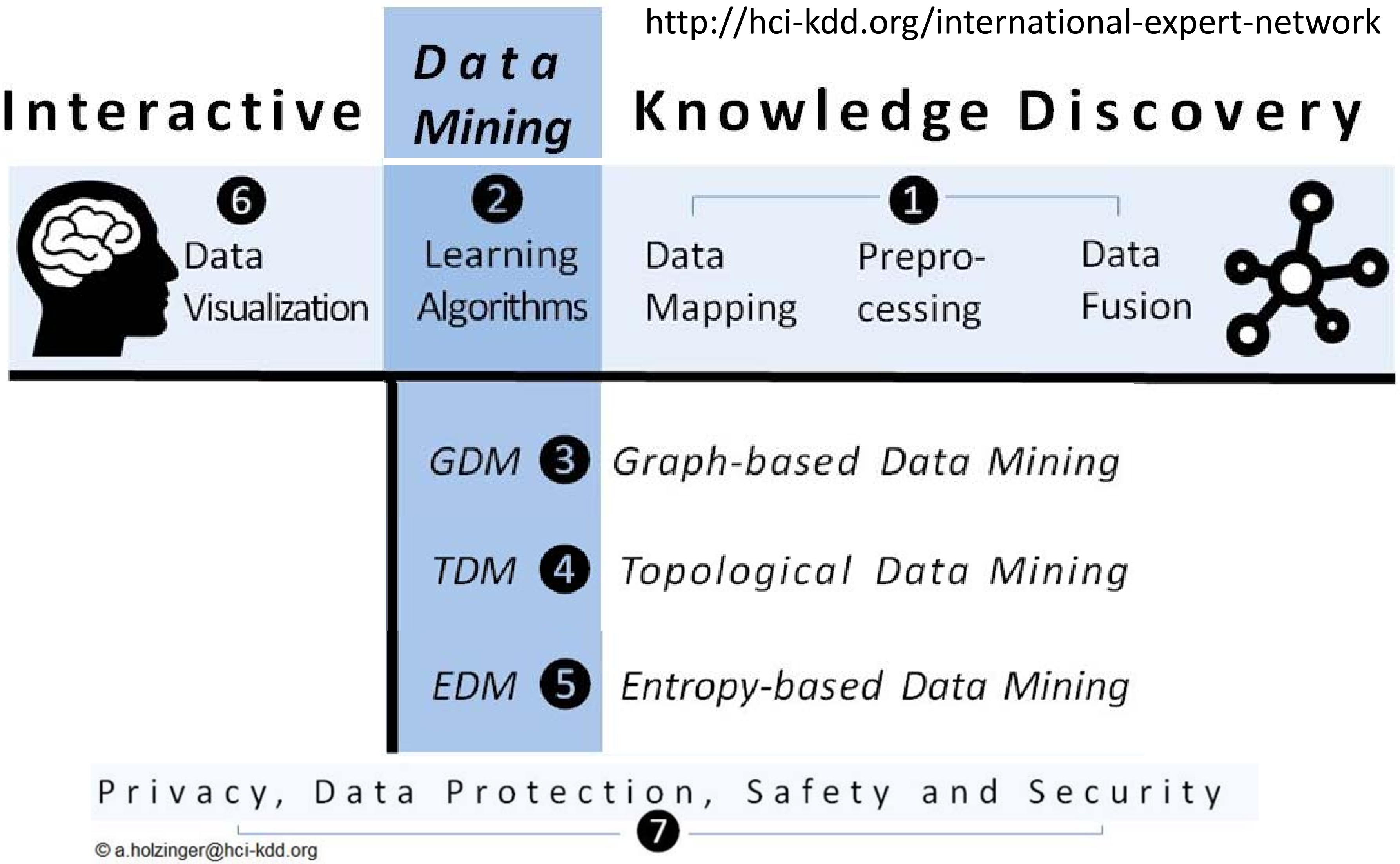
Workshop Machine Learning for Health Informatics at TU Vienna

Machine Learning (aML – iML) for Health Informatics

Andreas Holzinger

Holzinger Group, Research Unit HCI-KDD, Institute for Medical Informatics, Statistics and Documentation, Medical University Graz
&
Institute of Information Systems and Computer Media, Graz University of Technology

ML needs a concerted effort …



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

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

Discovery of causal relationships from data …

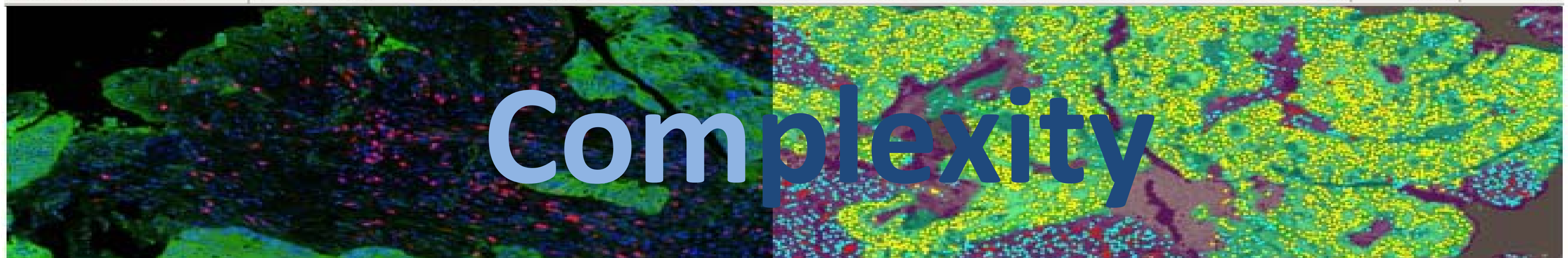
Hans Holbein d.J., 1533,
The Ambassadors,
London: National Gallery

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.

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

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.

Probabilistic Information p(x)

Thomas Bayes
1701 - 1761

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

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

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

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

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

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

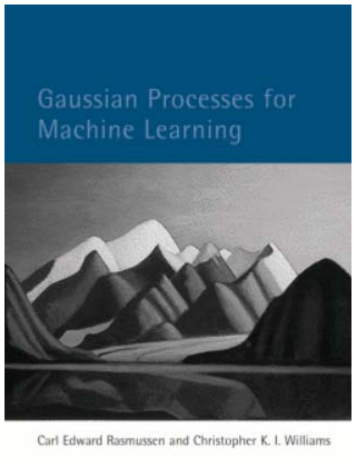
$$posterior\ p(x) = \frac{likelyhood * prior\ p(x)}{evidence}$$

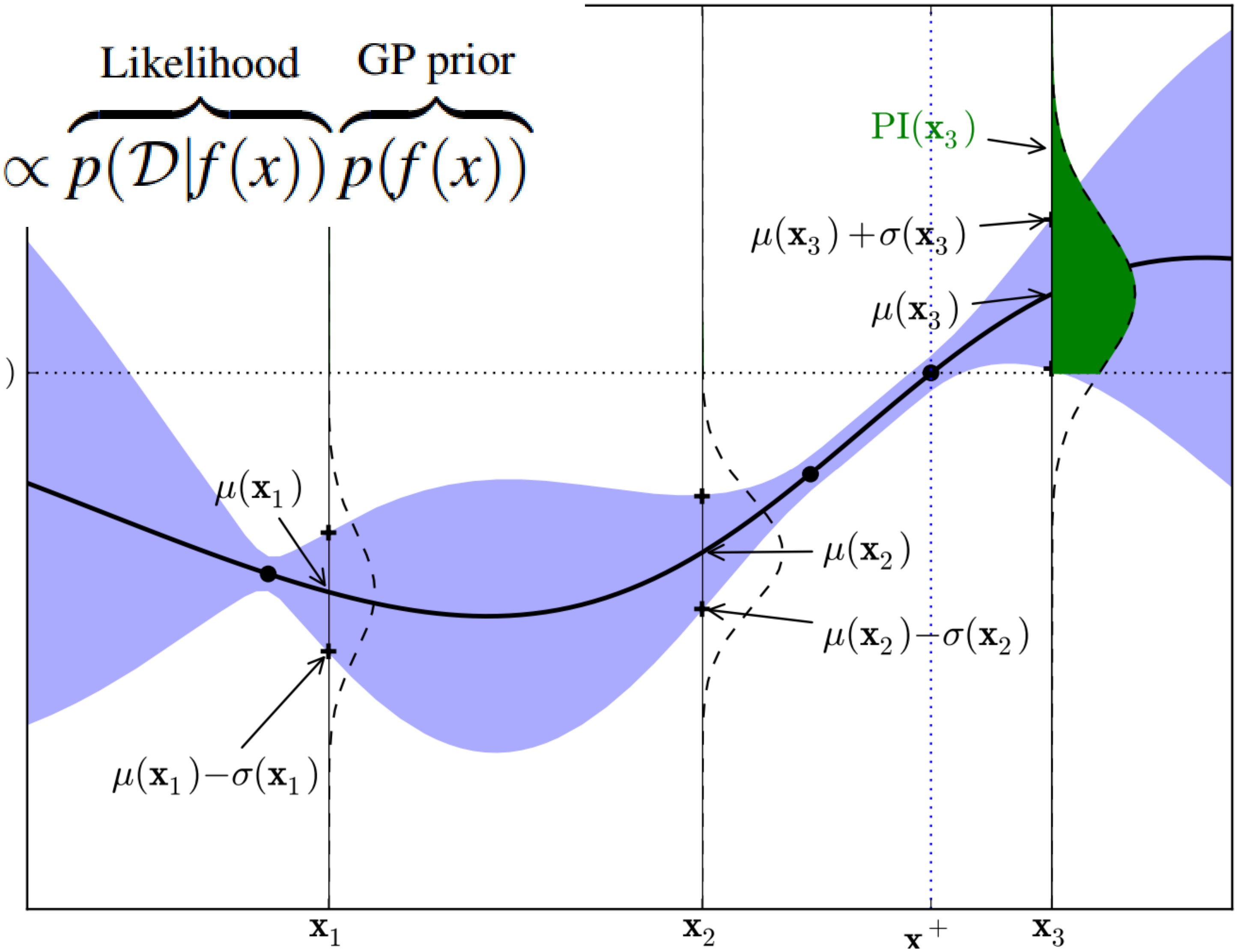
The inverse probability allows to infer unknowns, <u>learn from data</u> and make predictions …

… machine learning!

Simple 1-D Gaussian Process with three observations …

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

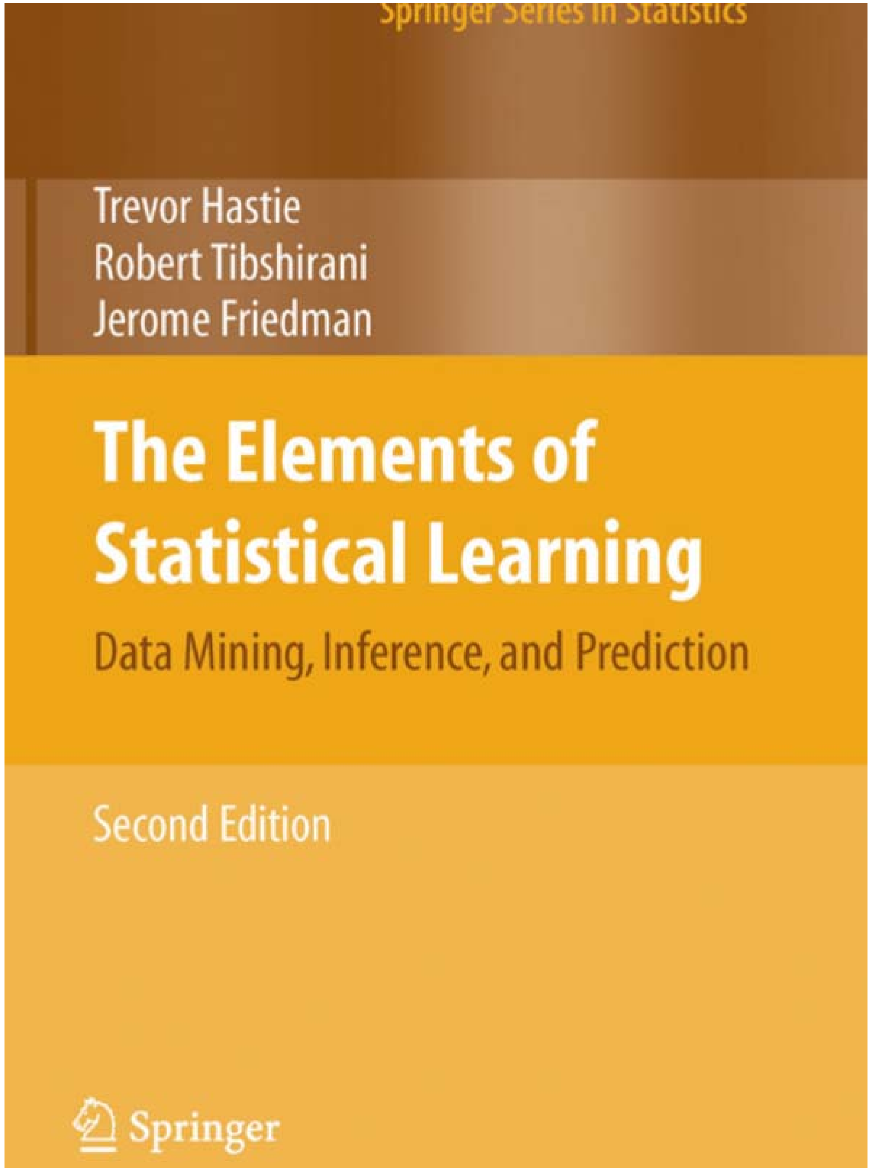




Brochu, E., Cora, V. M. & De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599.

- Machine Learning is the development of algorithms which can **learn from data**
- Pre-history in **statistical learning,**
- and the assessment of **uncertainty**

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



How do we build machines that learn from data?

How do we understand intelligence?

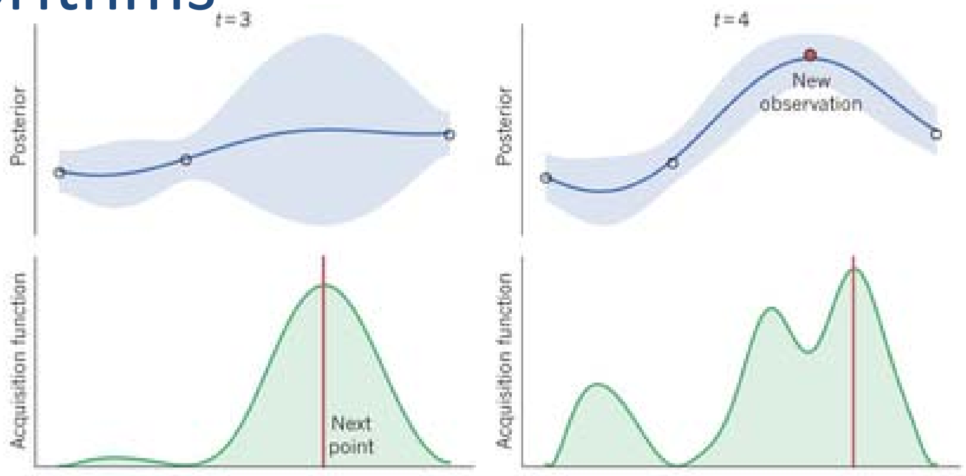
How can we extract knowledge from data?

Or as Demis Hassabis [1] from Google Deepmind points it out:

“Solve Intelligence!
… and then solve everything else … “

[1] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. & Hassabis, D. 2015. Human-level control through deep reinforcement learning. Nature, 518, (7540), 529-533.

- Many aspects of intelligence and learning depend on **probabilistic representation of uncertainty:**
- Forecasting
- Knowledge discovery
- Probabilistic programming
  e.g. Stochastic Python, Julia
- Universal inference algorithms
- Global optimization



Ghahramani, Z. 2015. Probabilistic machine learning and artificial intelligence. Nature, 521, (7553), 452-459.

- 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.

- 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 …**

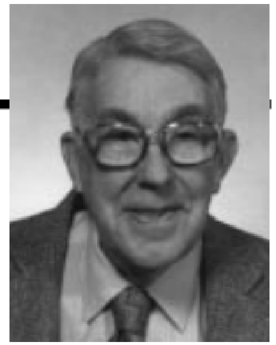
Introduction

The studies reported here have been cc
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 ar
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.

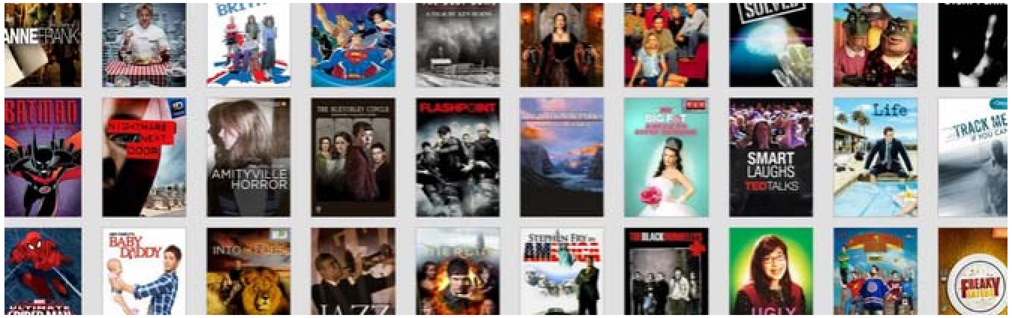
Samuel's papers on machine learn-

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.

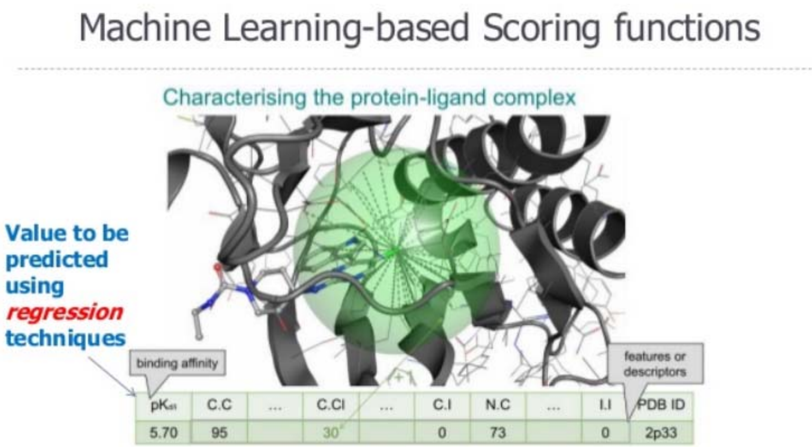
In 1949, Samuel joined IBM's Poughkeepsie Laboratory, where he worked on IBM's first stored program

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

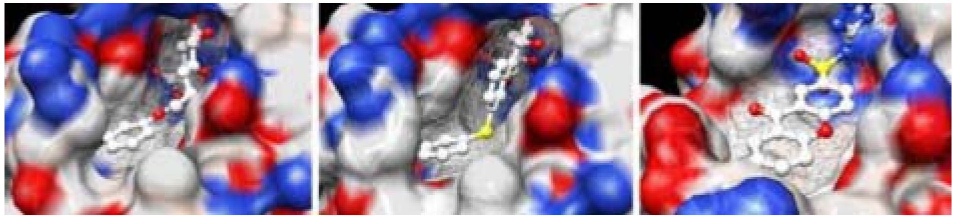
2015 everything is machine learning …



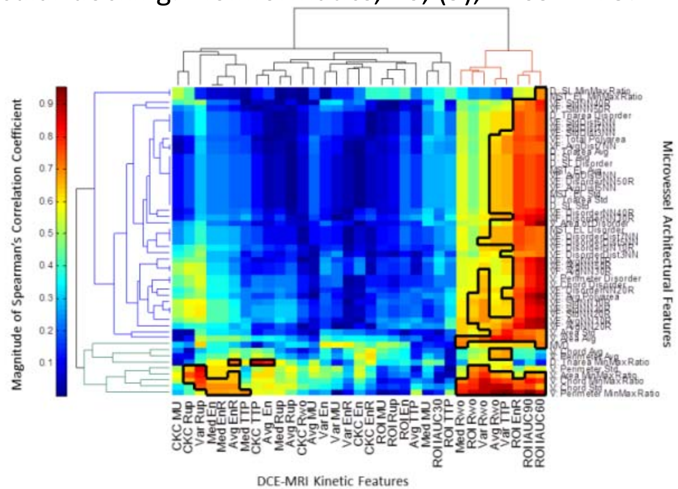
Takacs, G., Pilaszy, I., Nemeth, B., Tikk, D. & Acm 2008. Matrix Factorization and Neighbor Based Algorithms for the Netflix Prize Problem. Recsys'08: Proceedings of the 2008 ACM Conference on Recommender Systems, 267-274.



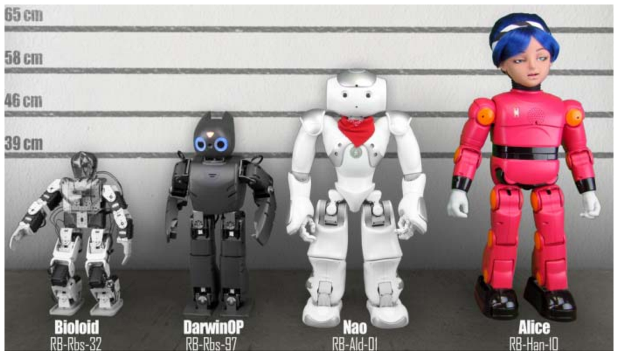
Ballester, P. J. & Mitchell, J. B. O. 2010. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics, 26, (9), 1169-1175.



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.

- most of it is automatic Machine Learning (aML)

- 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?

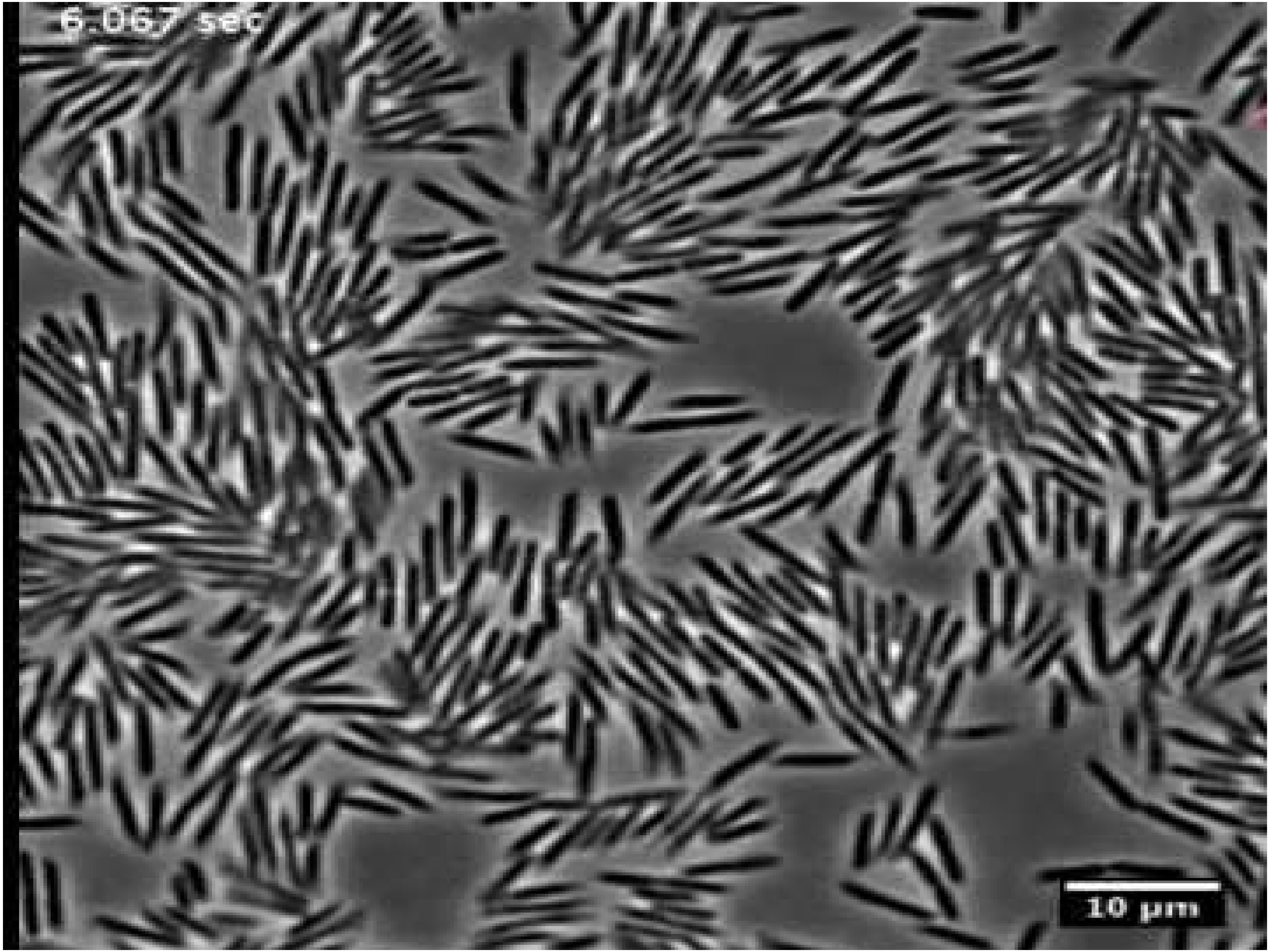
- Sometimes we **do not have “big data”,** where aML-algorithms benefit.
- Sometimes we have
  - **Small amount of data sets**
  - Rare Events – **no training samples**
  - **NP-hard problems, e.g.**
    - Subspace Clustering,
    - Protein-Folding,
    - k-Anonymization,
    - Graph Coloring, Category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

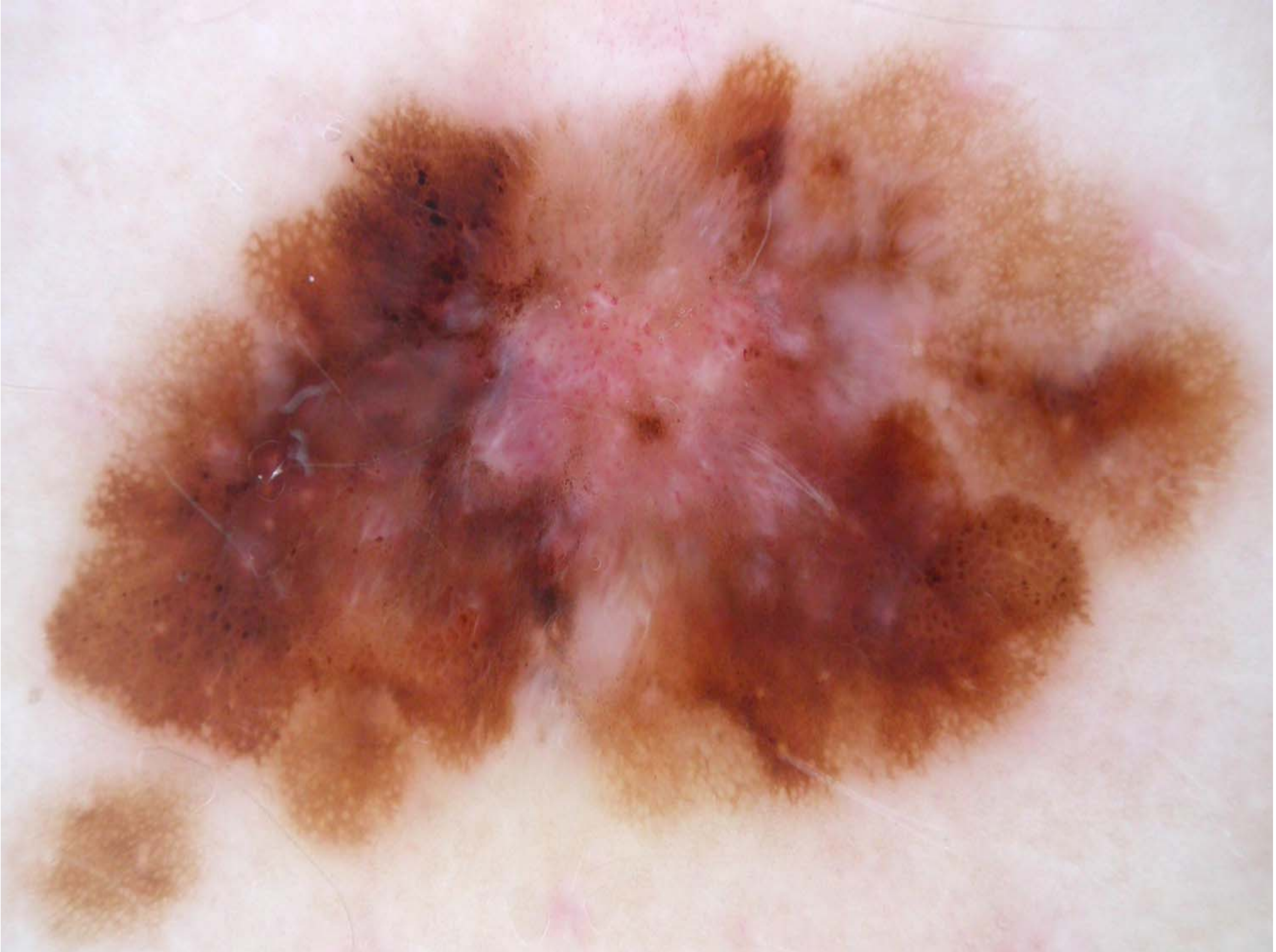
Sometimes we (still) need a human-in-the-loop


2
3
4
5
6

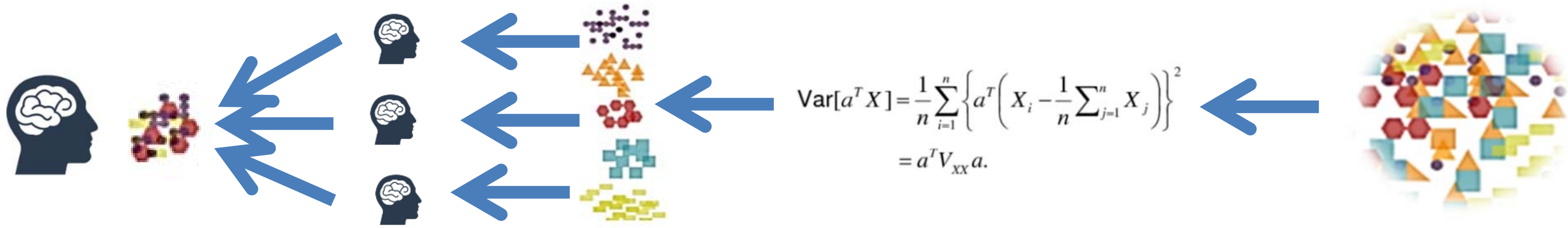
- interactive Machine Learning (iML) := algorithms which interact with agents*) and can optimize their learning behaviour trough this interaction
- *)**where the agents can be human**

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

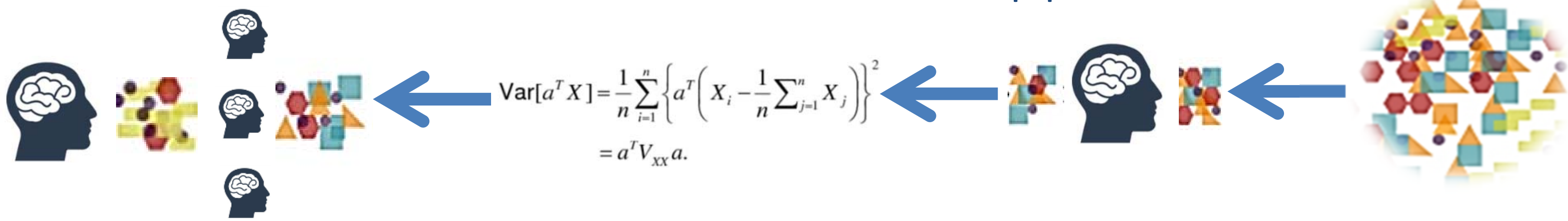

6.067 sec
10 μm



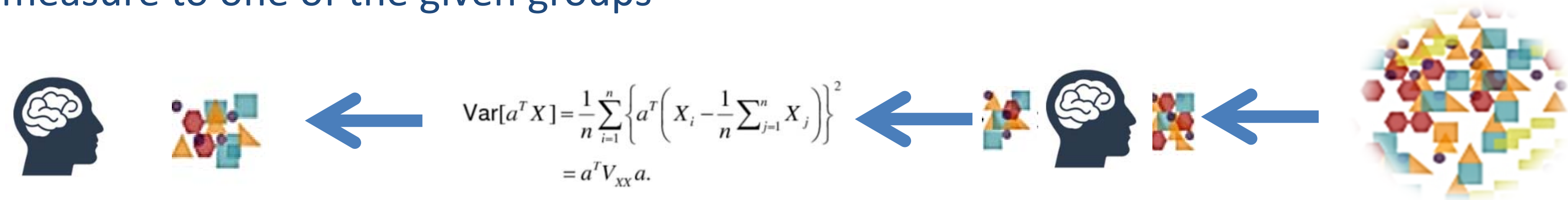
A) Unsupervised ML: Algorithm is applied on the raw data and learns fully automatic – Human can check results at the end of the ML-pipeline



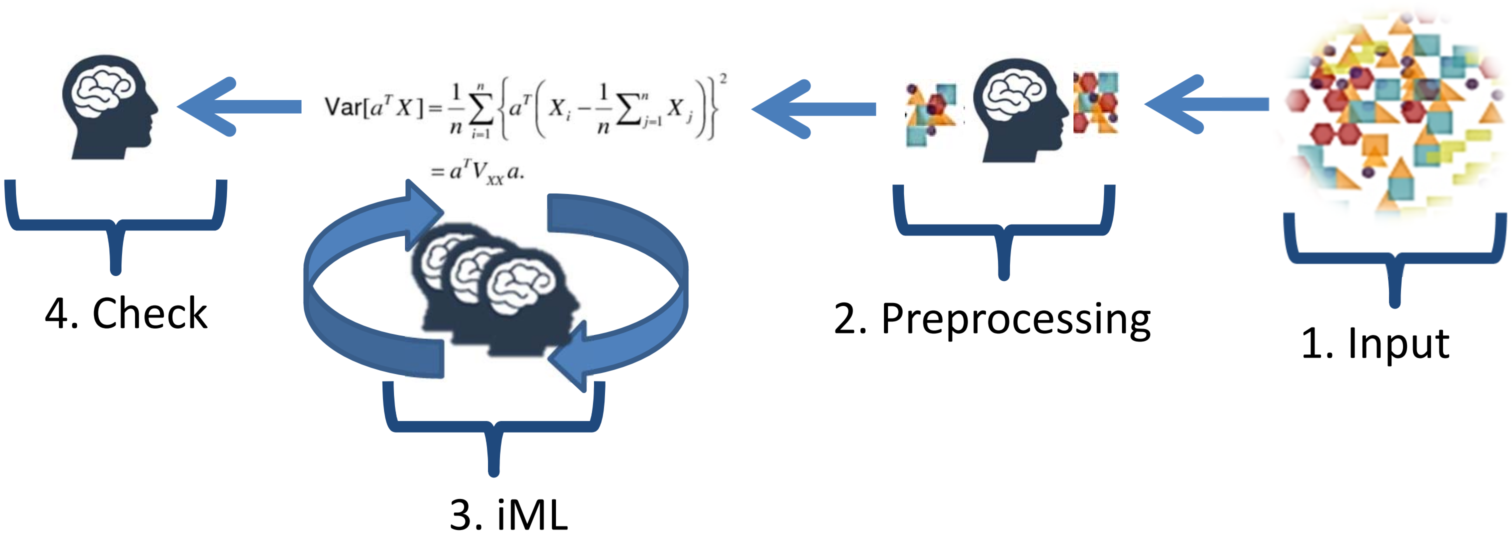
B) Supervised ML: Humans are providing the labels for the training data and/or select features to feed the algorithm to learn – the more samples the better – Human can check results at the end of the ML-pipeline



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, …

- 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

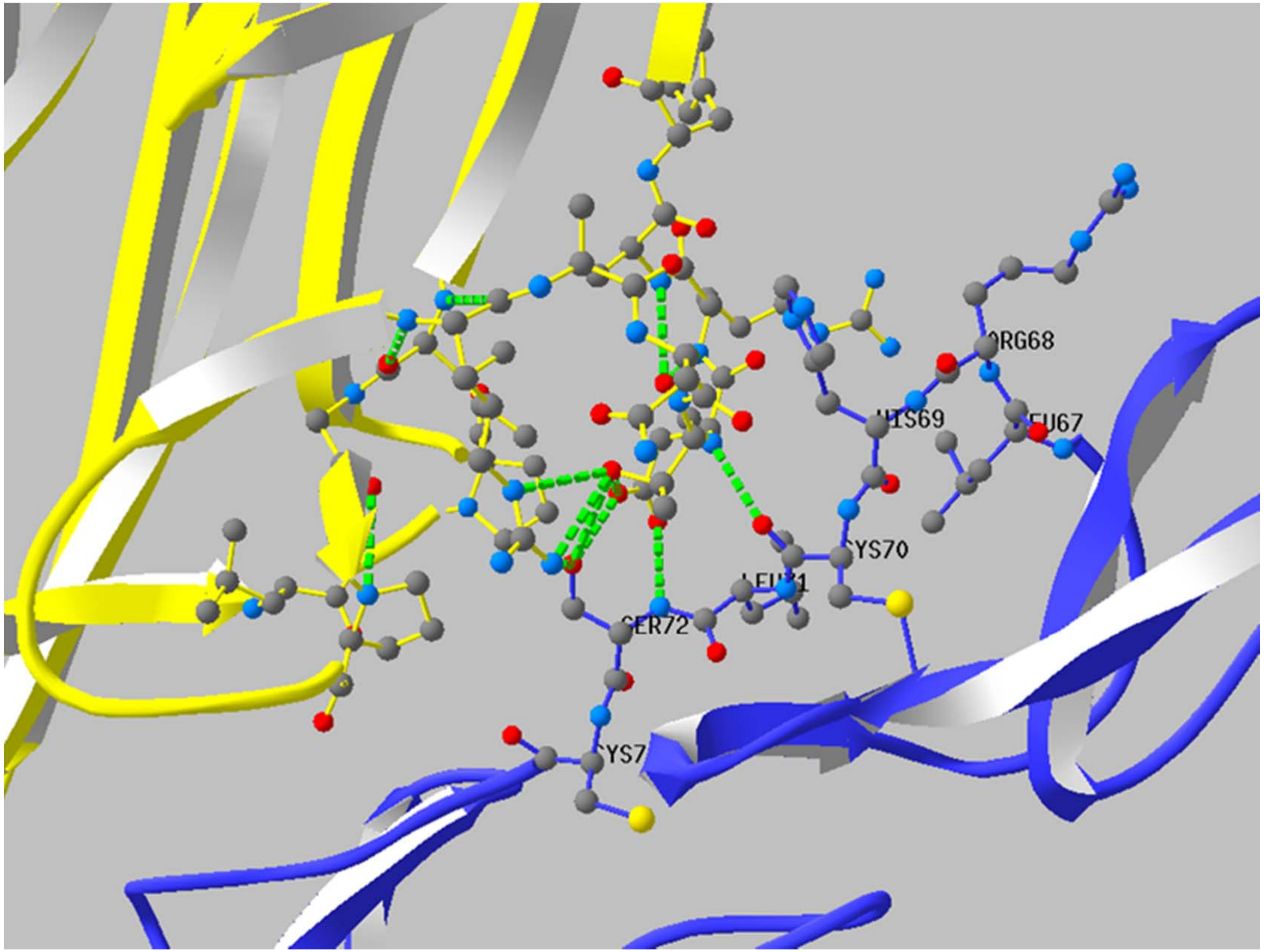
Birthdate	Sex	Zipcode	Disease
1/21/76	Male	53715	Flu
4/13/86	Female	53715	Hepatitis
2/28/76	Male	53703	Brochitis
1/21/76	Male	53703	Broken Arm
4/13/86	Female	53706	Sprained Ankle
2/28/76	Female	53706	Hang Nail

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

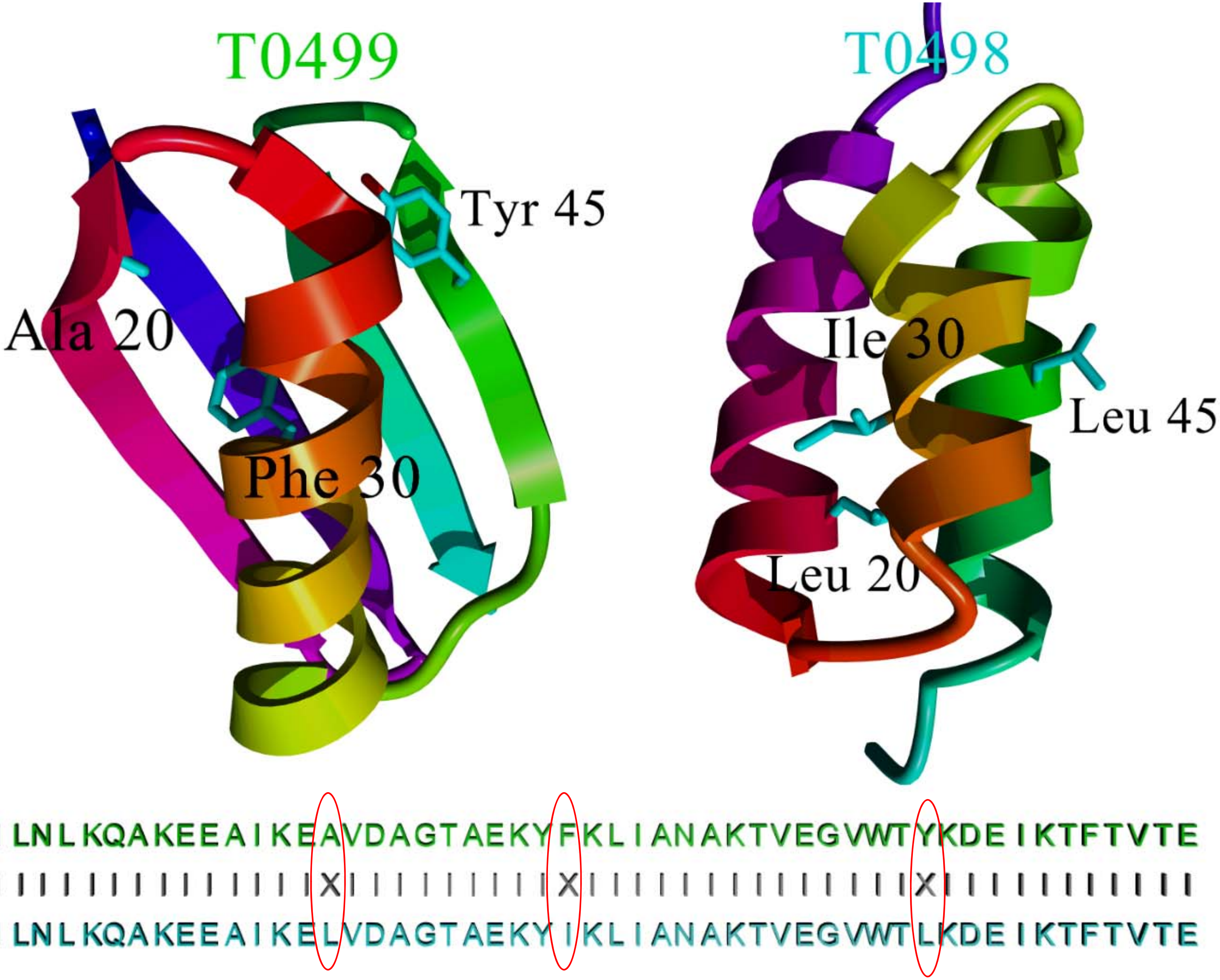
Sweeney, L. 2002. Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10, (05), 571-588.

Proteins are the building blocks of life …

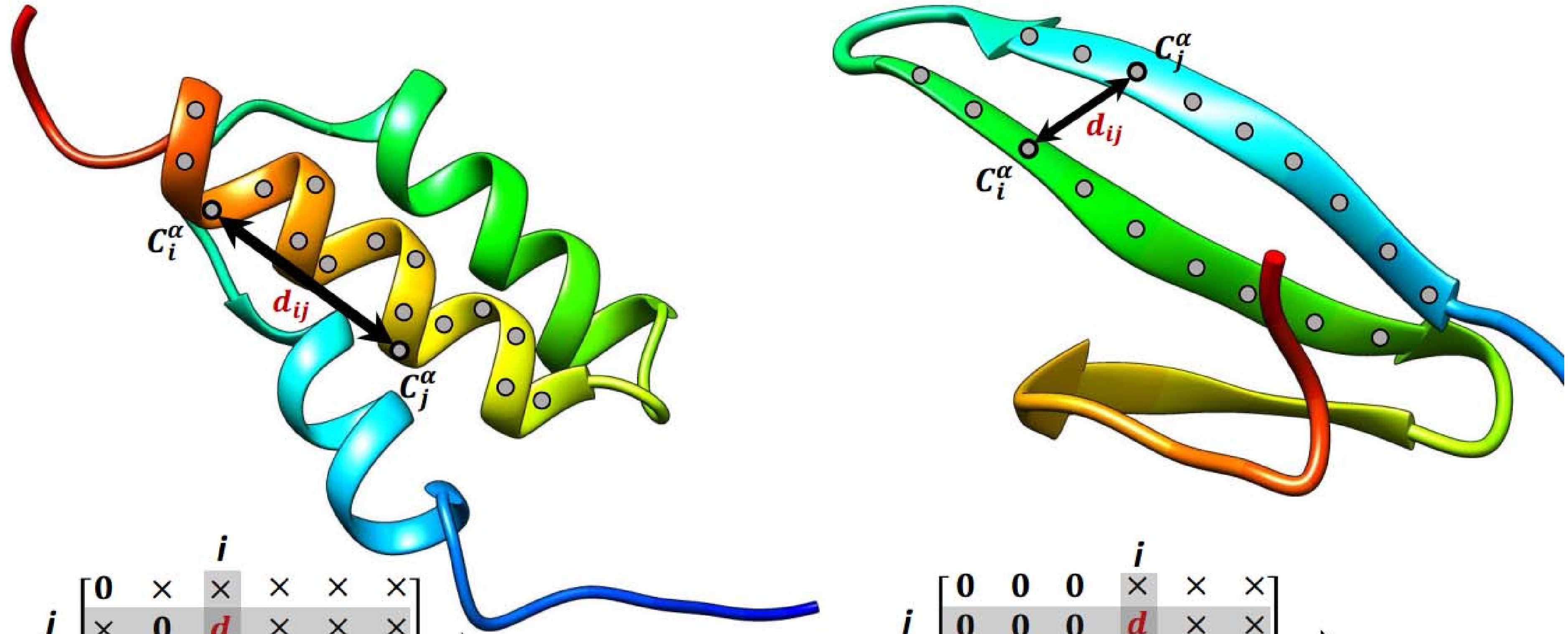


Wiltgen, M., Holzinger, A. & Tilz, G. P. (2007) Interactive Analysis and Visualization of Macromolecular Interfaces Between Proteins. In: *Lecture Notes in Computer Science (LNCS 4799). Berlin, Heidelberg, New York, Springer, 199-212.*

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.



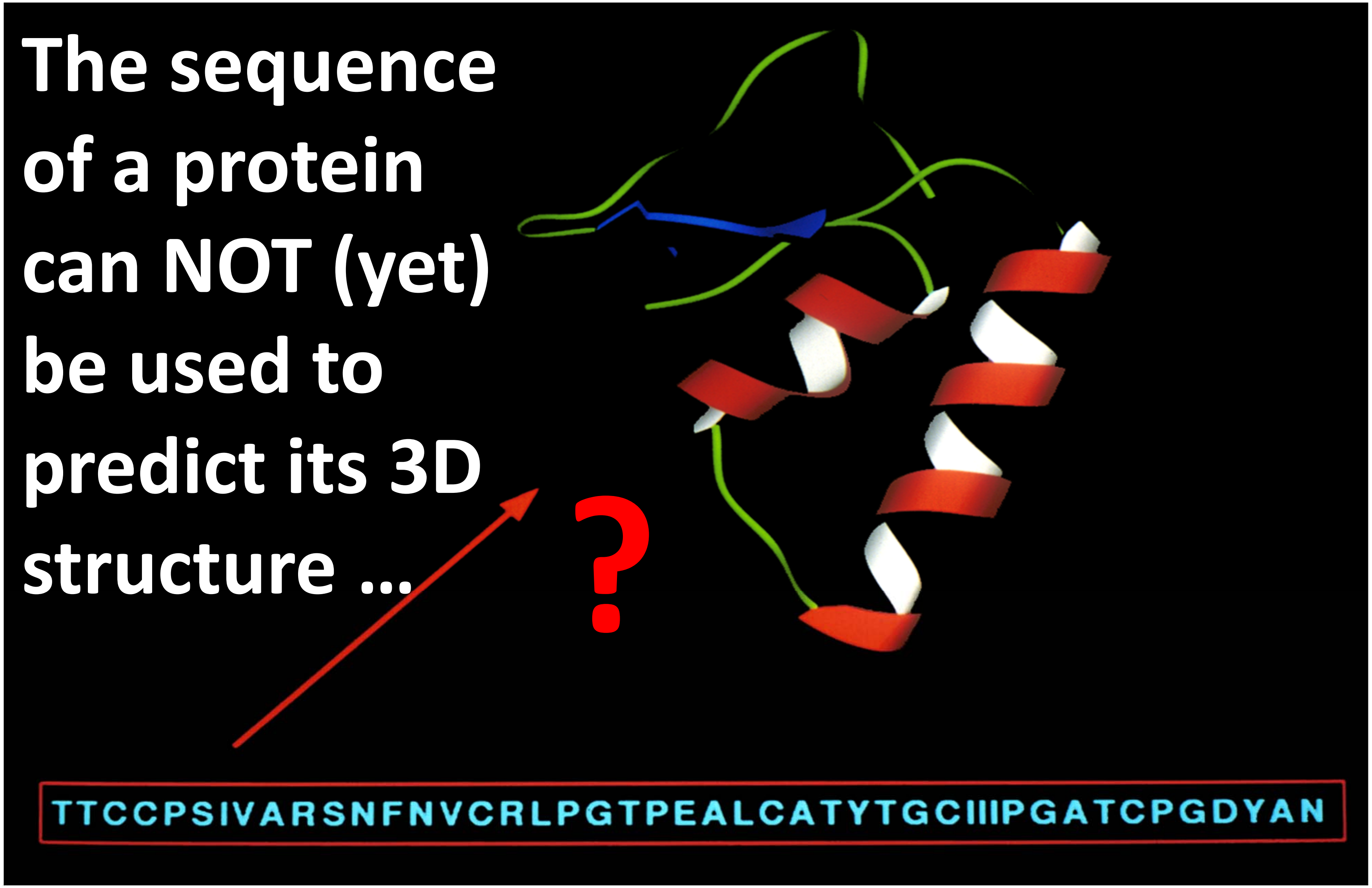
HCI-KDD



$$D = \begin{array}{cc} & \begin{array}{cccccc} & & i & & & \end{array} \\ \begin{array}{c} \\ j \\ \\ \\ \\ \\ \end{array} & \left[\begin{array}{cccccc} 0 & \times & \times & \times & \times & \times \\ \times & 0 & d & \times & \times & \times \\ \times & d & 0 & \times & \times & \times \\ \times & \times & \times & 0 & \times & \times \\ \times & \times & \times & \times & 0 & \times \\ \times & \times & \times & \times & \times & 0 \end{array}\right] \end{array} \Rightarrow \lambda_1$$

$$D = \begin{array}{cc} & \begin{array}{cccccc} & & & i & & \end{array} \\ \begin{array}{c} \\ j \\ \\ \\ \\ \\ \end{array} & \left[\begin{array}{cccccc} 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & d & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ \times & d & \times & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 \end{array}\right] \end{array} \Rightarrow \lambda_1$$

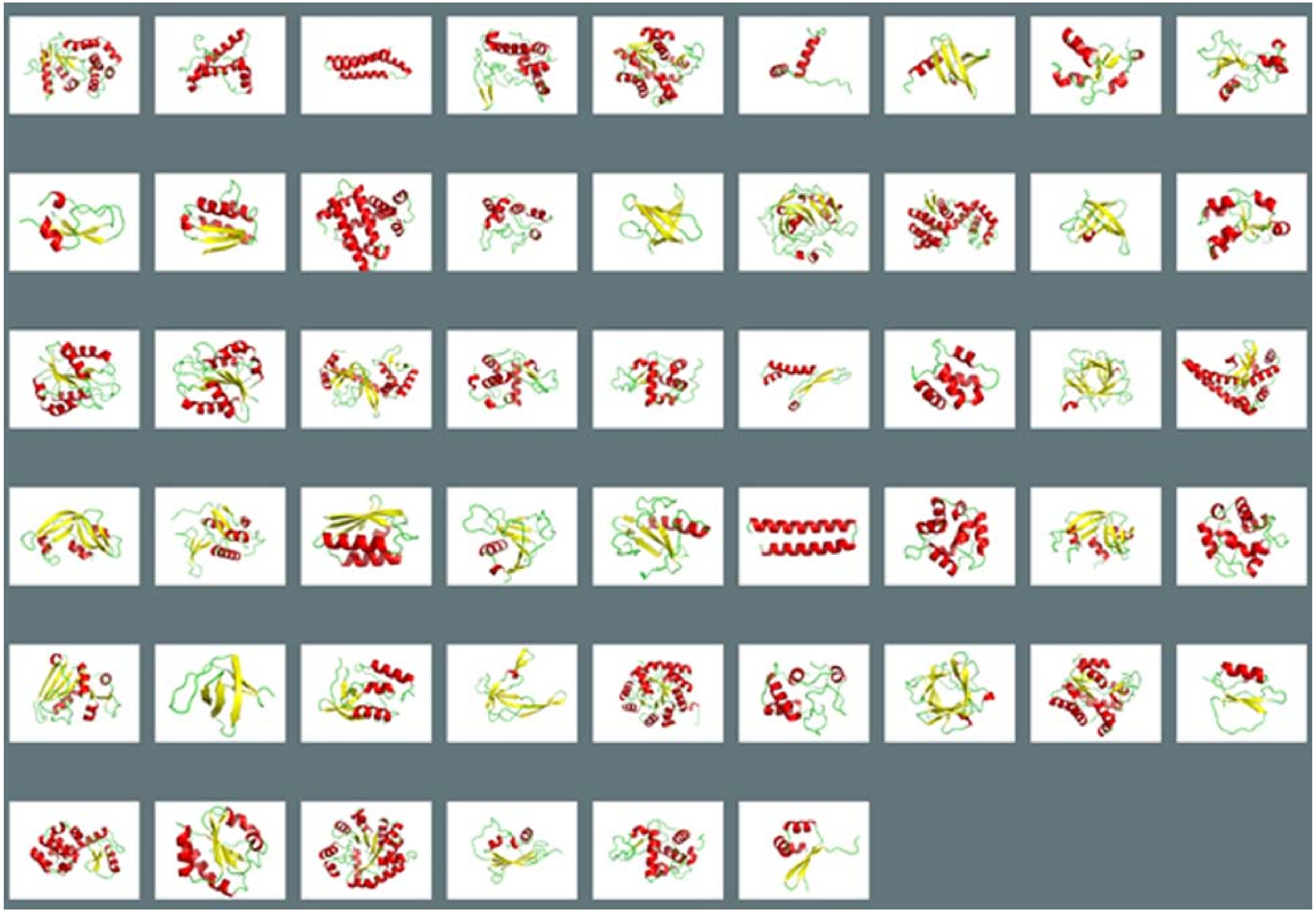
An old problem … (1973)



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

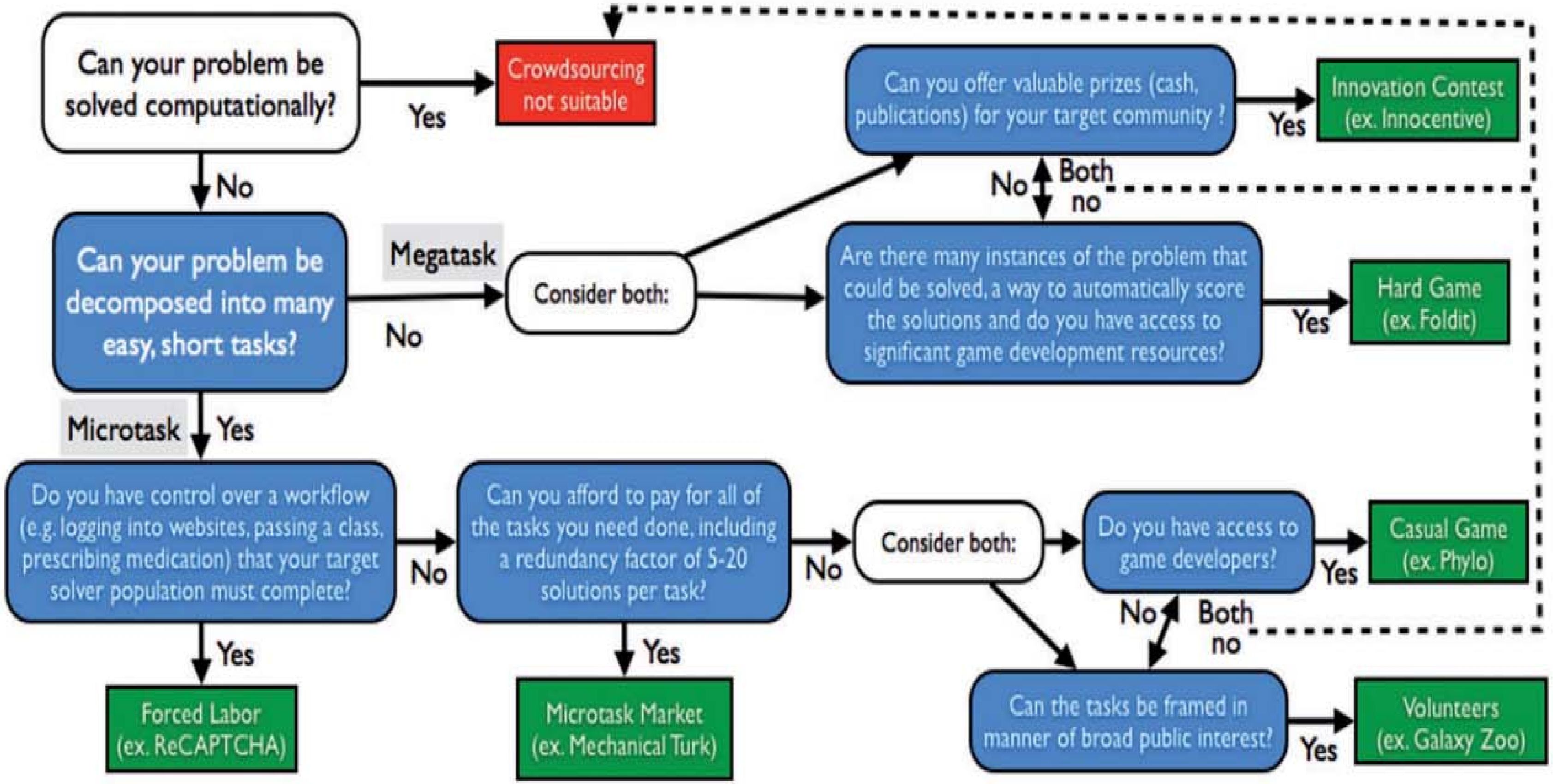
State-of-2015: aML still is not able to help sufficiently …

HCI-KDD



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).

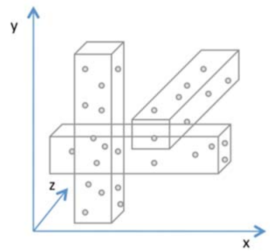
Humans can help here: Crowdsourcing

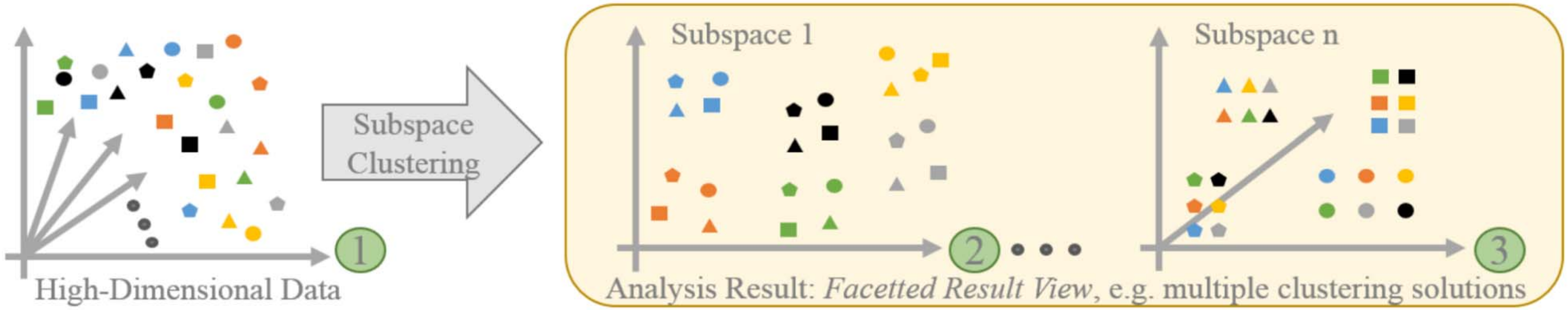


Good, B. M. & Su, A. I. 2013. Crowdsourcing for bioinformatics. *Bioinformatics,* 29, (16), 1925-1933.

Example 3: Subspace Clustering

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

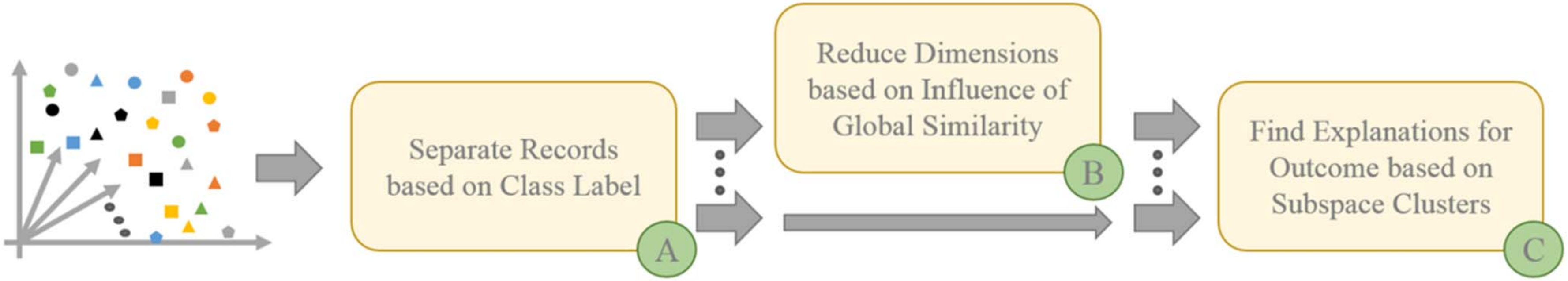




Hund, M., Sturm, W., Schreck, T., Ullrich, T., Keim, D., Majnaric, L. & Holzinger, A. 2015. Analysis of Patient Groups and Immunization Results Based on Subspace Clustering. In: Lecture Notes in Artificial Intelligence LNAI 9250, pp. 358-368.

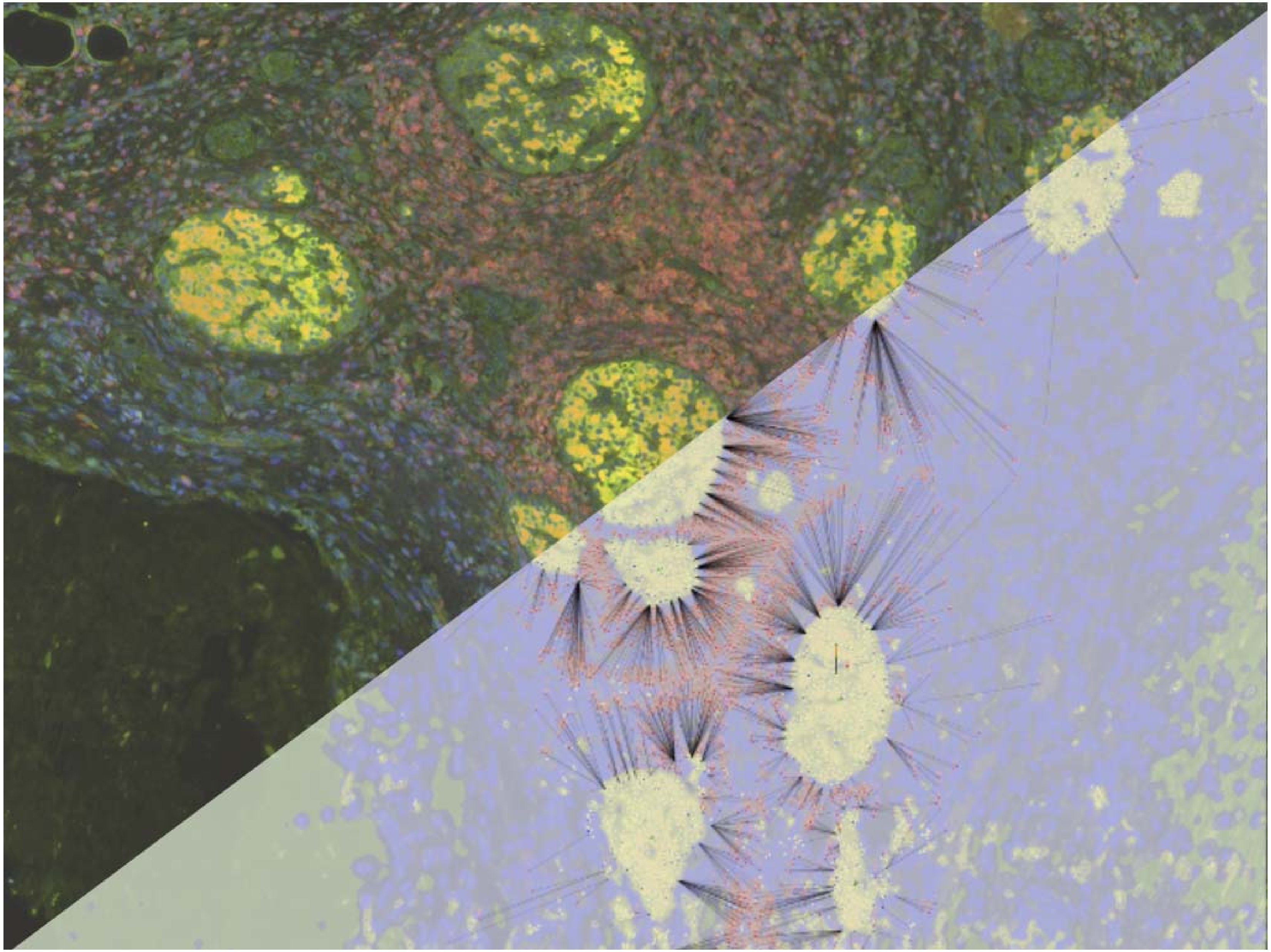
Nr.	Name	Type	missing	Description
1	age	numerical	0	age (years)
2	sex	binary	0	(M=Male, F=Female)
3	Hyper	binary	0	Hypertension (yes, no)
4	DM	nominal	0	Diabetes mellitus (yes, IGT=Impaired glucose tolerance, no)
5	F Glu	numerical	0	Fasting blood glucose (mmol/L)
6	HbA1c	numerical	0	Glycosilated Haemoglobin (%) (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 did 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)

Future Goal: Category Discovery

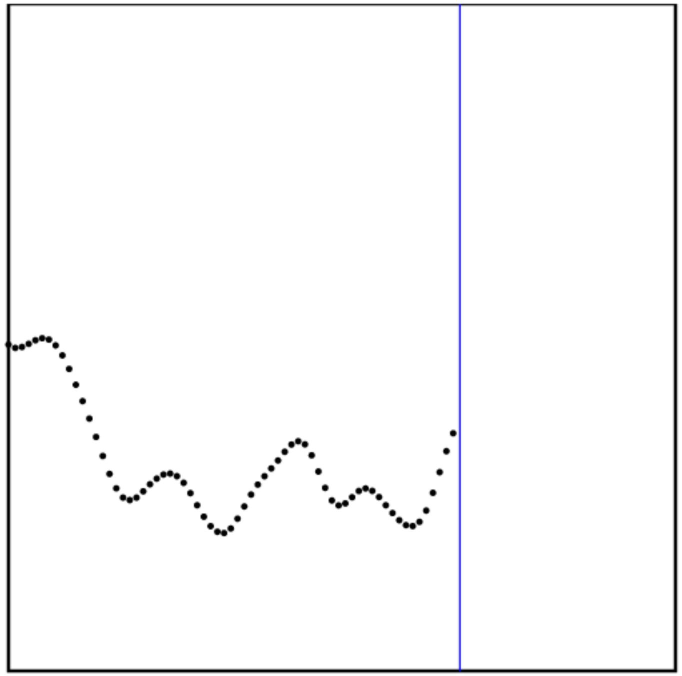


Judgment 1 out of 33

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.

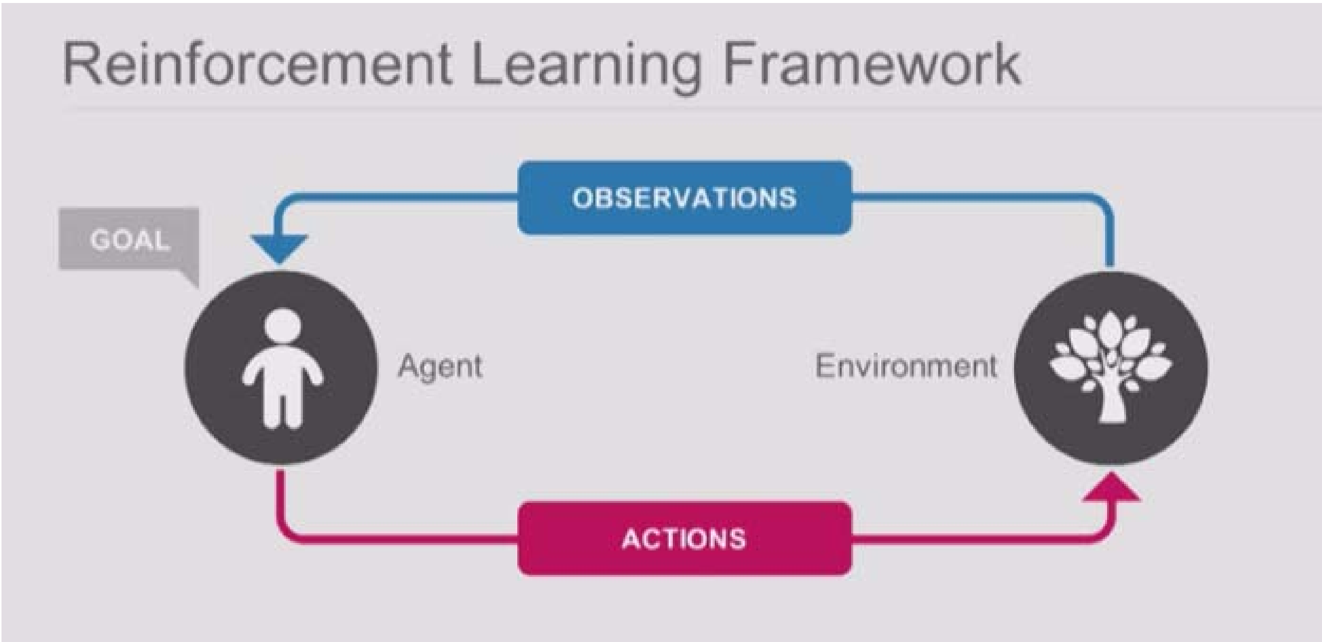
Please click along the blue line to say what you think the height of the point is for that location.

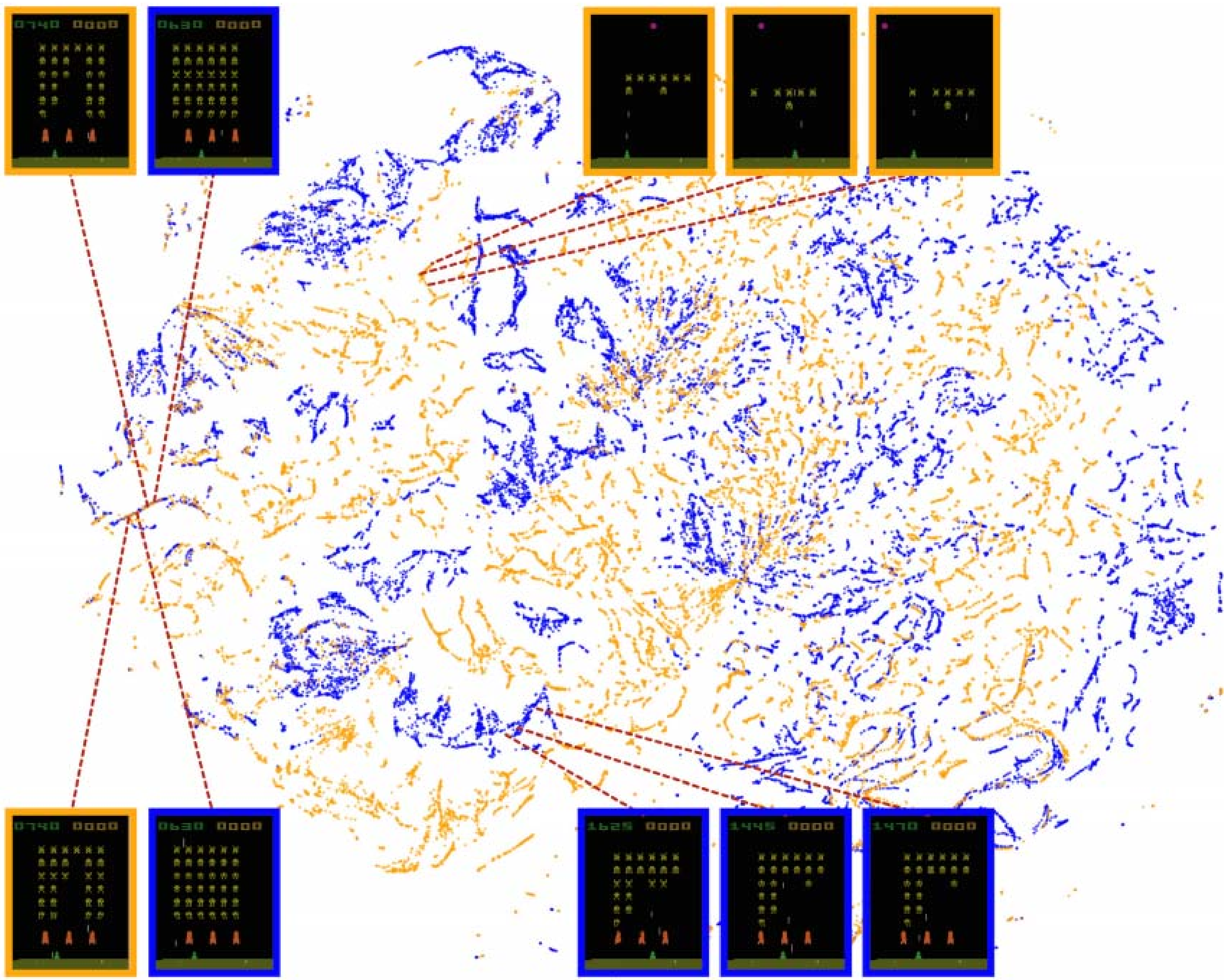
Once you have selected a position along the line, **hit the 's' key to submit the point**.



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

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



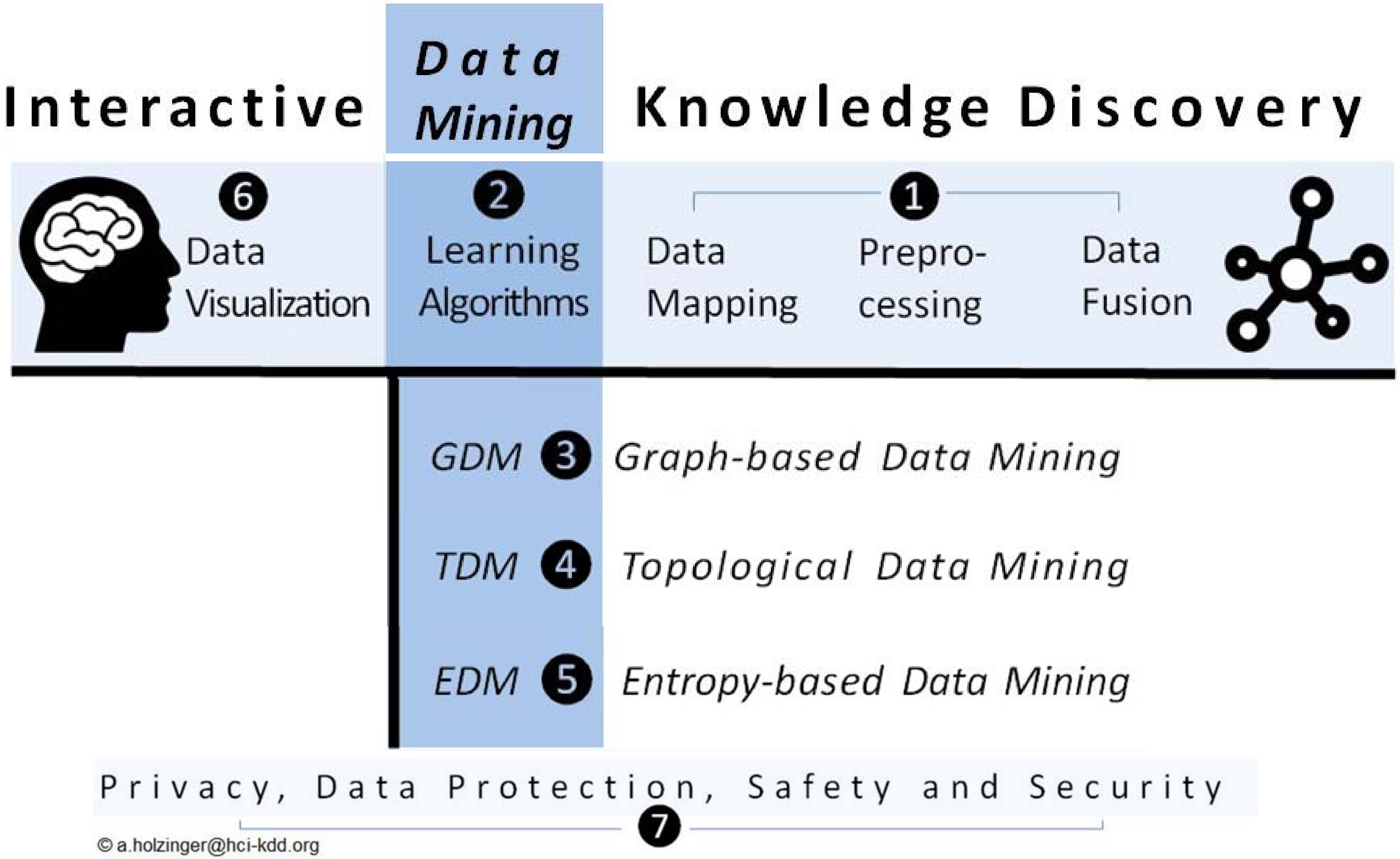


Mnih, V. et al. … Hassabis, D. 2015. Human-level control through deep reinforcement learning. Nature, 518, (7540), 529-533.

- 1 Heterogeneous data sources
  - need for data integration and data fusion
- 2 Complexity – reduction of search space
  - combining the best of Human & Computer
- 3 What is interesting? – and relevant!
  - need of effective mapping $\mathbb{R}^N \rightarrow \mathbb{R}^2$
- 4 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



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

concerted effort

international

without boundaries …


HCI-KDD
You

HCI-KDD

Thank you!