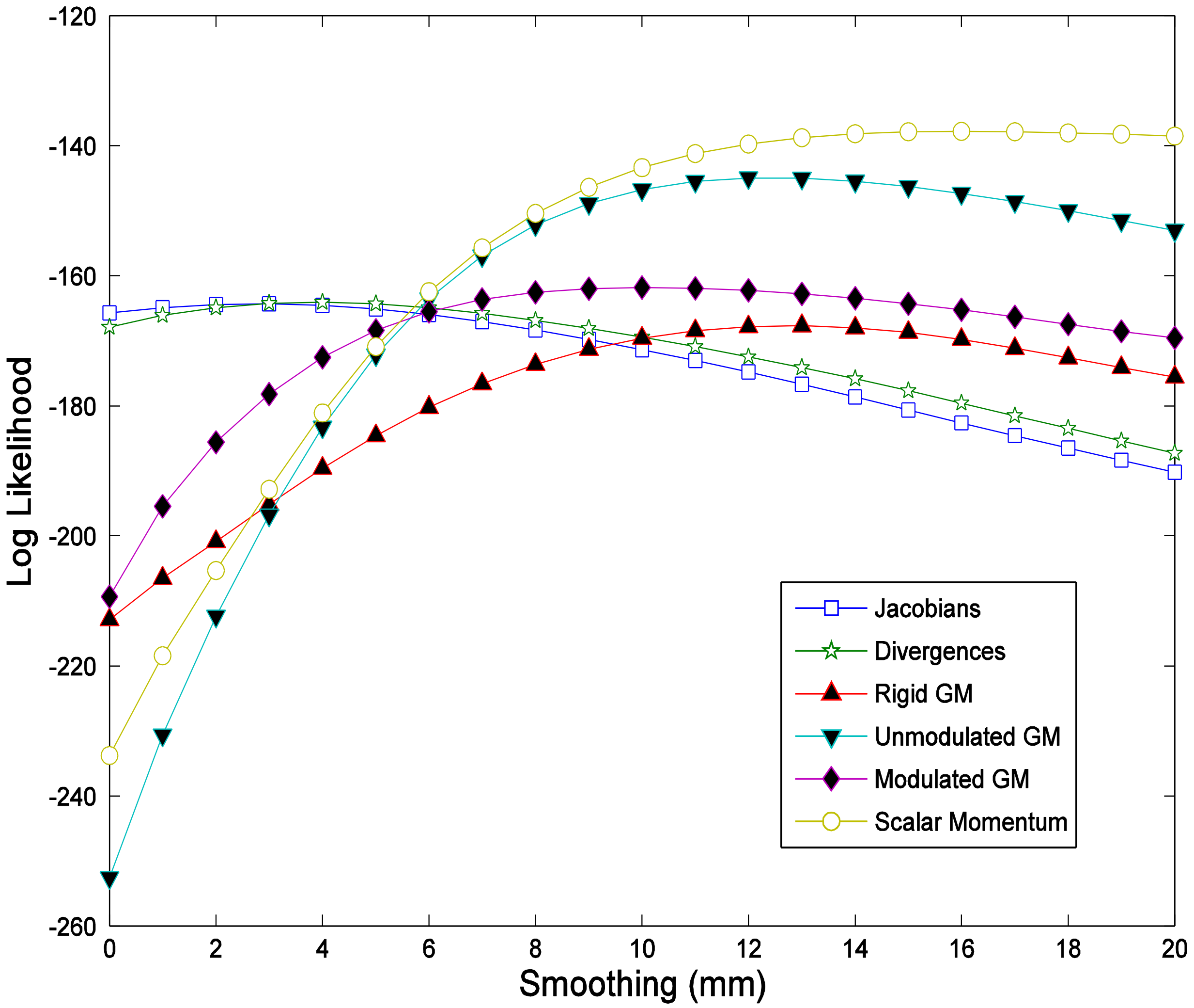


Sex Prediction – Comparison Among Features



Sex Prediction – Model Log Likelihoods

Differences > 4.6 indicate "decisive" evidence in favour of one approach over another.



References

- Singh, Fletcher, Preston, Ha, King, Marron, Wiener & Joshi (2010). *Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures*. T. Jiang et al. (Eds.): MICCAI 2010, Part III, LNCS 6363, pp. 529–537, 2010.

- Rasmussen & Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. ISBN-10 0-262-18253-X, ISBN-13 978-0-262-18253-9. http://www.gaussianprocess.org/gpml/