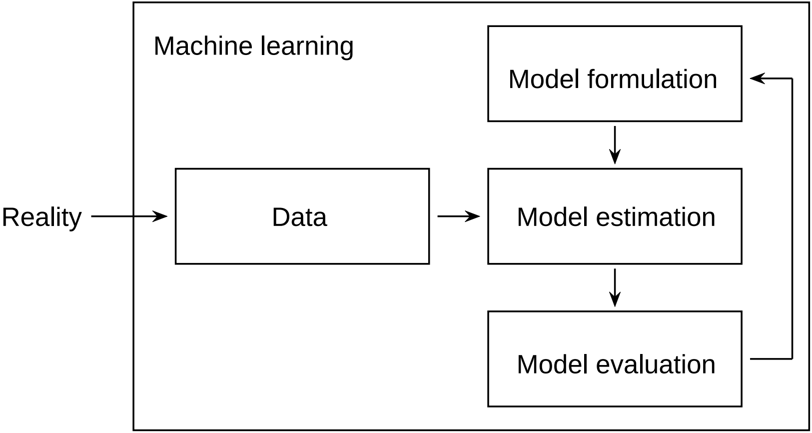
Machine learning

MSc Social, Cognitive, and Affective Neuroscience SoSe 2020

Prof. Dr. Dirk Ostwald

(1) Introduction

What is "machine learning"?



Data science: Statistics & Machine learning & Artificial intelligence

Statistics

- Probabilistic models
- Theoretical analysis
- Optimality
- Asymptotics
- Science philosophy

Machine learning

- Deterministic models
- Classification
- Bayesian models
- Benchmarking
- Applications

Artificial intelligence

- Deep learning
- Reinforcement
- Machine learning
- Data analysis
- Hype

Community nomenclatures

Statistic	Machine Learning	Meaning
Data	Training data	Data
Estimation	Learning, Training	Using data to estimate parameters
Frequentist inference	-	Optimal many samples methods
Bayesian inference	Bayesian inference	Data-based uncertainty updating
Covariates	Features	Structural and known data predictors

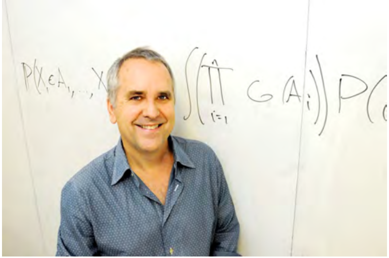
Some Studies in Machine Learning
Using the Game of Checkers
Abstract: Two machine-learning procedures have been investigated in some detail using the game of checkers. Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Furthermore, it can learn to do this in a remarkably short period of time (8 or 10 hours of machine-playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified. The principles of machine learning verified by these experiments are, of course, applicable to many other situations.

Samuel (1959)

1920s	Classical and mathematical statistics
1950s	Pioneering machine learning research (e.g. perceptron)
1960s	Bayesian methods are introduced for probabilistic inference in machine learning
1970s	AI Winter caused by pessimism about machine learning effectiveness
1980s	First rediscovery of backpropagation for neural networks (connectionism)
1990s	Support vector machines and recurrent neural networks popular
2000s	Kernel methods, unsupervised methods, and variational inference popular
2010s	Second rediscovery of backpropagation for neural networks as deep learning

Efron and Hastie (2016)

Michael I. Jordan



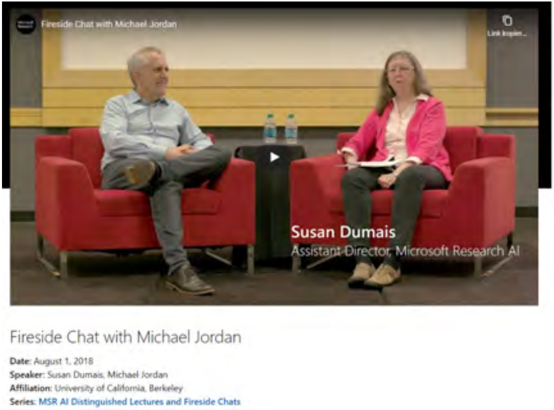
*1956, Professor at University of California, Berkeley

Research in statistics, machine learning, artificial intelligence

1985, Ph.D., Cognitive Science, UC San Diego

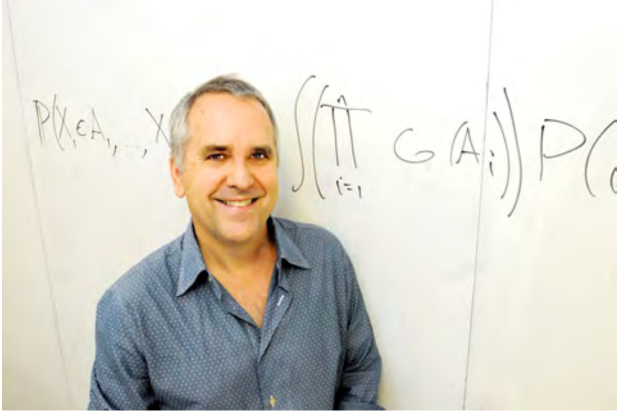
1980, M.S., Mathematics, Arizona State University

1978, B.S., Psychology, Louisiana State University



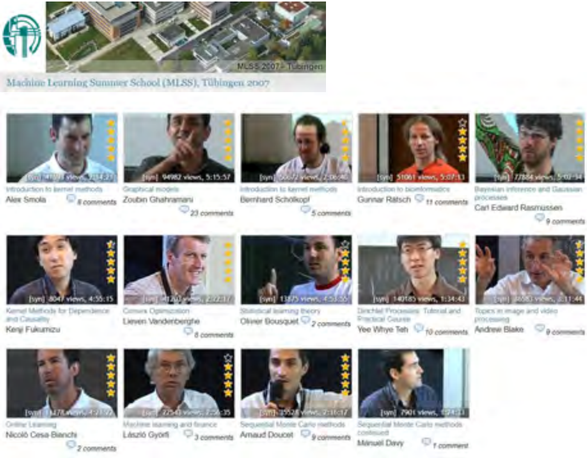
`www.microsoft.com/en-us/research/video/fireside-chat-with-michael-jordan/`

hdsr.mitpress.mit.edu/pub/wot7mkc1

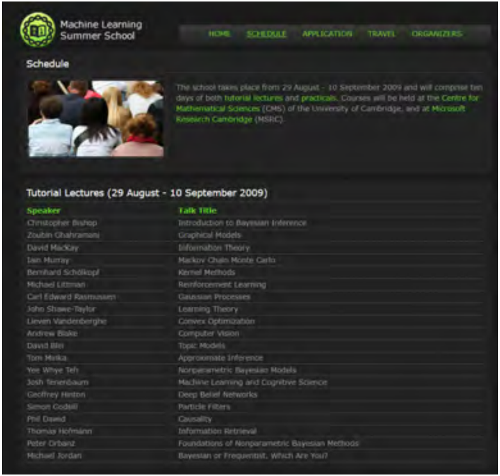


https://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan/

Topics in machine learning | MLSS 2007



Topics in machine learning | MLSS 2009



Machine Learning Summer School

HOME SCHEDULE APPLICATION TRAVEL ORGANIZERS

Schedule

The school takes place from 29 August - 10 September 2009 and will comprise ten days of both tutorial lectures and practicals. Courses will be held at the Centre for Mathematical Sciences (CMS) of the University of Cambridge, and at Microsoft Research Cambridge (MSRC).

Tutorial Lectures (29 August - 10 September 2009)

Speaker	Talk Title
Christopher Bishop	Introduction to Bayesian Inference
Zoubin Ghahramani	Graphical Models
David MacKay	Information Theory
Iain Murray	Markov Chain Monte Carlo
Bernhard Schölkopf	Kernel Methods
Michael Littman	Reinforcement Learning
Carl Edward Rasmussen	Gaussian Processes
John Shawe-Taylor	Learning Theory
Lieven Vandenberghe	Convex Optimization
Andrew Blake	Computer Vision
David Blei	Topic Models
Tom Minka	Approximate Inference
Yee Whye Teh	Nonparametric Bayesian Models
Josh Tenenbaum	Machine Learning and Cognitive Science
Geoffrey Hinton	Deep Belief Networks
Simon Godsill	Particle Filters
Phil Dawid	Causality
Thomas Hofmann	Information Retrieval
Peter Orbanz	Foundations of Nonparametric Bayesian Methods
Michael Jordan	Bayesian or Frequentist, Which Are You?

Topics in machine learning | MLSS 2017

Machine Learning Summer School 2017 - Schedule

	Sun, June 18	Mon, June 19	Tue, June 20	Wed, June 21	Thu, June 22	Fri, June 23	Sat, June 24
09:00 - 10:30		B. Schölkopf *What is ML?*	S. Ben-David *Learning Theory*	B. Schölkopf D. Janzing *Causality*	J. Lescovec *Network analysis*	Lescovec *Network analysis*	*Practical Session II* (9.30-12.00)
11:00 - 12:30		S. Ben-David *Learning Theory*	B. Schölkopf D. Janzing *Causality*	Max Welling *Large Scale Bayesian Inference*	Z. Ghahramani *Bayesian Inference*	Jan Peters *Reinforcement Learning in 90 min*	
12:30 - 14:00		*lunch*	*lunch*	*lunch*	*lunch*	*lunch*	*lunch*
14:00 - 15:00		B. Schölkopf D. Janzing *Causality*	cyberneum tour (13:45-15:45)	R. Salakhutdinov *Deep Learning*	Z. Ghahramani *Bayesian Inference*	J. Lescovec *Network analysis*	*free afternoon*
15:30 - 17:00	*registration & welcome reception* *(18.00-21.00)*	S. Ben-David *Learning Theory*	lab tours (15.30-18.30)	Z. Ghahramani *Bayesian Inference*	*R. Salakhutdinov* *Deep Learning*	*Practical Session I*	
17:00 - 18:00							
evening 19:00-21:00	*Ralf Herbrich* *Invited Talk* *(19.00–19.30)* *Open Bar*	*BBQ*	*Social event*	*dinner** poster session I Bar	*Conference dinner*	*dinner* poster session II Bar	*Night out in Town*

	Sun, June 25	Mon, June 26	Tue, June 27	Wed, June 28	Thu, June 29	Fri, June 30	Sat, July 1
09:00 - 10:30		S. Sra *Optimization*	S. Sra *Optimization*	B. Sriperumbudur *Kernel Methods*	M. Jordan *Distributed Architectures*	*rest*	*departure*
11:00 - 12:30	*Leisure activity* *High-wire garden*	I. Tolstikhin *Implicit generative modles*	Stefanie Jegelka *Submodularity*	M. Jordan *Distributed Architectures*	Stefan Schaal		
12:30 - 14:00	*or*	*lunch*	*lunch*	*lunch*	*lunch*	*lunch*	
14:00 - 15:00	*Hike*	S. Sra *Optimization*	Olivier Bousquet	Michael Black	M. Jordan *Distributed Architectures*	B. Sriperumbudur *Kernel Methods*	
15:30 - 17:00		*Practical Session III*	*Practical Session IV*	B. Sriperumbudur *Kernel Methods*	V. Mnih *Deep Reinforcement Learning*	V. Mnih *Deep Reinforcement Learning*	
17:00 - 18:00							
evening 19:00-21:00	*free evening*	*dinner* poster session III Bar	*Social event*	*dinner** poster session IV Bar	*dinner* *party*	*Social event*	

Practicals	Manuel Gomez-Rodriguez, Utkarsh Upadhyay and Isabel Valera Ilya Tolstikhin and Ruth Urner David Lopez-Paz Olivier Bousquet and Sylvain Gelly	*Social Network analysis* *Learning Theory* Causality *TensorFlow*
Notes	Meals will be served in the Max Planck Haus (except those marked with *)	

Topics in machine learning | MLSS 2019

Video recordings

	July 15	July 16	July 17	July 18	July 19	July 20	July 21	July 22	July 23	July 24	July 25	July 26
07:30	Registration and Welcome											
09:00												
09:30	Variational Inference	Optimization	Deep Learning	Reinforcement Learning	Gaussian Processes			Kernels	MCMC	Approximate Bayesian Computation	Speech Processing	ML in Computational Biology
10:00												
10:30												
11:00	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break			Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break
11:30												
12:00	Variational Inference	Optimization	Deep Learning	Tutorial Reinforcement Learning	Gaussian Processes			Kernels	MCMC	Tutorial Approximate Bayesian Computation	Tutorial Speech Processing	Submodularity
12:30												
13:00												
13:30											Lunch Break	
14:00	Lunch Break	Lunch Break	Lunch Break	Lunch Break	Lunch Break	Social Event		Lunch Break	Lunch Break	Lunch Break		Lunch Break
14:30												
15:00											Learning Theory	
15:30	Tutorial Variational Inference	Tutorial Optimization	Tutorial Deep Learning	Interpretability	Tutorial Gaussian Processes			Tutorial Kernels	Tutorial MCMC	Fairness		Tutorial Submodularity
16:00												
16:30											Coffee Break	
17:00				Coffee Break	Coffee Break				Coffee Break			
17:30			Coffee Break							Coffee Break	Learning Theory	
18:00		Poster Session	Reinforcement Learning	Interpretability	AI for Good + Panel Discussion			Industry Event and Poster Session	Approximate Bayesian Computation	Speech Processing		
18:30												
19:00	G-Research Welcome Reception											Farewell Reception at Bloomberg
					Microsoft Mid-course Reception at the British Library						Farewell Dinner	

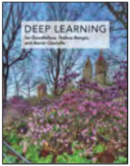
Topic	Content
(1) Introduction	General remarks
(2) Optimization	General importance
(3) Principal component analysis	Feature selection
(4) Gaussian models	Independent component analysis
(5) LDA and logistic regression	Probabilistic classification
(6) Neural networks	Deep learning
(7) Support vector machines	Non-probabilistic classification
(8) Variational autoencoders	Variational inference for deep learning

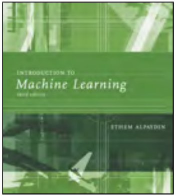
What is “machine learning”?

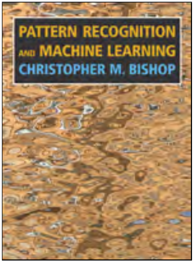
- Machine learning is bad (commonly not theoretically grounded) statistics.
- Machine learning is statistics as studied by computer scientists since the 1980s.
- For most people, machine learning seems to mean classification algorithms.
- Machine learning proper are support vector machines and neural networks.
- Since 2015, machine learning is also called artificial intelligence.

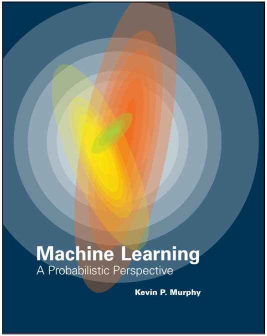
$\Rightarrow$ Just call statistics and machine learning plus x “Data science”.

Recommended reading



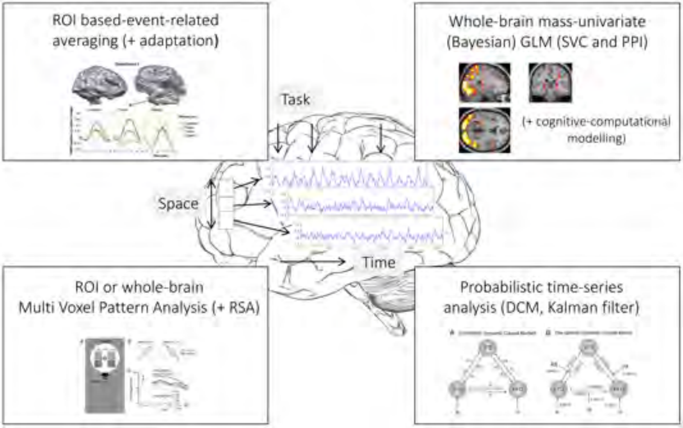




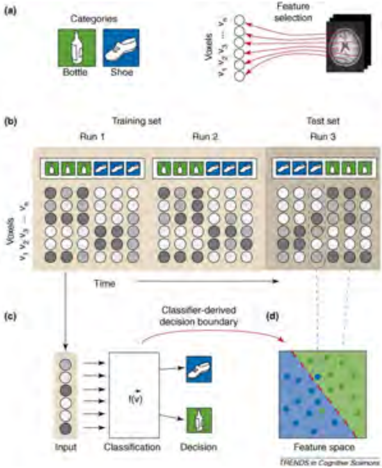


Goodfellow et al. (2017); Alpaydin (2014); Bishop (2006); Murphy (2012)

Machine learning in neuroimaging data analysis

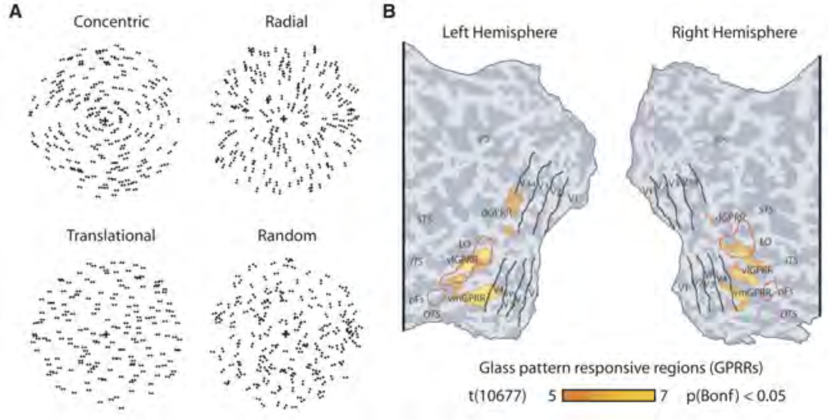


Multi-voxel pattern analysis (MVPA)



Norman et al. (2006)

Multi-voxel pattern analysis (MVPA)



Ostwald et al. (2008)

Multi-voxel pattern analysis (MVPA)

- Block design: 20 Glass patterns of a given type per 16 seconds.
- Voxel signal activation extraction from region of interest (ROI).
- Selection of 130 most activated voxels per ROI across conditions.
- Run-wise z-score normalization of each voxel's time course.
- Shifting the fMRI time series by 4 s to account for the hemodynamic response.
- Averaging all time series data points of one experimental block.
- Support Vector Machine for classification (SVMlight Matlab toolbox).
- Eightfold cross-validation leaving one run out.
- $\Rightarrow$ 112 patterns training patterns, 16 testing patterns.
- Average accuracy across cross-validations.
- Statistical evaluation using repeated-measures ANOVA.

Ostwald et al. (2008)

Multi-voxel pattern analysis (MVPA)

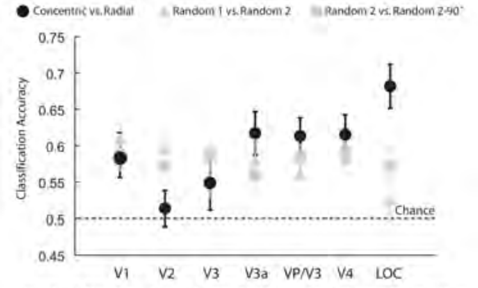


FIG. 4. Classification based on global structure vs. local orientations. Classification accuracies for similar local position and orientation changes in the presence (concentric vs. radial Glass patterns, black symbols) or absence (random 1 vs. random 2, random 2 vs. random 2–90°, gray symbols) of global stimulus structure. Mean classification accuracy (pattern size = 130 voxels per area) across observers is shown for all regions of interest; error bars indicate SE across observers ($n = 5$). The dashed line indicates the chance classification level (0.5).

Ostwald et al. (2008)

Multi-voxel pattern analysis (MVPA)

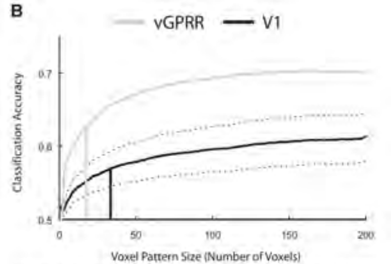


FIG. 6. Classification accuracy curves across voxel pattern size. Classification accuracy curves for classification of concentric vs. radial Glass patterns are plotted across voxel pattern size. Comparison of classification accuracy for V1 (black curve) with (*A*) LOC (gray curve) and (*B*) ventral GPRR (gray curve). The mean of classification accuracy for 500 MVPA iterations (randomized order of voxels in the pattern size across iterations) across subjects is shown. The dotted lines show SE across observers. Vertical lines indicate the voxel constant (υ) for V1 (υ = 33.38, black line) compared with LOC (υ = 15.06, gray line) and vGPRR (υ = 18.77, gray line), indicating that a smaller pattern size is necessary for discriminating concentric vs. radial Glass patterns in higher occipitotemporal than early visual areas.

Ostwald et al. (2008)

Topic	Neuroimaging application
(1) Introduction	General bullshitting
(2) Optimization	Data analysis software
(3) Principal component analysis	Voxel selection, electrode selection
(4) Gaussian models	M/EEG, fMRI data preprocessing
(5) LDA and logistic regression	MVPA
(6) Support vector machines	MVPA
(7) Neural networks	MVPA
(8) Variational autoencoders	To be applied

References

Alpaydin, E. (2014). *Introduction to Machine Learning*.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press, first edition. 00000.

Goodfellow, I., Bengio, Y., and Courville, A. (2017). *Deep Learning*. The Mit Press, Cambridge, Massachusetts. 02143.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning Series. MIT Press, Cambridge, MA. 00000.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9):424–430. 00000.

Ostwald, D., Lam, J. M., Li, S., and Kourtzi, Z. (2008). Neural Coding of Global Form in the Human Visual Cortex. *Journal of Neurophysiology*, 99(5):2456–2469. 00098.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. page 21.