



# MACHINE LEARNING METHODS FOR MEDICAL DATA

AMOS STORKEY  
SCHOOL OF INFORMATICS

# OVERVIEW

Overview

What Machine Learning Offers

Imaging Data

Sequential Data

Genomic Data

THEN:

Targeted specific individual disease studies involving a few thousand patients and well curated low dimensional data.

NOW:

Post-genomic. Interconnection of diseases, multiple disparate data sources, millions of records, high and very high dimensional, handling multiple measurements and multimodal integration. Less well curated data.



# WHAT MACHINE LEARNING OFFERS

Overview

What Machine Learning Offers

Imaging Data

Sequential Data

Genomic Data

- The term Big Data can be a distraction.
- Generalisation across problems. Commonality. Reduce Reinvention.
- Flexible methods for prediction.
- Multimodal integration
- High dimensional data analysis.
- Nonparametric methods – integrating over unknowns.
- Hierarchical, deep feature selection techniques.
- Causal analysis.
- BUT
  - Not automatic.
  - Requires people and requires design.
  - Handling large data involves knowing the data and knowing the problem.  
Finding a signal.



# IMAGING DATA

Overview

What Machine Learning Offers

Imaging Data

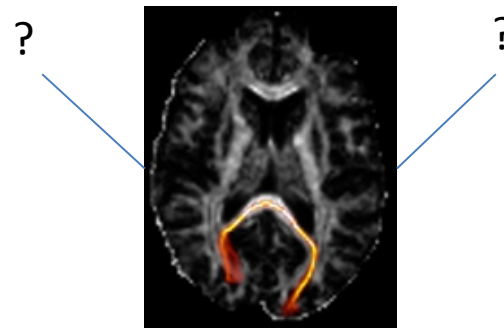
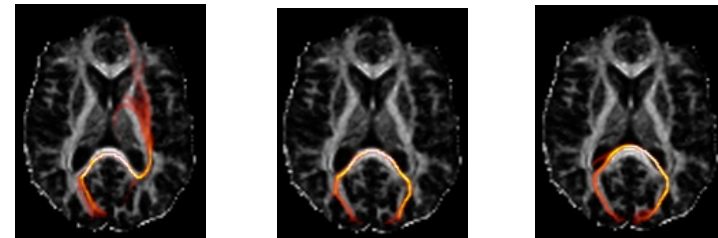
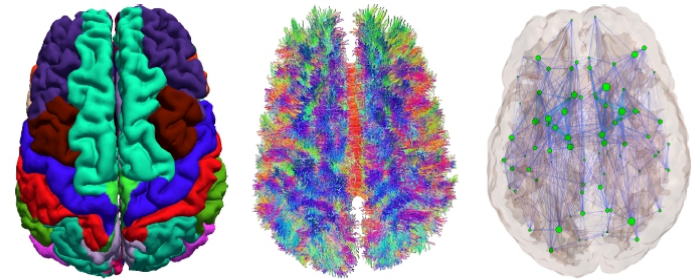
Sequential Data

Genomic Data

- Methods for obtaining accurate and relevant signal in images.
- Accurate methods of linking between subjects in imaging.
- Fitting of complex image models.
- Automated identification of image components.
- Identification of key image- based biomarkers.

**EXAMPLES:** Linking white matter fibre tracks across subjects. Network Analysis. Pre-surgical planning. Combining neuroanatomical and clinical variables for prediction of Schizophrenia.

Mark Bastin, Jon Clayden, Susana Munoz Maniega, Krzysztof Gorgolewski, Colin Buchanan, Cyril Pernet, Eleni Zarogianni, Stephen Lawrie



# SEQUENTIAL DATA

Overview

What Machine Learning Offers

Imaging Data

Sequential Data

Genomic Data

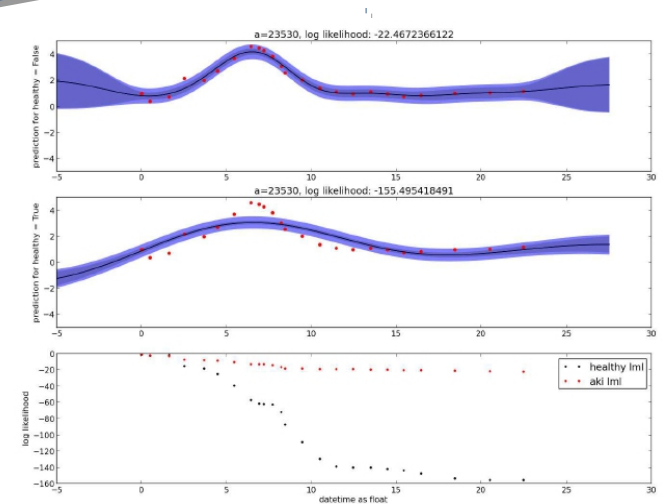
- Sequential patient data is one form of high dimensional data.
  - Rarely time series.
  - Can be multimodal.
  - Tapping clinical expertise.
  - But not always obvious where the signal lies.

## EXAMPLES: Renal Failure Analysis

with Richard Phelps, Jakub Piatkowski

Also

Chris Williams, Jon Quinn, Ioan Stanculescu, Factorial Temporal Models for Analysis in Neonatal Care.



# GENOMIC DATA

Overview

What Machine Learning Offers

Imaging Data

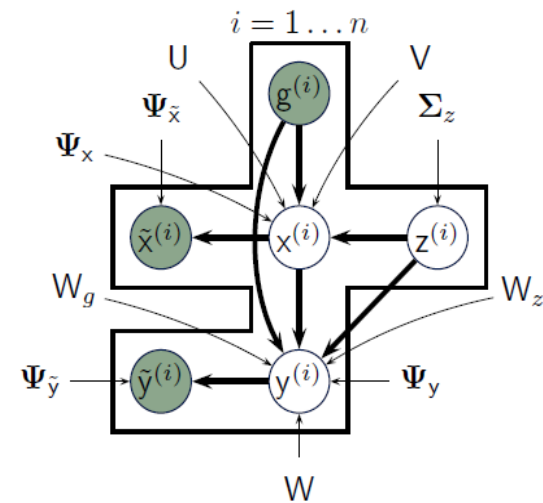
Sequential Data

Genomic Data

- Sparse methods for genomic analysis.
- Causal models connecting biomarkers to outcomes.
- Nonparametric approaches.
- Probabilistic feature discovery methods in disequilibrium studies.

EXAMPLES: Sparse Instrumental Variables (SPIV) for Genome-Wide Studies

Paul McKeigue, Felix Agakov



# THE WAY FORWARD

- Many opportunities.
- General methods across problems.
- More exemplar projects. Some quick wins? Predictive settings?
- Issues of privacy.
- Continued improvement in data sharing and linkage, curation. Coding of errors.
- Large multivariate longitudinal analyses from patient records.
- Automated linkage and aggregation of results, not just linkage of data.
- Distributed Analytics.

