185.A83 Machine Learning for Health Informatics
2016S, VU, 2.0 h, 3.0 ECTS
16.03.2016 17:00-18:30

A. Holzinger & R. Freund; Tutors: M. Bloice, B. Malle

Machine Learning for Health Informatics

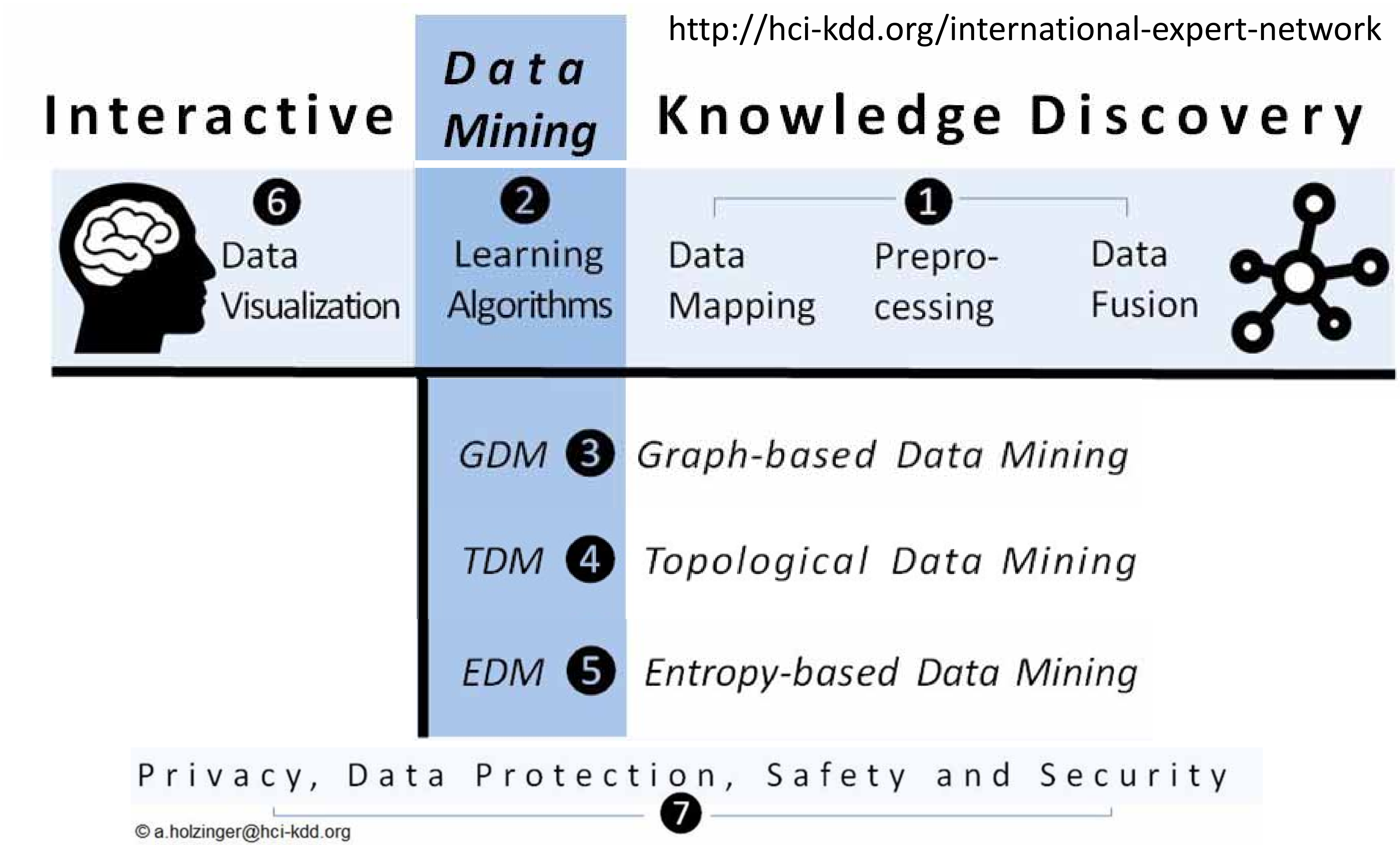
Introduction and Overview

{a.holzinger, m.bloice, b.malle}@hci-kdd.org

http://hci-kdd.org/machine-learning-for-health-informatics-course

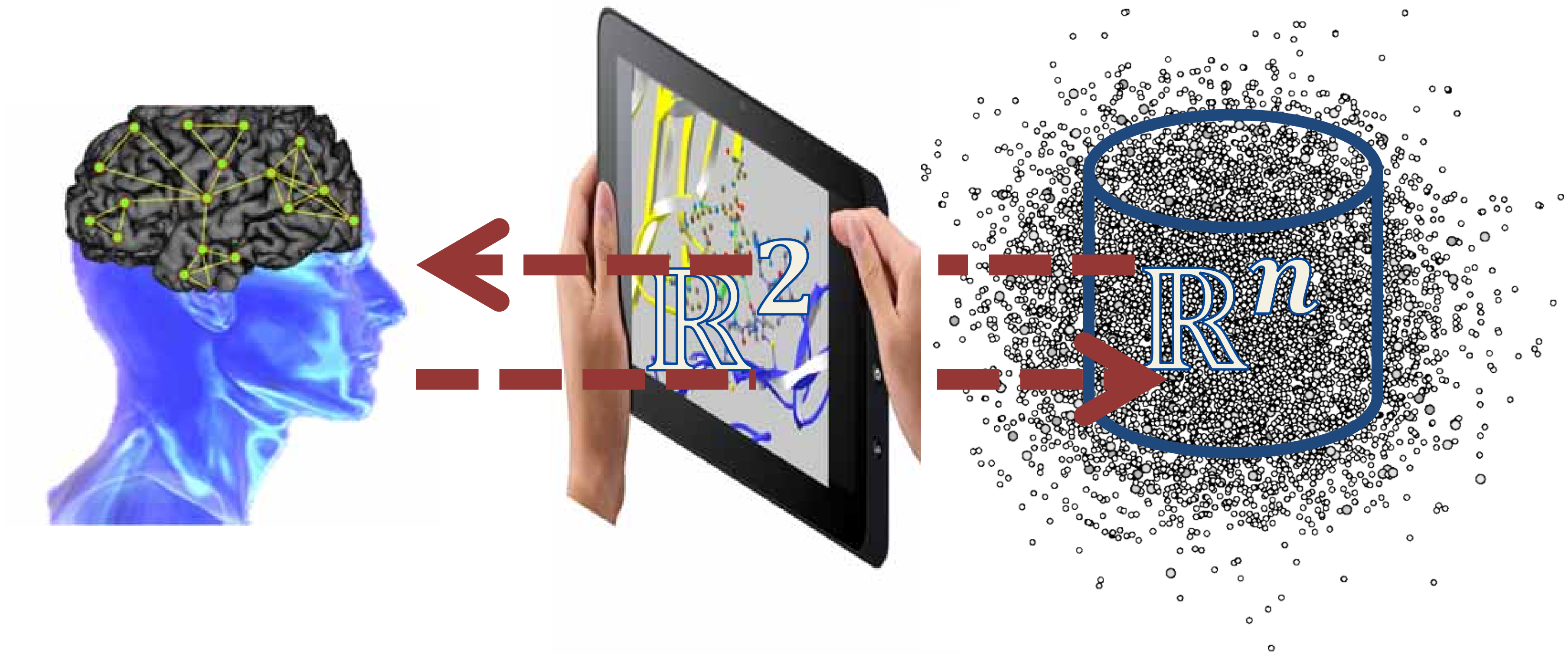
- What is the HCI-KDD approach?
- Why is ML and HI important ?
- What is aML versus iML ?
- A few examples
- What will you learn in this course ?
- How will you be assessed ?

ML needs a concerted effort fostering integrated research



Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: **Cognitive Science meets Machine Learning.** IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

Cognitive Science Computer Science



Human Interaction Computer

Holzinger, A. (2013). Human–Computer Interaction & Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? In: Lecture Notes in Computer Science 8127 (pp. 319-328)

- **Cognitive Science (human intelligence)**
  - Study the principles of human learning to understand intelligence –
  - "solve intelligence – then solve everything else" (Motto of Google Deepmind !)
- **Human-Computer Interaction (the bridge)**
  - Working with a computer that "learns" shall enhance user friendliness, concentrate on problem solving
  - Suitability of algorithms for iML
- **Computer Science (computational intelligence)**
  - Theoretical issues what learning procedures may achieve

- CS aims to reverse engineer **human intelligence**;
- ML provides powerful sources of insight into how **machine intelligence** is possible.
- CS therefore raises challenges for, and draws inspiration from ML;
- Insights about the human mind may help inspire **new directions for ML.**

Example: Discovery of causal relationships from data …

Hans Holbein d.J., 1533,
The Ambassadors,
London: National Gallery

https://www.youtube.com/watch?v=9KiVNIUMmCc

Lopez-Paz, D., Muandet, K., Schölkopf, B. & Tolstikhin, I. 2015. Towards a learning theory of cause-effect inference. Proceedings of the 32nd International Conference on Machine Learning, JMLR, Lille, France.

Salakhutdinov, R., Tenenbaum, J. & Torralba, A. 2012. One-shot learning with a hierarchical nonparametric Bayesian model. Journal of Machine Learning Research, 27, 195-207.

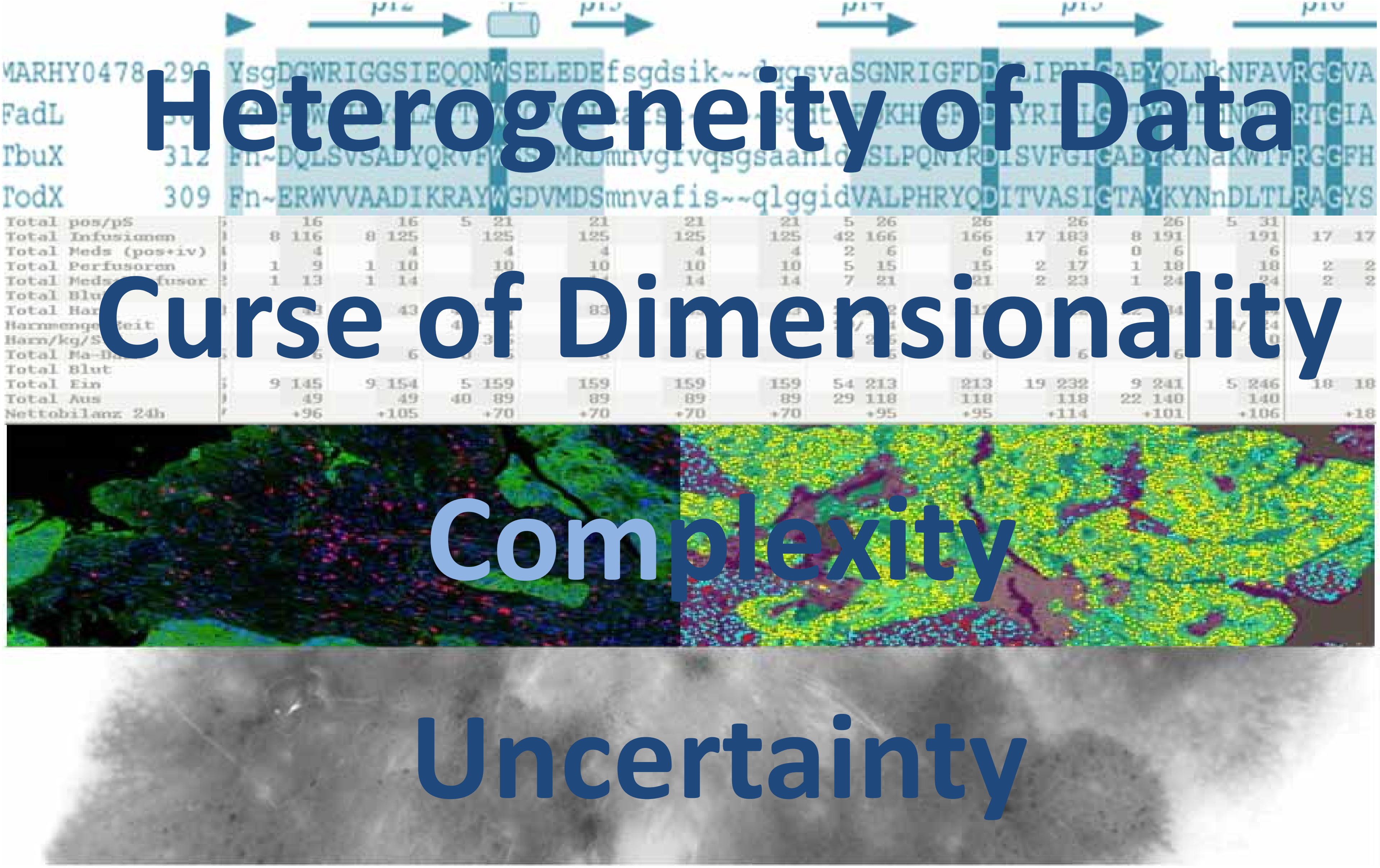
In medicine we have two different worlds …

Our central hypothesis:
Information bridges this gap

Holzinger, A. & Simonic, K.-M. (eds.) 2011. *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058, Heidelberg, Berlin, New York: Springer.*

Where is the problem in building this bridge

What are the problems?



Heterogeneity of Data

Curse of Dimensionality

Complexity

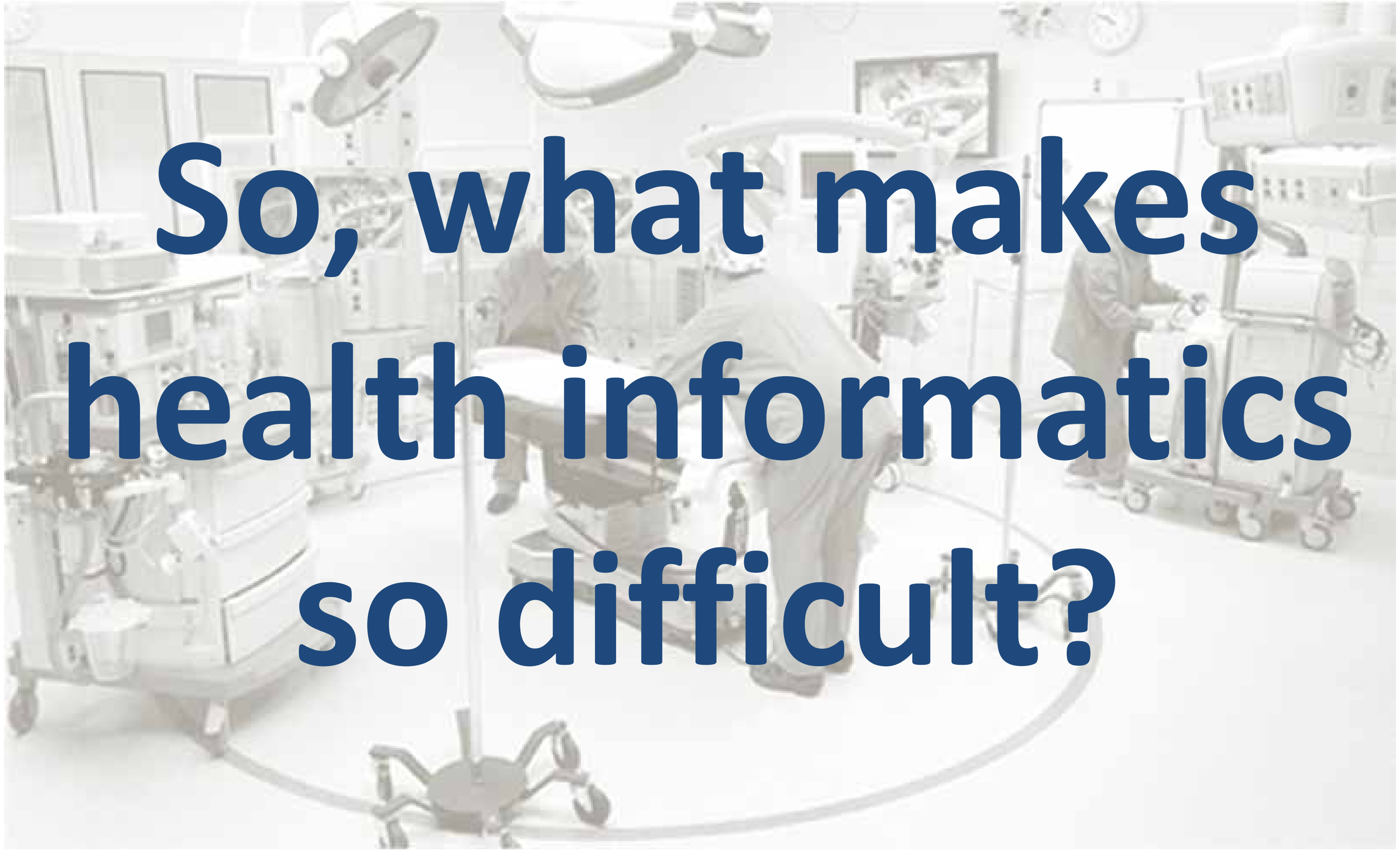
Uncertainty

Holzinger, A., Dehmer, M. & Jurisica, I. 2014. Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. BMC Bioinformatics, 15, (S6), I1.

A Look into the daily clinicial work

The beginning clinical clerk, the house officer and the practicing physician are all confronted with conditions that are frustrating in every phase of medical action. ... To deal effectively with these frustrations it will be necessary to develop a more organized approach to the medical record, a more rational acceptance and use of the paramedical personnel and a more positive attitude about the computer in medicine.

Weed, L. L. **1968.** Medical records that guide and teach. New England Journal of Medicine, 278, (12), 652-657.



So, what makes health informatics so difficult?

Probabilistic Information p(x)

Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances (Postum communicated by Richard Price). Philosophical Transactions, 53, 370-418.

$$p(x_i) = \sum P(x_i, y_j)$$

Thomas Bayes
1701 - 1761

$$p(x_i, y_j) = p(y_j|x_i)P(x_i)$$

Bayes' Rule is a corollary of the Sum Rule and Product Rule:

$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}$$

Barnard, G. A., & Bayes, T. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. Biometrika, 45(3/4), 293-315.

Bayes' Rule in words
d … data; h … hypothesis

$$p(h|d) = \frac{p(d|h)p(h)}{p(d)}$$

$$posterior\ p(x) = \frac{likelyhood * prior\ p(x)}{evidence}$$

The inverse probability allows to infer unknowns, <u>learn from data</u> and make predictions …

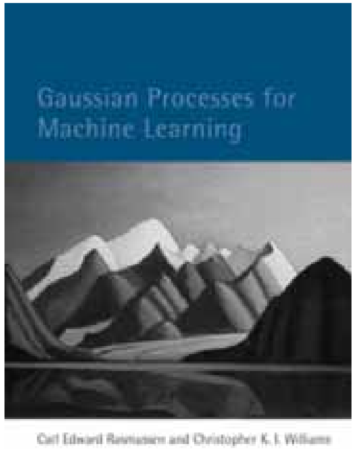
… machine learning!

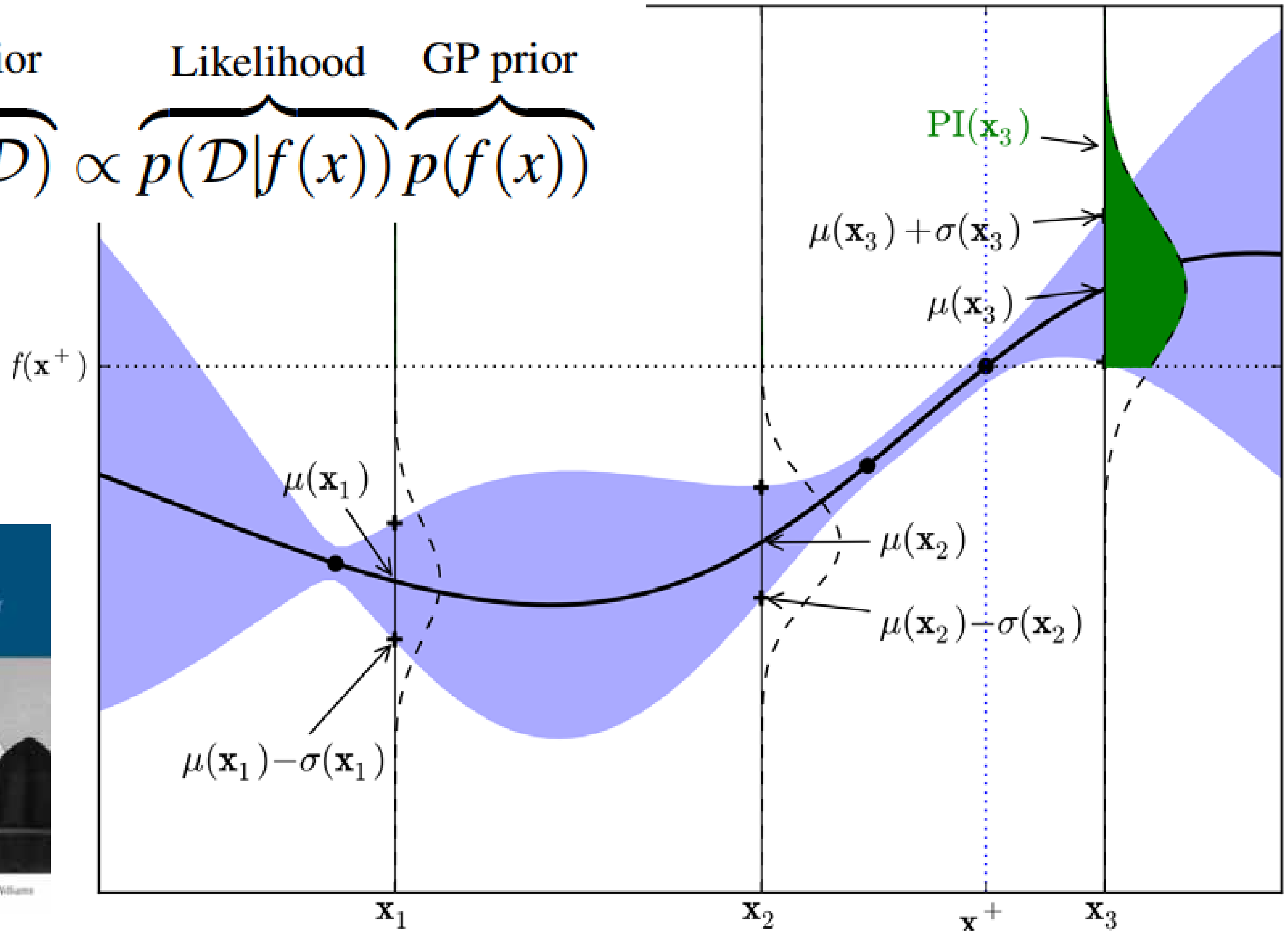
First hint:

$$\mathcal{N}(x;\mu,\Sigma) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^{\top}\Sigma^{-1}(x-\mu)\right]$$

GP = distribution, observations occur in a cont. domain, e.g. t or $\mathbb{R}$

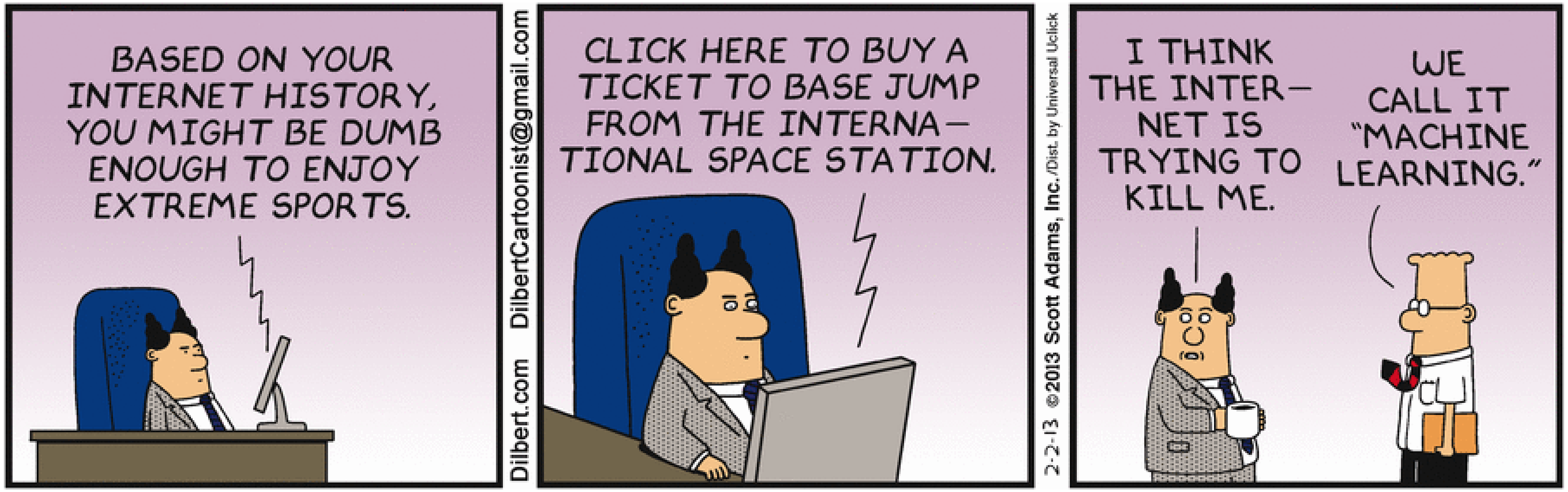
$$\overbrace{p(f(x)|\mathcal{D})}^{\text{GP posterior}} \propto \overbrace{p(\mathcal{D}|f(x))}^{\text{Likelihood}}\,\overbrace{p(f(x))}^{\text{GP prior}}$$

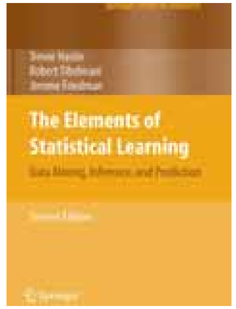




Brochu, E., Cora, V. M. & De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599.

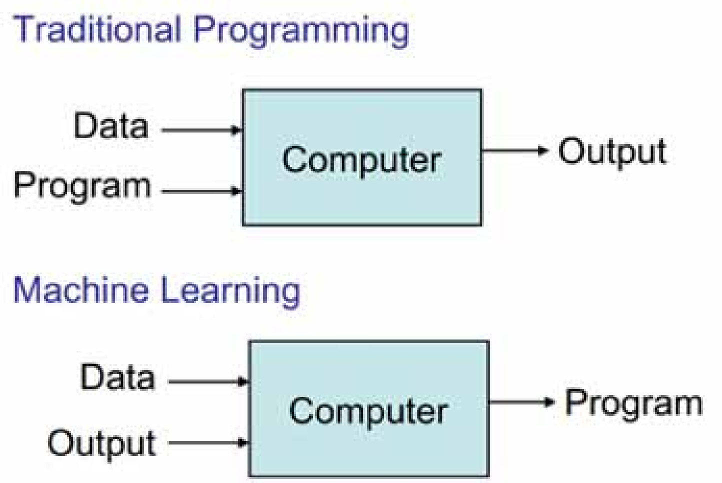
So, what is Machine Learning?





- Machine Learning is the development of algorithms which can **learn from data**
- and the assessment of **uncertainty,**
- Pre-history in **statistical learning.**
- **Automating automation - getting computers to program themselves – let the data do the work!**

Hastie, T., Tibshirani, R. & Friedman, J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, New York, Springer.



- Progress in ML is driven by the explosion in the availability of **“big” data** and **low-cost computation.**
- **Health is amongst the biggest challenges**

Jordan, M. I. & Mitchell, T. M. 2015. Machine learning: Trends, perspectives, and prospects. Science, 349, (6245), 255-260.

13

Graz, 26.01.2016

- Tom Mitchell: A scientific field is best defined by the **central question it studies.**
- ML seeks to answer the question
- *"How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?"*

Origin of ML: first def. by Arthur Samuel (1959)

SOME STUDIES IN
MACHINE LEARNING USING
THE GAME OF CHECKERS

by A. L. Samuel

Samuel, A. L. 1959. Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3, (3), 210-229.

Field of Study that gives computers the ability to **learn** [from Data] **without explicitly being programmed …**

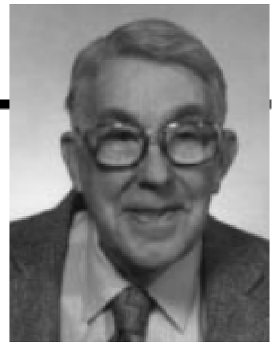
Introduction

The studies reported here have been co
a digital computer to behave in a way w
animals, would be described as involvin
this is not the place to dwell on the imp
cedures, or to discourse on the philosoph
very large amount of work, now done by
its demands on the intellect but does, ne
We have at our command computers wi
and with sufficient computational speed
techniques, but our knowledge of the ba
is still rudimentary. Lacking such know
methods of problem solution in minute a
and costly procedure. Programming com
should eventually eliminate the need for
ming effort.

Memoriam

AI Magazine Volume 11 Number 3 (1990) (© AAAI)

In Memoriam



Arthur Samuel: Pioneer in Machine Learning

Arthur Samuel (1901–1990) was a pioneer of artificial intelligence research. From 1949 through the late 1960s, he did the best work in making computers learn from their experience. His vehicle for this work was the game of checkers.

Programs for playing games often fill the role in artificial intelligence research that the fruit fly Drosophila

Samuel was a modest man, and the importance of his work was widely recognized only after his retirement from IBM in 1966, in part because he didn't relish the politics that were required to have his research more vigorously followed up on. He was also realistic about the large difference between what had been accomplished in understanding intellectual mechanisms and what would be required to reach human-level intelligence.

Samuel's papers on machine learn-

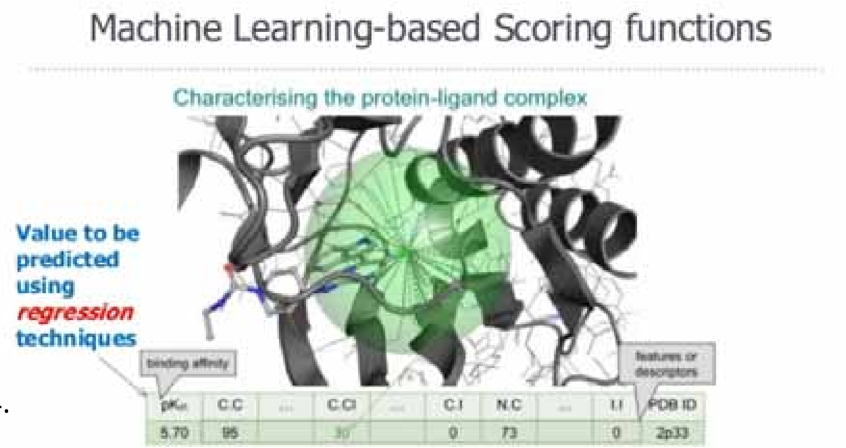
strate the power of electronic computers. He didn't finish the program while he was at the university of Illinois, perhaps because the computer wasn't finished in time.

In 1949, Samuel joined IBM's Poughkeepsie Laboratory, where he worked on IBM's first stored program

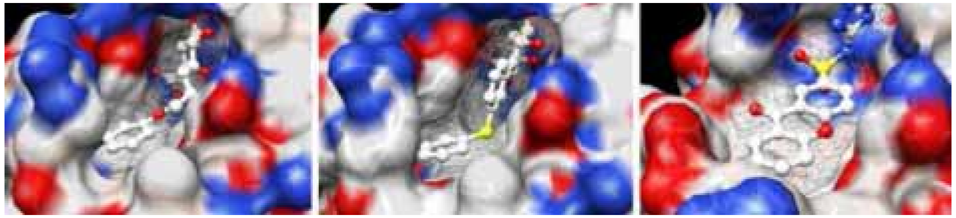
Mccarthy, J. & Feigenbaum, E. A. 1990. In Memoriam: Arthur Samuel: Pioneer in Machine Learning. AI Magazine, 11, (3), 10.

2015 everything is machine learning …

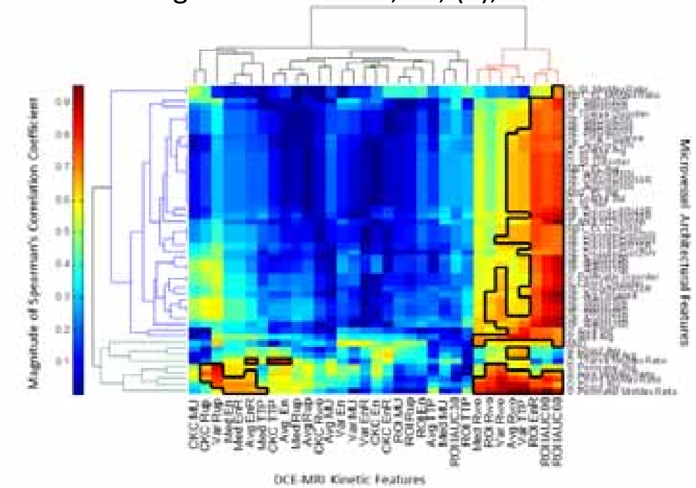
Takacs, G., Pilaszy, I., Nemeth, B., Tikk, D. & Acm 2008. Matrix Factorization and Neighbor Based Algorithms for the Netflix Prize Problem. Recsys'08: Proceedings of the 2008 ACM Conference on Recommender Systems, 267-274.



Ballester, P. J. & Mitchell, J. B. O. 2010. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics, 26, (9), 1169-1175.



Khamis, M. A., Gomaa, W. & Ahmed, W. F. 2015. Machine learning in computational docking. Artificial Intelligence in Medicine, 63, 3,135-152



Singanamalli, A. et al 2013: A radiohistomorphometric approach. SPIE Medical Imaging, 867604-867604-14.

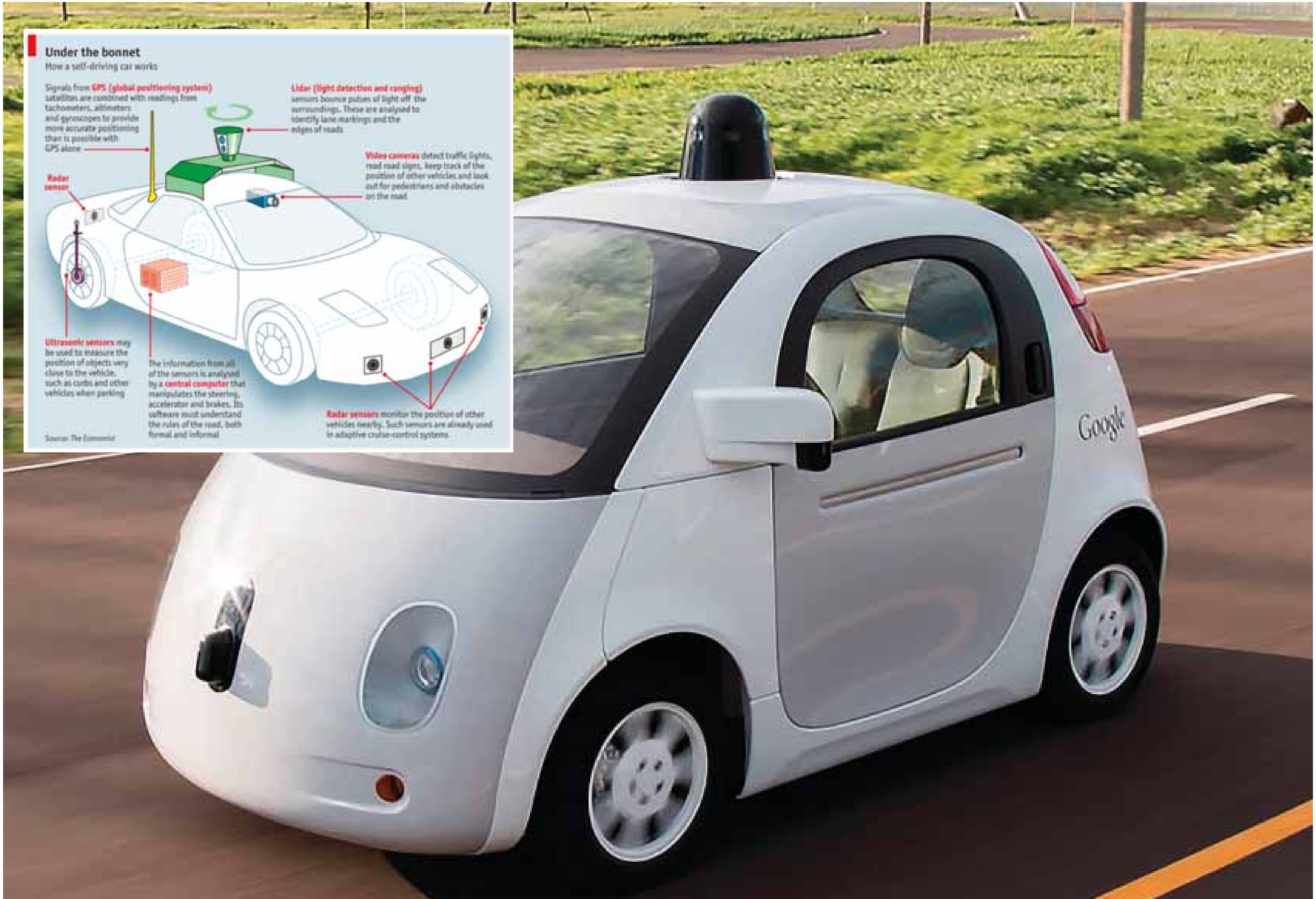
Schoenauer, M., Akrour, R., Sebag, M. & Souplet, J.-C. Programming by Feedback. Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014 Beijing. 1503-1511.

- **Supervised learning (function approximation)**
  - Given x, y pairs; find a f that map a new x to a proper y
  - $y = f(x)$
  - Regression, classification
- **Unsupervised learning (clustering)**
  - Given x (features only), find f that gives you a description of x
  - $f(x)$
  - Find similar points in high-dimensional spaces

- most of ML is automatic Machine Learning

- automatic Machine Learning (aML) := algorithms which interact with agents and can optimize their learning behaviour trough this interaction

What is a best practice example of aML …



Dietterich, T. G. & Horvitz, E. J. 2015. Rise of concerns about AI: reflections and directions. Communications of the ACM, 58, (10), 38-40.


Does this work here
as well?

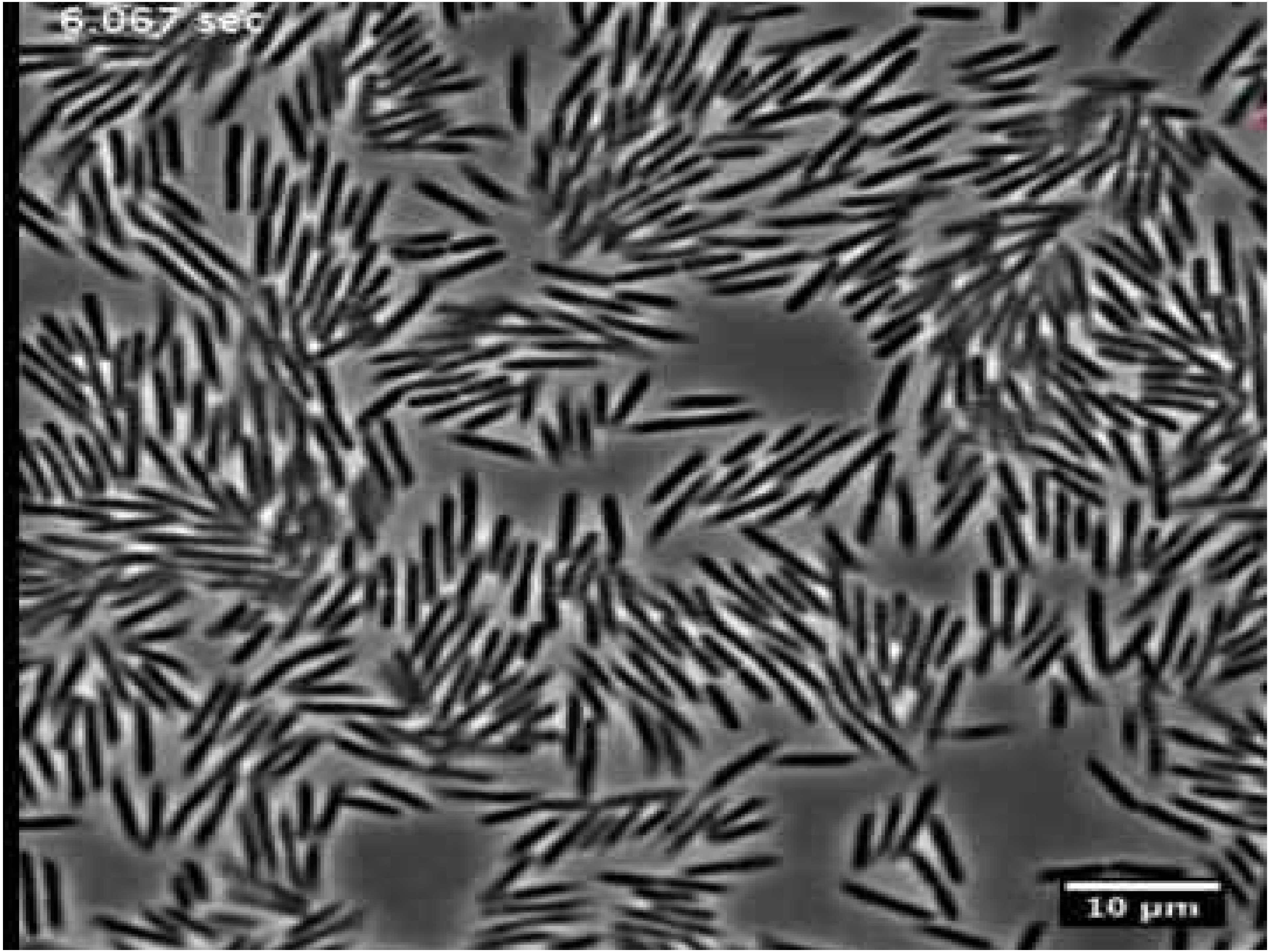
- Sometimes we **do not have “big data”,** where aML-algorithms benefit.
- Sometimes we have
  - **Small amount of data sets**
  - Rare Events – **no training samples**
  - **NP-hard problems, e.g.**
    - Subspace Clustering,
    - Protein-Folding,
    - k-Anonymization,
    - Graph Coloring, Category Discovery, etc. etc.…

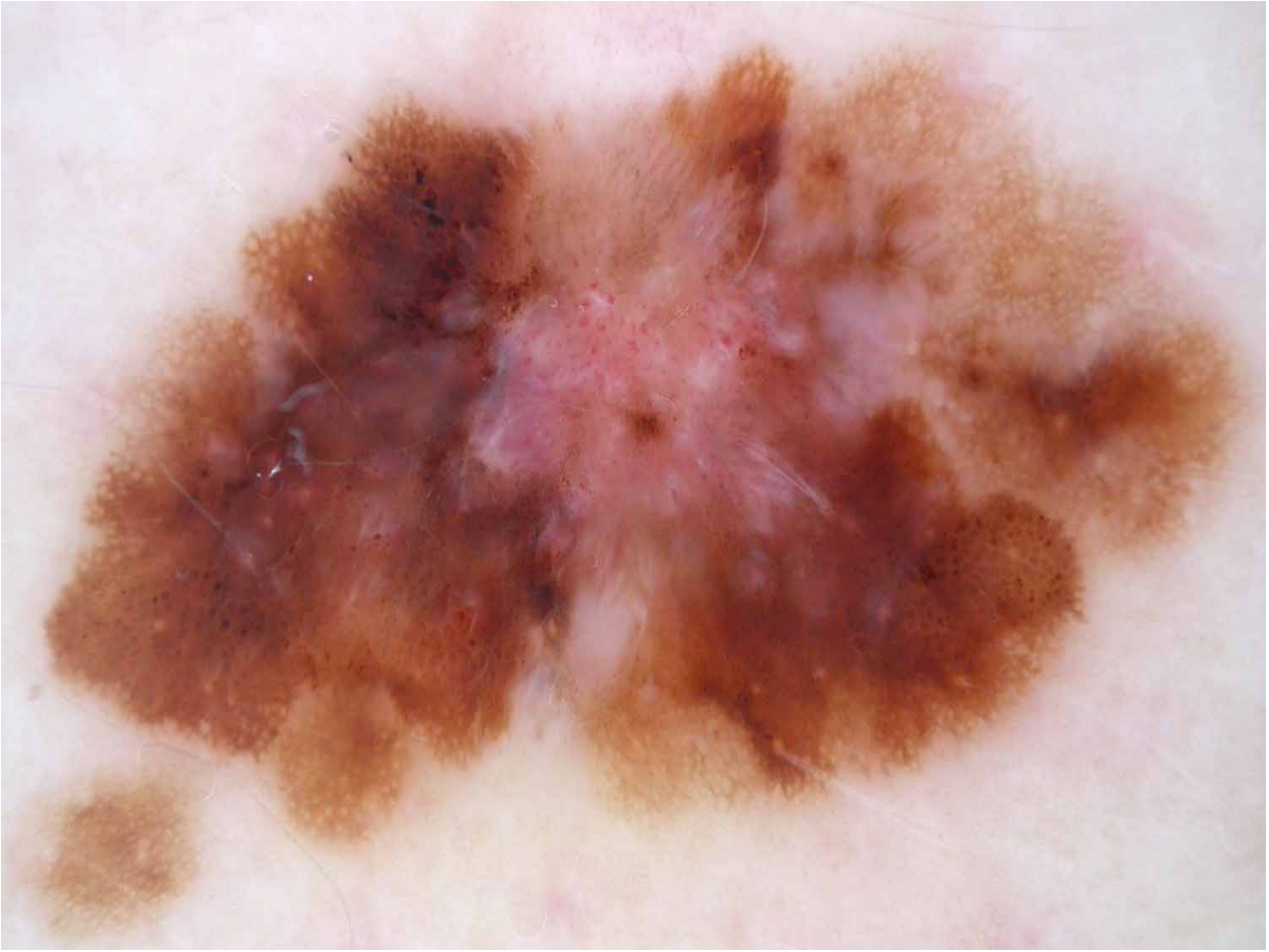
Sometimes we (still) need a human-in-the-loop


2
3
4
5
6

- interactive Machine Learning (iML) := algorithms which interact with agents*) and can optimize their learning behaviour trough this interaction
- *)**where the agents can be human**

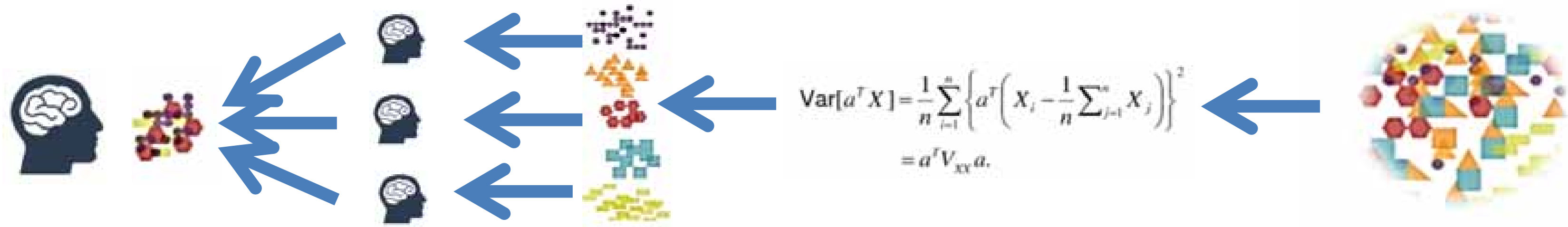
Holzinger, A. 2015. Interactive Machine Learning (iML). Informatik Spektrum
DOI: 10.1007/s00287-015-0941-6


6.067 sec
10 µm

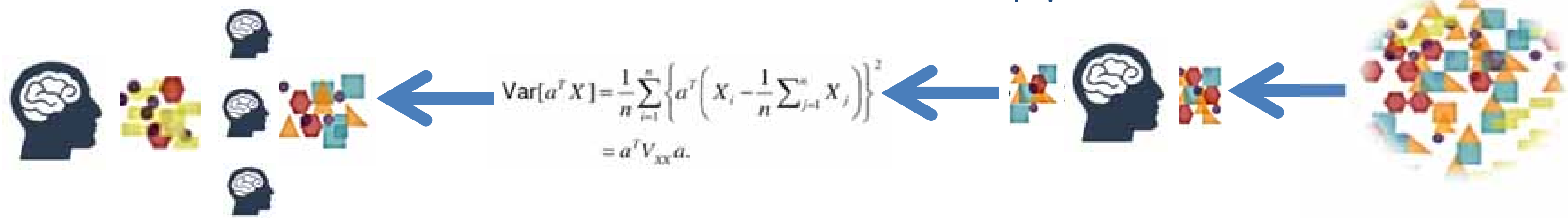


aML: Unsupervised – Supervised – Semi-supervised

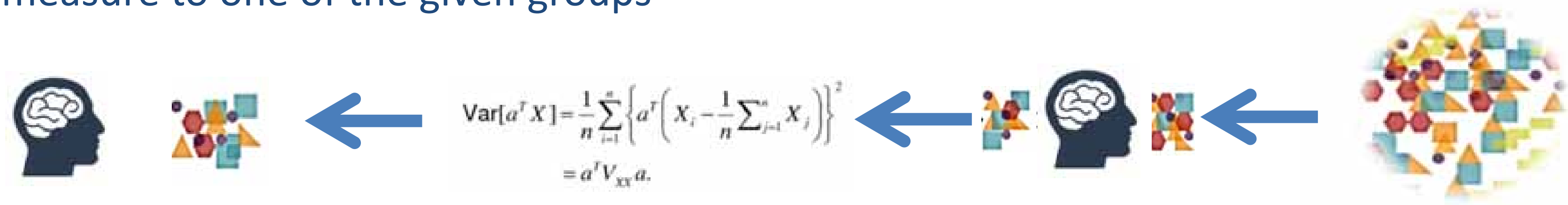
A) Unsupervised ML: Algorithm is applied on the raw data and learns fully automatic – Human can check results at the end of the ML-pipeline



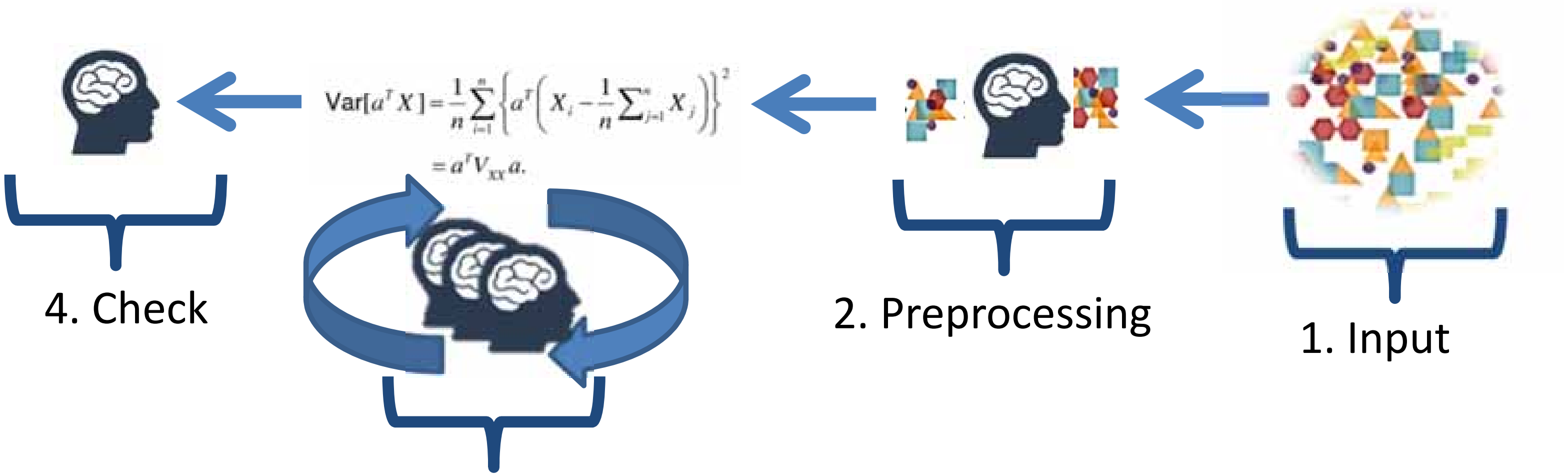
B) Supervised ML: Humans are providing the labels for the training data and/or select features to feed the algorithm to learn – the more samples the better – Human can check results at the end of the ML-pipeline



C) Semi-Supervised Machine Learning: A mixture of A and B – mixing labeled and unlabeled data so that the algorithm can find labels according to a similarity measure to one of the given groups



D) **Interactive Machine Learning:** Human is seen as an agent involved in the actual learning phase, step-by-step influencing measures such as distance, cost functions …



Constraints of humans: Robustness, subjectivity, transfer?

Open Questions: Evaluation, replicability, …

Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Springer Brain Informatics (BRIN), 3, 1-13. http://link.springer.com/article/10.1007/s40708-016-0042-6

Holzinger, A. 2016. Interactive Machine Learning (iML). Informatik Spektrum, 39, (1), 64-68.

- Example 1: **k-Anonymity**
- Example 2: **Protein Folding**
- Example 3: **Subspace Clustering**

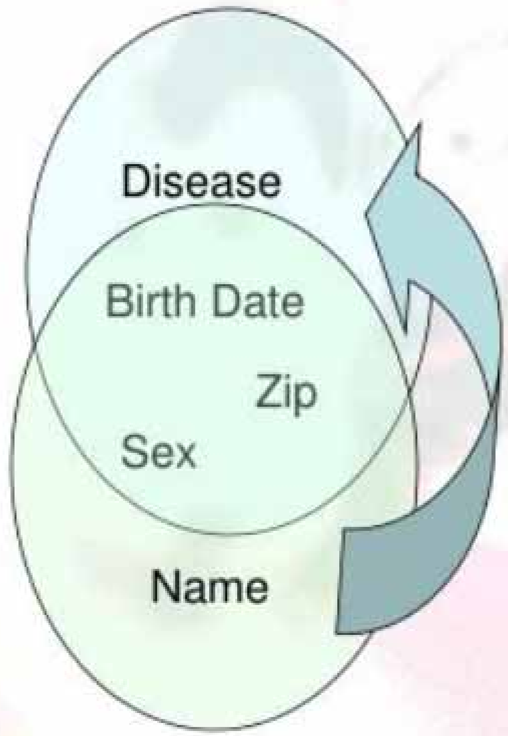
87 % of the population in the USA can be uniquely re-identified by Zip-Code, Gender and date of birth

Hospital Patient Data

Birthdate	Sex	Zipcode	Disease
1/21/76	Male	53715	Flu
4/13/86	Female	53715	Hepatitis
2/28/76	Male	53703	Brochitis
1/21/76	Male	53703	Broken Arm
4/13/86	Female	53706	Sprained Ankle
2/28/76	Female	53706	Hang Nail

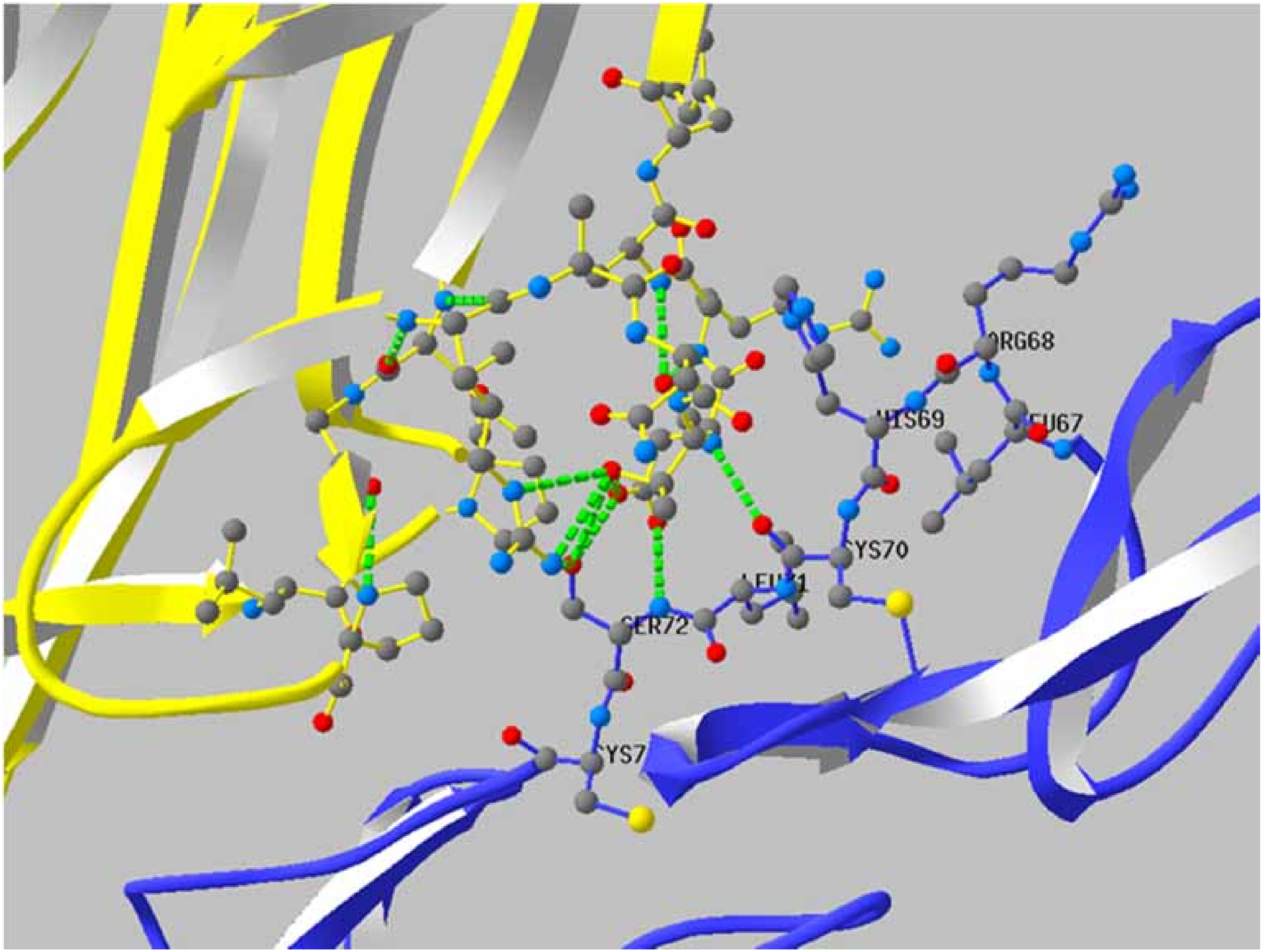
Voter Registration Data

Name	Birthdate	Sex	Zipcode
Andre	1/21/76	Male	53715
Beth	1/10/81	Female	55410
Carol	10/1/44	Female	90210
Dan	2/21/84	Male	02174
Ellen	4/19/72	Female	02237



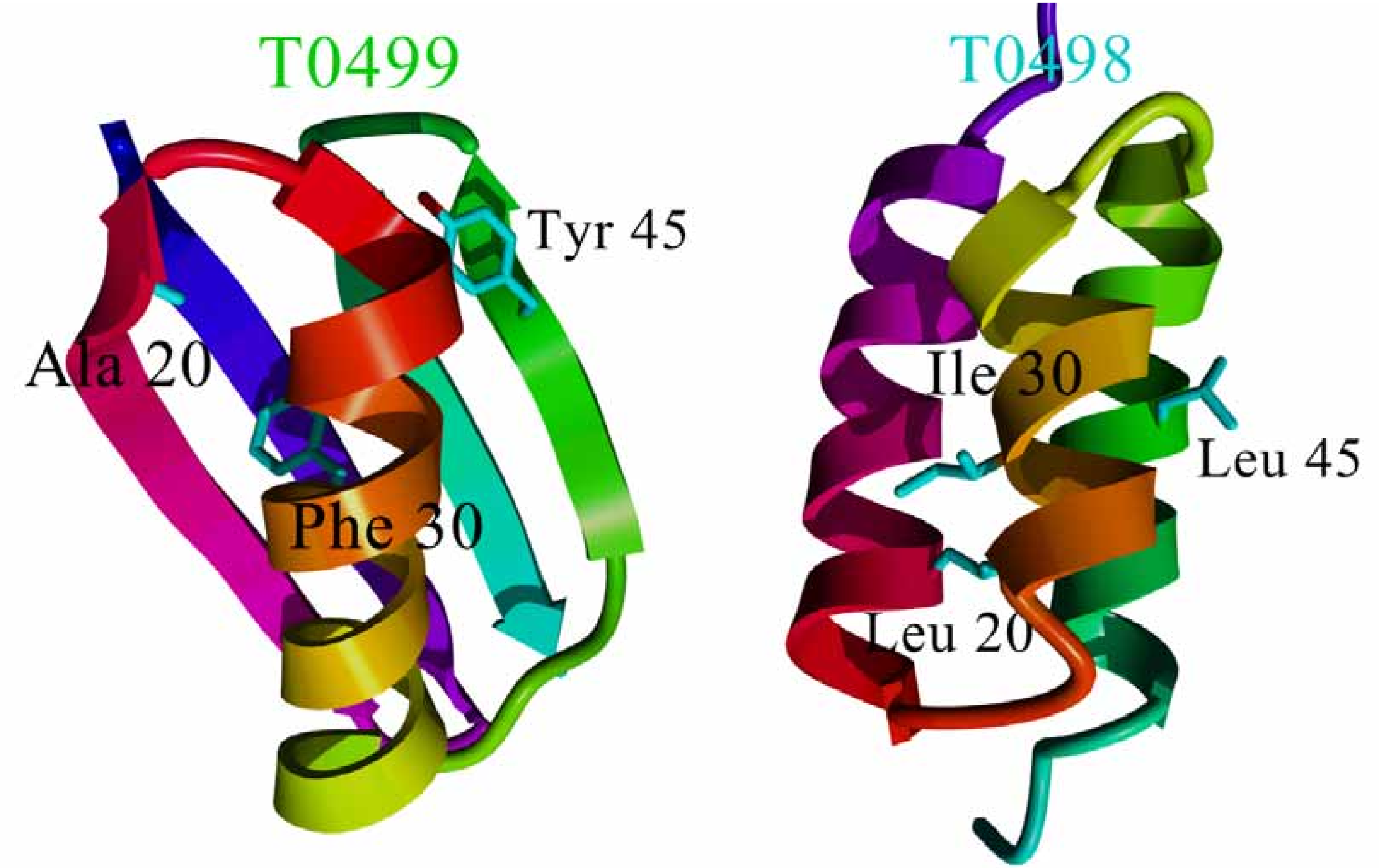
Sweeney, L. 2002. Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10, (05), 571-588.

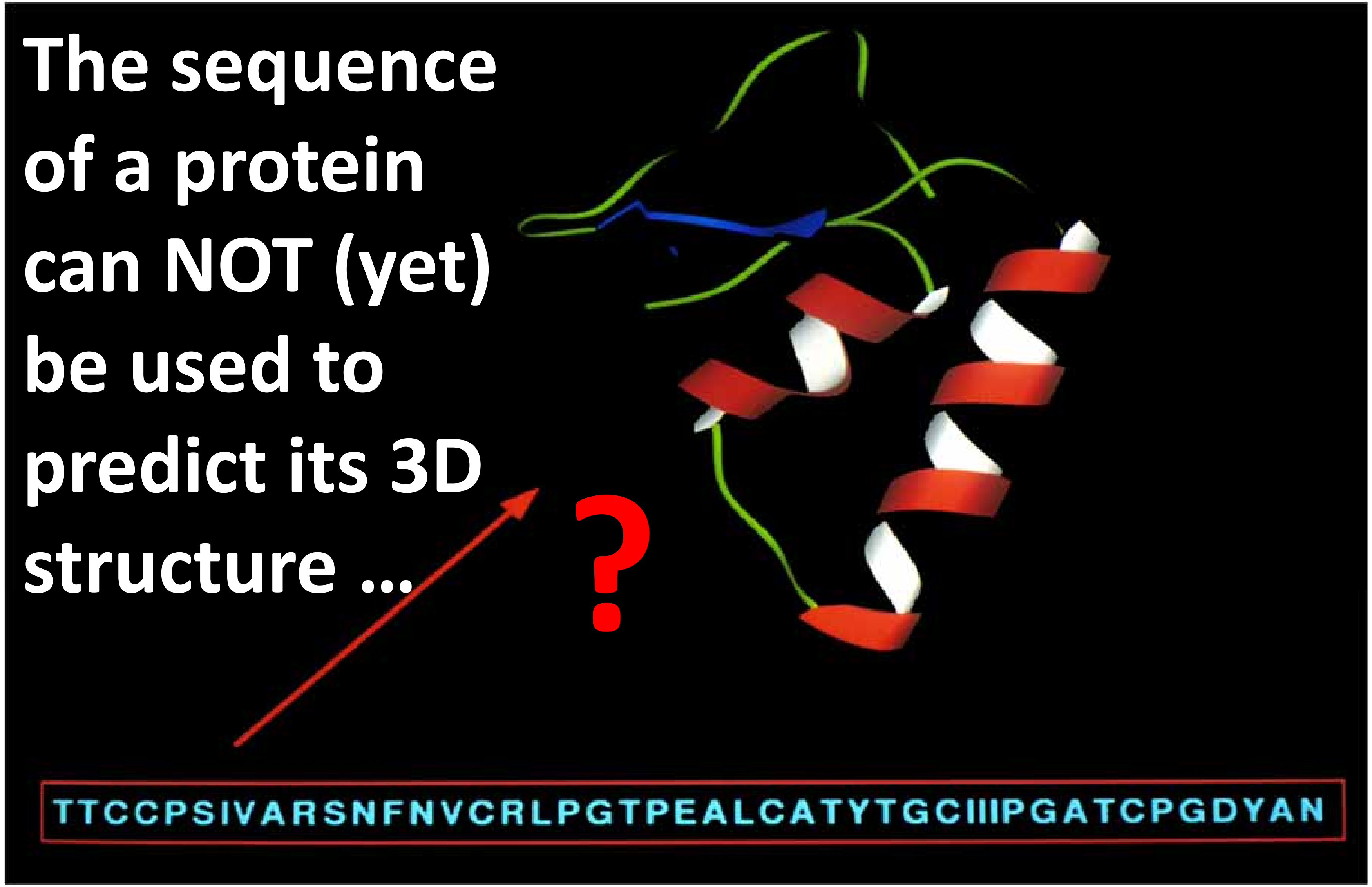
Proteins are the building blocks of life …



Wiltgen, M., Holzinger, A. & Tilz, G. P. (2007) Interactive Analysis and Visualization of Macromolecular Interfaces Between Proteins. In: *Lecture Notes in Computer Science (LNCS 4799). Berlin, Heidelberg, New York, Springer, 199-212.*

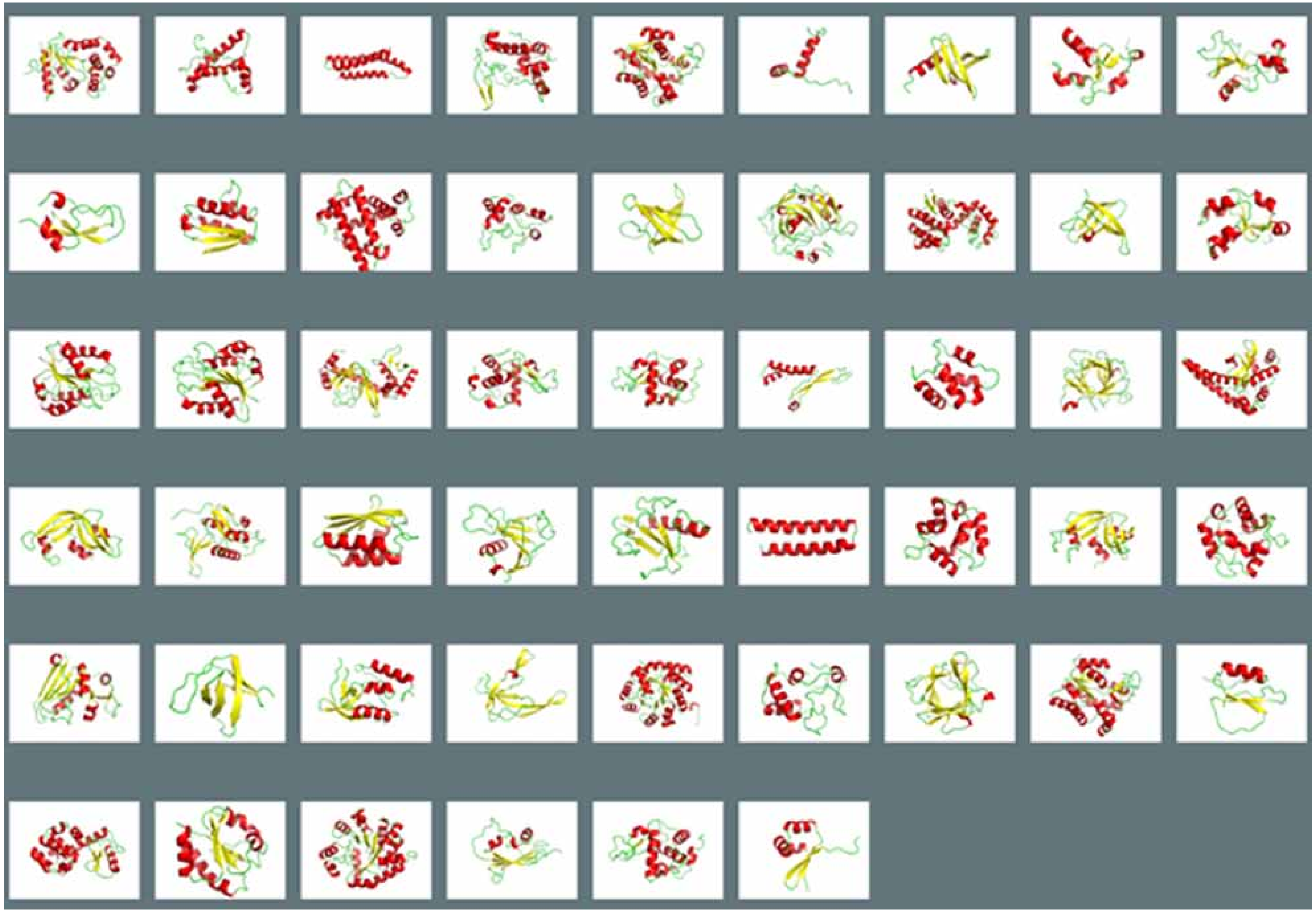
He, Y., Chen, Y., Alexander, P., Bryan, P. N. & Orban, J. (2008) NMR structures of two designed proteins with high sequence identity but different fold and function. Proceedings of the National Academy of Sciences, 105, 38, 14412.





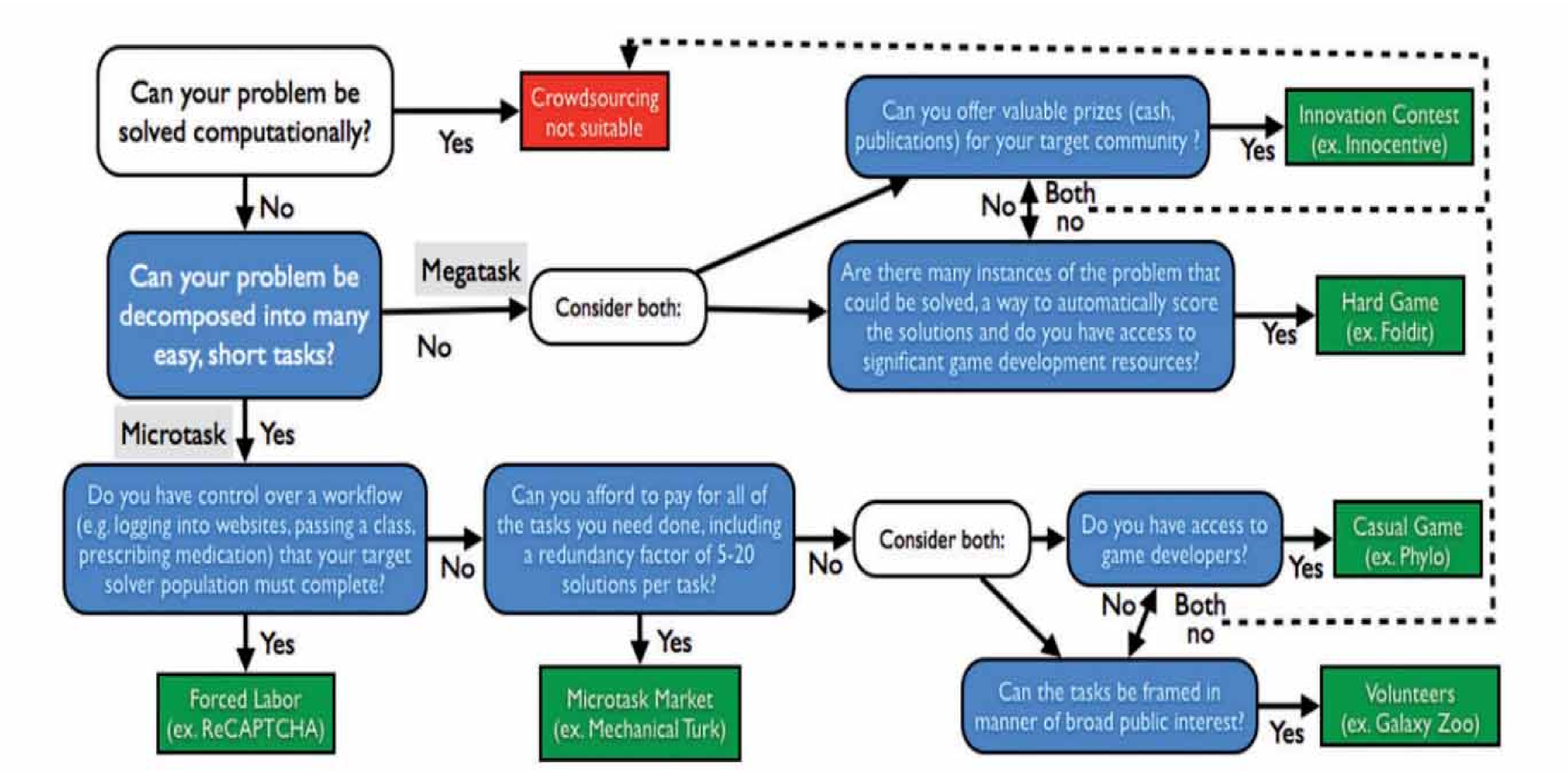
Anfinsen, C. B. **1973.** Principles that Govern the Folding of Protein Chains. Science, 181, (4096), 223-230.

State-of-2015: aML still is not able to help sufficiently …



Jia, L., Yarlagadda, R. & Reed, C. C. 2015. Structure Based Thermostability Prediction Models for Protein Single Point Mutations with Machine Learning Tools. Plos One, 10, (9).

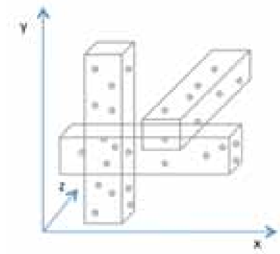
Humans can help here: Crowdsourcing

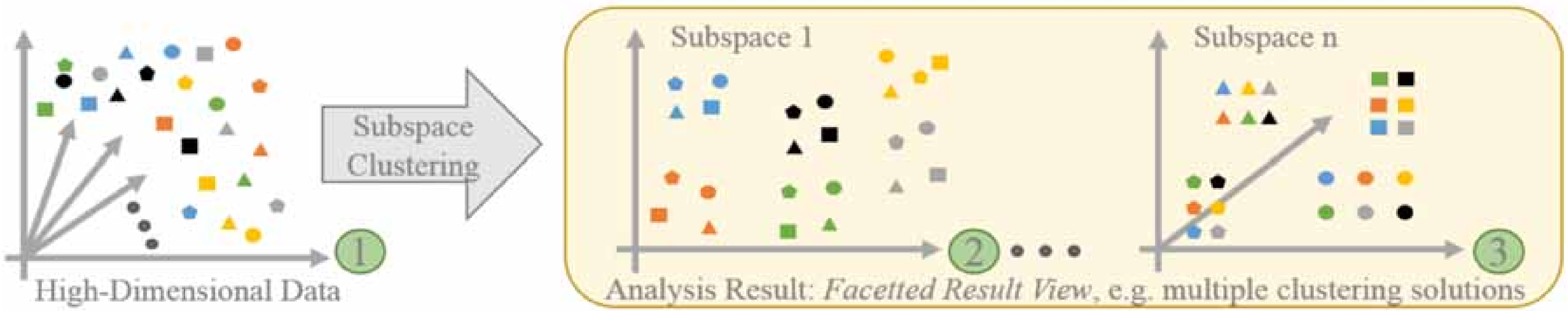


Good, B. M. & Su, A. I. 2013. Crowdsourcing for bioinformatics. *Bioinformatics,* 29, (16), 1925-1933.

Example 3: Subspace Clustering

- Patterns may be found in subspaces (dimension combinations)
- Clustering and subset selection: Non-convex & NP-hard problem
- Real data are often noisy and corrupted
- Little prior knowledge about low-dim structures
- Data points in different subgroups can be very close

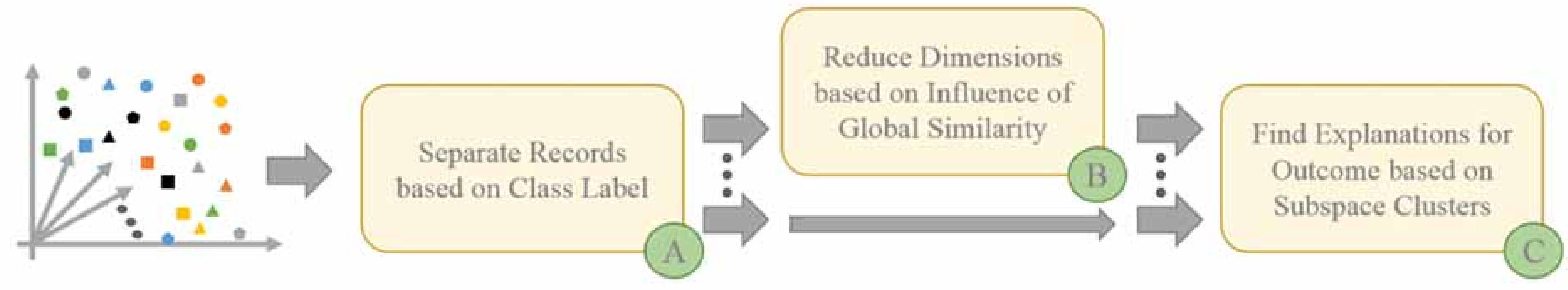




Hund, M., Sturm, W., Schreck, T., Ullrich, T., Keim, D., Majnaric, L. & Holzinger, A. 2015. Analysis of Patient Groups and Immunization Results Based on Subspace Clustering. In: Lecture Notes in Artificial Intelligence LNAI 9250, pp. 358-368.

Nr.	Name	Type	missing	Description
1	age	numerical	0	age (years)
2	sex	binary	0	(M=Male, F=Female)
3	Hyper	binary	0	Hypertension (yes, no)
4	DM	nominal	0	Diabetes mellitus (yes, IGT=Impaired glucose tolerance, no)
5	F Glu	numerical	0	Fasting blood glucose (mmol/L)
6	HbA1c	numerical	0	Glycosilated Haemoglobin (%) (showing average blood glucose during last three months)
7	Chol	numerical	0	Total Cholesterol (mmol/L)
8	TG	numerical	0	Triglycerides (mmol/L)
9	HDL	numerical	0	HDL-cholesterol (mmol/L)
10	Statins	binary	0	Therapy with statins (yes,no)
11	Anticoag	binary	1	Therapy with anticoagulant/antiaggregant drugs (yes,no)
12	CVD	binary	0	Cardiovascular diseases (yes, no) (myocardial infarction, angina, history of revascularisation, stroke, transient ischaemic cerebral event, peripheral vascular disease)
13	BMI	numerical	0	Body Mass Index (kg/m^2)
14	w/h	numerical	0	Waist/hip ratio
15	Arm cir	numerical	1	Mid arm circumference (mm)
16	skinf	numerical	0	Triceps skinfold thickness (mm)
17	gastro	binary	0	Gastroduodenal disorders (yes,no) (gastritis, ulcer)
18	uro	binary	1	Chronic urinary tract disorders (yes,no) (recurrent cystitis in women, symptoms of prostatism in men)
19	COPB	binary	0	Chronic obstructive pulmonary disease (yes,no)
20	Aller d	binary	0	Allergy (Rhinitis and/or Asthma) (yes,no)
21	dr aller	binary	0	Drugs allergy (yes, no)
22	analg	binary	0	Therapy with analgetics/NSAR (yes,no)
23	derm	binary	0	Chronic skin disorders (yes,no) (chronic dermatitis, dermatomycosis)
24	neo	binary	0	Malignancy (yes,no)
25	OSP	binary	18	Osteoporosis (yes, no)
26	Psy	binary	0	Neuropsychiatric disorders (yes,no) (anxiety/depression, Parkinson‘s disease, cognitive impairments)
27	MMS	numerical	0	Mini Mental Score - test for screening on cognitive dysfunction Max Score=30 Score ¡24 indicates cognitive impairment
28	CMV	numerical	0	Cytomegalovirus specific IgG antibodies (IU/ml)
29	EBV	numerical	0	Epstein-Barr virus specific IgG (IU/ml)
30	HPG	numerical	0	Helicobacter pylori specific IgG (IU/ml)
31	HPA	numerical	0	Helicobacter pylori specific IgA (IU/ml)
32	LE	numerical	0	Leukocytes Number x10^9/L
33	NEU	numerical	0	Neutrophils $ in White Blood Cell differential
34	EO	numerical	0	Eosinophils % in White Blood Cell differential
35	MO	numerical	0	Monocytes % in White Blood Cell differential
36	LY	numerical	0	Lymphocytes % in White Blood Cell differential
37	CRP	numerical	1	C-reactive protein (mg/L)
38	E	numerical	0	Erythrocytes number x10^{12}/L
39	HB	numerical	0	Haemoglobin (g/L)
40	HTC	numerical	0	Haematocrite (erythrocyte volume blood fraction)
41	MCV	numerical	0	Mean cell Volume (fL)
42	FE	numerical	0	Iron (g/L)
43	PROT	numerical	2	Total serum proteins (g/L)
44	ALB	numerical	0	Serum albumin (g/L)
45	clear	numerical	1	Creatinine clearance (ml/s/1.73m^2)
46	HOMCIS	numerical	0	Homocistein (μmol/L)
47	ALFA1	numerical	0	Serum protein electrophoresis (g/L)
48	ALFA2	numerical	0	Serum protein electrophoresis (g/L)
49	BETA	numerical	0	Serum protein electrophoresis (g/L)
50	GAMA	numerical	0	Serum protein electrophoresis (g/L)
51	RF	numerical	0	Rheumatoid Factor level (IU/ml)
52	VITB12	numerical	0	Vitamin B12 (pmol/L)
53	FOLNA	numerical	0	Folic acid (mM/L)
54	INS	numerical	0	Insulin (μIU/L)
55	CORTIS	numerical	0	Cortisol in the morning (nmol/L)
56	PRL	numerical	0	Prolactin in the morning (mIU/L)
57	TSH	numerical	1	Thyroid-stimulating hormone (IU/ml)
58	FT3	numerical	0	Free triiodothyronine (pmol/L)
59	FT4	numerical	0	Free thyroxine (pmol/L)

What can the “doctor-in-the-loop” do?



- **Positive** subspace clusters
  - One homogeneous cluster (healthy patients)
    - `hyper, CVD, neoplasm, psy.disorder, drug allergy`
    - No medications: `statins, anticoagulants, analgesics` and `clear` (preserved renal function)
- **Negative** subspace clusters
  - Cluster with obvious reasons for neg. outcome
    - Impairment of certain pathophysiologic mechanism increased MCV, decreased `VITB12, FOLNA, CORTIS`) despite no: `DM, drug allergy, Fglu, E/HB` (anaemia)

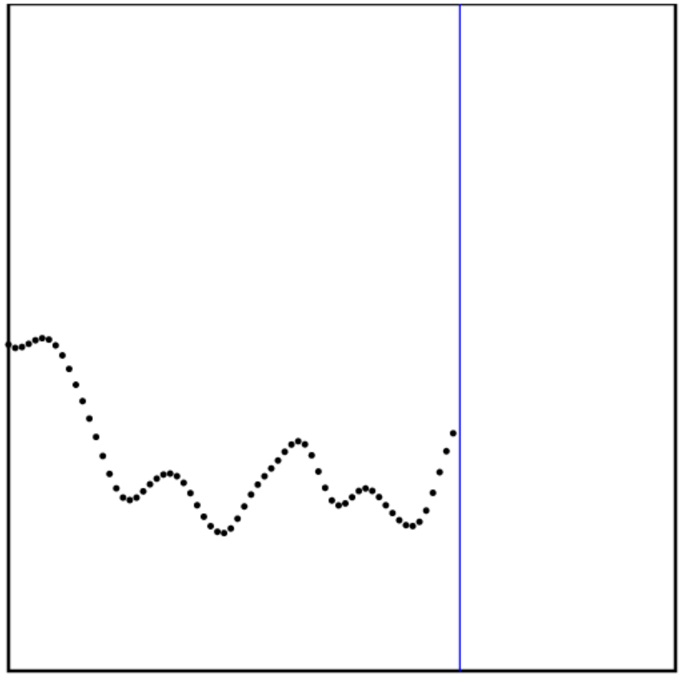
Human kernel ...

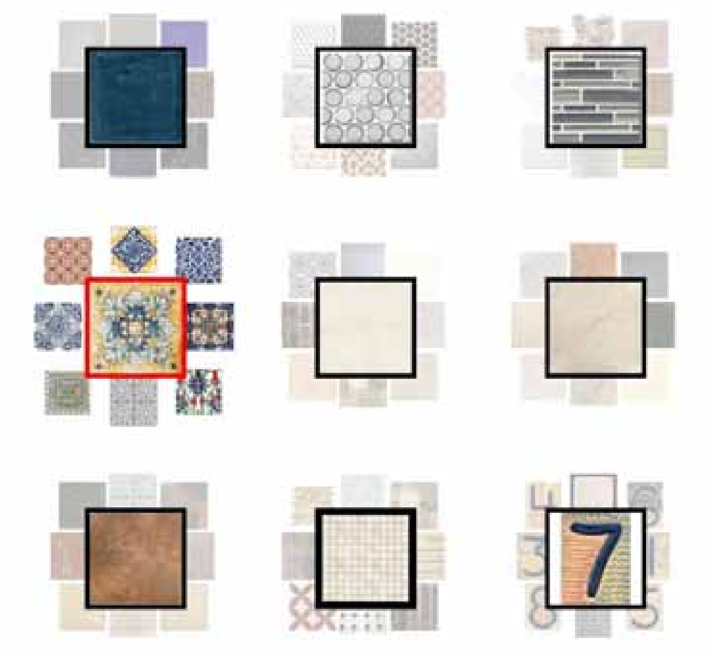
Judgment 1 out of 33

This is the first function from the system. Please try to predict the new points as well as you can based on the points you can see.

Please click along the blue line to say what you think the height of the point is for that location.

Once you have selected a position along the line, **hit the 's' key to submit the point**.

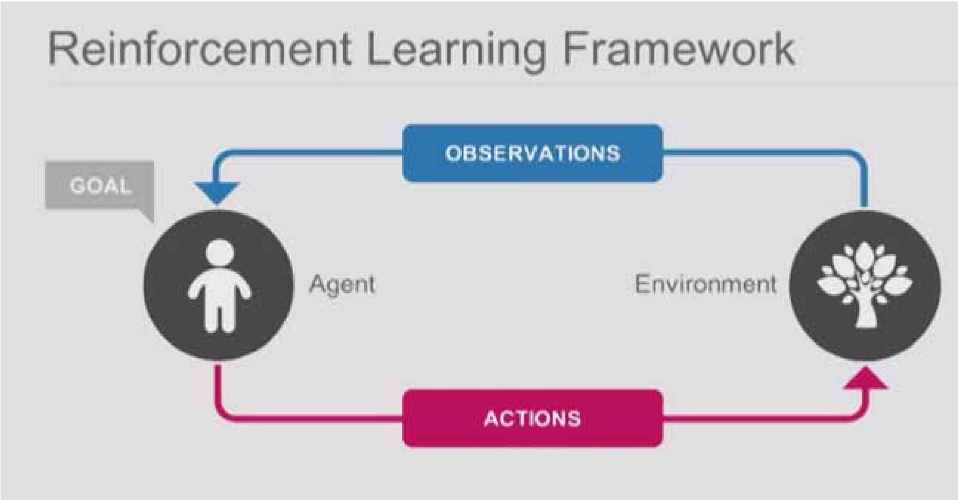




Wilson, A. G., Dann, C., Lucas, C. & Xing, E. P. The Human Kernel. Advances in Neural Information Processing Systems, 2015. 2836-2844.

$$\hat{p}^{a}_{bc} = \frac{\mu + \delta_{ac}}{2\mu + \delta_{ab} + \delta_{ac}} \quad \text{and} \quad K_{ii} = 1,$$

- **Reinforcement Learning** (1950)
- **Preference Learning** (1987)
- **Active Learning** (1996)



Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. & Hassabis, D. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature,* 529, (7587), 484-489.

- ❶ Heterogeneous data sources
  - need for data integration and data fusion
- ❷ Complexity – reduction of search space
  - combining the best of Human & Computer
- ❸ What is interesting? – and relevant!
  - need of effective mapping $\mathbb{R}^N \rightarrow \mathbb{R}^2$
- ❹ Clinical time limits “5 Minutes”
  - need of efficient solutions

Holzinger, A. & Jurisica, I. 2014. Knowledge Discovery and Data Mining in Biomedical Informatics: The future is in Integrative, Interactive Machine Learning Solutions In: LNCS 8401. Heidelberg, Berlin: Springer, pp. 1-18.

- **Multi-Task Learning (MTL)**
  - for improving prediction performance, help to reduce catastrophic forgetting
- **Transfer learning (TL)**
  - is not easy: learning to perform a task by exploiting knowledge acquired when solving previous tasks:
  - a solution to this problem would have major impact to AI research generally and machine learning specifically.
- **Multi-Agent-Hybrid Systems (MAHS)**
  - To include swarm-intelligence and crowdsourcing

HW LV185.A83 S2016

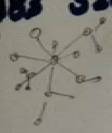
MATHEMATICS BASICS
- Linear Algebra
- Probability
- Statistical Inference
- Optimization
- Measure Theory
- Spaces
- Complexity
- Algorithm

BIOMEDICAL SPECIALITIES / CHALL
- noise
- sparsity
- complexity
- dirty data
- missing data
- real-world

Bayesian Statistics
Occam
Jeffreys-Lindley
Decision Theory

Cognition
cognitive science
perception
decision making

R^2 R^n



- Physics of data
- databases
- data structures
- data management
- data preprocessing

data mapping

HCI ⇄ KDD
visualization
interaction

CHALLENGES GENERALLY
- big data sets
- by high-dimensional problem
- missing data, dirty data,
- complicated set of factors
- local, global, ensemble
- NP-hard, small data sets
- GENERALIZATION
- TRANSFER

CONCEPTS

▷ Data visualization

ICA

non-parametric

- feature selection
- feature extraction

PCA, FA, MDS — linear
LDA — supervised

Curse of R^n

Dimensionality reduction

Isomap, LLE, autoencoder — non-linear

linear regression
non-linear
Logistic regression

Laplacian Eigenmaps

over fitting
model selection

▷ Nearest-Neighbor
▷ Neighborhood
▷ ARENKOTT
▷ ...
▷ ...

DISTANCES
▷ NearestNeighbors ARENKOTT ?

no free lunch theorem

THEORIES

- Information theory (Entropy) - mutual info
- KL-divergence
- BAYESIAN $p(x)$
- COMPLEXITY (Runtime) → NP-completeness
- PAC-Learning
- Vapnik Chervonenkis
- Bias-variance
- Monte-Carlo

TECHNIQUES

MODELS

- Gaussian Processes p. 545 (max. Entropy of the Gaussian)
- Graphical models (Bayes Nets, graphs, Naïve Bayes, Markov Blanket) MEP p. 313
- NN
- SVM
- linear — Factor Analysis, PCA, pPCA, ICA
- Decisiontree — C4.5, Random forests, ID3, CART

METHODS
- regularization
- validation
- aggregation
- hyper parameters

PARADIGMS

Given x (feature only) find f that gives a description of x

Unsupervised — discovery, clustering — K-means

function approx.

Supervised — classification, regression
$f: X \to Y$, $y = f(x)$
find a f that maps x to a property

regression — Stochastic Gradient Descent, Lasso, Elastic Net, SVR Kernel, Ensemble Regressors

semi-supervised
Online
interactive ML

learn from interaction
RL (Turing, 1950)
Kaelbling et al (1996)
Littmann (2015) NATURE

- Single agent RL
- n-armed bandits
- multiplayer stochastic games
- MARL → swarm
personal traits

Busoni et al (2008)

Multi-Task Feature learning (MTFL)
Thrun & Pratt (1998), Vert, Argyriou, Evgeniou, Pontil (2007)
transfer learning → "catastrophic forgetting"

The Human Kernel
Wilson, Dann, Lucas, Xing (2015)
J. Tenenbaum

KERNELS
SCHÖLKOPF MLSS 2013

Human learning and machine learning

Deep Learning. cs.cmu.edu
- Brain & Cognition — Bain (1873)
- Newson, Perceptron, Backpropagation
- Kohonen, BOLTZMANN
- Long-Short-Term Memory — Hochreiter, Schmidhuber

TOOLS: Torch, Theano, AWS, Numpy, Scipy, matplotlib
- WEKA
- Mathematica
- Matlab
- R
- Python
- Julia

Privacy Aware ML
- Features / Feature Spaces
- learning based intrusion detection
- automatic signature generation
- clustering methods for malware analysis
- pattern based vulnerability discovery
- privacy attacks

Experiments & Evaluation
What to measure?
How to measure?
How to interpret?

- Accuracy (Precision, Recall) - avoid overfitting
- ROC (Area under the curve)
- Training time
- Linearity

- Occam's Razor (Sir William Hamilton, 1852)

Cognition as probabilistic inference
▷ TENENBAUM MLSS 09

- visual perception
- language processing
- motor learning
- Associative learning
- memory
- Attention
- Concept learning
- Reasoning
- Causal inference

Symbiotic Multi-Robot Organism
Levi / Kernbach (2010)

PL — Medin et al. (1987) cognitive science
Sirmann & Kellermeier (2010)

CONCEPT LEARNING Angluin (1988)

AL — Cohn, Ghahramani, Jordan (1996)

ILO → ILASTIK: Interactive Learning and Segmentation Toolkit (2011)

APL — Akrour / Sebag
Grachen, Fürnkranz (2007)

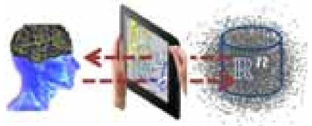
ML-Jungle Top Level View

Maths

Challenges

Cognition | Visualization | Data structure

Perception



Preprocessing

Decision | Interaction | Integration

Always with a focus/application in health informatics

CONCEPTS	THEORIES	PARADIGMS	MODELS	METHODS	TOOLS
Curse of Dim	Bayesian p(x)	unsupervised	Gaussian P.	Regularization	Python
DR					
NfL-Theorem	Complexity	supervised	Graphical M.	Validation	Julia
Overfitting	KL-Divergence	Semi-supv.	NN	Aggregation	Etc.
			DL		
Non-Parametric	Info Theory	online	SVM	Input Processes	Azure
		iML	Linear Models		
		RL	PL		
		AL	D. Trees		

Exp. & Eval.

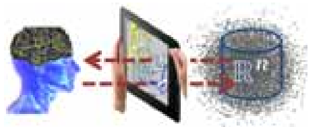
Privacy ML

Our plan

Cognition

Perception

Decision



Data structure

Preprocessing

Integration

Challenges

Always with a focus/application in health informatics

CONCEPTS

THEORIES

PARADIGMS

MODELS

METHODS

TOOLS

Curse of Dim

DR

Bayesian p(x)

Python

Graphical M.

NN

DL

Azure

iML

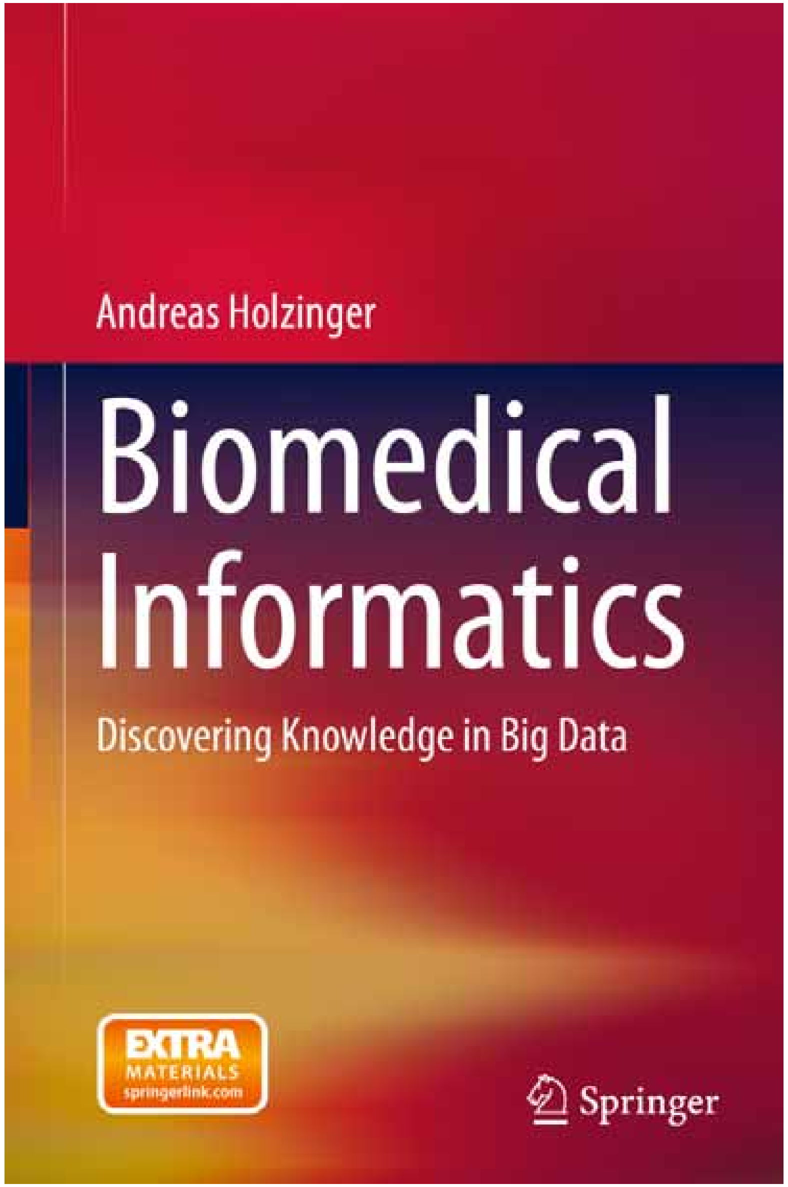
RL

PL

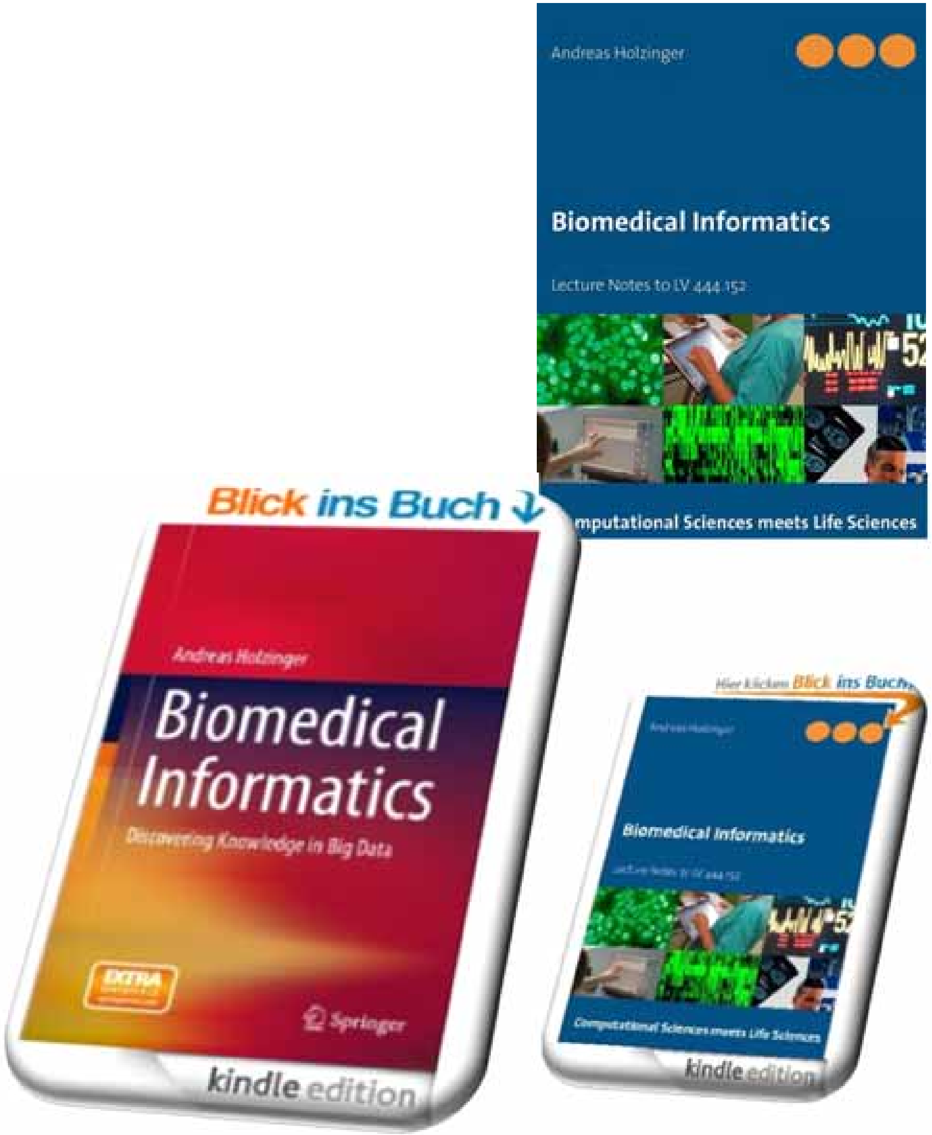
AL

Exp. & Eval.

Privacy ML



Andreas Holzinger
Biomedical Informatics
Discovering Knowledge in Big Data
EXTRA MATERIALS
springerlink.com
Springer



Andreas Holzinger
Biomedical Informatics
Lecture Notes to LV 444.152
mputational Sciences meets Life Sciences
Blick ins Buch
Andreas Holzinger
Biomedical Informatics
Discovering Knowledge in Big Data
Springer
kindle edition
Hier klicken Blick ins Buch
Biomedical Informatics
Computational Sciences meets Life Sciences
kindle edition

nature International weekly journal of science

Journal home > Archive > Books and Arts > Opinion > Full Text

Journal content
- Journal home
- Advance online publication
- Current issue
- Nature News
- Archive
- Supplements
- Web focuses
- Podcasts
- Videos
- News Specials

Journal information
- About the journal
- For authors
- Online submission
- Nature Awards
- Nature history

NPG services
- Advertising
- work@npg
- Reprints & permissions
- For librarians
- Authors & referees

NPG resources
- Gateways & databases
- Nature Reports

Books and Arts

Nature **464**, 680 (1 April 2010) | doi:10.1038/464680a; Published online 31 March 2010

A reality check for personalized medicine

Muin J. Khoury[1], James Evans[2] & Wylie Burke[3]

Bringing genetic information into health care is welcome but its utility in the clinic needs to be rigorously reviewed, caution Muin J. Khoury, James Evans and Wylie Burke.

BOOK REVIEWED
Personal Genomics and Personalized Medicine
by Hamid Bolouri
Imperial College Press: 2010. 280 pp. £34

T. FLACH/STONE/GETTY

Genomic information: should it be treated in the same way as X-ray results?

Why do this?

"…10 years from now, each cancer patient is going to want to get a genomic analysis of their cancer and will expect customized therapy based on that information."

Director, The Cancer Genome Atlas (TCGA), Time Magazine, 6/13/11

HCI-KDD

Thank you!