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
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

\[
p(x, y=0) = p(x|y=0) \, p(y=0)
\]

\[
p(x, y=1) = p(x|y=1) \, p(y=1)
\]

\[
p(x) = p(x, y=0) + p(x, y=1)
\]

\[
p(y=0|x) = \frac{p(x, y=0)}{p(x)}
\]
**Discriminative Classification Approaches**

- **Ground truth**
- **FLDA**
- **SVC**
- **Simple Logistic Regression**

Each figure represents a different classification approach applied to a dataset with two features. The graphs illustrate the decision boundaries and the distribution of data points for each approach.
Bayesian classification

Simple Logistic Regression

Hyperplane Uncertainty

Bayesian Logistic Regression

John Ashburner  Anatomical Features
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.

\[
P(\text{Disease}|\text{Pred}+) = \frac{P(\text{Pred}+|\text{Disease})P(\text{Disease})}{P(\text{Pred}+|\text{Disease})P(\text{Disease}) + P(\text{Pred}+|\text{Healthy})P(\text{Healthy})} = \frac{\text{Sensitivity} \times P(\text{Disease})}{\text{Sensitivity} \times P(\text{Disease}) + (1 - \text{Specificity}) \times P(\text{Healthy})}
\]

Better decision-making by accounting for utility functions.

---

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.

\[
P(\text{Disease}|\text{Pred}+) = \frac{P(\text{Pred}+|\text{Disease})P(\text{Disease})}{P(\text{Pred}+|\text{Disease})P(\text{Disease}) + P(\text{Pred}+|\text{Healthy})P(\text{Healthy})} = \frac{\text{Sensitivity} \times P(\text{Disease})}{\text{Sensitivity} \times P(\text{Disease}) + (1 - \text{Specificity}) \times P(\text{Healthy})}
\]

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**

\[ ((x<0.3) \& (y<2)) \mid ((x<0.75) \& (y<1.25)) \mid (y<0.4) \]

- Not nice smooth separations.
- Lots of sharp corners.
Corners matter in high-dimensions
Corners matter in high-dimensions

Circle area = $\pi r^2$

Sphere volume = $\frac{4}{3} \pi r^3$

Volume of hyper-sphere ($r=1/2$)

Number of dimensions

John Ashburner  Anatomical Features
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.

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.
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

A suitable model would reduce these data to two dimensions.
Many methods are based on similarity measures.

A common similarity measure is the dot product.

\[
\text{Similarity: } k(x, y) = \sum_{k} x_k y_k
\]

Nonlinear methods are often based on distances.

\[
\text{Distance: } \quad d(x, y) = \sqrt{\sum_{k} (x_k - y_k)^2}
\]

\[
\text{Similarity: } \quad k(x, y) = \exp(-\lambda d(x, y)^2)
\]

How do we best measure distances between brain images?
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

Empirical Statistics and Stochastic Models for Visual Signals

![Images of shapes A, B, C, D, E]

**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

John Ashburner
Anatomical Features
Distances need to satisfy the properties of a *metric*:

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

Satisfying (3) requires inverse-consistent image registration. Satisfying (4) requires a specific family of image registration algorithm.
Distances are not always measured along a straight line.

“Shapes are the ultimate non-linear sort of thing”
Some example (non-brain) images.
We could register the images to their average shape...
...and study the deformations...
Jacobian Determinants

...or the relative volumes...
Scalar Momentum

... or “scalar momentum”
Reconstructed Images

Reconstructions from template and scalar momenta.
Used 550 T1w brain MRI from IXI (Information eXtraction from Images) 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.

Diffeomorphic Alignment

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


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.

<table>
<thead>
<tr>
<th>Jacobian Determinants</th>
<th>Initial Velocity Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Images of brain scans]</td>
<td>![Images of brain scans]</td>
</tr>
</tbody>
</table>

John Ashburner  
Anatomical Features
Grey Matter Features

Rigidly Registered GM

Nonlinearly Registered GM

Registered and Jacobian Scaled GM
“Scalar momentum” actually has two components because GM was matched with GM and WM was matched with WM.
Age Regression

Linear Gaussian Process Regression to predict subject ages.

**Sex Classification**

Linear Gaussian Process Classification (EP) to predict sexes.

**Predictive Accuracies**

**Age**

Scalar Momentum (10mm FWHM)

<table>
<thead>
<tr>
<th>Actual Age (years)</th>
<th>Predicted Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

**Sex**

ROC Curve (AUC=0.9769)

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
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.
Introduction
Geometric Variability
Similarity Measures
Real data

Data
Features
Results

John Ashburner
Anatomical Features