Machine Learning I: basics

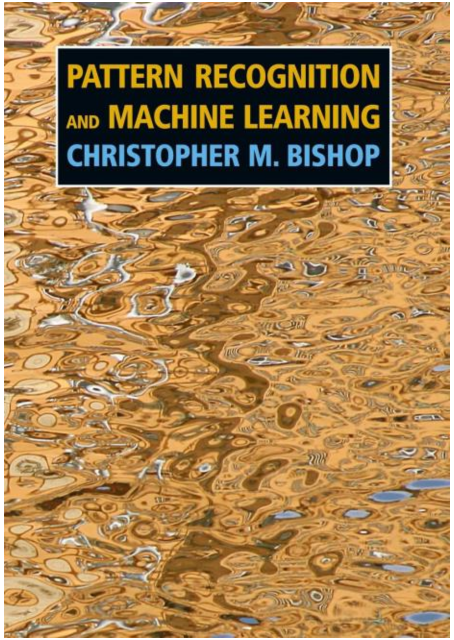
Barbora Rehák Bučková

Barbora.Rehak-Buckova@radboudumc.nl

Machine Learning and Pattern Recognition

"The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories"

-Bishop 2006-



- Pattern recognition has its origins in engineering, whereas machine learning grew out of computer science
- Automatically learns patterns

Machine Learning and Pattern Recognition

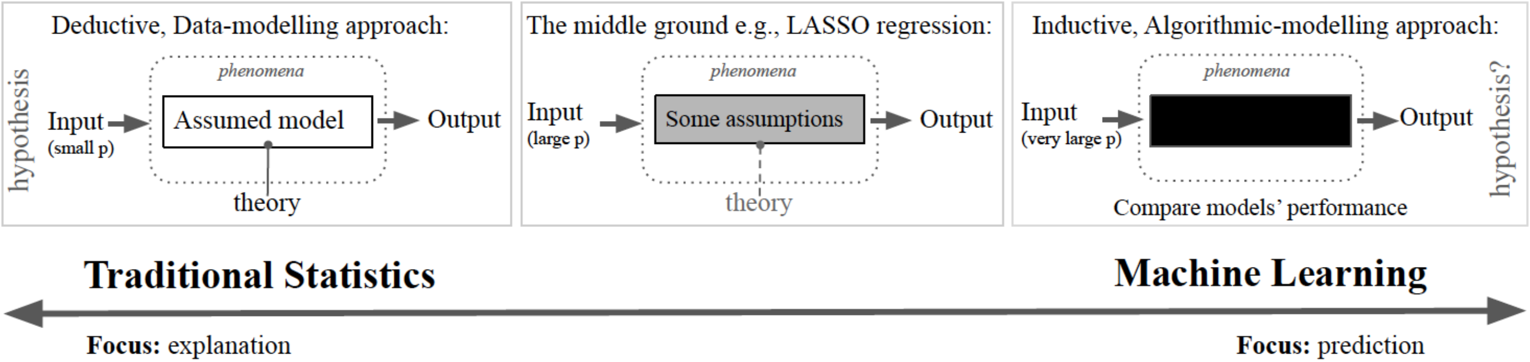


Figure 2: The continuum between traditional statistics and machine learning. p = set of possible predictors.

Traditional statistical approach:

the focus is on applying a theoretically conceived model to data and understanding the data generation mechanisms.

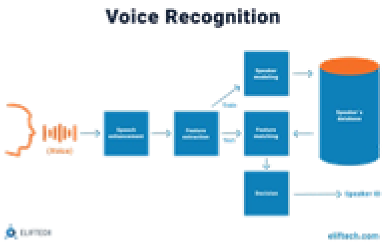
Machine learning approach:

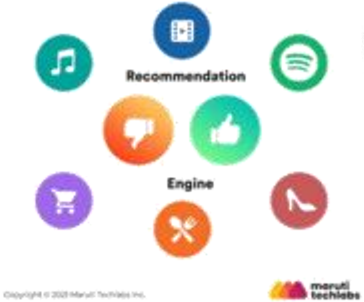
where the focus is to make a prediction from the input to the output without assuming a theoretically motivated model."

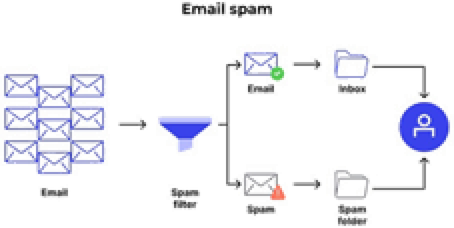
What is pattern recognition used for?

Historically, has been applied in many domains:

- Speech Recognition
- Automatic Character recognition / handwriting recognition
- Document classification (e.g. spam filters)
- Analysis of genetic microarray data
- Recommender systems / online shopping
- Natural language processing (NLP)







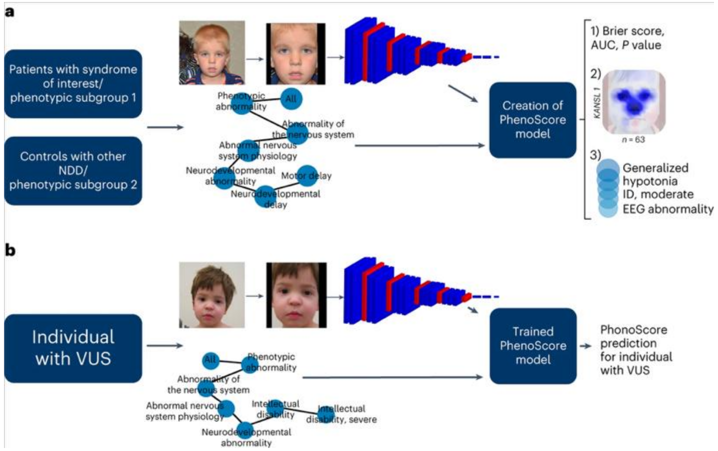
Machine Learning and Pattern Recognition

Increasingly used in computational psychiatry for:

1. Predicting clinical variables (diagnosis or treatment response)

2. Stratifying psychiatric disorders

3. Learning mappings between behaviour and brain systems

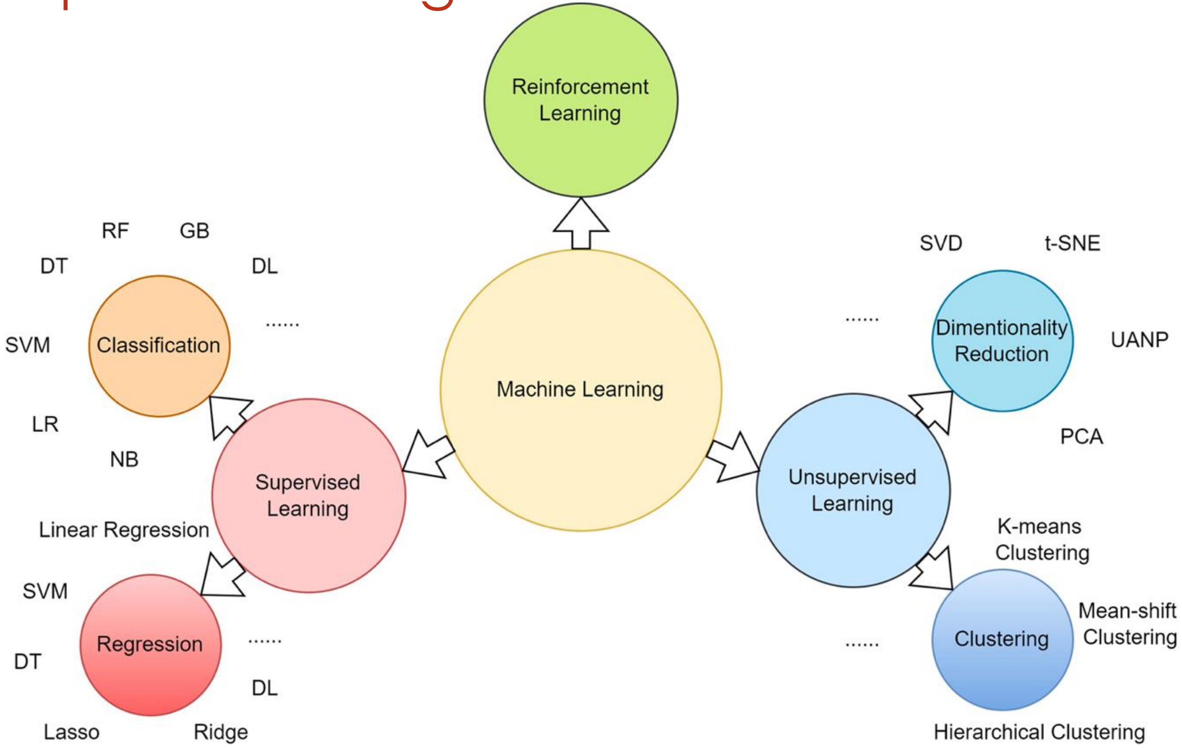


Outline

1. Introduction to Machine Learning
2. Basics of Pattern Recognition Analyses
3. Applications in Psychiatry
4. Conclusions

Introduction to Machine Learning

Types of pattern recognition



Types of pattern recognition

Supervised learning

learning a mapping between input and output:

Input Training examples $\{x_1, x_2, \dots, x_N\}$

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

Output Target variables $\{y_1, y_2, \dots, y_N\}$

$y_i \in \mathbb{R}$ *Regression*

$y_i \in \{0,1\}$ *Classification*

Unsupervised learning

algorithms are not provided with output labels and must learn to structure the data by applying heuristics:

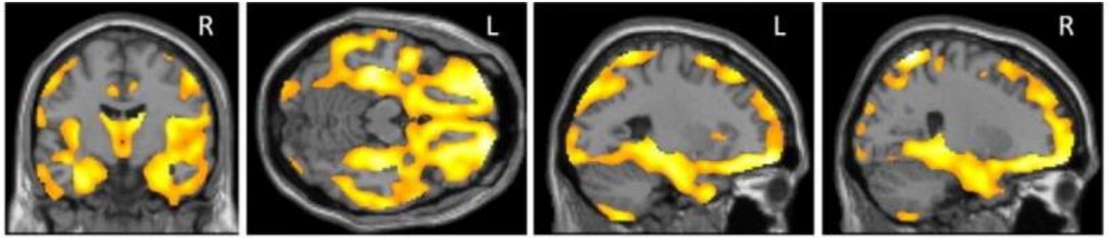
Input Training examples $\{x_1, x_2, \dots, x_N\}$

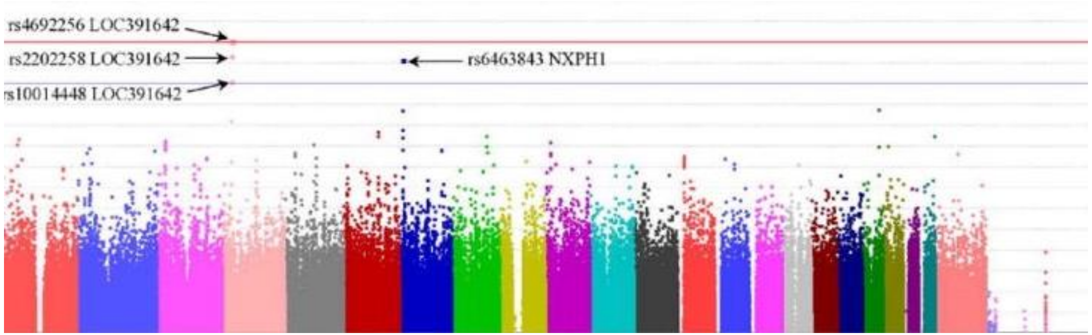
Output NO Target variables

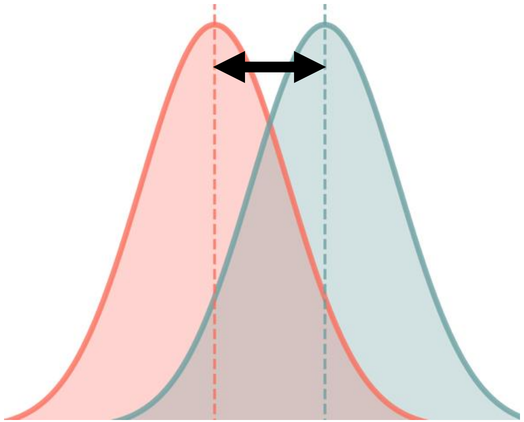
Biomarkers for prediction

Shen et al. (2010)

Mass univariate association testing (SPM, GWAS)







testing for group differences

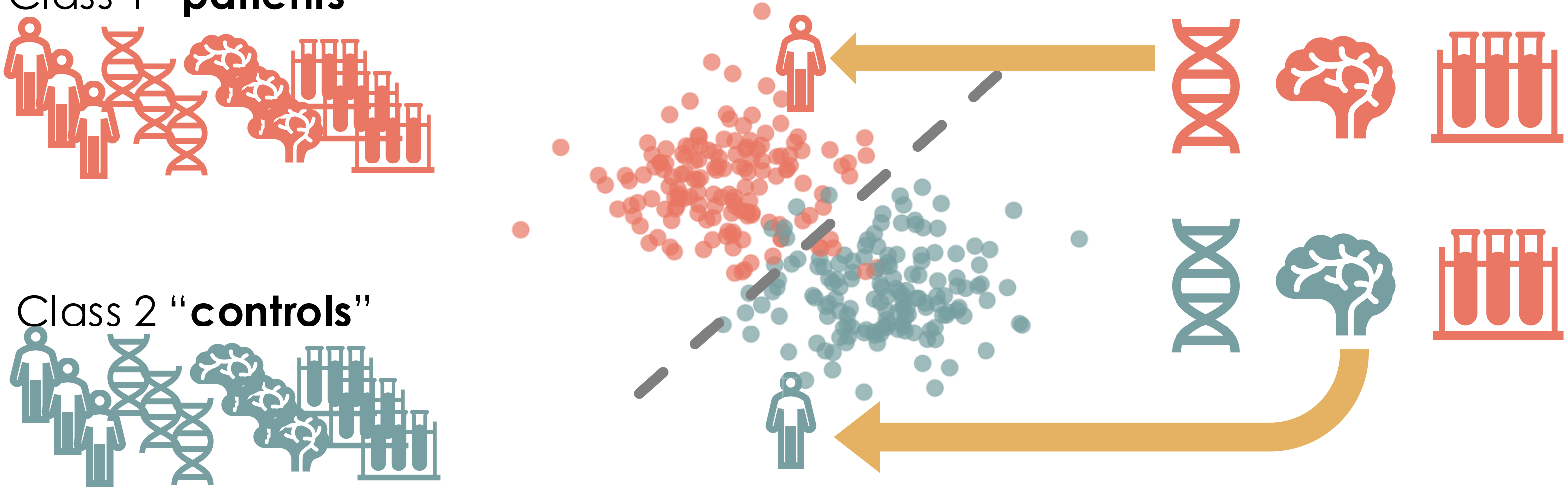
Predicting disease state

Making subject level predictions of diagnosis and outcome

Useful to find a measure of overall group separation

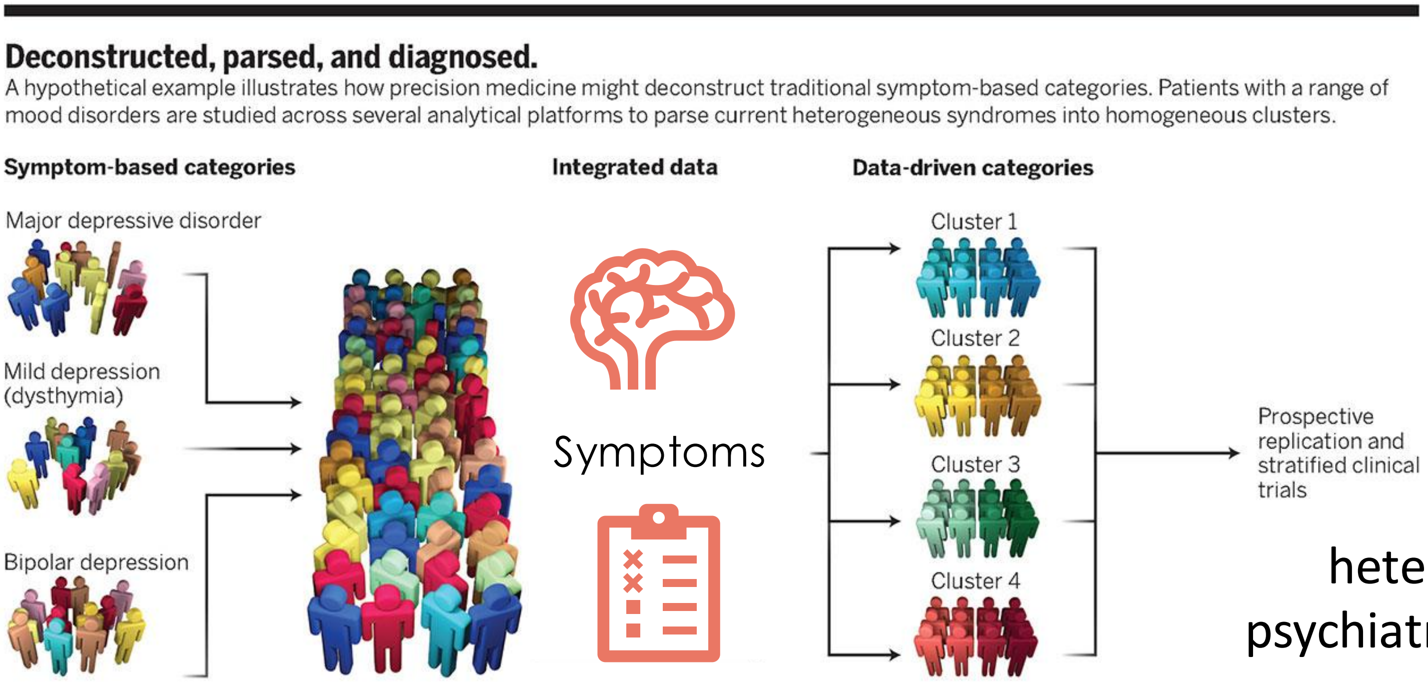
Class 1 "**patients**"

Class 2 "**controls**"



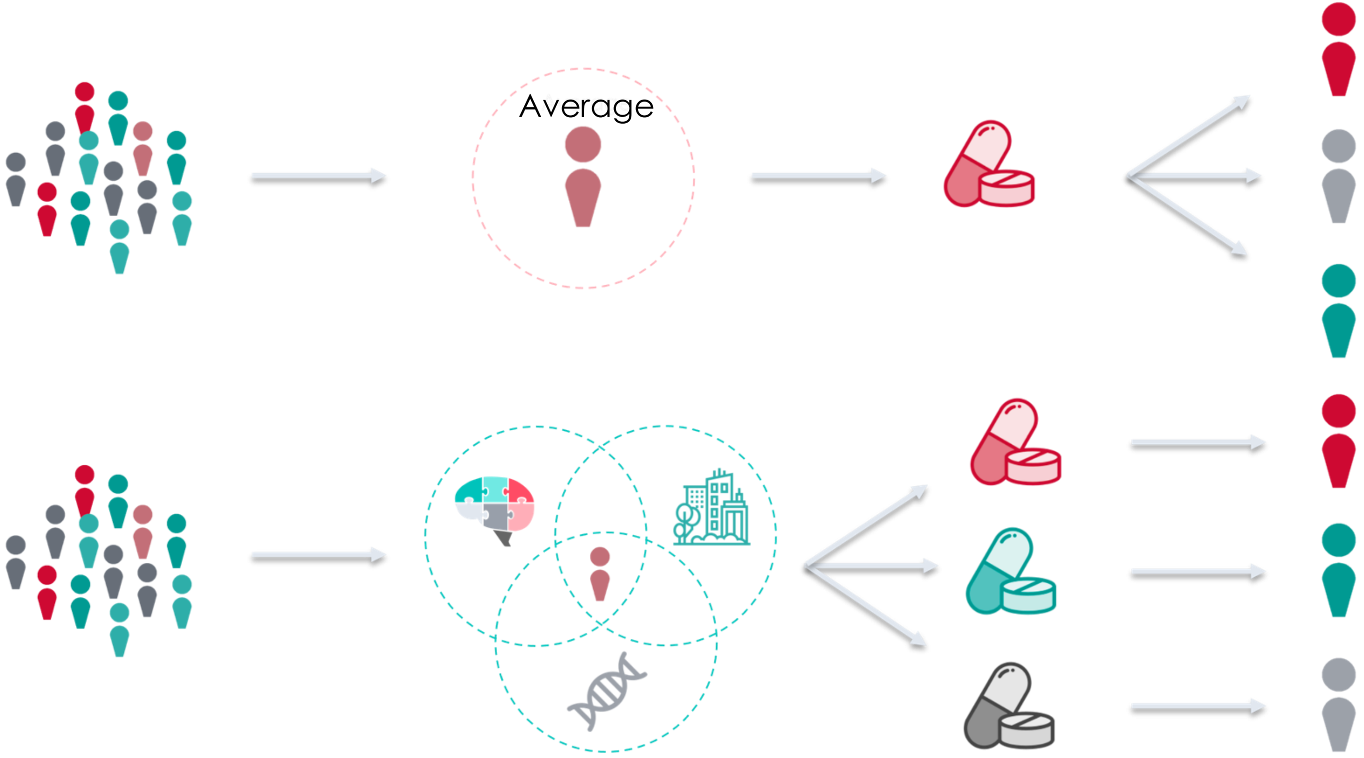
Subject level inference

Machine Learning in Psychiatry: Stratification



Tackling the clinical and biological heterogeneity of psychiatric disorders

Goal: Precision Medicine



Outline

1. Introduction to Machine Learning
2. Basics of Pattern Recognition Analyses
3. Applications in Psychiatry
4. Conclusions

Naselaris et al. (2011)

Basics of Pattern Recognition Analyses

Encoding and Decoding

Cognitive state	Model	Brain activity
	Encoding $Y = f(X)$	Which brain regions are activated in certain task condition?
	$X = f(Y)$ **Decoding**	To which clinical group does the subject belong?

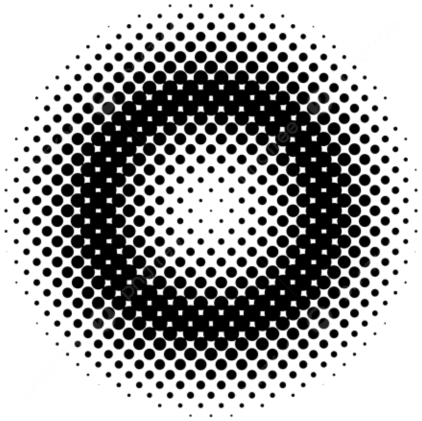
Cognitive state:

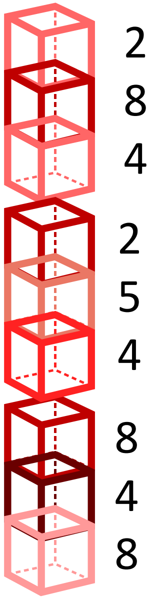
- Response to stimulus
- Cognitive scores
- Diagnostic labels
- Symptom scores
- …

- **Encoding** models use stimuli to predict activity
- **Decoding** models use activity to predict information about the stimuli

Multivariate models

Sensitivity for spatially distributed (or multivariate) effects

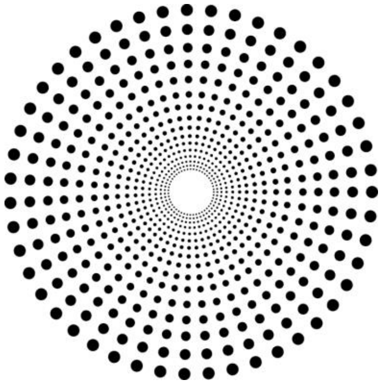


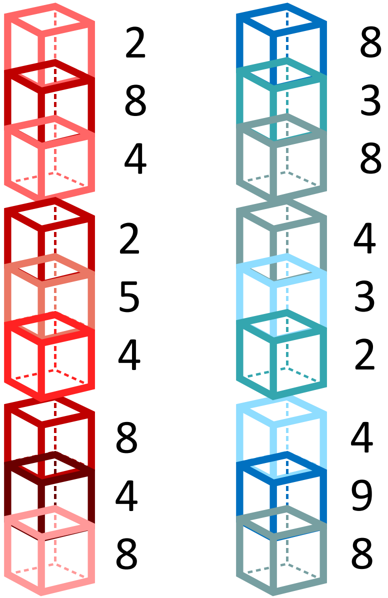


Basics of Pattern Recognition Analyses

Multivariate models

Sensitivity for spatially distributed (or multivariate) effects



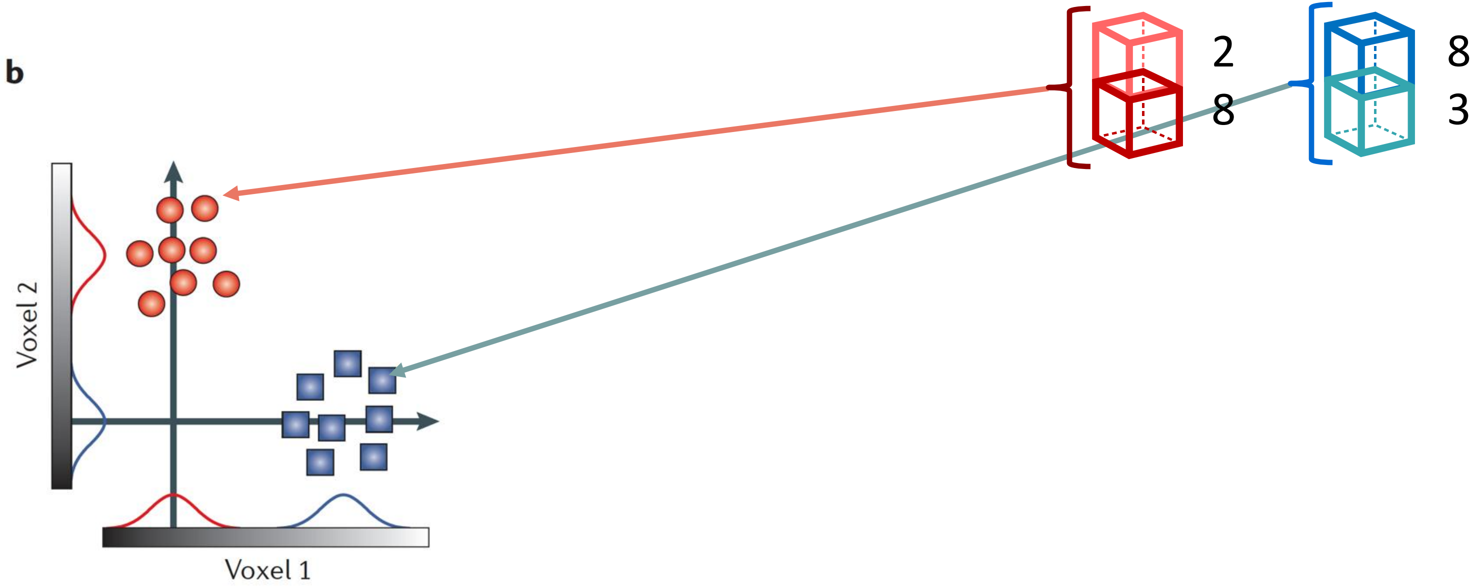


Haynes & Rees (2006)

Basics of Pattern Recognition Analyses

Multivariate models

Sensitivity for spatially distributed (or multivariate) effects

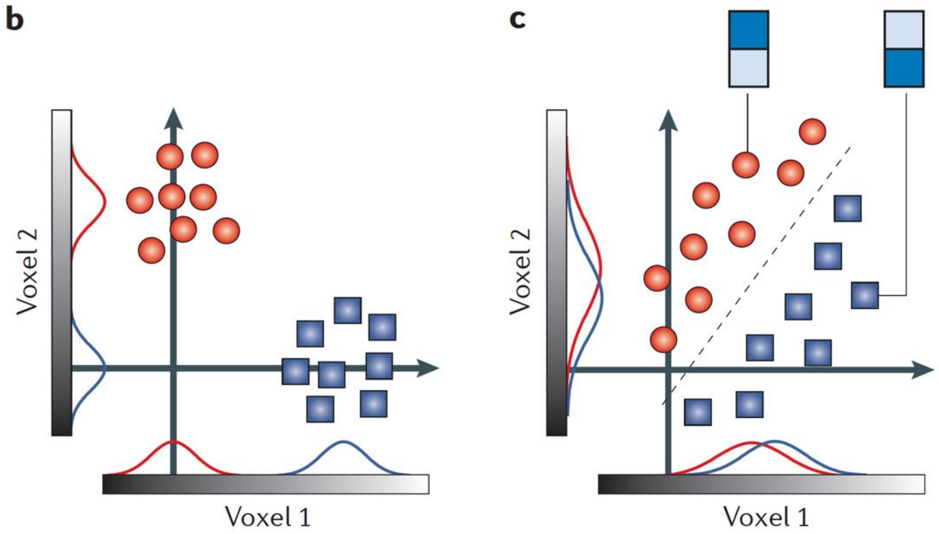


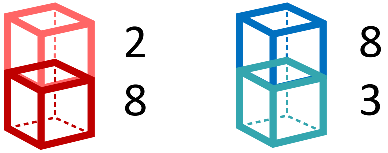
Haynes & Rees (2006)

Basics of Pattern Recognition Analyses

Multivariate models

Sensitivity for spatially distributed (or multivariate) effects





Stages of supervised pattern recognition analysis

Feature **extraction** and/or **selection**

Classification/ Regression using cross-validation

Performance evaluation

Stages of supervised pattern recognition analysis

Feature **extraction** and/or **selection**

Curse of dimensionality

- When the dimensionality increases, the volume of the space increases so fast that the available data become **sparse**
- To obtain a statistically significant result, the amount of the data needed often grows **exponentially**

- **100 voxels -> 5 false positives**
- **1,000,000 voxels -> 50,000 false positives**

Stages of supervised pattern recognition analysis

Feature **extraction** and/or **selection**

Feature **construction** — PCA, ICA

Feature **learning** — deep learning

Whole brain

Searchlight

Region of interest

Feature selection

Stages of supervised pattern recognition analysis

- Feature **extraction** and/or **selection**
- Feature **construction** — PCA, ICA
- Feature **learning** — deep learning

TRAINING DATA ONLY

Stages of supervised pattern recognition analysis

Feature **extraction** and/or **selection**

Classification/ Regression using cross-validation

Preliminaries

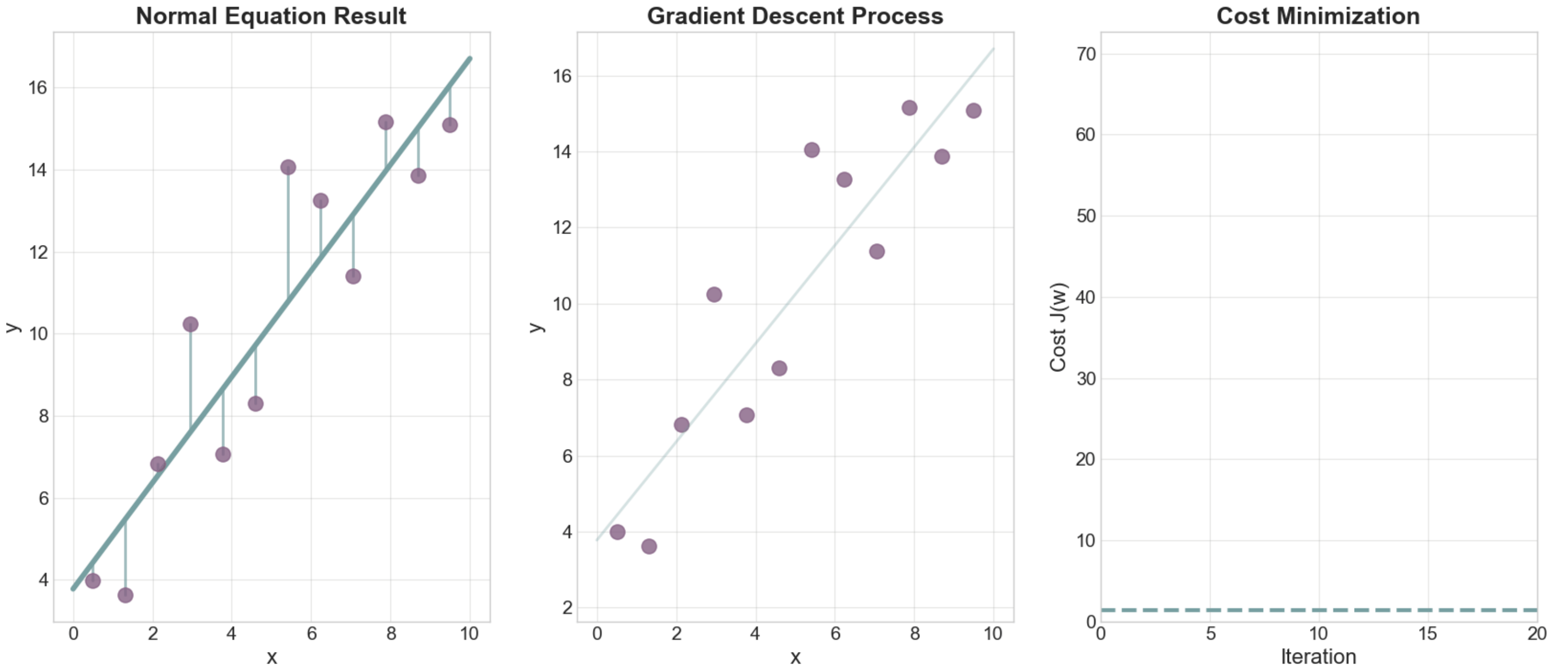
Notation

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

Approaches to Supervised Pattern Recogniton: Linear Regression

- Model Representation $\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_D x_D$
- Cost Function (Mean Squared Error) $$\mathrm{l}(\mathbf{w}) = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2$$
- Method 1: Normal equation $\mathbf{w} = (X^T X)^{-1} X^T y$
- Method 2: Optimization Update Rule (Gradient Descent) $$w_j := w_j - \alpha \frac{1}{m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})x_j^{(i)}$$

Approaches to Supervised Pattern Recogniton: Linear Regression



Bayesian Approach

- Bayesian models also repend on multiple variance/noise hyperparameters

Conditional probability (likelihood)

Prior

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

Posterior

Bayesian Approach

- Bayesian models also repend on multiple variance/noise hyperparameters

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

Given weights w and noise level σ, how probable is this data y

What do we believe about w before seeing data

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

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

Normal equation $\quad \mathbf{w} = (X^T X)^{-1} X^T y$

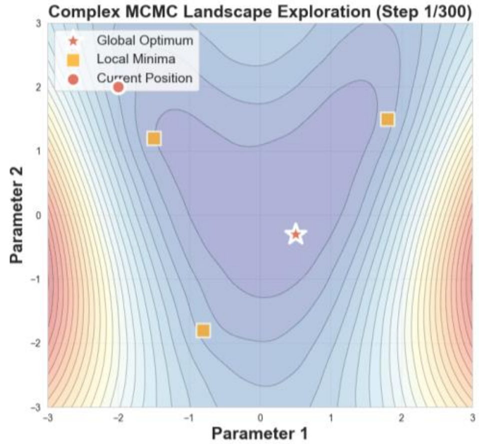
Bayesian parameter optimization

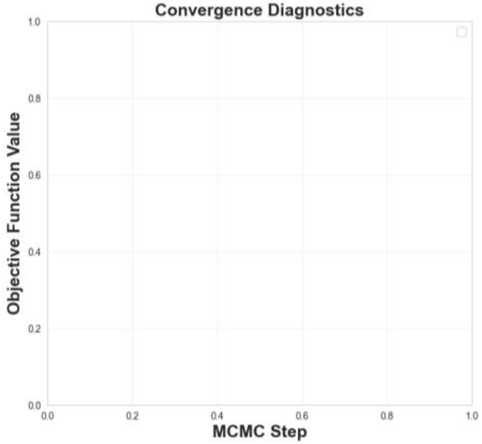
- Bayesian model also depend on multiple variance/noise hyperparameters

$$p(\boldsymbol{w}|\boldsymbol{y},\theta,\sigma) = \frac{p(\boldsymbol{y}|\boldsymbol{w},\sigma)p(\boldsymbol{w}|\theta)}{p(\boldsymbol{y}|\theta,\sigma)},$$

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

- Many approaches: nested cross validation, Empirical Bayes, Monte Carlo MC





A smorgasbord of different approaches

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

- 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,…

$$\max(\mathrm{p}(\boldsymbol{w}|\boldsymbol{y})) = \max_w (\mathrm{p}(\boldsymbol{y}|\boldsymbol{w})\mathrm{p}(\boldsymbol{w}))$$

$$= \max_w \left(\ln\left(\mathrm{p}(\boldsymbol{y}|\boldsymbol{w})\right) + \ln\left(\mathrm{p}(\boldsymbol{w})\right) \right)$$

$$= \min_w \left(-\ln\left(\mathrm{p}(\boldsymbol{y}|\boldsymbol{w})\right) - \ln\left(\mathrm{p}(\boldsymbol{w})\right)\right)$$

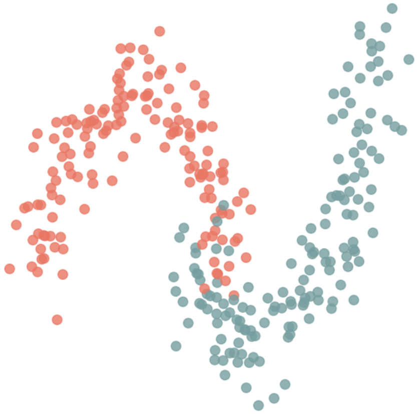
$$= \min_w \sum_{i=1}^{n} -\ln\left(\mathrm{p}(y_i|\boldsymbol{w})\right) - \ln(\mathrm{p}(\boldsymbol{w}))$$

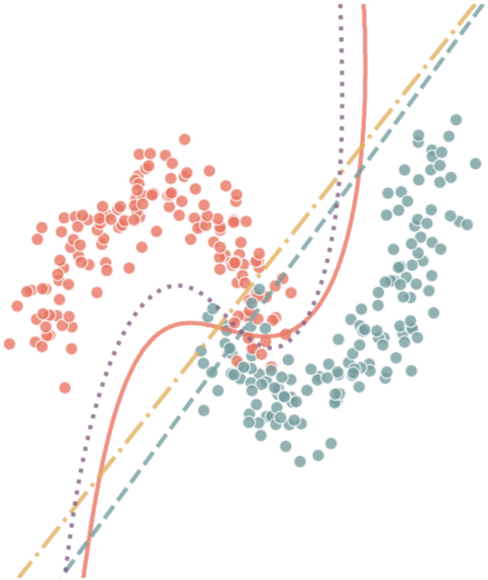
$$= \min_w - \left(\sum_{i=1}^{\mathrm{n}} \mathrm{loss}(y_i|\boldsymbol{w}) + \lambda J(\boldsymbol{w})\right)$$

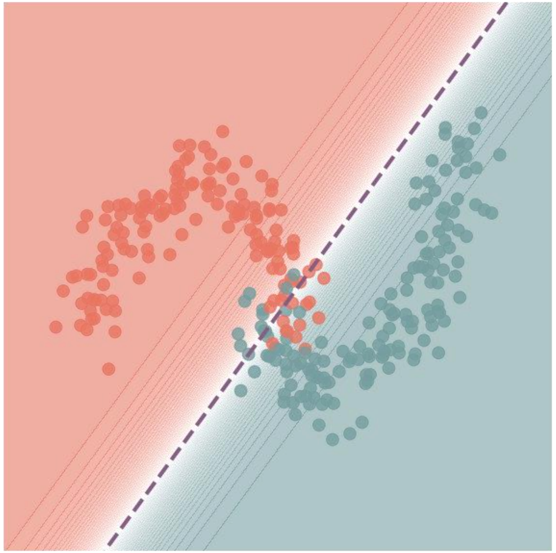
When you do regularized regression, you're secretly placing priors on weights!

Basics of Pattern Recognition Analyses

Choice of pattern recognition algorithm





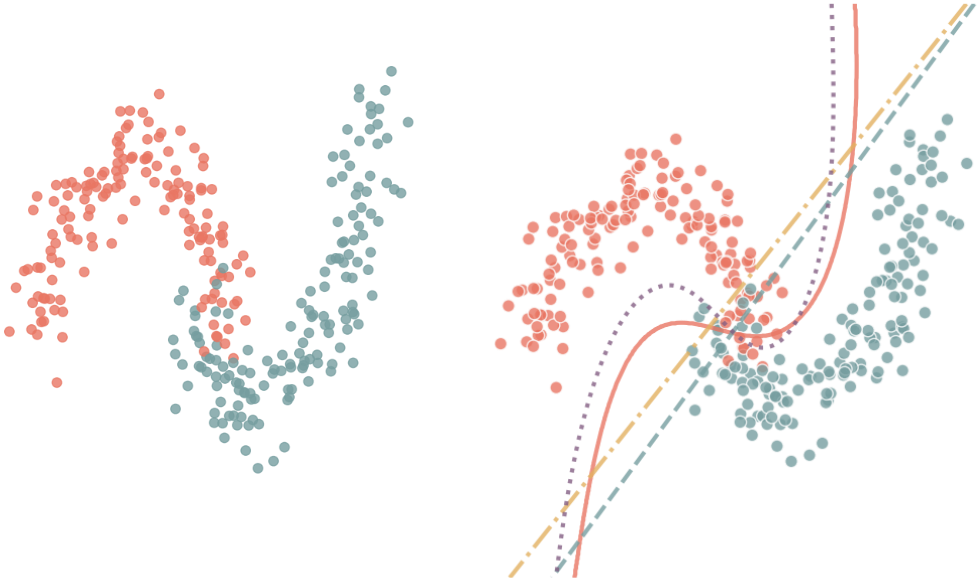


The relationship is linear, weights shouldn't be too large

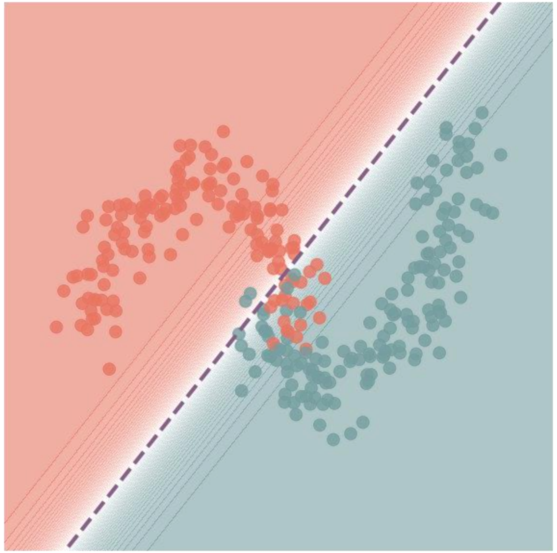
Logistic Regression

Model log-odds ratio between classes

Choice of pattern recognition algorithm



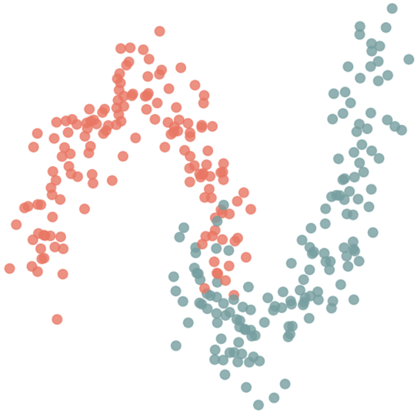
Classes are Gaussian clouds with same shape

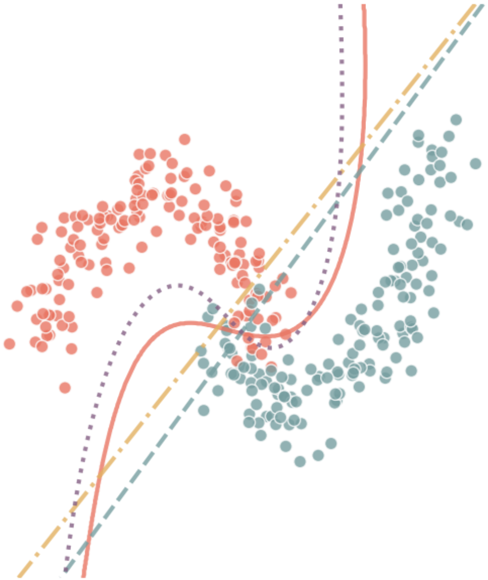


Linear Discriminant Analysis

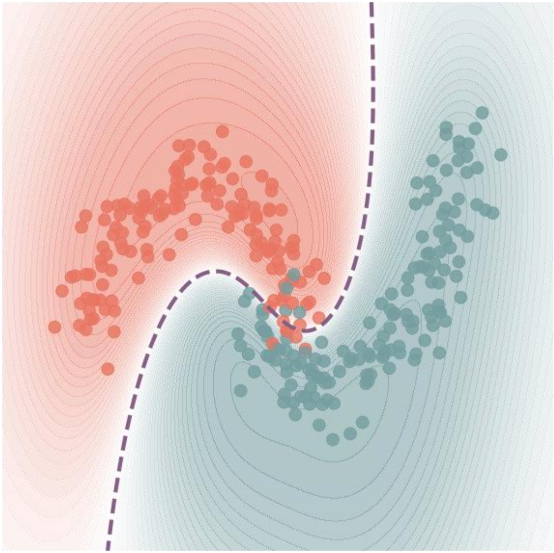
Maximise between to within class variance

Choice of pattern recognition algorithm





The decision function is smooth in a way defined by the kernel

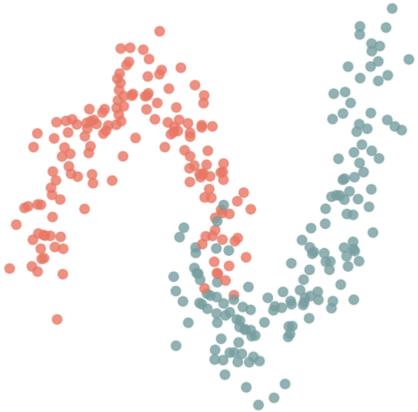


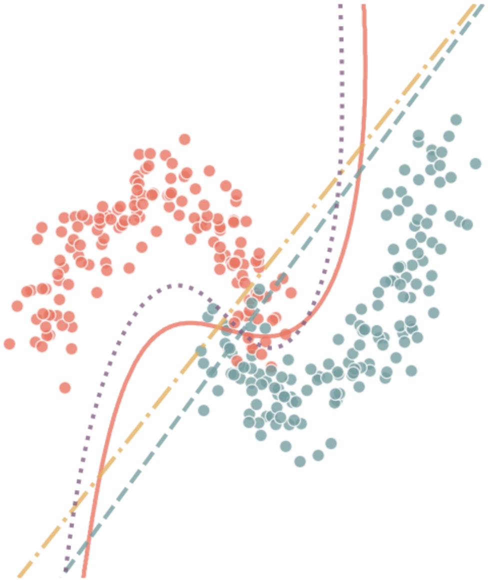
Gaussian Process Classification

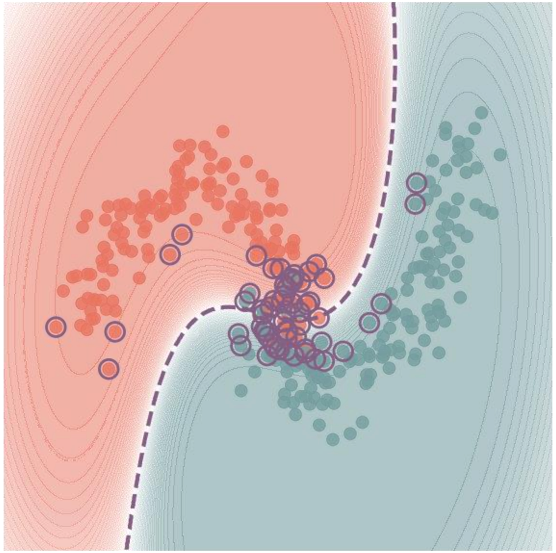
Integrate over all possible decision functions

Basics of Pattern Recognition Analyses

Choice of pattern recognition algorithm







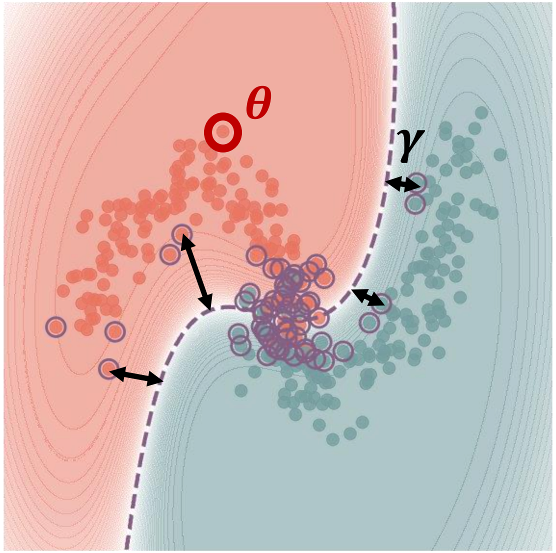
There's a clear margin between classes

Support Vector Machine

Maximise the margin between the classes

Choice of pattern recognition algorithm

- 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 (θ)
- 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 Machine

Maximise the margin between the classes

Deep learning

- "Deep" neural networks have seen an enormous surge in popularity over the last few years
- Extend 1950's-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

Human-level control through deep reinforcement learning

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg & Demis Hassabis

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

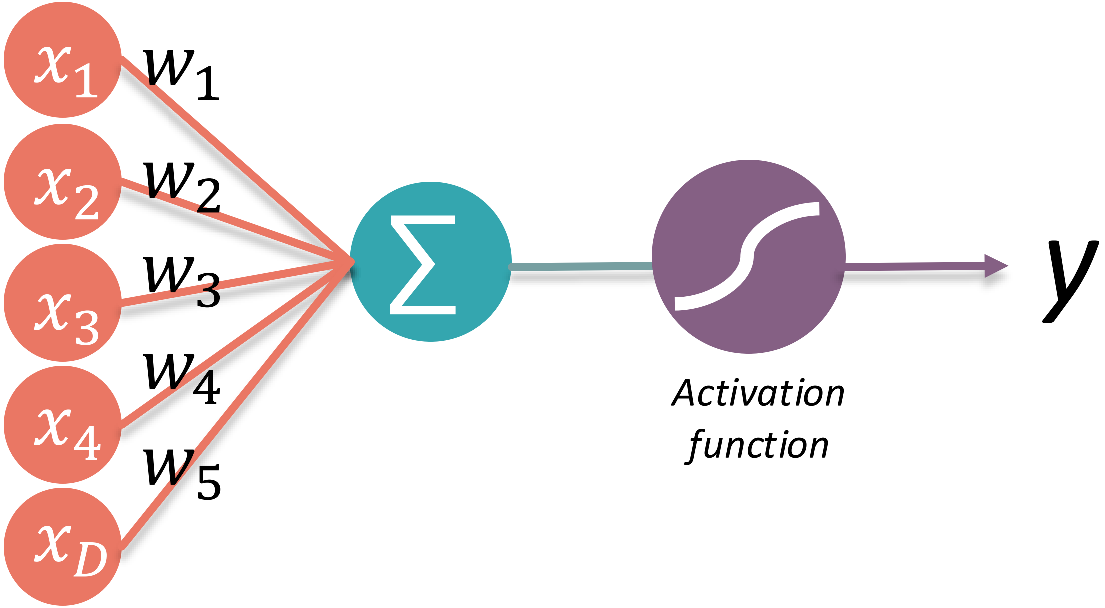
David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis

Scalable and accurate deep learning with electronic health records

Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, Peter J. Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, Patrik Sundberg, Hector Yee, Kun Zhang, Yi Zhang, Gerardo Flores, Gavin E. Duggan, Jamie Irvine, Quoc Le, Kurt Litsch, Alexander Mossin, Justin Tansuwan, De Wang, James Wexler, Jimbo Wilson, ... Jeffrey Dean + Show authors

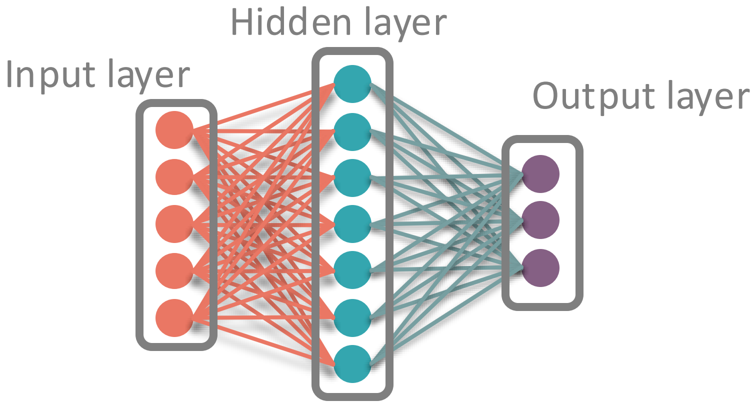
Deep learning

- Activation functions introduce non-linearity (Sigmoid – Logistic Function)
- Predominantly supervised learning
- Usually many parameters to optimise (more in lecture 2)



Deep learning

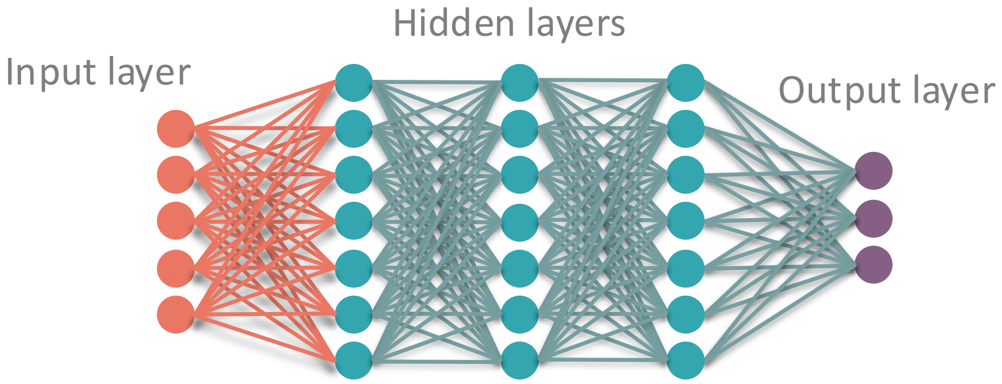
- Many variants but "convolutional" networks are popular
- Predominantly supervised learning
- Usually many parameters to optimise (more in lecture 2)



Shallow neural network

Deep learning

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

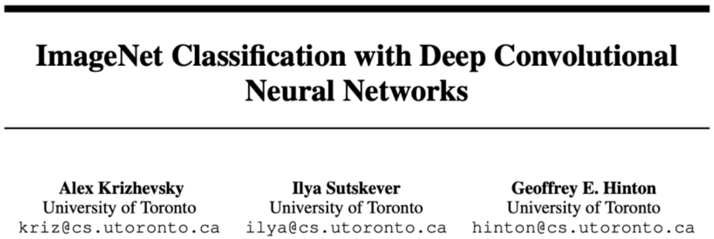


Deep neural network

Krizhevsky et al. (2012)

Basics of Pattern Recognition Analyses

Deep learning

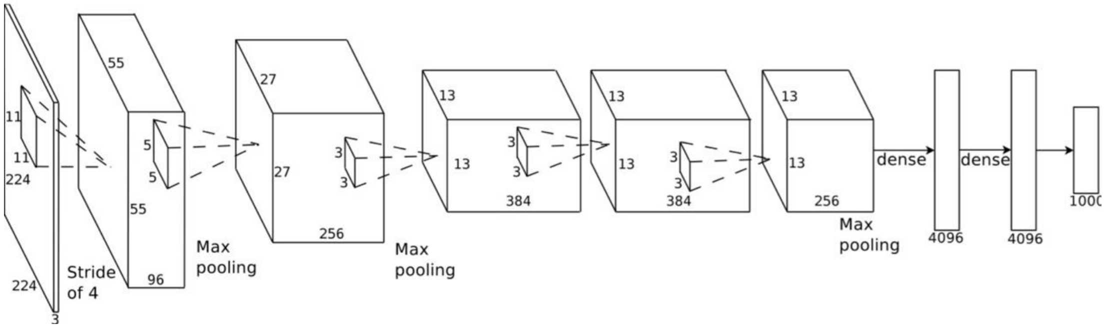


ImageNet Classification with Deep Convolutional Neural Networks

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 won 2012 ImageNet large-scale visual recognition challenge by 10%
- Trained the network on 15 milion annotated images from over 22,000 categories
- More than 93,000 citations since 2012!

Dangers of Deep Learning

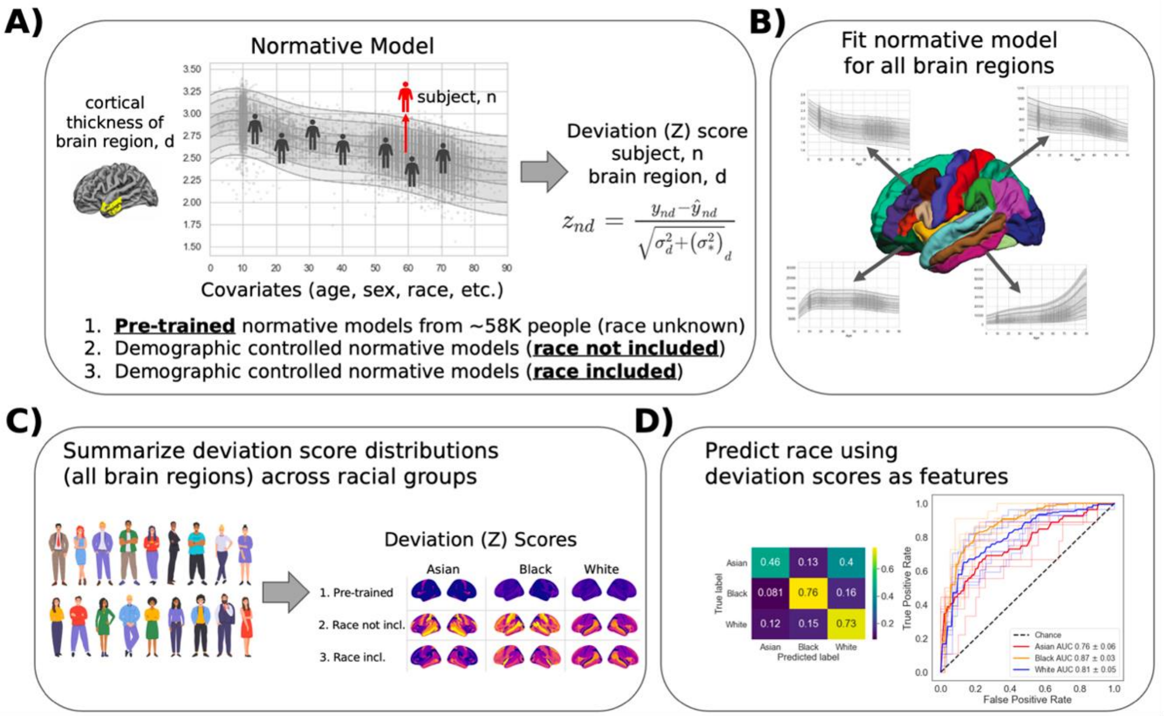
Henrich et al. (2010)

1. Data Bias
   - Western, Educated, Industrialized, Rich and Democratic (WEIRD)
2. Black Box: lack of transparency
3. Overfitting: perform poorly on unseen data
4. High Complexity
   - Resource intensive and hard to interpret
5. Adversarial Attacks
   - Small changes in data can cause large prediction errors
6. Ethical concerns
7. Environmental Impact
8. Data Privacy

Basics of Pattern Recognition Analyses

Data Bias

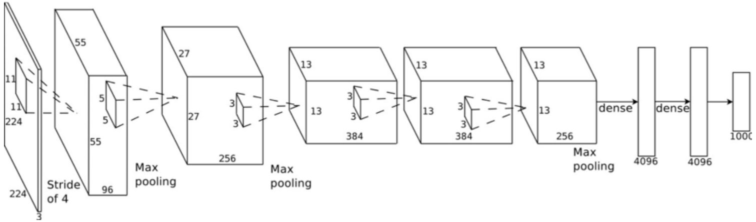
- Persistent racial biases are present in brain models, even when race is accounted for, suggesting that more flexible and representative modeling techniques are necessary.



Basics of Pattern Recognition Analyses

Dangers of Deep Learning

1. Data Bias
   1. Western, Educated, Industrialized, Rich and Democratic (WEIRD)
2. Black Box: lack of transparency
3. Overfitting: perform poorly on unseen data



4. High Complexity
   4. Resource intensive and hard to interpret
5. Adversarial Attacks
   4. Small changes in data can cause large prediction errors
6. Ethical concerns

7. Environmental Impact
8. Data Privacy

Dangers of Deep Learning

Henrich et al. (2010)

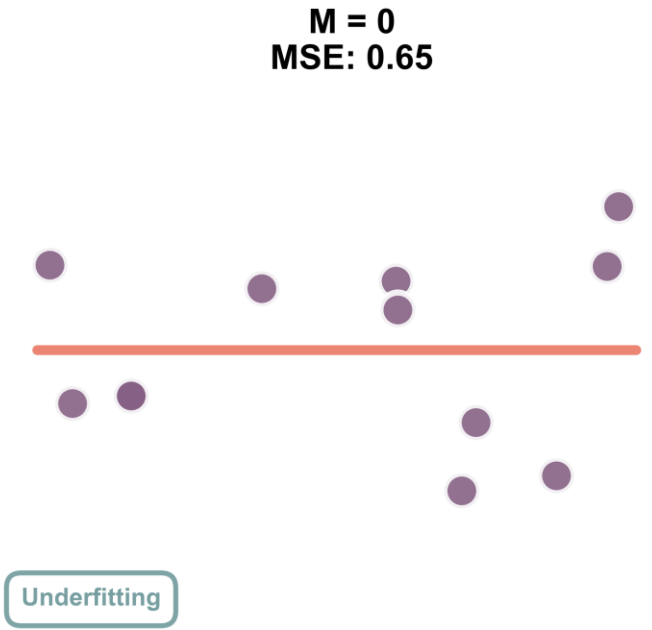
1. Data Bias
    1. Western, Educated, Industrialized, Rich and Democratic (WEIRD)
2. Black Box: lack of transparency
3. Overfitting: perform poorly on unseen data
4. High Complexity
    4. Resource intensive and hard to interpret
5. Adversarial Attacks
    4. Small changes in data can cause large prediction errors
6. Ethical concerns
7. Environmental Impact
8. Data Privacy

Basics of Pattern Recognition Analyses

Underfitting and Overfitting

- Occurs when a 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 optimization or feature selection

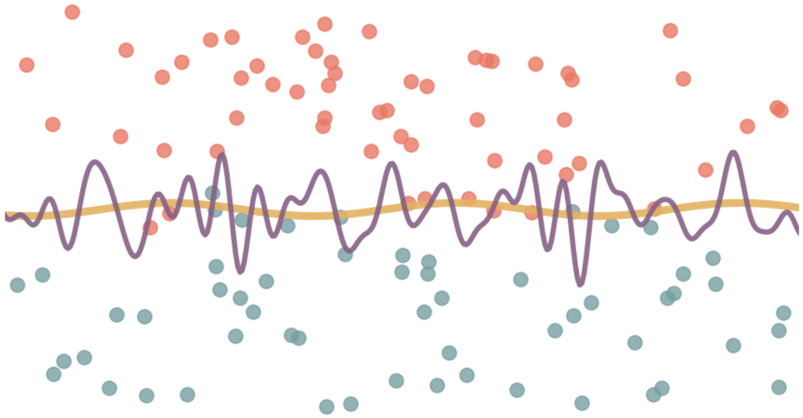


Basics of Pattern Recognition Analyses

Underfitting and Overfitting

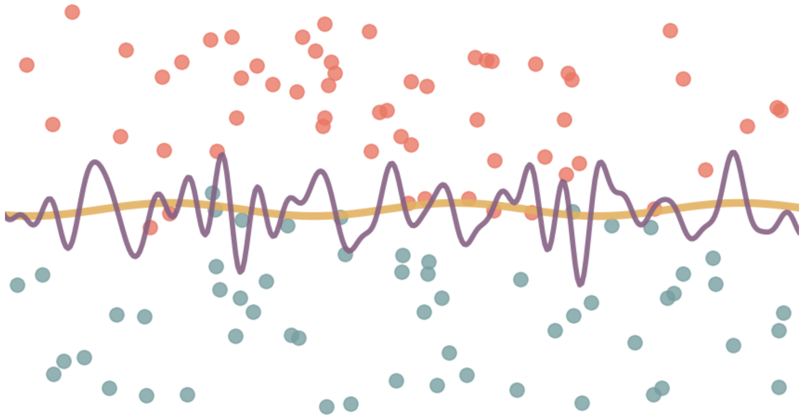
- Occurs when a model performs well on the data that it was estimated or trained on, but poorly on new data

- Can arise in very many ways including improper parameter optimization or feature selection



Underfitting and Overfitting

1. Increase/decrease model complexity
2. Feature selection
3. Regularization
4. Early stopping
5. Data augmentation
6. Cross-validation



Stages of supervised pattern recognition analysis

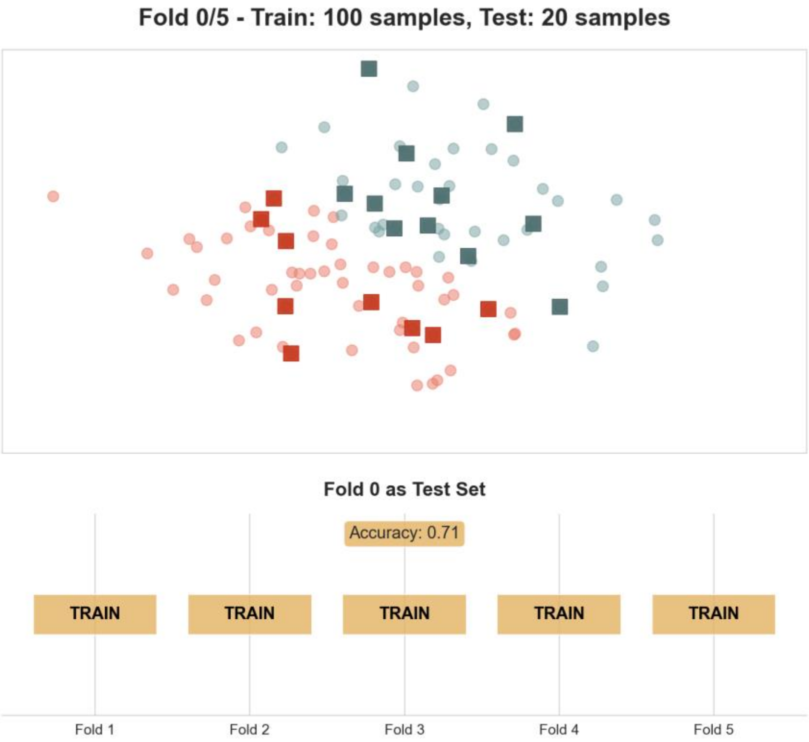
1. Feature **extraction** and/or **selection**
2. **Classification/ Regression** using cross-validation
3. **Performance** evaluation

Cross validation

- Testing on unseen data is essential to assess generalizability
- Cross validation is one popular way to do this

- K-fold CV split
  - Split the data into K approximately equal chunks
- Leave one out
  - One sample is left out at a time (K=N)

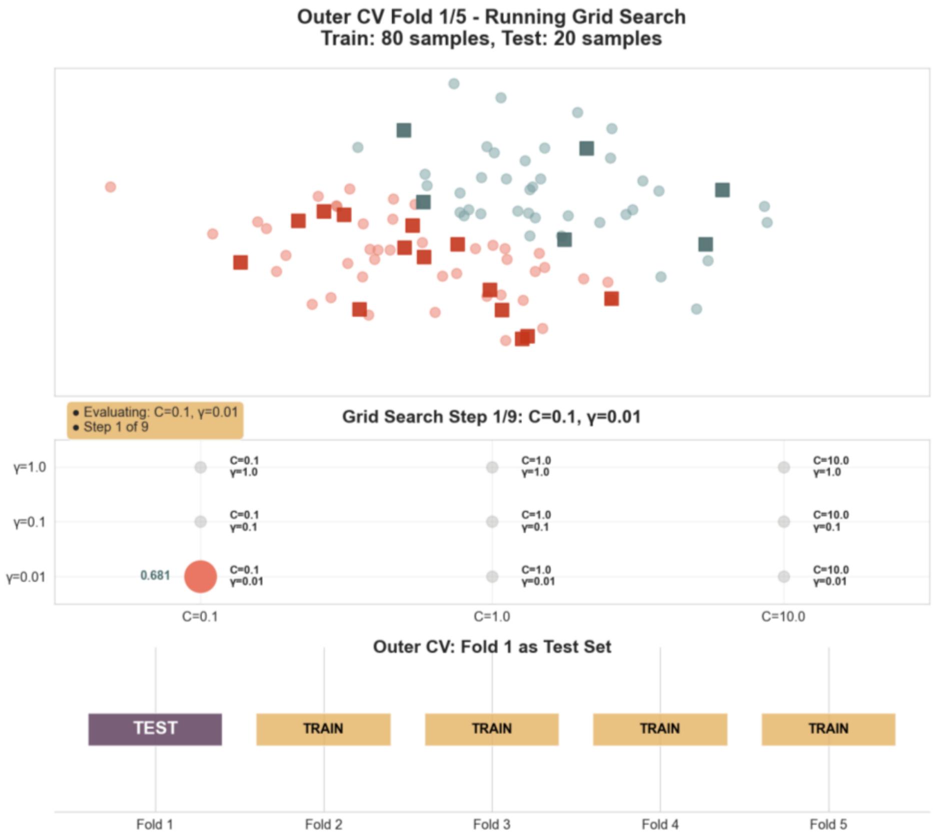
$$accuracy = \frac{1}{K}\sum_{i=i}^{K} acc_i$$



Parameter optimisation

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

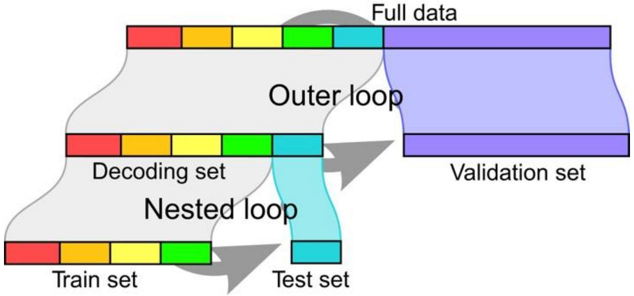
$$\hat{\boldsymbol{w}} = \min w \sum_{i=1}^{n} l(y_i, f_i) + \lambda J(\boldsymbol{w})$$

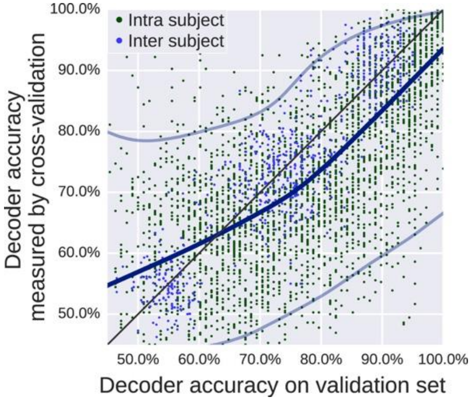


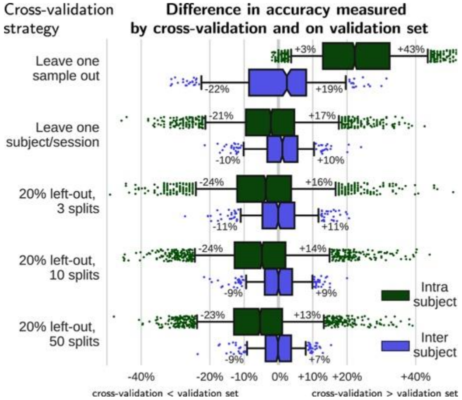
Varoquax (2017)

Multi-stage validation

- Despite being theoretically unbiased, cross-validation can still overfit
- A multistage validation approach protects against this
- Cross-validation also invalidates parametric tests (more later)





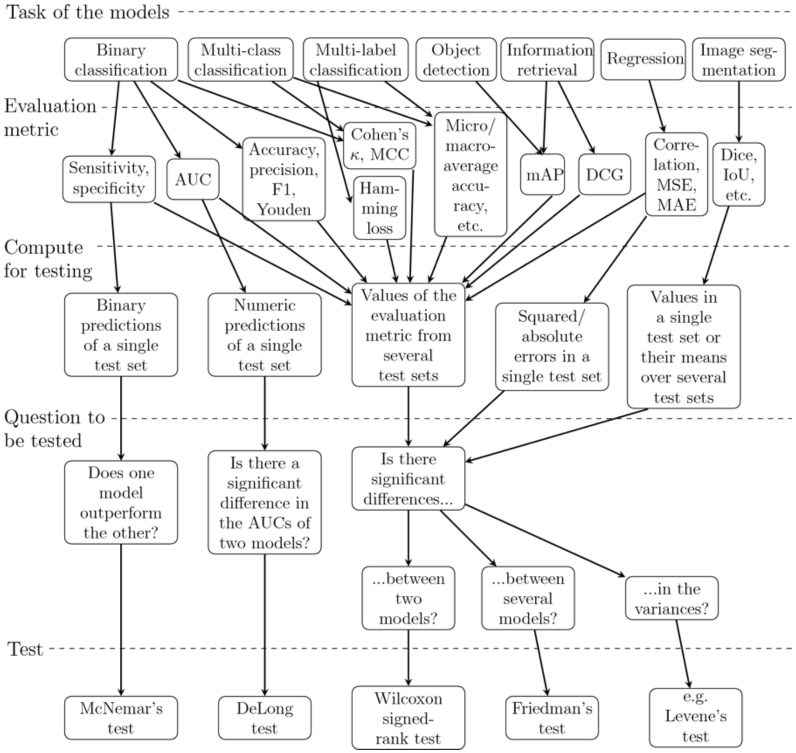


Basics of Pattern Recognition Analyses

Statistical assessment of model performance

Choice of error measure

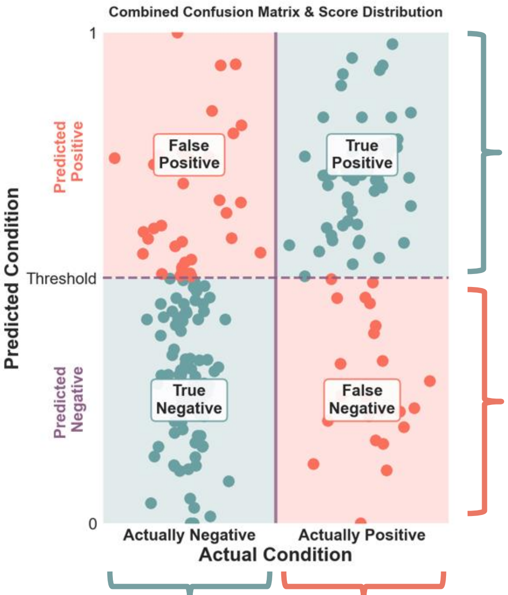
- Different error metrics are sensitive to different aspects (e.g. MSE depends on the scale of the data)
- Accuracy≠ Clinical utility



Basics of Pattern Recognition Analyses

Statistical assessment of model performance

Choice of error measure



PPV

$$\frac{\textit{True Positive}}{\textit{Predicted Positive}}$$

NPV

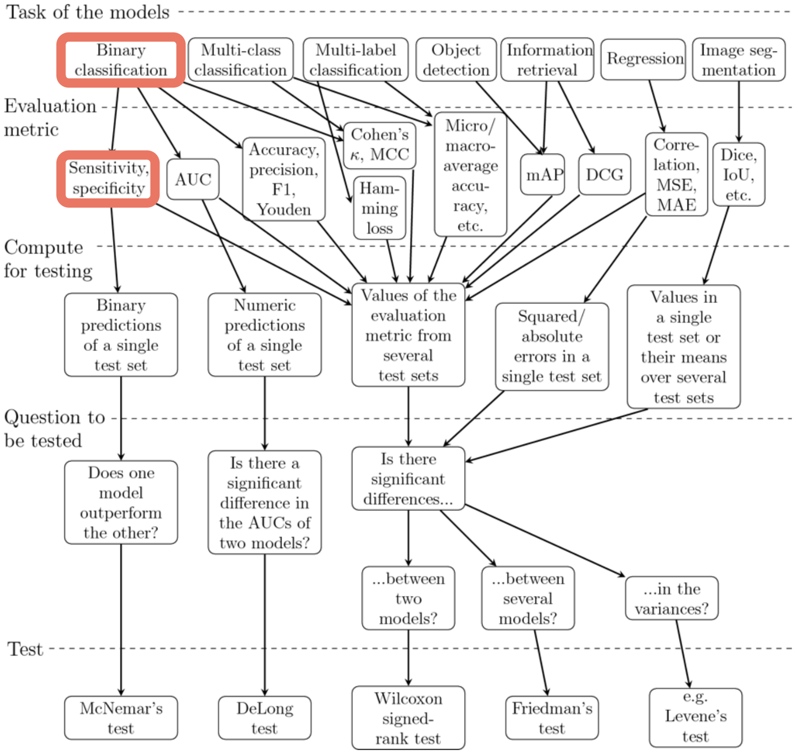
$$\frac{\textit{True Negative}}{\textit{Predicted Negative}}$$

Specificity

$$\frac{\textit{True Negative}}{\textit{Negative}}$$

Sensitivity

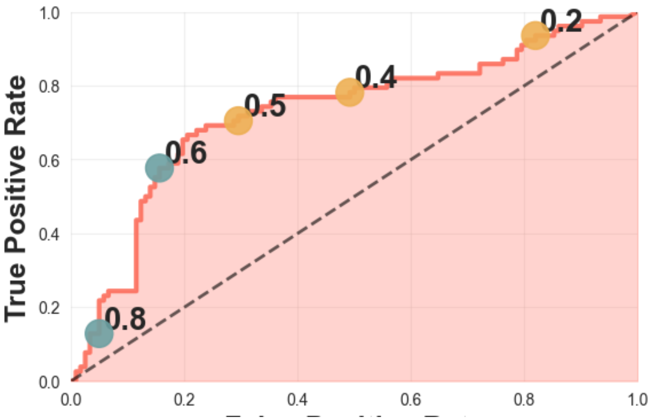
$$\frac{\textit{True Positive}}{\textit{Positive}}$$

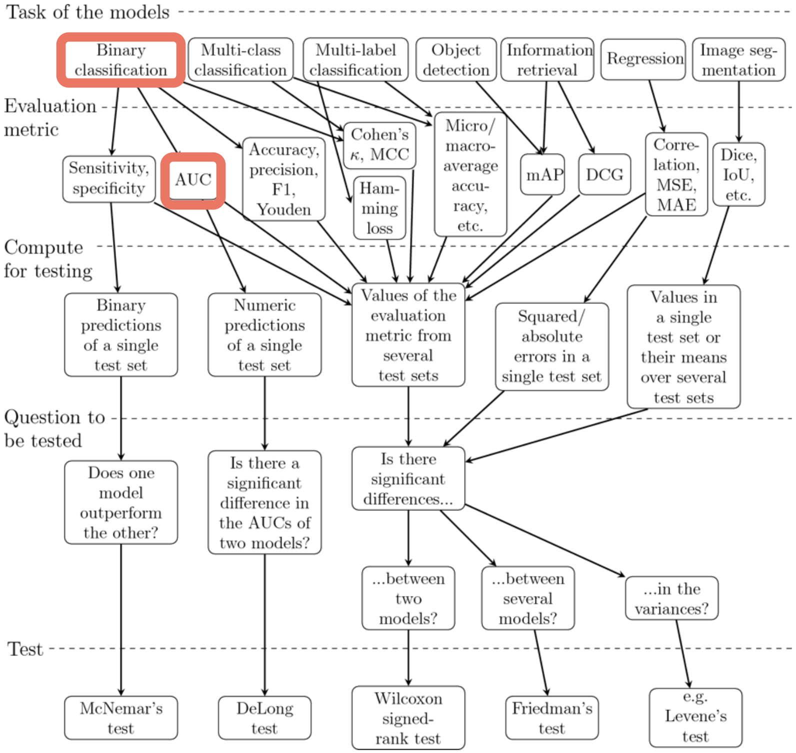


Basics of Pattern Recognition Analyses

Statistical assessment of model performance

Choice of error measure





Stelzer et al. (2013), Winkler et al. (2016)

Basics of Pattern Recognition Analyses

Statistical assessment of model performance

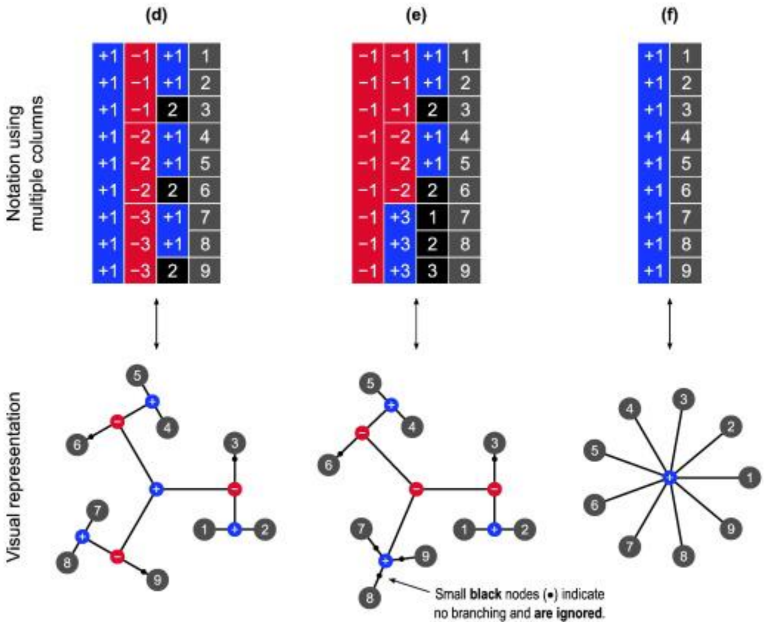
How to test the model performance?

- Parametric tests (e.g. binomial test, t-test)

BUT!
Cross-validation induces dependency between the folds invalidating parametric statistics

- Parametric assumptions may not be met
  - *(e.g. interval data)*

- Permutation tests must respect *exchangeability*
  - e.g. site effects, family structure…



Outline

1. Introduction to Machine Learning
2. Basics of Pattern Recognition Analyses
3. Applications in Psychiatry
4. Conclusions

Applications in Psychiatry

Supervised learning for automated diagnosis and prognosis

Contents lists available at ScienceDirect
Neuroscience and Biobehavioral Reviews
journal homepage: www.elsevier.com/locate/neubiorev

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]

Contents lists available at ScienceDirect
NeuroImage
journal homepage: www.elsevier.com/locate/ynimg

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]

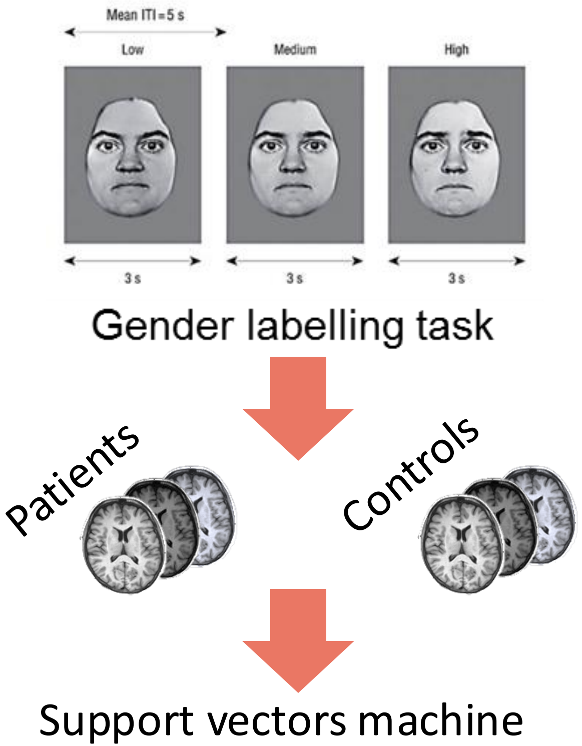
Building better biomarkers: brain models in translational neuroimaging

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

Applications in Psychiatry

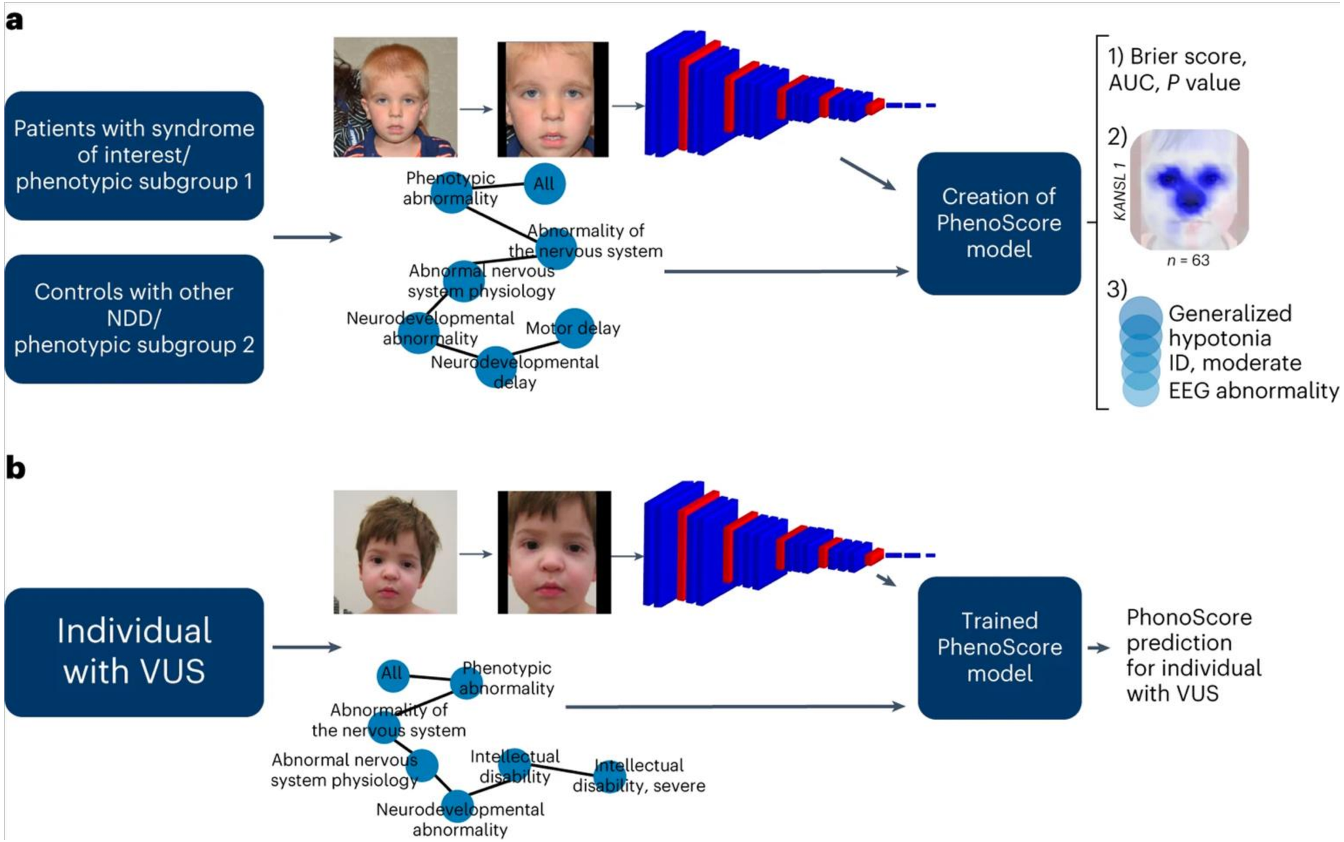
Supervised learning for depression diagnosis

- 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



Applications in Psychiatry

Supervised learning for depression diagnosis



Applications in Psychiatry

Unsupervised learning

Validation of clusters is difficult:

- Clustering always gives a result and there is no clear measure of success (e.g. stability? separability? predictive ability?)
- Rarely test against the "null" hypothesis that there are no clusters in the data
- Clustering using symptoms may not map onto biology

PNAS

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

Damien A. Fair[a,b,c,1], Deepti Bathula[a,d], Molly A. Nikolas[e], and Joel T. Nigg[a,b]

Departments of [a]Behavioral Neuroscience and [b]Psychiatry and [c]Advanced Imaging Research Center, Oregon Health and Science University, Portland, OR 97239; [d]Department of Computer Science and Engineering, Indian Institute of Technology Ropar, Rupagnar, India 140001; and [e]Department of Psychology, University of Iowa, Iowa City, IA 52242

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]

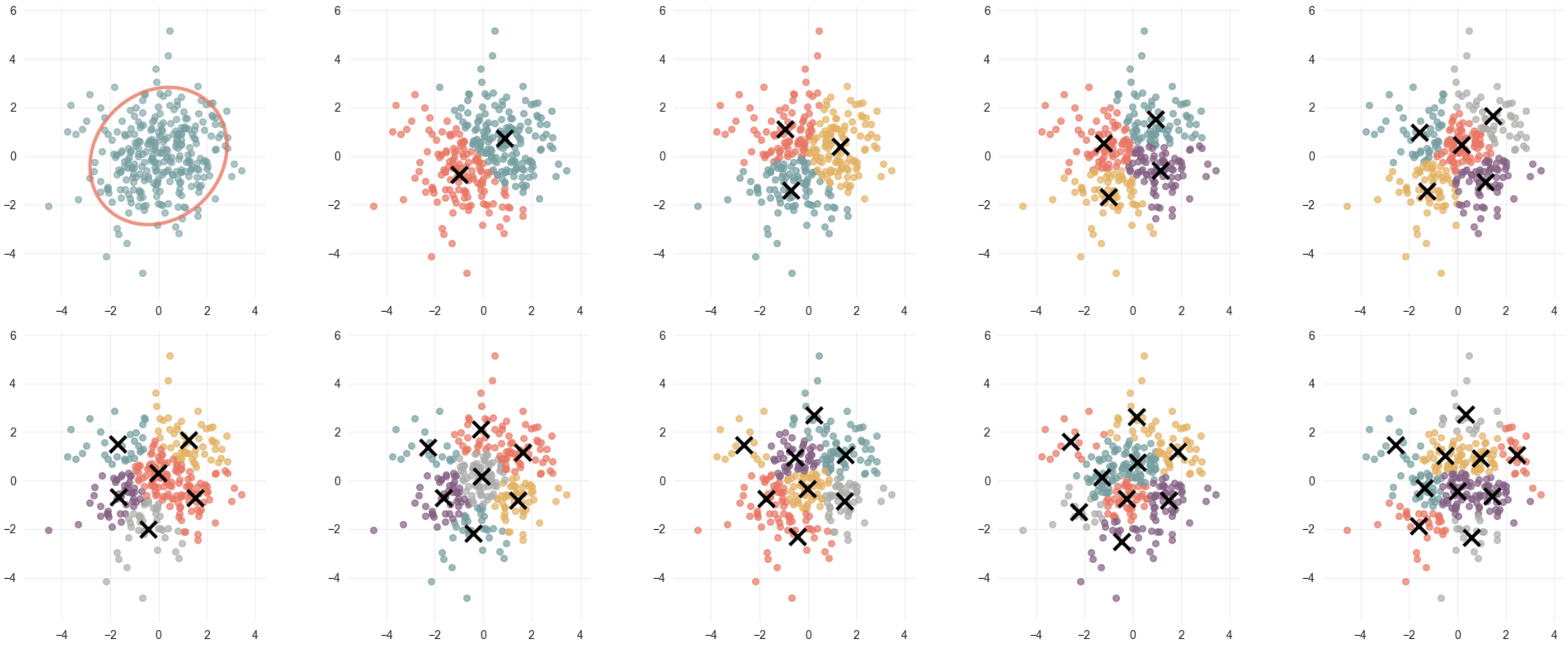
BRAIN
A JOURNAL OF NEUROLOGY

Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning

Ganesh B. Chand,[1,2,*] Dominic B. Dwyer,[3,*] Guray Erus,[1,2] Aristeidis Sotiras,[1,2,4] Erdem Varol,[1,2,5] Dhivya Srinivasan,[1,2] Jimit Doshi,[1,2] Raymond Pomponio,[1,2] Alessandro Pigoni,[3,6] Paola Dazzan,[7] Rene S. Kahn,[8] Hugo G. Schnack,[9] Marcus V. Zanetti,[10,11] Eva Meisenzahl,[12] Geraldo F. Busatto,[10] Benedicto Crespo-Facorro,[13,14] Christos Pantelis,[15] Stephen J. Wood,[16,17,18] Chuanjun Zhuo,[19,20] Russell T. Shinohara,[2,21] Haochang Shou,[2,21] Yong Fan,[1,2] Ruben C. Gur,[1,22] Raquel E. Gur,[1,22] Theodore D. Satterthwaite,[2,22] Nikolaos Koutsouleris,[3] Daniel H. Wolf[2,22,#] and Christos Davatzikos[1,2,#]

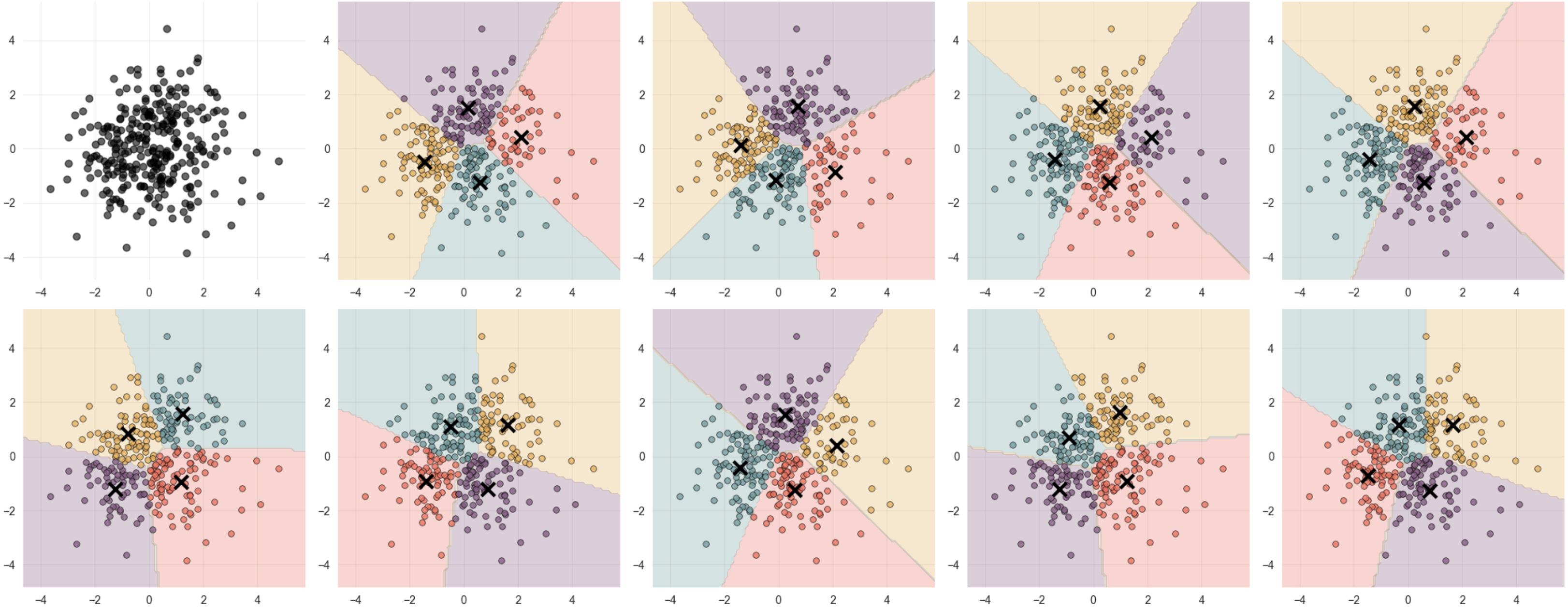
Applications in Psychiatry

Issues with clustering



Applications in Psychiatry

Issues with clustering

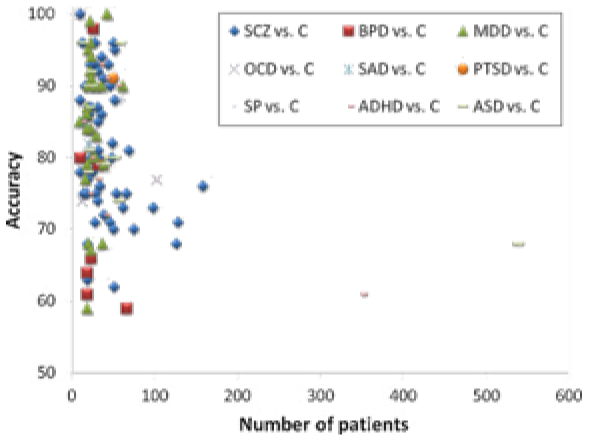


Wolfers al. (2015)

Applications in Psychiatry

Where are we now?

- Moderate accuracy, 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

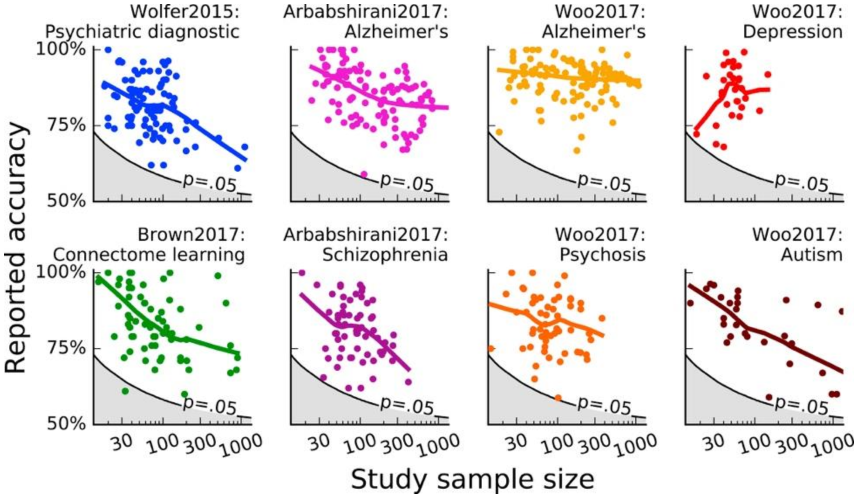


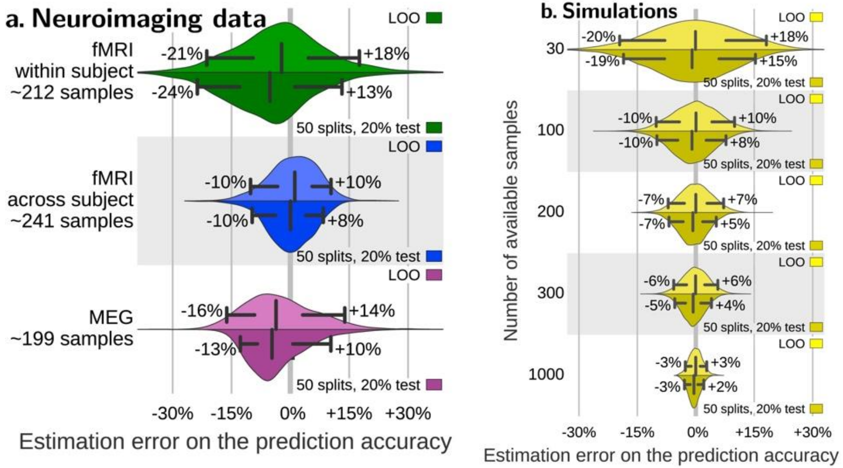
- 116 studies
- Most disorders
- Many data modalities

Varoquaux al. (2018)

Applications in Psychiatry

Where are we now?





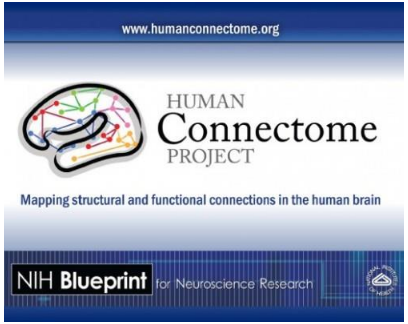
Applications in Psychiatry

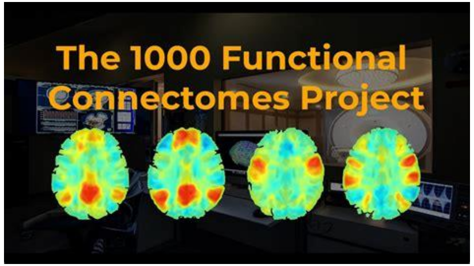
Where are we now?

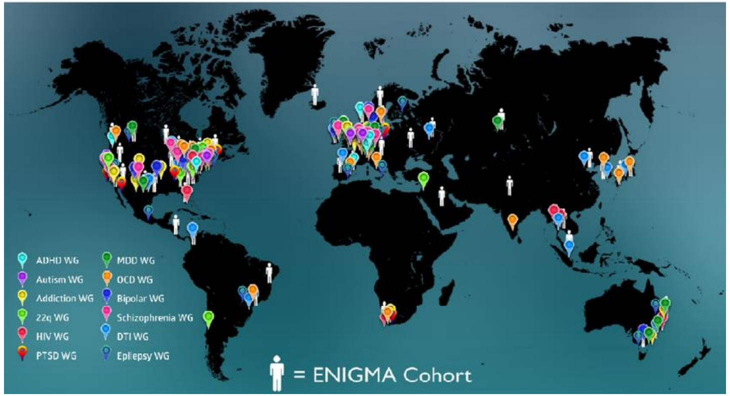
uk biobank

Parkinson's Progression Markers Initiative

ADNI Alzheimer's Disease Neuroimaging Initiative







Applications in Psychiatry

Where are we now?

Parkinson's Progression Markers Initiative

ADNI Alzheimer's Disease Neuroimaging Initiative

Article | Published: 26 June 2023

Replicable brain–phenotype associations require large-scale neuroimaging data

Shu Liu, Abdel Abdellaoui, Karin J. H. Verweij & Guido A. van Wingen

Nature Human Behaviour **7**, 1344–1356 (2023) | Cite this article

NIH Blueprint

Article | Open access | Published: 16 March 2022

Reproducible brain-wide association studies require thousands of individuals

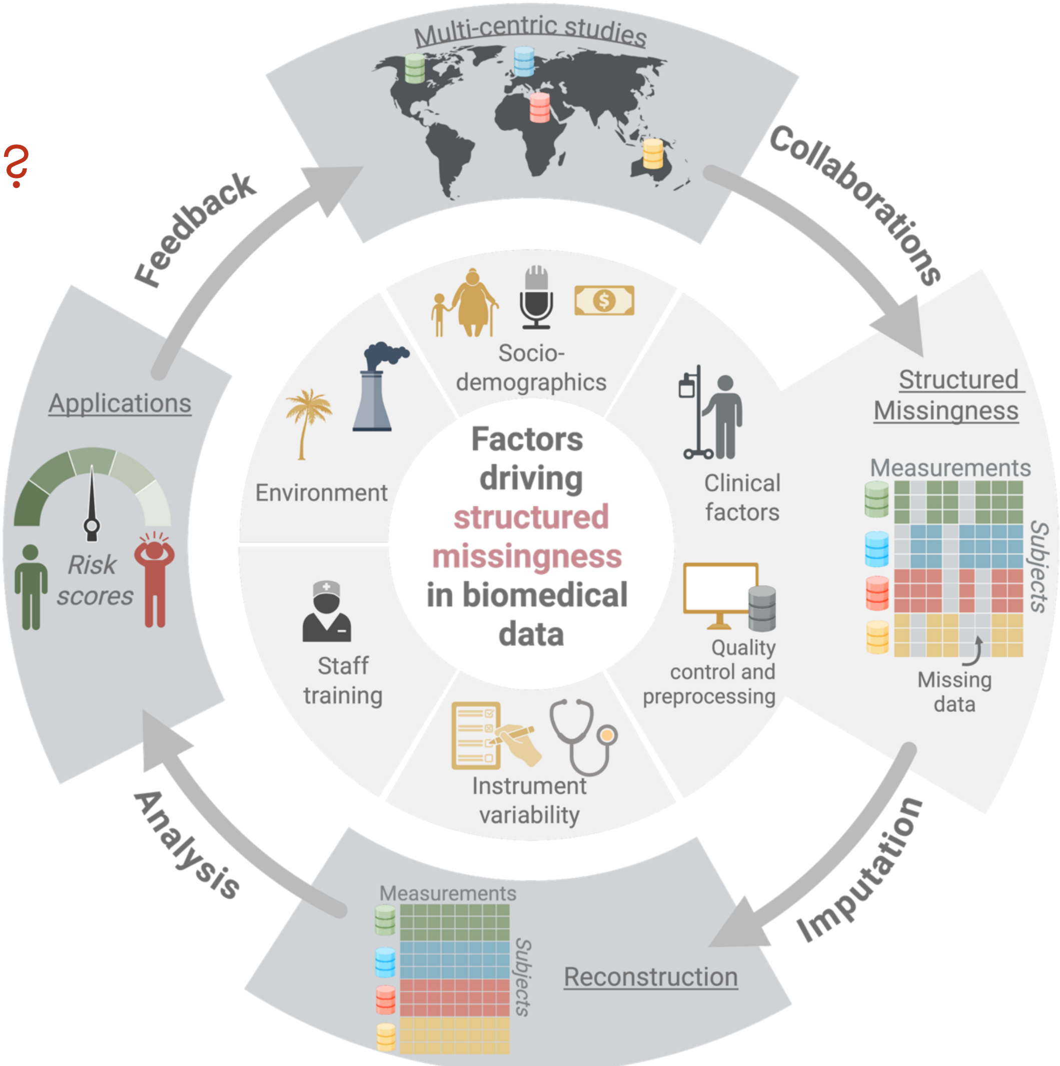
Scott Marek, Brenden Tervo-Clemmens, Finnegan J. Calabro, David F. Montez, Benjamin P. Kay, Alexander S. Hatoum, Meghan Rose Donohue, William Foran, Ryland L. Miller, Timothy J. Hendrickson, Stephen M. Malone, Sridhar Kandala, Eric Feczko, Oscar Miranda-Dominguez, Alice M. Graham, Eric A. Earl, Anders J. Perrone, Michaela Cordova, Olivia Doyle, Lucille A. Moore, Gregory M. Conan, Johnny Uriarte, Kathy Snider, Benjamin J. Lynch, ... Nico U. F. Dosenbach + Show authors

Nature **603**, 654–660 (2022) | Cite this article

Rehak Buckova al. (2025)

Applications in Psychiatry

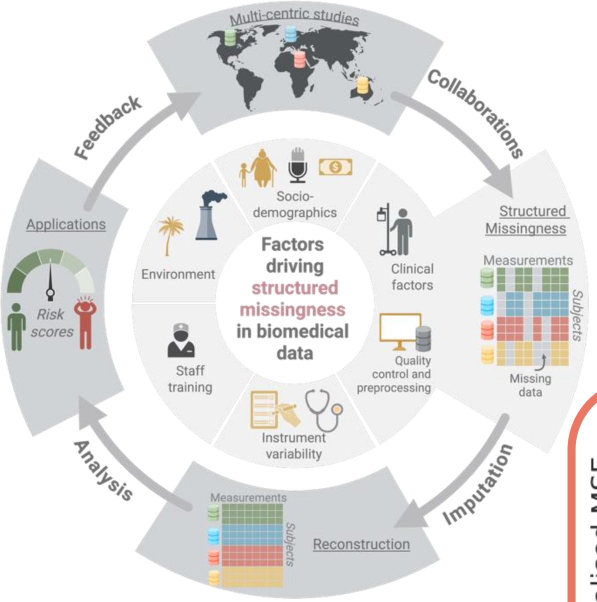
Where are we now?

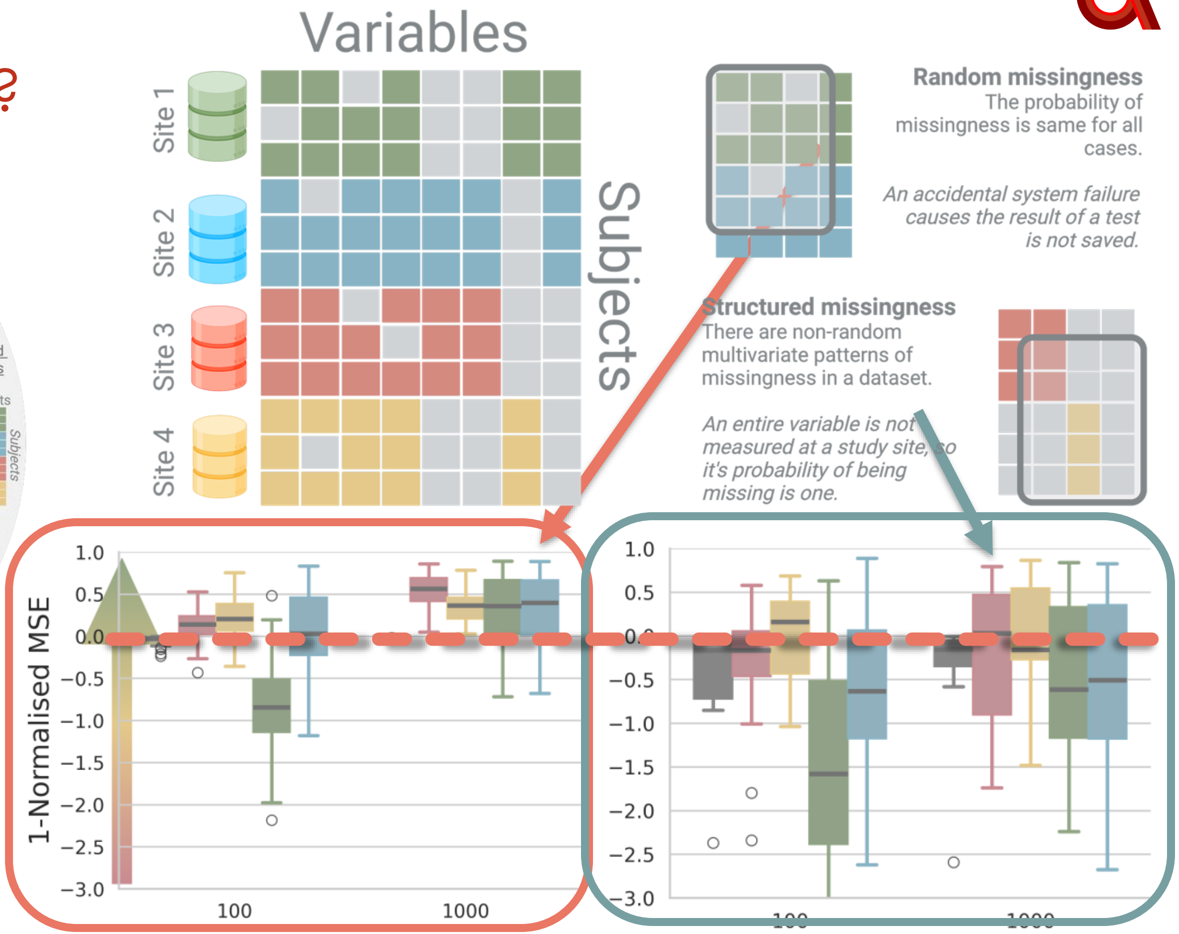


Rehak Buckova et al. (2025)

Applications in Psychiatry

Where are we now?





Ethical Challenges

- **Interpretability:** Balancing model complexity with interpretability for clinical decision-making.

- **Overreliance:** Avoid overreliance on algorithmic predictions.

- **Long-term Impact:** Consider potential psychological and societal impacts of diagnosis and treatment based on machine learning.

- **Equity:** Ensure equal access to computational psychiatry tools.

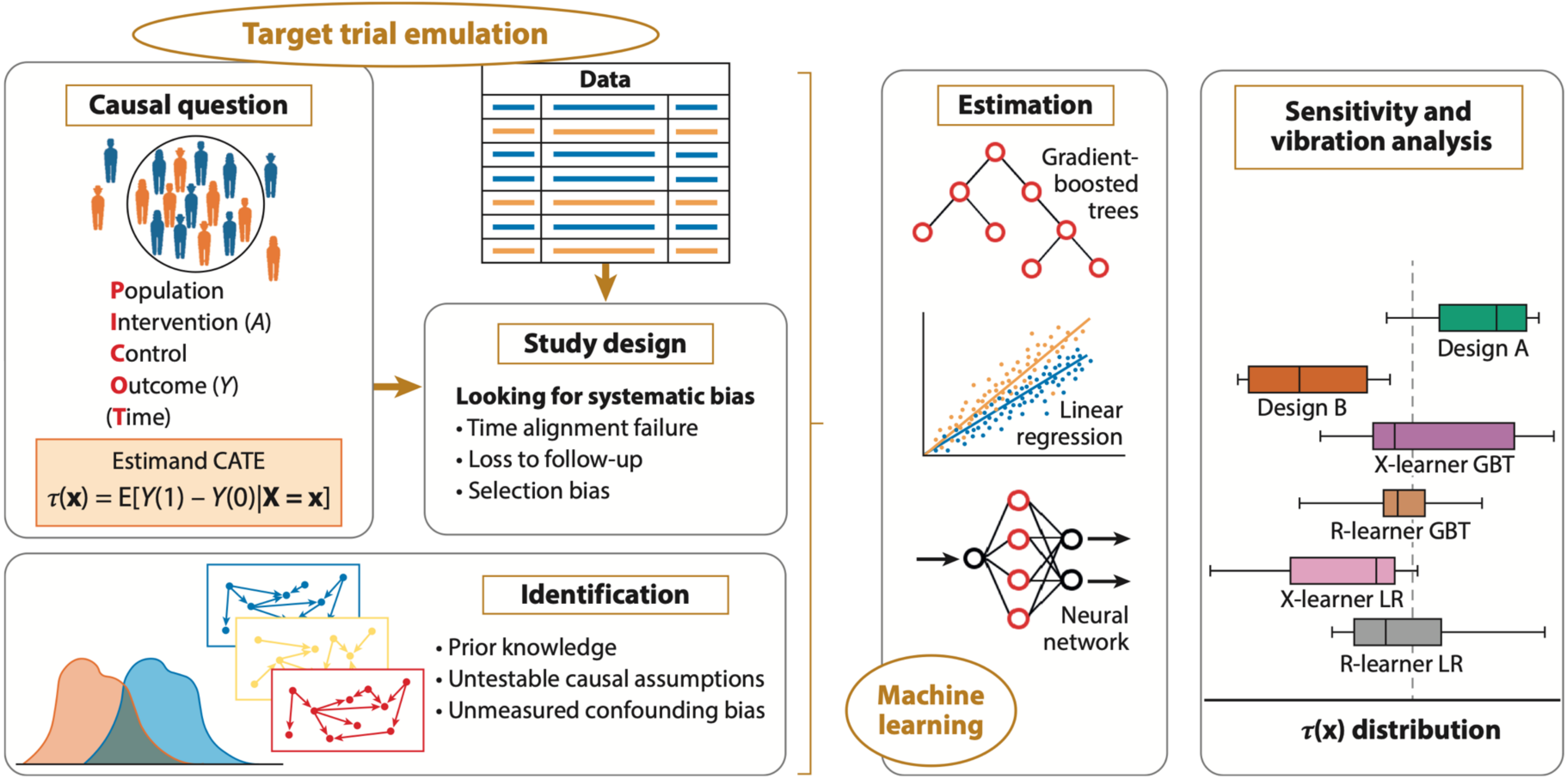
*'While 'off the shelf' **machine learning tools**, like TensorFlow, are becoming **more accessible** from the point of view of setting up your own system, the **underlying logics of those systems**, and the datasets for training them **are accessible** to and controlled by very **few entities**. In the dynamic of dataset collection through platforms like Facebook, users are feeding and training the neural networks with behavioral data, voice, tagged pictures and videos or medical data. In an era of extractivism, the real value of that data is controlled and exploited by the very few at the top of the pyramid.'*

Outline

1. Introduction to Machine Learning
2. Basics of Pattern Recognition Analyses
3. Applications in Psychiatry
4. Conclusions

Applications in Psychiatry

Interim summary



Interim summary

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
3. Estimating mappings between brain and behaviour

- Validation of models is extremely important to ensure generalisability
- More on that in Lecture 2 ...