Machine Learning I: basics

Andre F. Marquand
a.marquand@donders.ru.nl

- Pattern Recognition (PR) is a subfield of machine learning, which relates to the automatic discovery of patterns of statistical regularity in data
- Aim to learn from empirical data rather than following fixed rules
- **Learn by example**
- Increasingly used in computational psychiatry for:

- Pattern Recognition (PR) is a subfield of machine learning, which relates to the automatic discovery of patterns of statistical regularity in data
- Aim to learn from empirical data rather than following fixed rules
- **Learn by example**
- Increasingly used in computational psychiatry for:
    1. Predicting clinical variables (diagnosis or treatment response)
    2. Stratifying psychiatric disorders

Outline

Outline

What is pattern recognition used for?

Historically, has been applied in many application domains:

Example Applications

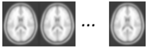
- Speech Recognition
- Automatic Character recognition / handwriting recognition
- Document classification (e.g. spam filters)
- Analysis of genetic microarray data
- Self-driving cars
- Recommender systems / online shopping
- ...

Types of pattern recognition

Supervised learning involves learning a mapping between input and output:

Input

Training examples
$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$



$\Rightarrow$

No mathematical
Model available

$\Rightarrow$

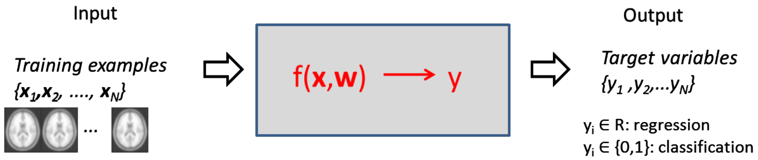
Output

Target variables
$\{y_1, y_2, \ldots y_N\}$

$y_i \in \mathbb{R}$: regression
$y_i \in \{0,1\}$: classification

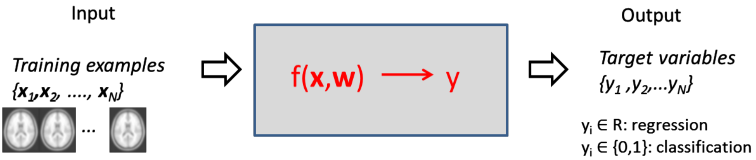
Types of pattern recognition

Supervised learning involves learning a mapping between input and output:

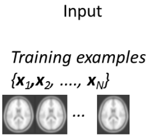


Types of pattern recognition

Supervised learning involves learning a mapping between input and output:



In **Unsupervised** learning, algorithms are not provided with output labels and must learn to structure the data by applying heuristics



Types of pattern recognition

Supervised learning involves learning a mapping between input and output:

Input

Training examples
$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$

$$f(\mathbf{x}, \mathbf{w}) \longrightarrow y$$

Output

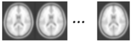
Target variables
$\{y_1, y_2, \ldots y_N\}$

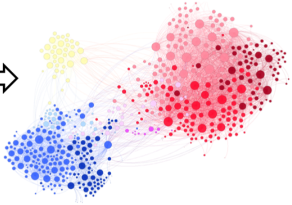
$y_i \in R$: regression
$y_i \in \{0,1\}$: classification

In **Unsupervised** learning, algorithms are not provided with output labels and must learn to structure the data by applying heuristics

Input

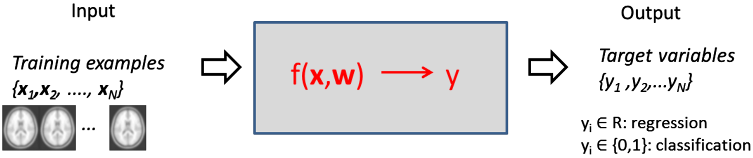
Training examples
$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$



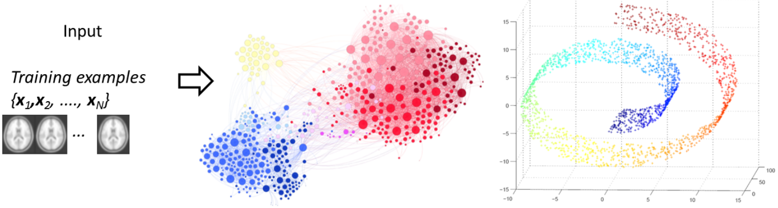


Types of pattern recognition

Supervised learning involves learning a mapping between input and output:

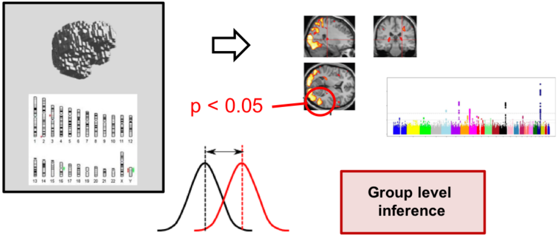


In **Unsupervised** learning, algorithms are not provided with output labels and must learn to structure the data by applying heuristics

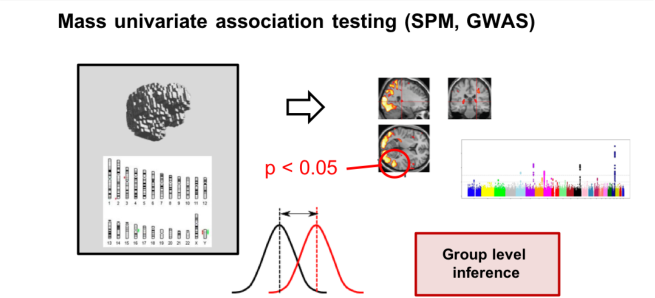


Biomarkers for prediction

Mass univariate association testing (SPM, GWAS)

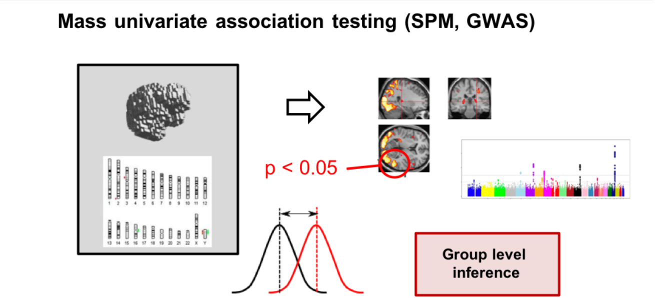


Biomarkers for prediction



- Useful for understanding mechanisms

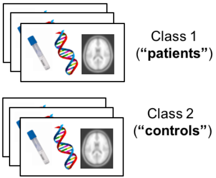
Biomarkers for prediction



- Useful for understanding mechanisms
- For clinical decision making this does not suffice. It is necessary to make predictions about individuals

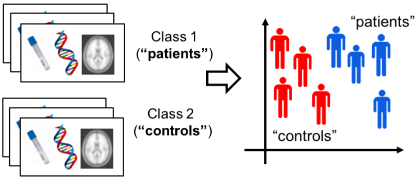
Predicting disease state

Making subject level predictions of diagnosis and outcome



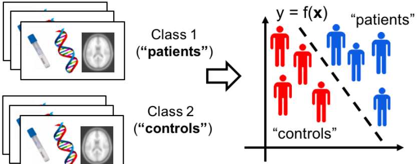
Predicting disease state

Making subject level predictions of diagnosis and outcome

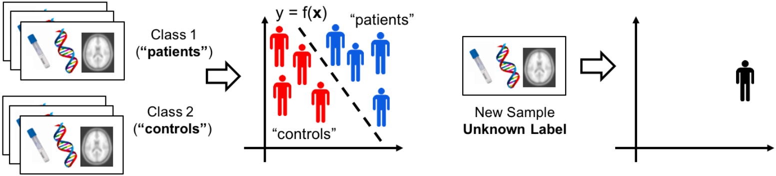


Predicting disease state

Making subject level predictions of diagnosis and outcome

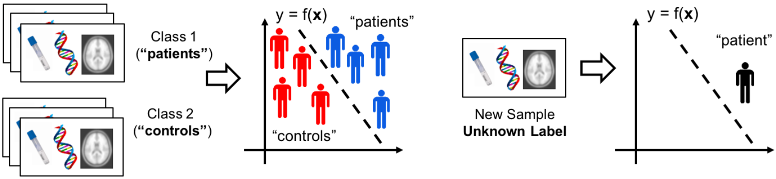


Making subject level predictions of diagnosis and outcome



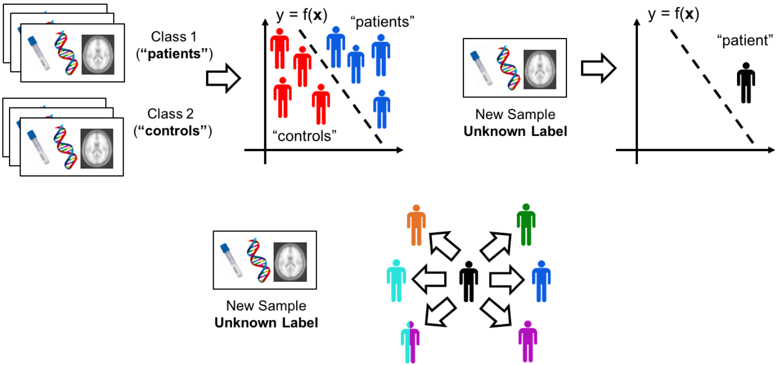
Predicting disease state

Making subject level predictions of diagnosis and outcome



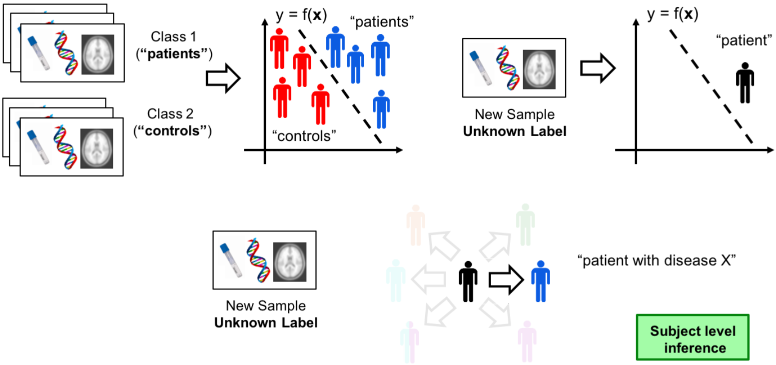
Predicting disease state

Making subject level predictions of diagnosis and outcome



Predicting disease state

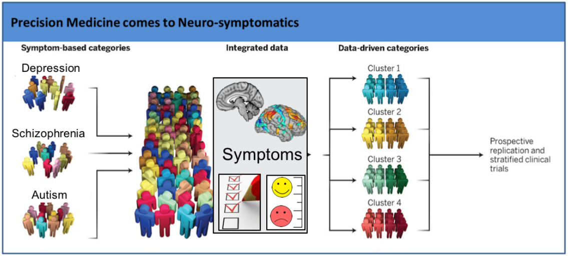
Making subject level predictions of diagnosis and outcome



Useful to find a measure of overall group separation

Machine Learning in Psychiatry: Stratification

Tackling the clinical and biological heterogeneity of psychiatric disorders



NIH National Institute of Mental Health

RDoC Research Domain Criteria Initiative

Insel et al. (2015)

Outline

Encoding and Decoding

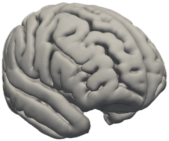
Cognitive state

- Response to stimulus
- Cognitive scores
- Diagnostic label
- Symptom scores
- ...

Encoding model

$$Y = f(X)$$

Brain activity (e.g. BOLD)



Encoding and Decoding

Which brain regions are activated in a certain task condition?

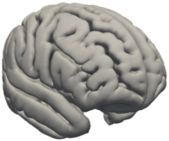
Cognitive state

- Response to stimulus
- Cognitive scores
- Diagnostic label
- Symptom scores
- ...

Encoding model

$$Y = f(X)$$

Brain activity (e.g. BOLD)



Encoding and Decoding

Which brain regions are activated in a certain task condition?

Cognitive state

- Response to stimulus
- Cognitive scores
- Diagnostic label
- Symptom scores
- ...

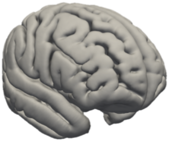
Encoding model

$$Y = f(X)$$

$$X = f(Y)$$

Decoding model

Brain activity (e.g. BOLD)



Encoding and Decoding

Which brain regions are activated in a certain task condition?

Cognitive state

- Response to stimulus
- Cognitive scores
- Diagnostic label
- Symptom scores
- ...

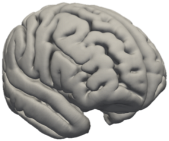
Encoding model

$$Y = f(X)$$

$$X = f(Y)$$

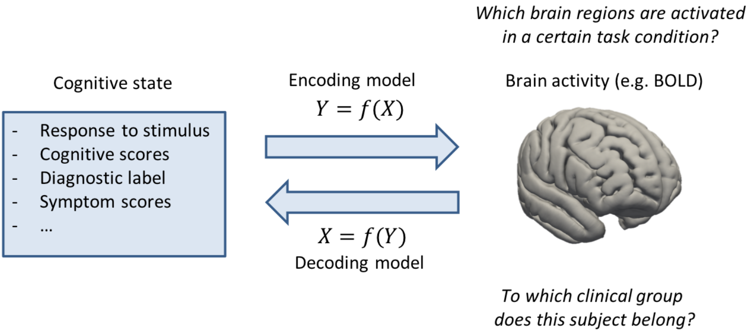
Decoding model

Brain activity (e.g. BOLD)



To which clinical group does this subject belong?

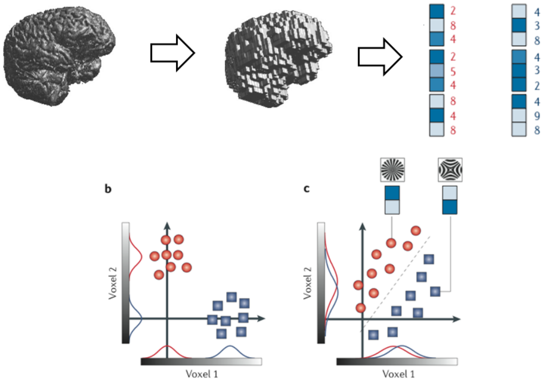
Encoding and Decoding



- Also called Generative/Recognition models
- $\neq$ Generative/Discriminative models in machine learning
- This distinction relates to the brain, not to the methods

Multivariate models

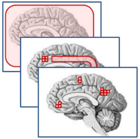
Sensitivity for spatially distributed (or multivariate) effects:



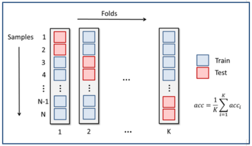
Haynes and Rees (2006)

Stages of supervised pattern recognition analysis

1. Feature extraction and/or feature selection

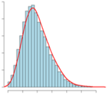


2. Classification / Regression using cross-validation



3. Performance evaluation

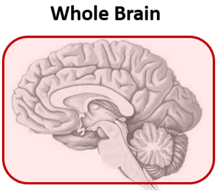
$$acc = \frac{1}{K}\sum_{i=1}^{K} acc_i$$

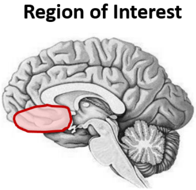


Feature selection and feature construction

Whole Brain

Feature selection and feature construction

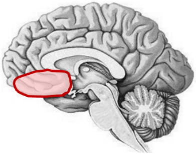




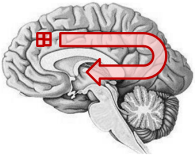
Feature selection and feature construction

Whole Brain

Region of Interest



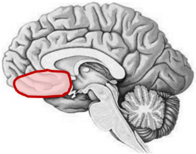
Searchlight



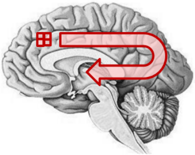
Feature selection and feature construction

Whole Brain

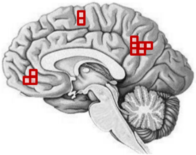
Region of Interest



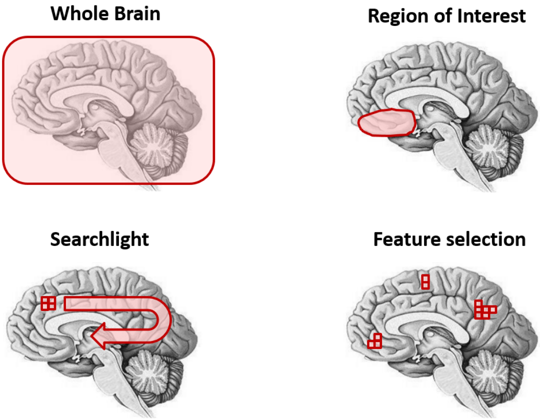
Searchlight



Feature selection

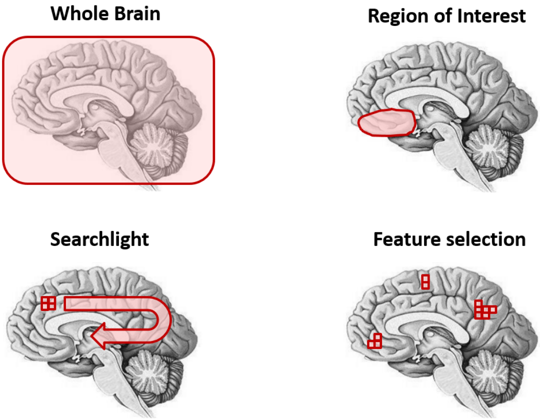


Feature selection and feature construction



- Can also construct features (e.g. using ICA/ PCA,...)

Feature selection and feature construction



- Can also construct features (e.g. using ICA/ PCA,...)
- Feature selection should be performed on training data only!

Notation

$\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ or $\{\mathbf{X}, \mathbf{Y}\}$	Dataset
$\mathbf{X}_{N \times D} = [\mathbf{x}_1, ..., \mathbf{x}_N]^T$	N samples, D features
$\mathbf{y} = [y_1, ..., y_N]^T$	Targets
$\mathbf{w} = [w_1, ..., w_D]^T$	Weights

A smorgasbord of different approaches

- Regularisation methods (Penalized linear models, support vector machines, ...)
- Probabilistic approaches (Linear discriminant analysis, Gaussian processes, ...)
- Ensemble methods (Random forests, boosting, ...)
- Neural networks (multi-layer perceptrons, deep learning, ...)

A smorgasbord of different approaches

- Regularisation methods (Penalized linear models, support vector machines, ...)
- Probabilistic approaches (Linear discriminant analysis, Gaussian processes, ...)
- Ensemble methods (Random forests, boosting, ...)
- Neural networks (multi-layer perceptrons, deep learning, ...)

Most methods aim to trade-off data fit with complexity

$$f(\mathbf{x}_i, \mathbf{w}) = f_i = \mathbf{x}_i^T \mathbf{w}$$

A smorgasbord of different approaches

- Regularisation methods (Penalized linear models, support vector machines, ...)
- Probabilistic approaches (Linear discriminant analysis, Gaussian processes, ...)
- Ensemble methods (Random forests, boosting, ...)
- Neural networks (multi-layer perceptrons, deep learning, ...)

Most methods aim to trade-off data fit with complexity

$$f(\mathbf{x}_i, \mathbf{w}) = f_i = \mathbf{x}_i^T \mathbf{w} \quad \Rightarrow \hat{\mathbf{w}} \; = \min_{\mathbf{w}} \sum_{i=1}^{n} \ell(y_i, f_i) + \lambda J(\mathbf{w})$$

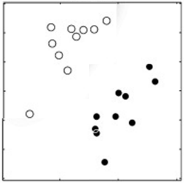
A smorgasbord of different approaches

- Regularisation methods (Penalized linear models, support vector machines, ...)
- Probabilistic approaches (Linear discriminant analysis, Gaussian processes, ...)
- Ensemble methods (Random forests, boosting, ...)
- Neural networks (multi-layer perceptrons, deep learning, ...)

Most methods aim to trade-off data fit with complexity

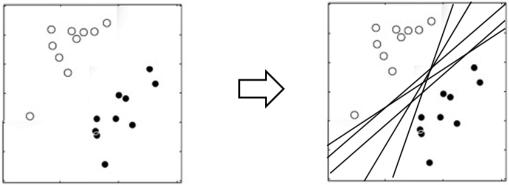
$$
\begin{aligned}
f(\mathbf{x}_i, \mathbf{w}) = f_i = \mathbf{x}_i^T \mathbf{w} \qquad \Rightarrow \hat{\mathbf{w}} &= \min_{\mathbf{w}} \sum_{i=1}^{n} \ell(y_i, f_i) + \lambda J(\mathbf{w}) \\
\Rightarrow \hat{\mathbf{w}} &= \min_{\mathbf{w}} - \ln p(\mathbf{w}|\mathbf{y}) \\
&= \min_{\mathbf{w}} \sum_{i=1}^{n} \ln p(y_i|f_i) + \ln p(\mathbf{w}|\boldsymbol{\theta})
\end{aligned}
$$

Choice of pattern recognition algorithm



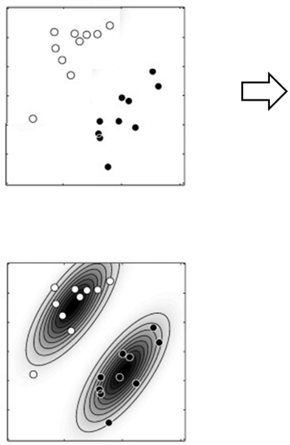
Ashburner and Klöppel (2011)

Choice of pattern recognition algorithm



Ashburner and Klöppel (2011)

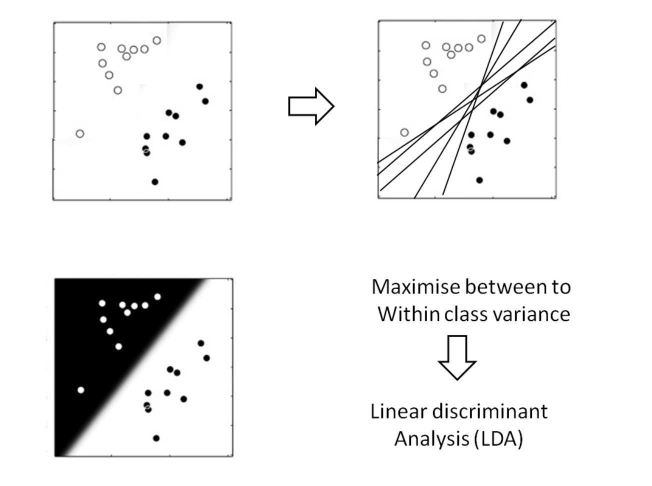
Choice of pattern recognition algorithm



Maximise between to
Within class variance

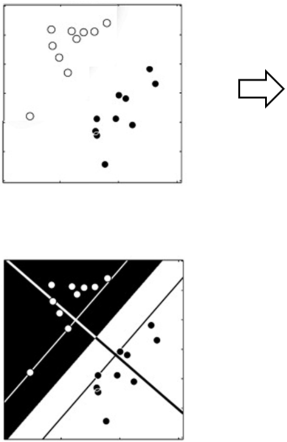
Ashburner and Klöppel (2011)

Choice of pattern recognition algorithm



Ashburner and Klöppel (2011)

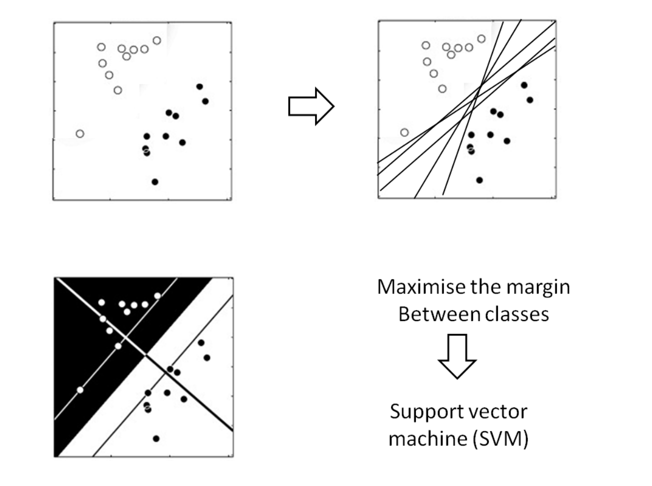
Choice of pattern recognition algorithm



Maximise the margin
Between classes

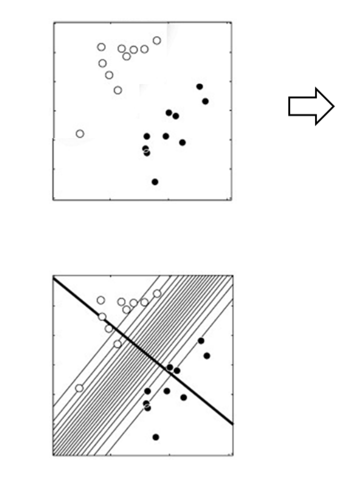
Ashburner and Klöppel (2011)

Choice of pattern recognition algorithm



Ashburner and Klöppel (2011)

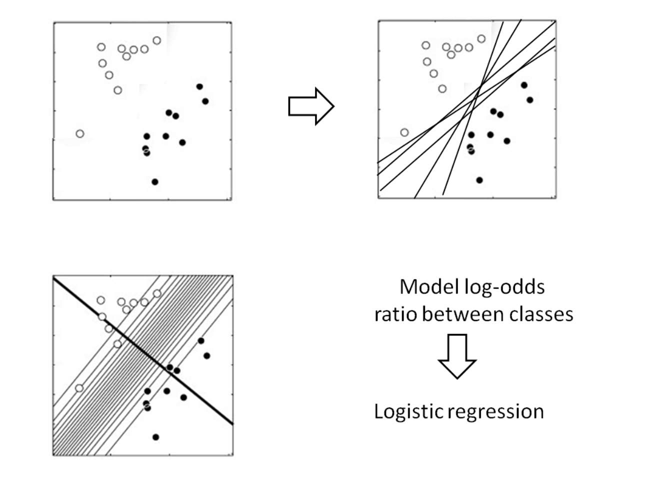
Choice of pattern recognition algorithm



Model log-odds
ratio between classes

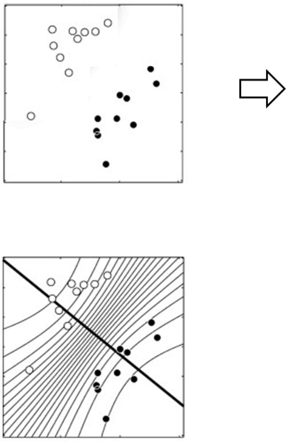
Ashburner and Klöppel (2011)

Choice of pattern recognition algorithm



Ashburner and Klöppel (2011)

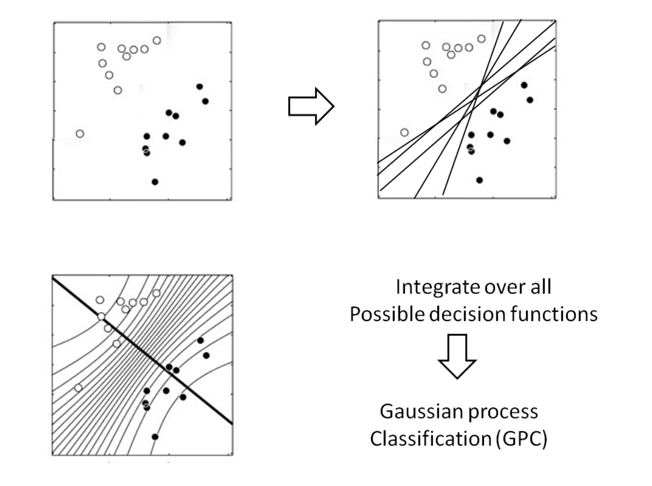
Choice of pattern recognition algorithm



Integrate over all
Possible decision functions

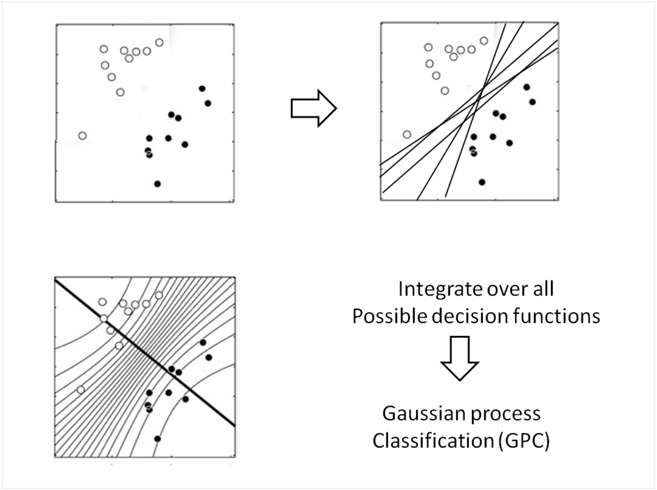
Ashburner and Klöppel (2011)

Choice of pattern recognition algorithm



Ashburner and Klöppel (2011)

Choice of pattern recognition algorithm

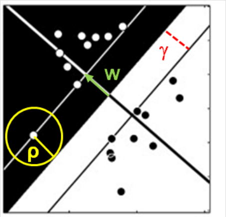


All methods make assumptions!

Ashburner and Klöppel (2011)

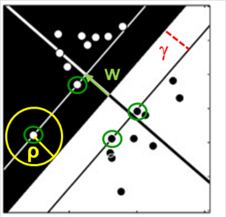
Support Vector Machines

- Finds a separating hyperplane that is “optimal” in that it leads to the largest margin between classes (γ)
- Based on the assumption that each point is bounded by unknown noise (ρ)
- New points will be well classified if $\gamma > \rho$
- The hyperplane is uniquely defined by a subset of the most ambiguous data points (“support vectors”)

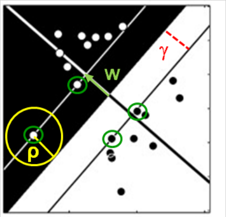


Support Vector Machines

- Finds a separating hyperplane that is "optimal" in that it leads to the largest margin between classes (γ)
- Based on the assumption that each point is bounded by unknown noise (ρ)
- New points will be well classified if $\gamma > \rho$
- The hyperplane is uniquely defined by a subset of the most ambiguous data points ("support vectors")



- Finds a separating hyperplane that is "optimal" in that it leads to the largest margin between classes (γ)
- Based on the assumption that each point is bounded by unknown noise (ρ)
- New points will be well classified if $\gamma > \rho$
- The hyperplane is uniquely defined by a subset of the most ambiguous data points ("support vectors")



$$\min_{\mathbf{w},\xi,b} \quad -\gamma + C\sum_{i=1}^{N}\xi_i$$

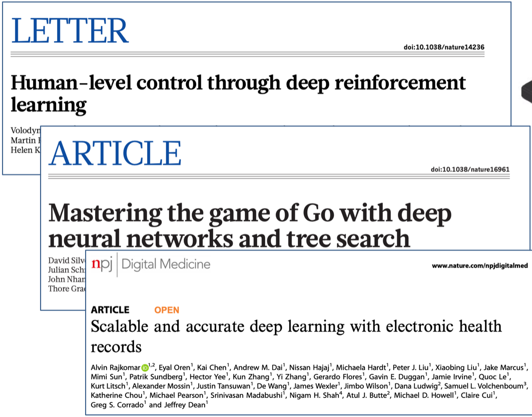
$$\text{s.t.:} \quad y_i(\mathbf{w}^T\phi(\mathbf{x}) + b) > \gamma - \xi_i$$

$$\xi_i > 0$$

$$||\mathbf{w}||^2 = 1$$

Deep Learning

- ‘Deep’ neural networks have seen an enormous surge in popularity over the last few years
- Extend 1950s-era neural networks to have many hidden layers
- Now provide state of the art performance in many domains, e.g. computer vision, game playing and perception


LETTER

doi:10.1038/nature14236

Human-level control through deep reinforcement learning

Volodym
Martin I
Helen K

ARTICLE

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

David Silv
Julian Sch
John Nhan
Thore Grae

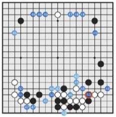
npj | Digital Medicine

www.nature.com/npjdigitalmed

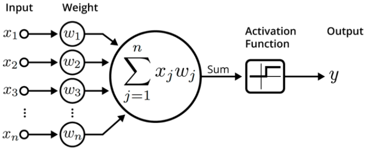
ARTICLE OPEN

Scalable and accurate deep learning with electronic health records

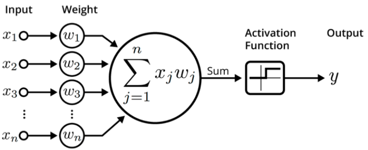
Alvin Rajkomar[1,2], Eyal Oren[1], Kai Chen[1], Andrew M. Dai[1], Nissan Hajaj[1], Michaela Hardt[1], Peter J. Liu[1], Xiaobing Liu[1], Jake Marcus[1], Mimi Sun[1], Patrik Sundberg[1], Hector Yee[1], Kun Zhang[1], Yi Zhang[1], Gerardo Flores[1], Gavin E. Duggan[1], Jamie Irvine[1], Quoc Le[1], Kurt Litsch[1], Alexander Mossin[1], Justin Tansuwan[1], De Wang[1], James Wexler[1], Jimbo Wilson[1], Dana Ludwig[2], Samuel L. Volchenboum[3], Katherine Chou[1], Michael Pearson[1], Srinivasan Madabushi[1], Nigam H. Shah[4], Atul J. Butte[2], Michael D. Howell[1], Claire Cui[1], Greg S. Corrado[1] and Jeffrey Dean[1]



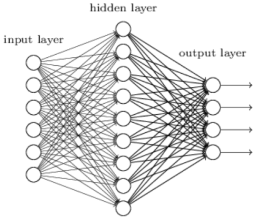
Deep Learning



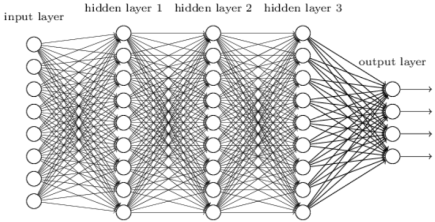
Deep Learning



"Non-deep" feedforward neural network



Deep neural network



- Many variants but "convolutional" networks are popular
- Predominantly supervised learning
- Usually many parameters to optimise (more in lecture 2)

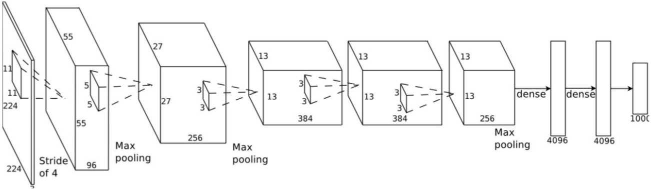
Deep Learning

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
`kriz@cs.utoronto.ca`

Ilya Sutskever
University of Toronto
`ilya@cs.utoronto.ca`

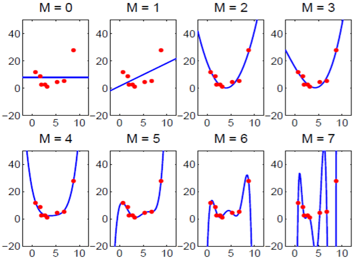
Geoffrey E. Hinton
University of Toronto
`hinton@cs.utoronto.ca`

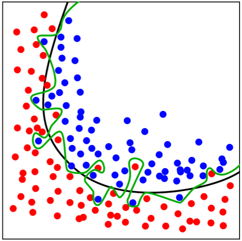


- 7 layer network that won 2012 ImageNet large-scale visual recognition challenge by 10%
- Trained the network on 15 million annotated images from over 22,000 categories
- more than 45,000 citations since 2012!

Overfitting

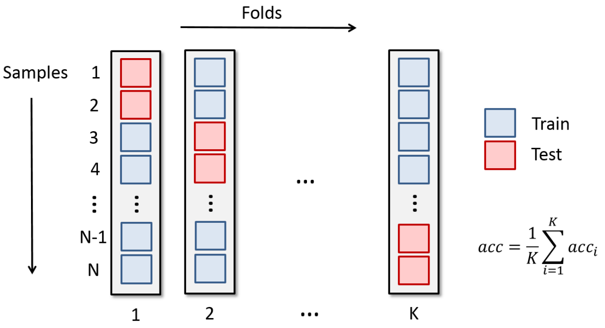
- Occurs when an model performs well on the data that it was estimated or trained on, but poorly on new data
- Can arise in very many ways including improper parameter optimisation or feature selection





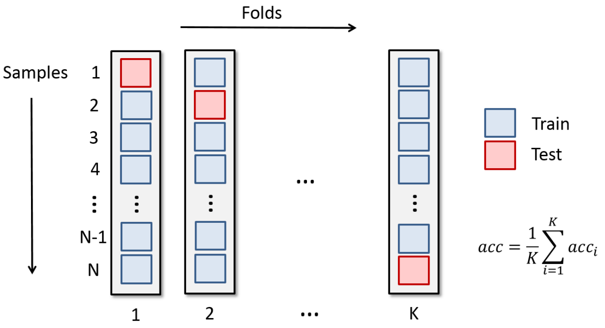
Cross-validation

- Testing on unseen data is essential to assess generalizability
- Cross-validation is the standard way to do this
- 'K-fold CV': split the data into K approximately equal chunks
- 'Leave-one-out': one sample is left out at at a time ($K = N$)



Cross-validation

- Testing on unseen data is essential to assess generalizability
- Cross-validation is the standard way to do this
- 'K-fold CV': split the data into K approximately equal chunks
- 'Leave-one-out': one sample is left out at at a time ($K = N$)



Parameter optimisation

- Most approaches depend on multiple (hyper)parameters
- e.g. regularization parameters in penalized linear models

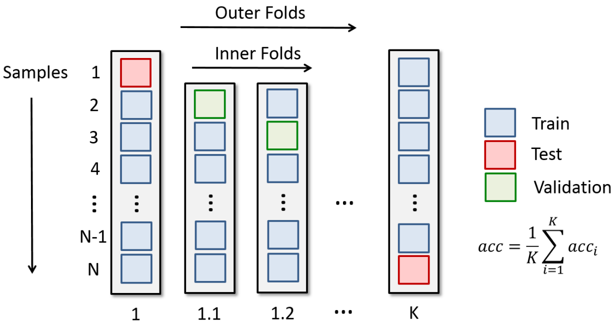
$$\hat{\mathbf{w}} = \min_{\mathbf{w}} \sum_{i=1}^{n} \ell(y_i, f_i) + \lambda J(\mathbf{w})$$

Parameter optimisation

- Most approaches depend on multiple (hyper)parameters
- e.g. regularization parameters in penalized linear models

$$\hat{\mathbf{w}} = \min_{\mathbf{w}} \sum_{i=1}^{n} \ell(y_i, f_i) + \lambda J(\mathbf{w})$$

- Standard approach is *nested* cross-validation with a grid search



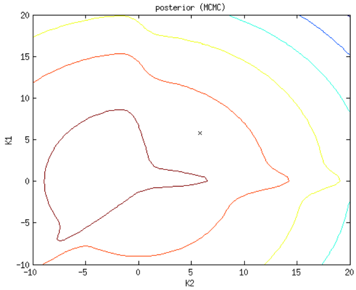
$$acc = \frac{1}{K}\sum_{i=1}^{K} acc_i$$

Parameter optimisation

- Bayesian models also depend on multiple variance/noise hyperparameters

$$p(\mathbf{w}|\mathbf{y},\theta,\sigma) = \frac{p(\mathbf{y}|\mathbf{w},\sigma)p(\mathbf{w}|\theta)}{p(\mathbf{y}|\theta,\sigma)}, \quad p(\mathbf{y}|\theta,\sigma) = \int p(\mathbf{y}|\mathbf{w},\sigma)p(\mathbf{w}|\theta)d\mathbf{w}$$

- Many approaches: nested CV, Empirical Bayes, MCMC ...

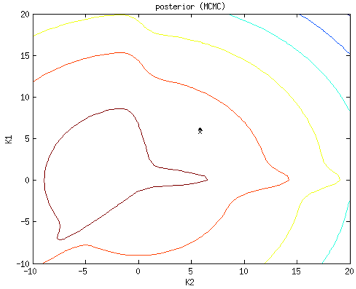


Parameter optimisation

- Bayesian models also depend on multiple variance/noise hyperparameters

$$p(\mathbf{w}|\mathbf{y},\theta,\sigma) = \frac{p(\mathbf{y}|\mathbf{w},\sigma)p(\mathbf{w}|\theta)}{p(\mathbf{y}|\theta,\sigma)}, \quad p(\mathbf{y}|\theta,\sigma) = \int p(\mathbf{y}|\mathbf{w},\sigma)p(\mathbf{w}|\theta)d\mathbf{w}$$

- Many approaches: nested CV, Empirical Bayes, MCMC ...

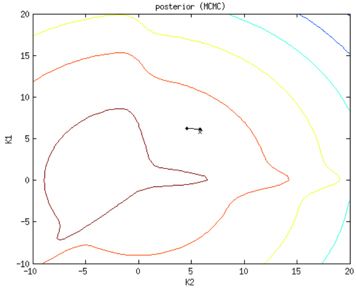


Parameter optimisation

- Bayesian models also depend on multiple variance/noise hyperparameters

$$p(\mathbf{w}|\mathbf{y}, \theta, \sigma) = \frac{p(\mathbf{y}|\mathbf{w}, \sigma)p(\mathbf{w}|\theta)}{p(\mathbf{y}|\theta, \sigma)}, \quad p(\mathbf{y}|\theta, \sigma) = \int p(\mathbf{y}|\mathbf{w}, \sigma)p(\mathbf{w}|\theta)d\mathbf{w}$$

- Many approaches: nested CV, Empirical Bayes, MCMC ...

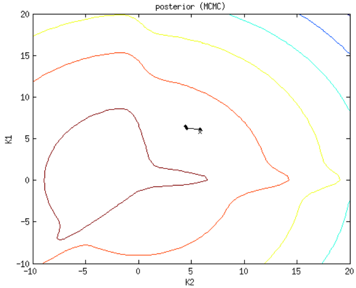


Parameter optimisation

- Bayesian models also depend on multiple variance/noise hyperparameters

$$p(\mathbf{w}|\mathbf{y},\theta,\sigma) = \frac{p(\mathbf{y}|\mathbf{w},\sigma)p(\mathbf{w}|\theta)}{p(\mathbf{y}|\theta,\sigma)}, \quad p(\mathbf{y}|\theta,\sigma) = \int p(\mathbf{y}|\mathbf{w},\sigma)p(\mathbf{w}|\theta)d\mathbf{w}$$

- Many approaches: nested CV, Empirical Bayes, MCMC ...

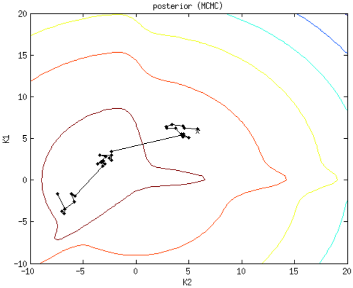


Parameter optimisation

- Bayesian models also depend on multiple variance/noise hyperparameters

$$p(\mathbf{w}|\mathbf{y}, \theta, \sigma) = \frac{p(\mathbf{y}|\mathbf{w}, \sigma)p(\mathbf{w}|\theta)}{p(\mathbf{y}|\theta, \sigma)}, \quad p(\mathbf{y}|\theta, \sigma) = \int p(\mathbf{y}|\mathbf{w}, \sigma)p(\mathbf{w}|\theta)d\mathbf{w}$$

- Many approaches: nested CV, Empirical Bayes, MCMC ...

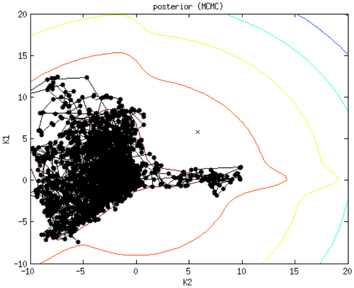


Parameter optimisation

- Bayesian models also depend on multiple variance/noise hyperparameters

$$p(\mathbf{w}|\mathbf{y},\theta,\sigma) = \frac{p(\mathbf{y}|\mathbf{w},\sigma)p(\mathbf{w}|\theta)}{p(\mathbf{y}|\theta,\sigma)}, \quad p(\mathbf{y}|\theta,\sigma) = \int p(\mathbf{y}|\mathbf{w},\sigma)p(\mathbf{w}|\theta)d\mathbf{w}$$

- Many approaches: nested CV, Empirical Bayes, MCMC ...



Statistical assessment of model performance

- Aims to determine if accuracy exceeds "chance" (e.g. 50%)

Statistical tests for binary classification

1. Binomial test
2. T-test based on individual subject accuracies
3. Permutation test

Statistical assessment of model performance

- Aims to determine if accuracy exceeds "chance" (e.g. 50%)

Statistical tests for binary classification

1. Binomial test
2. T-test based on individual subject accuracies
3. Permutation test

- Cross-validation induces dependency between the folds
- The assumptions of the first two methods (e.g. normality) are also usually not met by neuroimaging data

NeuroImage 65 (2013) 69–82

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

ELSEVIER

Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control

Johannes Stelzer [a,*], Yi Chen [a,b], Robert Turner [a]

Outline

Supervised learning for automated diagnosis and prognosis

Neuroscience and Biobehavioral Reviews 57 (2015) 328–349

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

ELSEVIER

Review

From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics

Thomas Wolfers[a,b,*], Jan K. Buitelaar[c,d], Christian F. Beckmann[b,c,e], Barbara Franke[a,f], Andre F. Marquand[b,g]

NeuroImage 145 (2017) 137–165

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

ELSEVIER

Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls

Mohammad R. Arbabshirani[a,b,*], Sergey Plis[a], Jing Sui[a,c], Vince D. Calhoun[a,d]

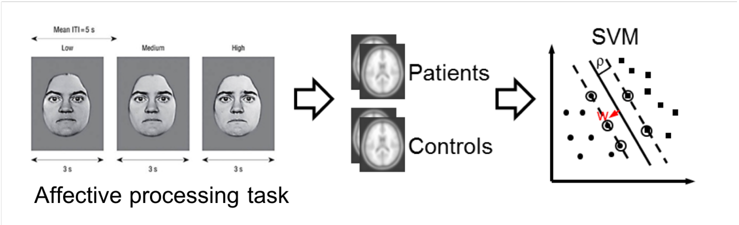
nature neuroscience

Building better biomarkers: brain models in translational neuroimaging

Choong-Wan Woo[1–4], Luke J Chang[5], Martin A Lindquist[6] & Tor D Wager[3,4]

Supervised learning for depression diagnosis

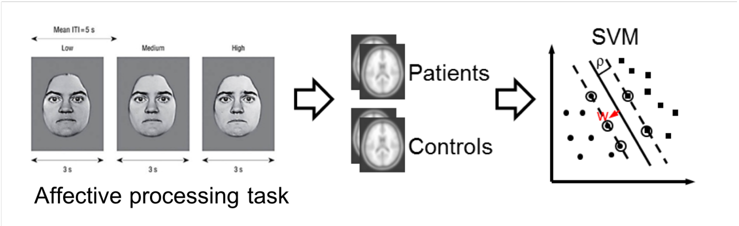
This study was an early application of Pattern recognition to predict disease state in major depression



Fu et al. (2008)

Supervised learning for depression diagnosis

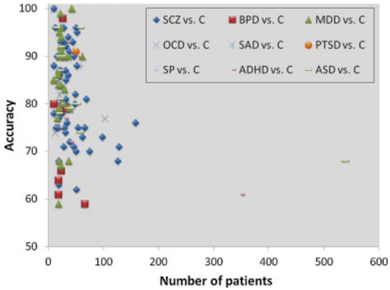
This study was an early application of Pattern recognition to predict disease state in major depression



- Patients could be discriminated from controls with 87% accuracy
- Patients who responded well to fluoxetine could be discriminated from non-responders with 67% accuracy

Fu et al. (2008)

Where are we now?

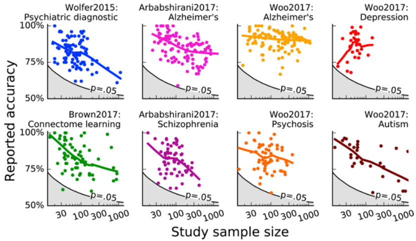


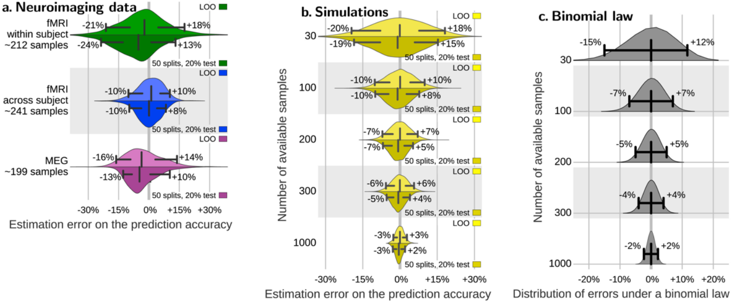
- 116 studies
- Most disorders
- Many data modalities

- Moderate acccuracy, highly variable across studies
- Mostly small samples, minimal validation across cohorts
- Accuracy in small samples is extremely variable
- **Heterogeneity** is a major challenge in clinical cohorts

Wolfers et al. (2015)

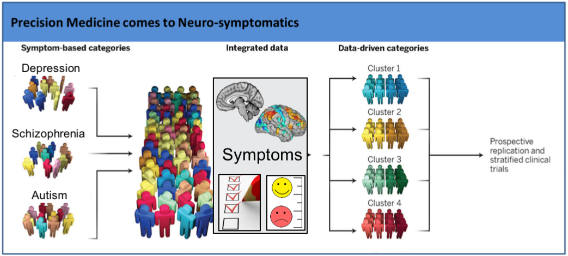
Cross-validation with small samples





Varoquaux (2017)

Research Domain Criteria

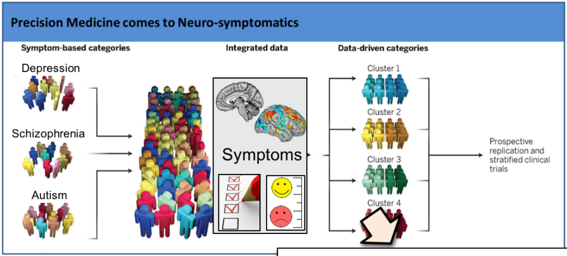


NIH National Institute of Mental Health

RDoC Research Domain Criteria Initiative

Insel and Cuthbert (2015)

Research Domain Criteria



Review

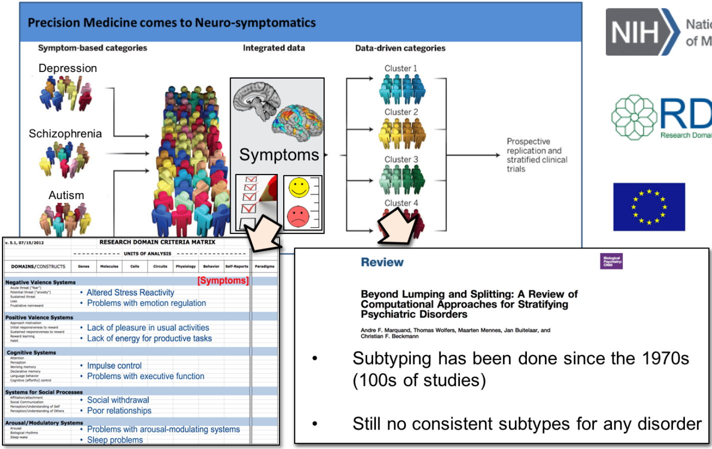
Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders

Andre F. Marquand, Thomas Wolfers, Maarten Mennes, Jan Buitelaar, and Christian F. Beckmann

- Subtyping has been done since the 1970s (100s of studies)
- Still no consistent subtypes for any disorder

Insel and Cuthbert (2015)

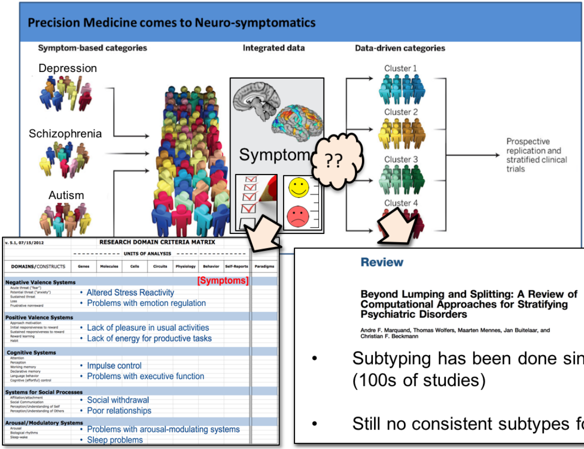
Research Domain Criteria



- Subtyping has been done since the 1970s (100s of studies)
- Still no consistent subtypes for any disorder

Insel and Cuthbert (2015)

Research Domain Criteria



Review

Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders

Andre F. Marquand, Thomas Wolfers, Maarten Mennes, Jan Buitelaar, and Christian F. Beckmann

- Subtyping has been done since the 1970s (100s of studies)
- Still no consistent subtypes for any disorder

Insel and Cuthbert (2015)

Subtyping psychiatric disorders

PNAS

Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD

Damien A. Fair[a,b,c,1], Deep

Departments of [a]Behavioral Neur
97239; [d]Department of Computer
University of Iowa, Iowa City, IA 5

The American Journal of Psychiatry
Current Issue | Archive | About | Residents' Journal | AIP In Advance | Podcast | CME | Author Resources

Back to table of contents Previous Next

Articles Full Access

Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers

Brett A. Clementz, Ph.D.,
Godfrey D. Pearlson, M.D., M

Published Online: 7 Dec 2

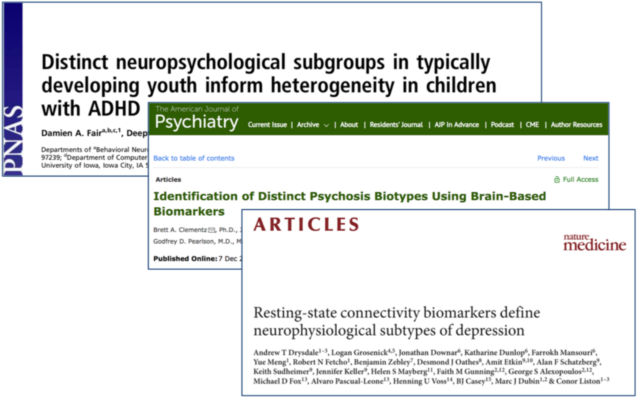
ARTICLES

nature medicine

Resting-state connectivity biomarkers define neurophysiological subtypes of depression

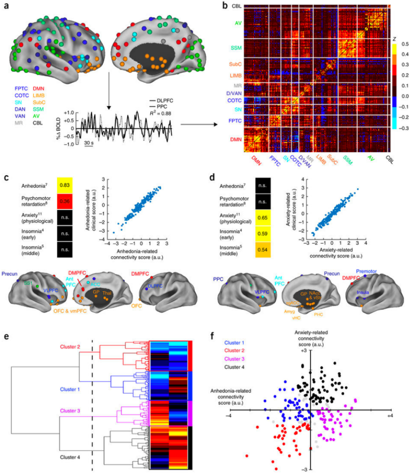
Andrew T Drysdale[1–3], Logan Grosenick[4,5], Jonathan Downar[6], Katharine Dunlop[6], Farrokh Mansouri[6], Yue Meng[1], Robert N Fetcho[1], Benjamin Zebley[7], Desmond J Oathes[8], Amit Etkin[9,10], Alan F Schatzberg[9], Keith Sudheimer[9], Jennifer Keller[9], Helen S Mayberg[11], Faith M Gunning[2,12], George S Alexopoulos[2,12], Michael D Fox[13], Alvaro Pascual-Leone[13], Henning U Voss[14], BJ Casey[15], Marc J Dubin[1,2] & Conor Liston[1–3]

Subtyping psychiatric disorders



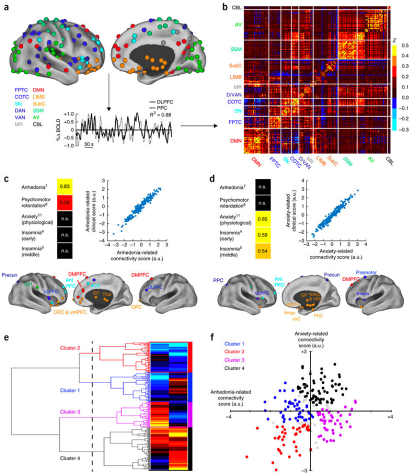
- Generally minimal validation or replication
- Do not test against the 'null' hypothesis that there are no clusters in the data
- Clustering on the basis of symptoms may not map onto the underlying biology

Stratification of major depression

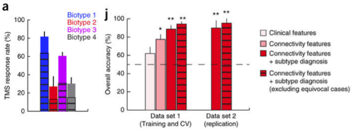


Drysdale et al. (2017)

Stratification of major depression

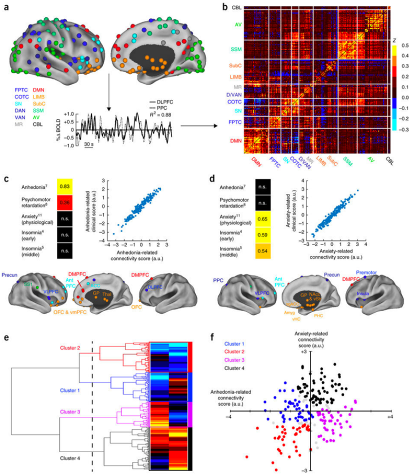


- Extensive validation
- Predict treatment response (TMS)

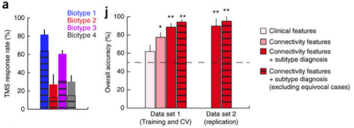


Drysdale et al. (2017)

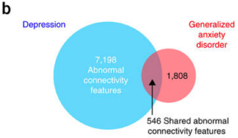
Stratification of major depression



- Extensive validation
- Predict treatment response (TMS)



- Cut across diagnoses



- Sounds good, but we will examine some problems with this study in lecture 2

Drysdale et al. (2017)

Outline

Take home messages

- PR is a powerful tool to perform single subject inference and detect spatially distributed effects
- Useful in clinical neuroscience for:
  1. Making predictions at the subject level (e.g. prognosis)
  2. Stratifying psychiatric disorders
- Validation of models is extremely important to ensure generalisability

References

J. Ashburner and S. Klöppel. Multivariate models of inter-subject anatomical variability. *NeuroImage*, 56(2): 422–439, 2011.

Richard Dinga, Lianne Schmaal, Brenda W.J.H. Penninx, Marie Jose van Tol, Dick J. Veltman, Laura van Velzen, Maarten Mennes, Nic J.A. van der Wee, and Andre F. Marquand. Evaluating the evidence for biotypes of depression: Methodological replication and extension of drysdale et al. (2017). *NeuroImage: Clinical*, 22: 101796, 2019. ISSN 2213-1582. doi: https://doi.org/10.1016/j.nicl.2019.101796. URL http://www.sciencedirect.com/science/article/pii/S2213158219301469.

Andrew T Drysdale, Logan Grosenick, Jonathan Downar, Katharine Dunlop, Farrokh Mansouri, Yue Meng, Robert N Fetcho, Benjamin Zebley, Desmond J Oathes, Amit Etkin, Alan F Schatzberg, Keith Sudheimer, Jennifer Keller, Helen S Mayberg, Faith M Gunning, George S Alexopoulos, Michael D Fox, Alvaro Pascual-Leone, Henning U Voss, BJ Casey, Marc J Dubin, and Conor Liston. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, (23):28–38, 2017.

C. H. Fu, J. Mourao-Miranda, S. G. Costafreda, A. Khanna, A. F. Marquand, S. C. Williams, and M. J. Brammer. Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. *Biol Psychiatry*, 63(7):656–62, 2008.

J. D. Haynes and G. Rees. Decoding mental states from brain activity in humans. *Nat Rev Neurosci*, 7(7):523–34, 2006.

Thomas R. Insel and Bruce N. Cuthbert. Brain disorders? precisely. *Science*, 348(6234):499–500, 2015. ISSN 0036-8075.

Gael Varoquaux. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage*, (In press), 2017.

T. Wolfers, J. K. Buitelaar, C. F. Beckmann, B. Franke, and A. F. Marquand. From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neuroscience and Biobehavioral Reviews*, in press, 2015.