Andreas Holzinger

706.315 Selected Topics on Knowledge Discovery: Interactive Machine Learning

2015W, SE, 2.0 h, 3.0 ECTS

Week 42 - 16.10.2015 10:00-11:30

Introduction to Machine Learning (ML): automatic ML - interactive ML

a.holzinger@hci-kdd.org

http://hci-kdd.org/lv-706-315-interactive-machine-learning

Agenda

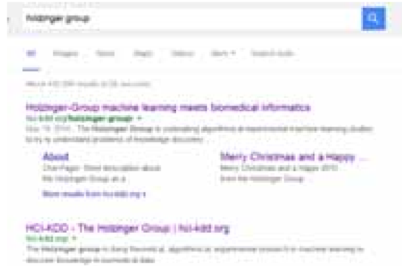
- 01 The HCI-KDD approach
- 02 What is Machine Learning ?
- 03 Application Area: Health Informatics
- 04 Probabilistic Information
- 05 Gaussian Processes
- 06 Automatic Machine Learning (aML)
- 07 Interactive Machine Learning (iML)
- 08 Conclusion and Future Outlook
- 09 Questions
- 10 Appendix

Quiz: Which of the following tasks include ML

1

2

3

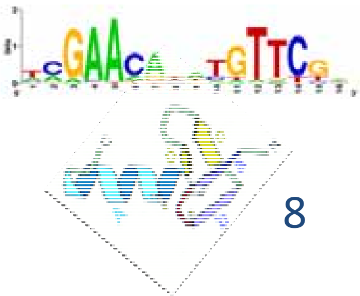


4

5

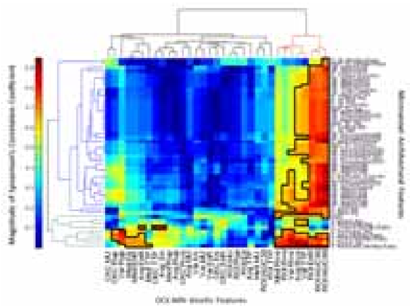
6

7

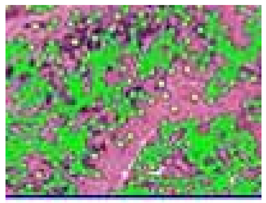


8

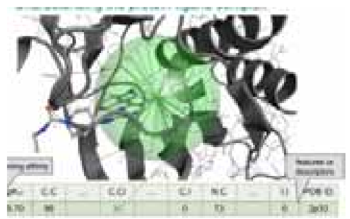
9



10



11

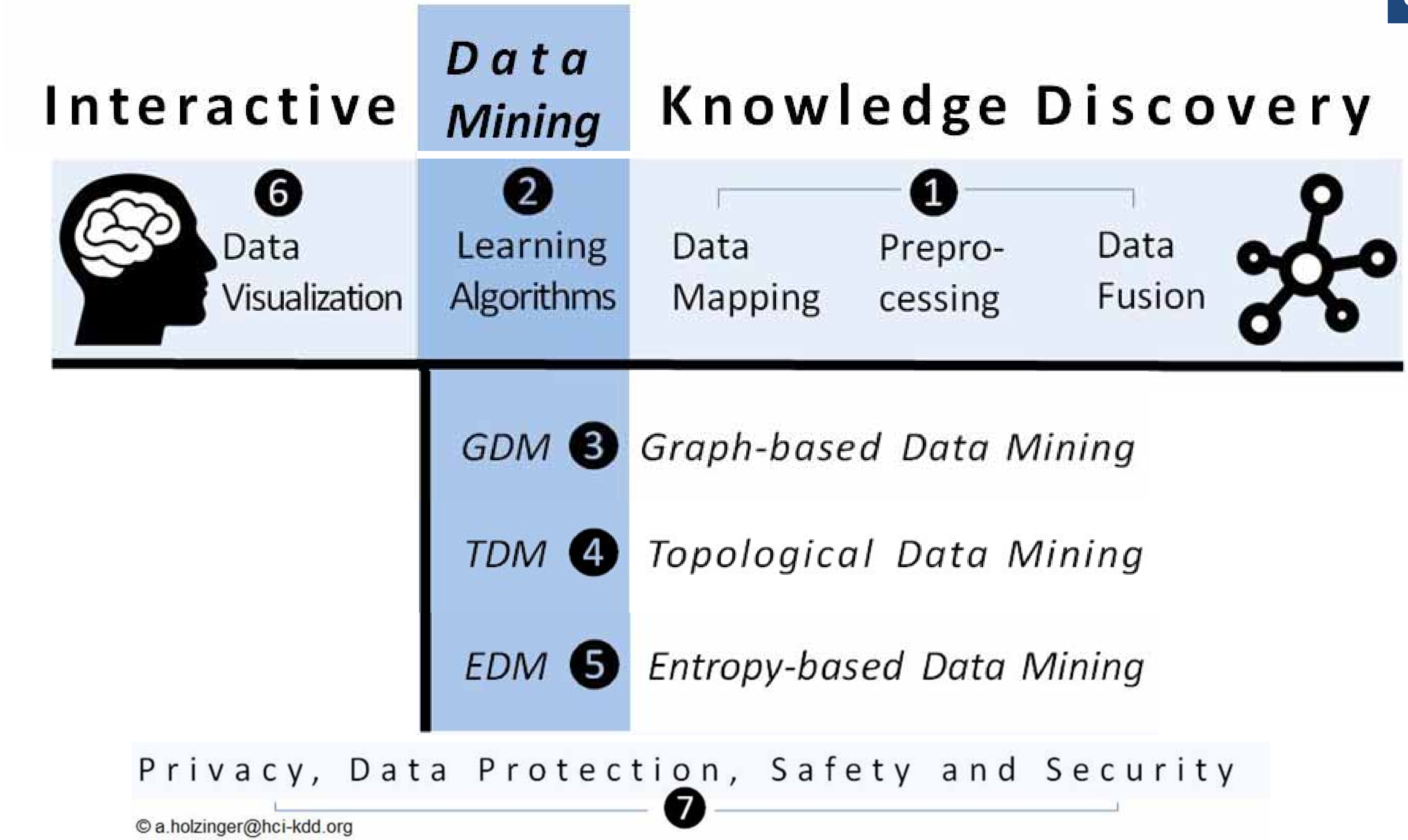


12

01 HCI-KDD

In the Knowledge Discovery pipeline ML is the heart ...

01



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

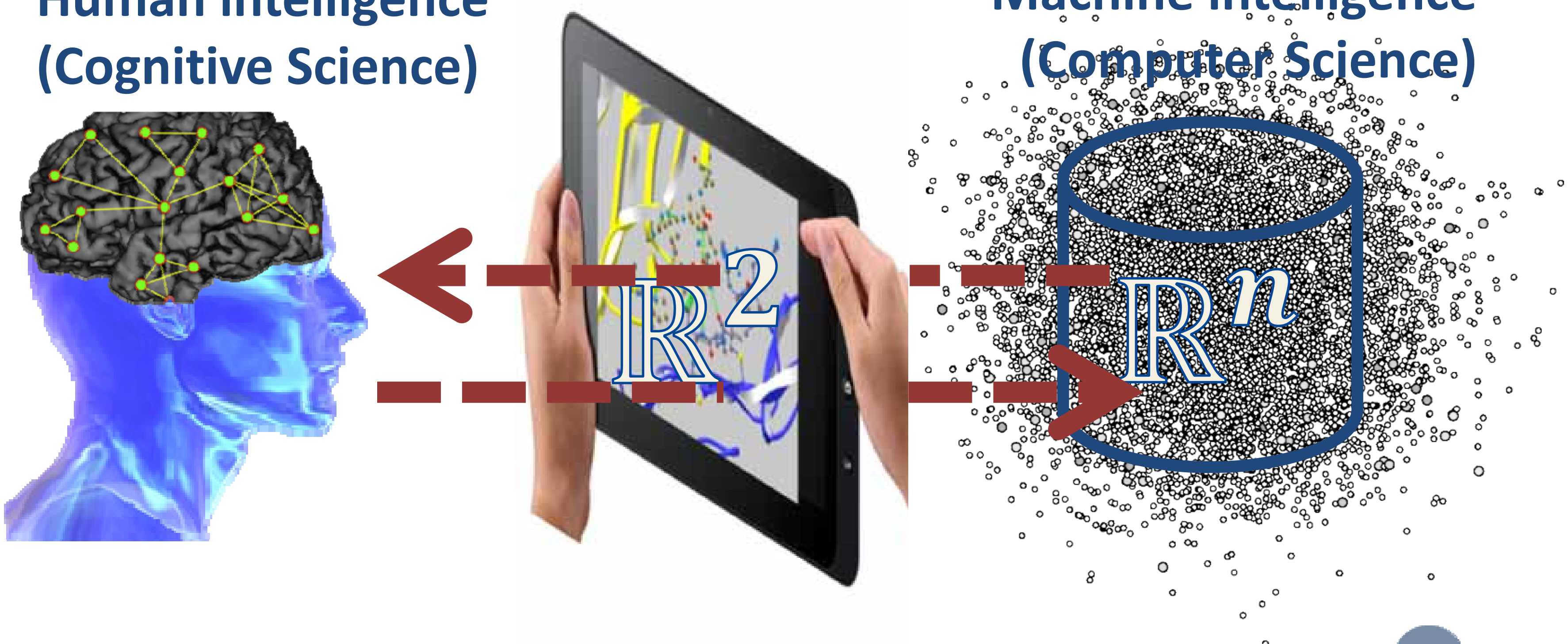
without boundaries …


HCI-KDD
You
http://hci-kdd.org/international-expert-network

Grand challenge: Transfer results from $\mathbb{R}^n$ to $\mathbb{R}^2$

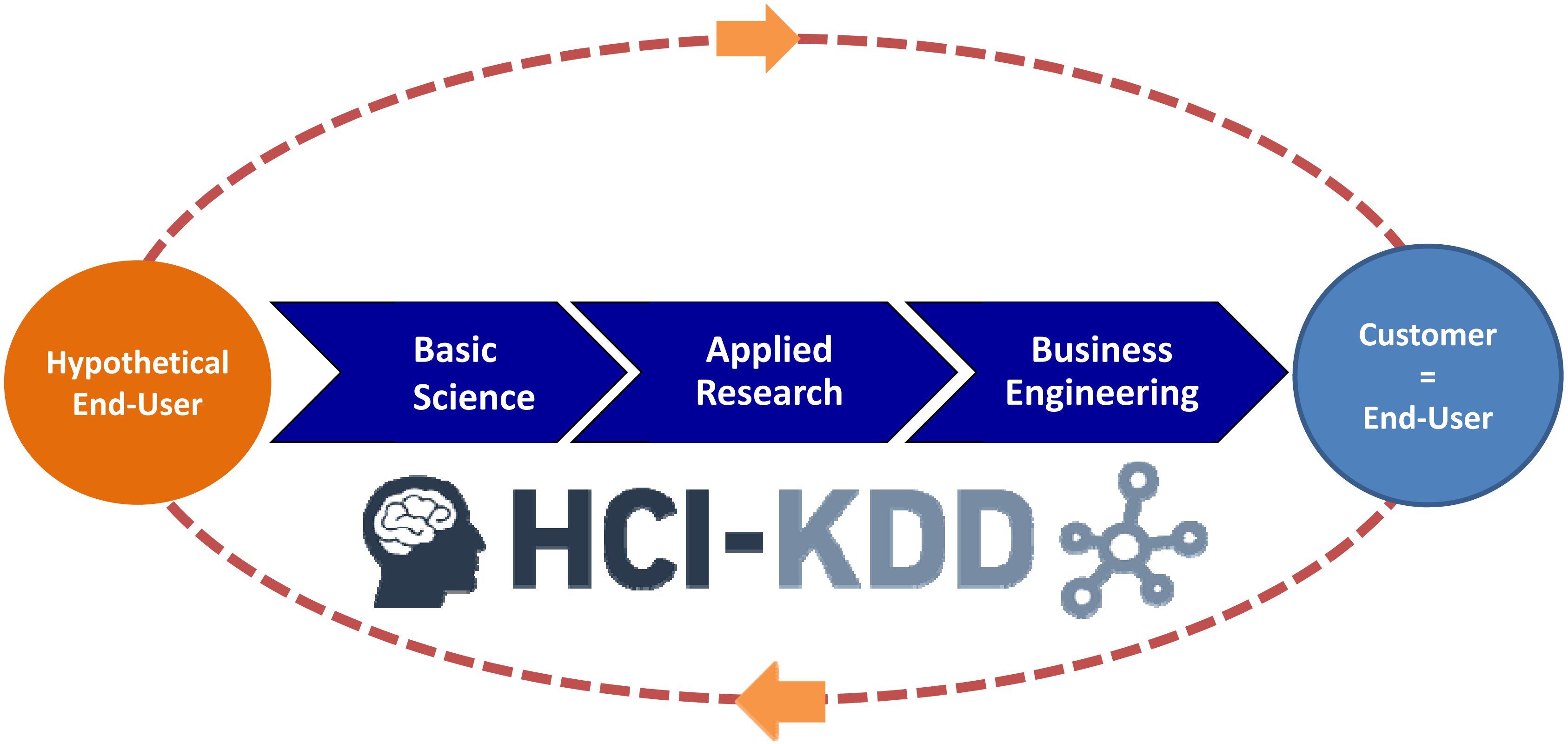
Human intelligence (Cognitive Science)

Machine intelligence (Computer Science)



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)

Science is to test crazy ideas – Engineering is to put these ideas into Business



Holzinger, A. 2011. Successful Management of Research and Development, Norderstedt: BoD.

"solve intelligence – then solve everything else"

Demis Hassabis, 22 May 2015

The Royal Society,
Future Directions of Machine Learning Part 2

https://youtu.be/XAbLn66iHcQ?t=1h28m54s

- **Cognitive Science $\rightarrow$ human intelligence**
  - Study the principles of ***human learning*** to understand biological intelligence
- **Human-Computer Interaction $\rightarrow$ the bridge**
  - Interacting with algorithms that learn shall enhance user friendliness and let concentrate on problem solving - Opening the “black-box” to a “glass-box”
- **Computer Science $\rightarrow$ computational intelligence**
  - Study the principles of ***machine learning*** to understand artificial intelligence

- 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 …**

Is the human brain a inference engine ?

- Learning concepts from examples (babies!)
- Causal inference and reasoning
- Predicting everyday events
- Even little children solve complex problems unconsciously, effortlessly, and … successfully
- Compare your best Machine Learning algorithm with a seven year old child!

Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.1192788.

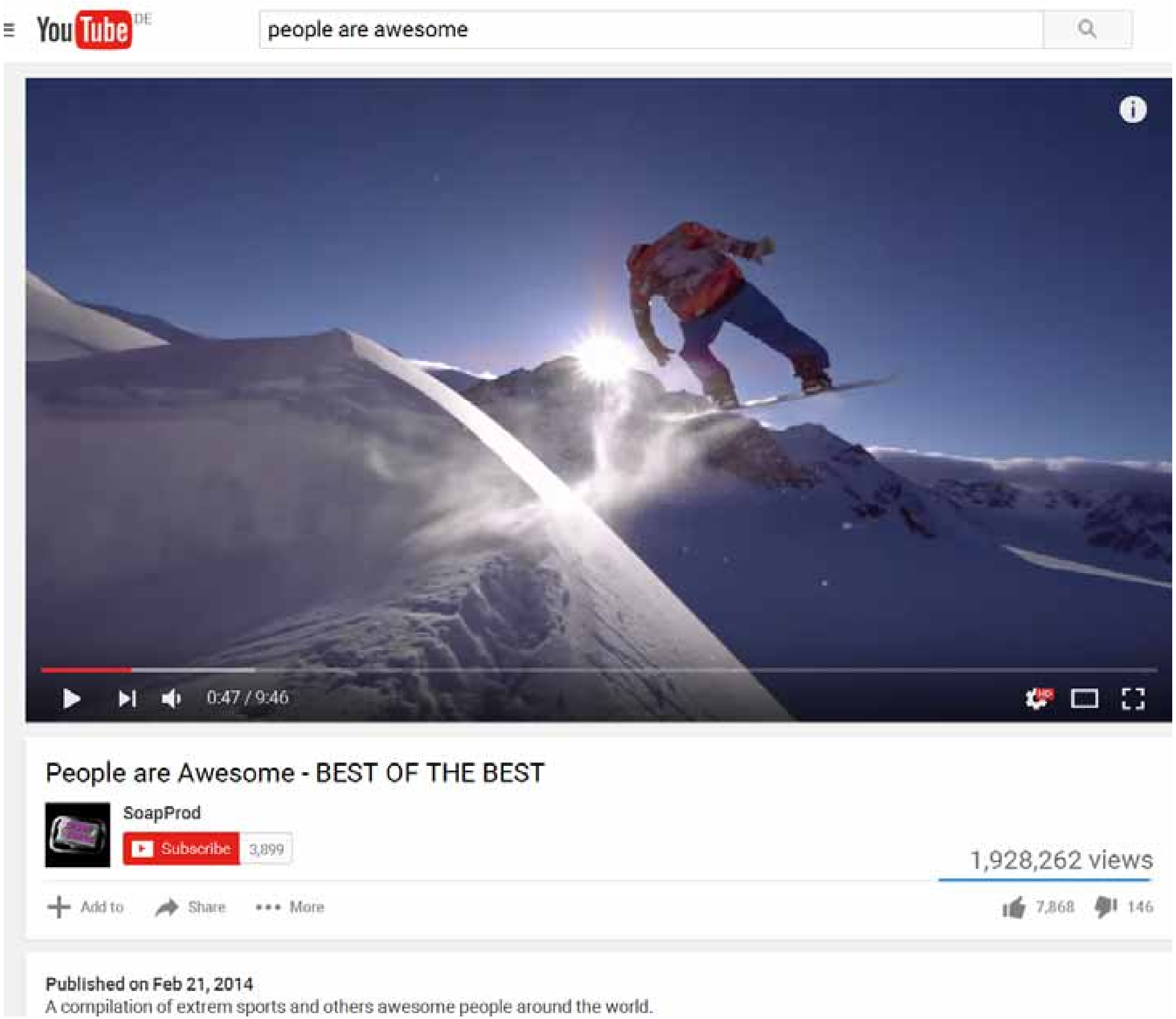
Griffiths, T. L. Connecting human and machine learning via probabilistic models of cognition. Interspeech 2009, 2009 Brighton (UK). ISCA, 9-12. available online via: https://cocosci.berkeley.edu/tom/papers/probmods.pdf

02 What is ML?

- No hand-crafted pre-programmed solutions – adapting/modifying behavior on experience
- Arthur Samuel (1959) [1]: "Field of study that gives computers the ability to learn without being explicitly programmed"

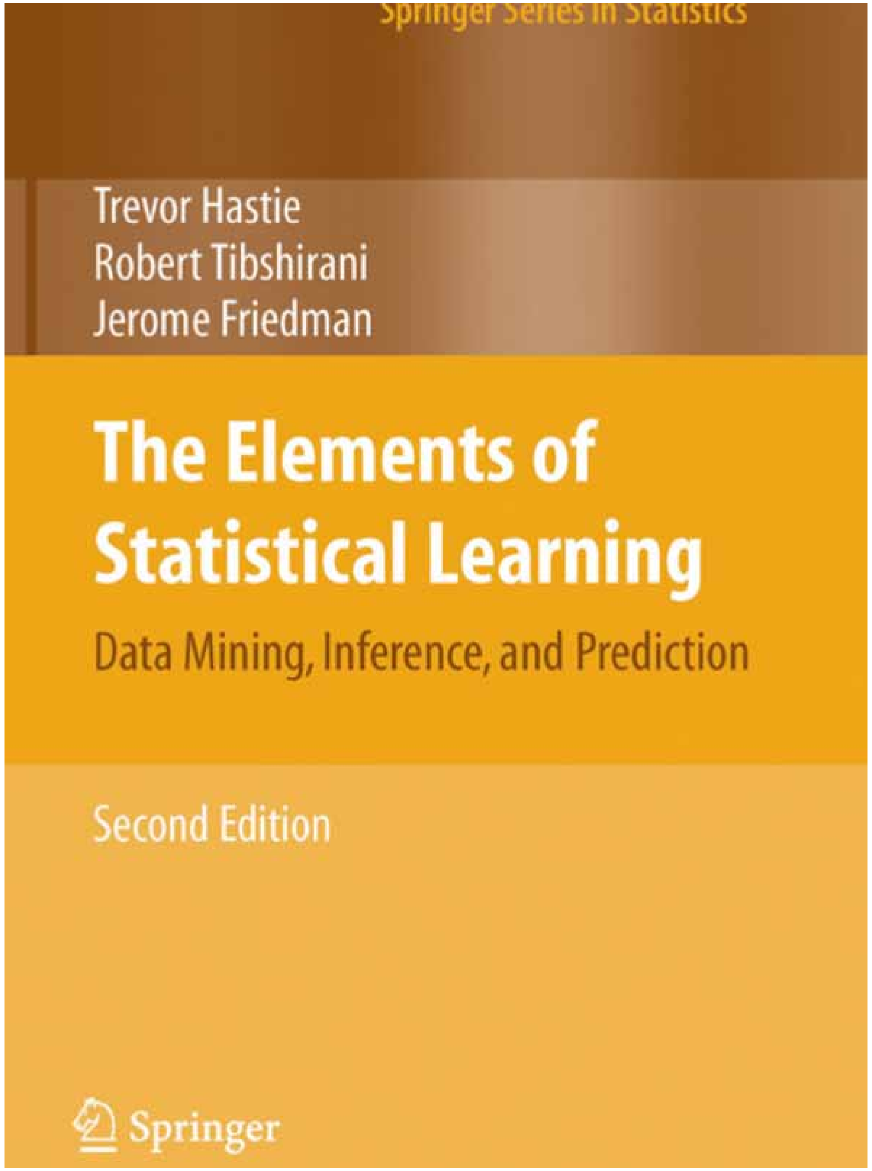
[1] Samuel, A. L. 1959. Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3, (3), 210-229, doi:10.1147/rd.33.0210.

People are awesome …

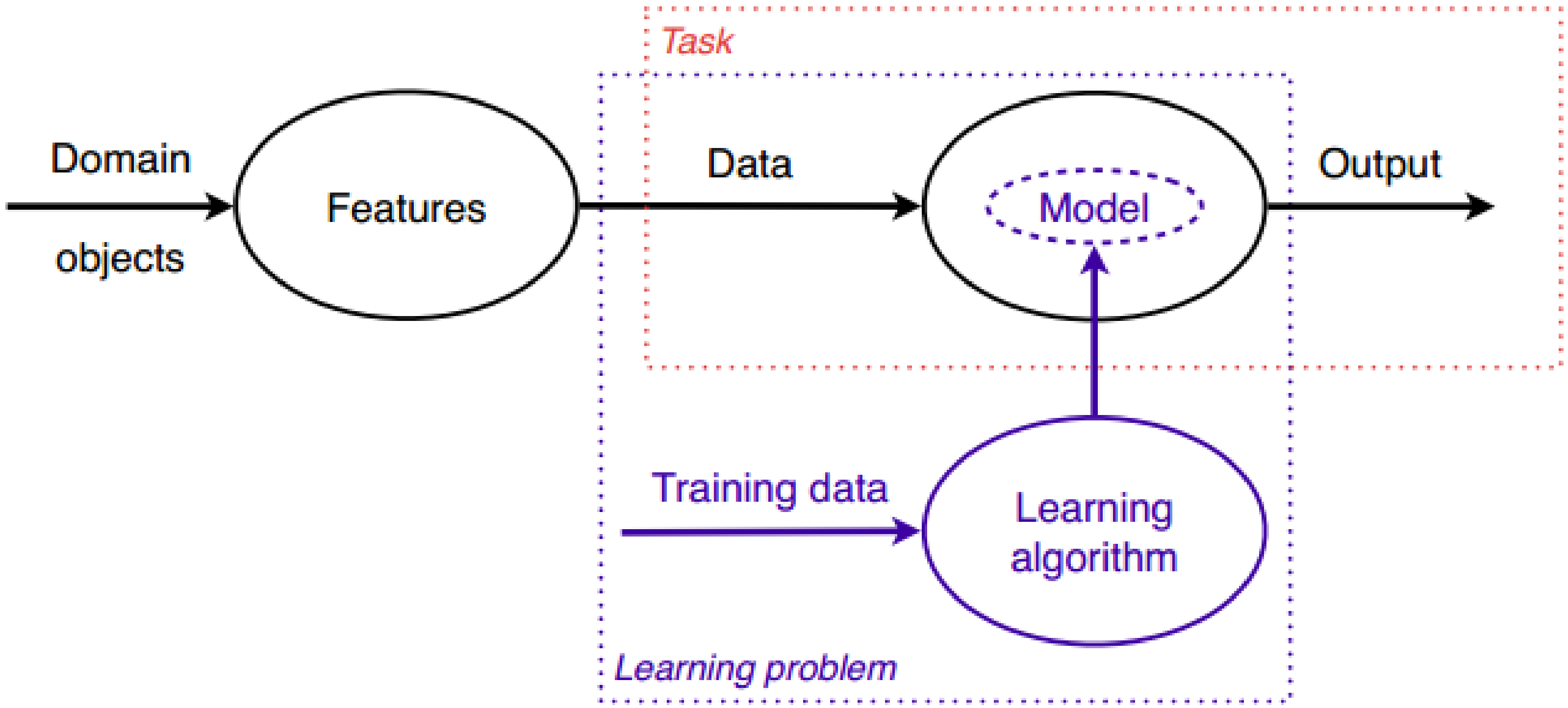


See Youtube: "people are awesome" … hundreds of examples

- Machine Learning is the development of algorithms which can **learn from data**
- Machine Learning has a pre-history in **statistical learning,** which is the application of statistical models and the assessment of **uncertainty**



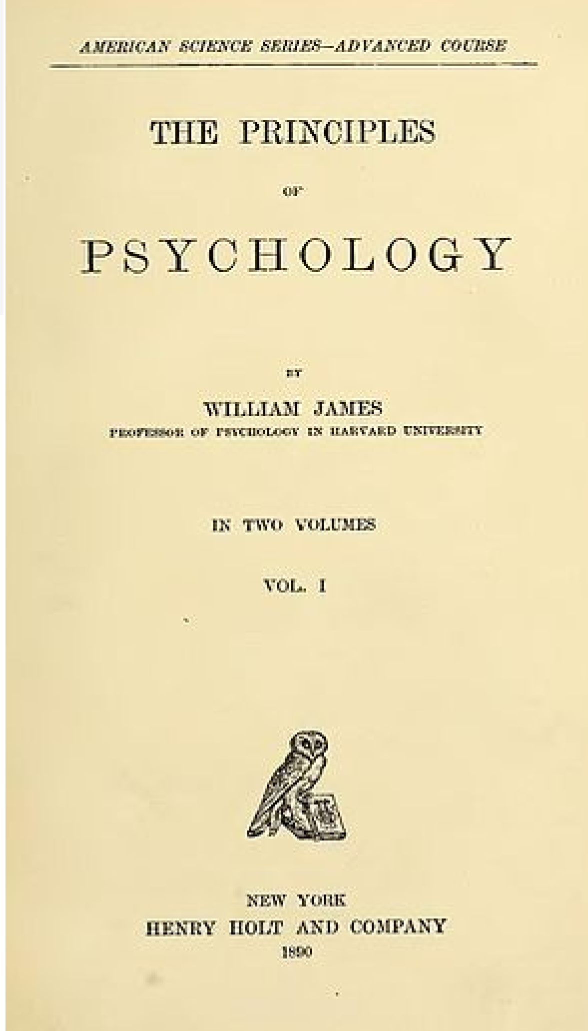
Tasks are addressed by models, whereas learning problems are solved by **learning algorithms that produce models.** Important: Using the right **features** to build the right models that achieve the right tasks!



Flach, P. 2012. Machine learning: the art and science of algorithms that make sense of data, Cambridge University Press, p.11, Figure 3

When do we need machine learning?

- ML is necessary …
  - when human expertise does not yet exist or is absent (bioinformatics, biomarkers, …)
  - when solutions change over time and/or solutions need to adapt to new situations (personalization, …)
  - when humans are unable (or unwilling) to explain their knowledge
  - When the problem space is so enormous large and data are of very high-dimensions which cannot processed manually
- ML is useful … for many, many applications!



AMERICAN SCIENCE SERIES—ADVANCED COURSE

THE PRINCIPLES

OF

PSYCHOLOGY

BY

WILLIAM JAMES

PROFESSOR OF PSYCHOLOGY IN HARVARD UNIVERSITY

IN TWO VOLUMES

VOL. I

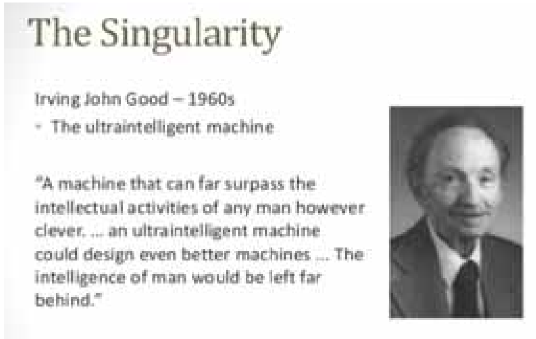
NEW YORK

HENRY HOLT AND COMPANY

1890

- “… The baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion; and to the very end of life, our location of all things in one space is due to the fact that the original extents or bignesses of all the sensations which came to our notice at once, coalesced together into one and the same space …”

James, W. 1890. The principles of psychology, New York, Henry Holt, http://psychclassics.yorku.ca/James/Principles/index.htm.



"An ultra-intelligent machine could design even better machines; there would then unquestionably be an **"intelligence explosion*"** and the intelligence of man would be left far behind … It is curious that this point is made so seldom outside of science fiction."

Irving John Good, Trinity College, Oxford, 1965
Colleague of Alan Turing in Bletchley Park

Good, I. J. 1966. Speculations Concerning the First Ultraintelligent Machine*. In: Franz, L. A. & Morris, R. (eds.) Advances in Computers. Elsevier, pp. 31-88, doi:10.1016/S0065-2458(08)60418-0

https://web.archive.org/web/20010527181244/http://www.aeiveos.com/~bradbury/Authors/Computing/Good-IJ/SCtFUM.html

*) https://intelligence.org/ie-faq/

A ultrashort history of ML

- McCulloch, W. S. & Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biology, 5, (4), 115-133, doi:10.1007/BF02459570.
- Rosenblatt, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65, (6), 386-408, doi:10.1037/h0042519.

Deep Learning is a very old concept …

Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN [1]

F. ROSENBLATT

Cornell Aeronautical Laboratory

If we are eventually to understand the capability of higher organisms for perceptual recognition, generalization, recall, and thinking, we must first have answers to three fundamental questions:

1. How is information about the physical world sensed, or detected, by the biological system?
2. In what form is information stored, or remembered?
3. How does information contained in storage, or in memory, influence recognition and behavior?

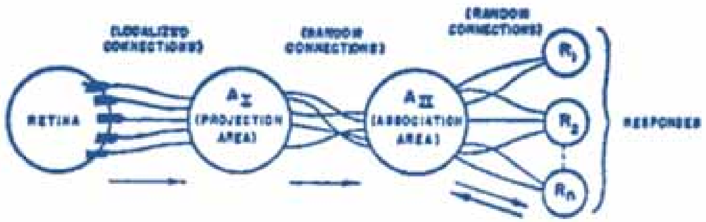
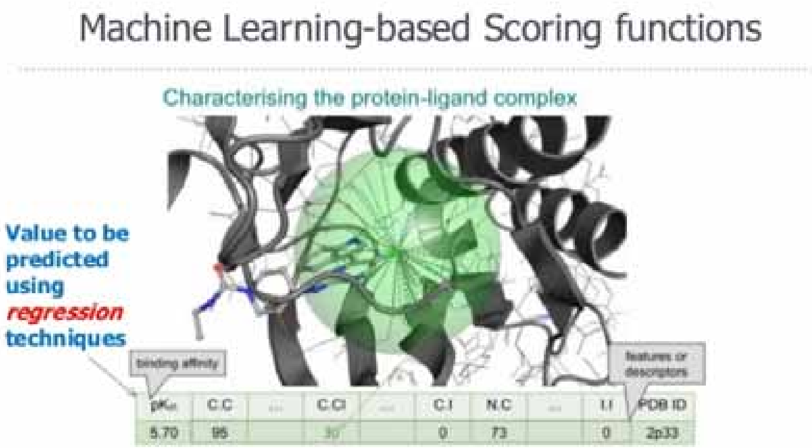


FIG. 1. Organization of a perceptron.

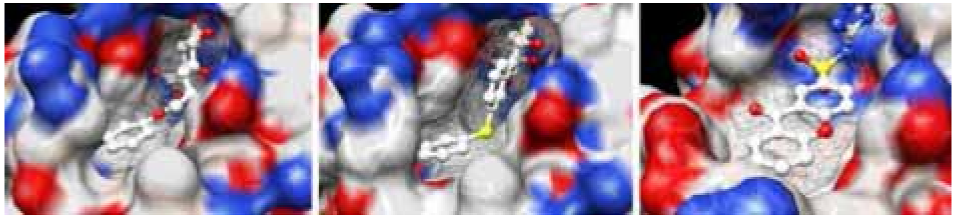
Rosenblatt, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65, (6), 386-408, doi:10.1037/h0042519.

2015 everything is machine learning …

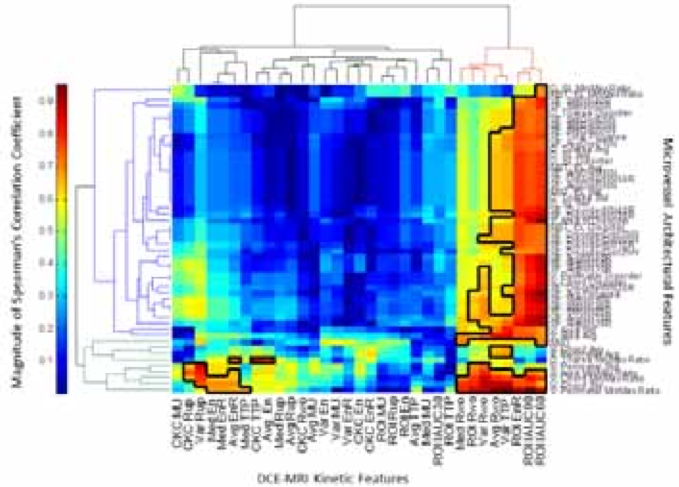
Takacs, G., Pilaszy, I., Nemeth, B., Tikk, D. & Acm 2008. Matrix Factorization and Neighbor Based Algorithms for the Netflix Prize Problem. Recsys'08: Proceedings of the 2008 ACM Conference on Recommender Systems, 267-274.



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.

Fundamental Questions of ML

- 1) Which hypothesis space $\mathcal{H}$ to choose?
- 2) How to measure the degree of fit?
- 3) How to trade-off degree of fit vs. complexity?
- 4) How to find a good hypothesis h ?
- 5) How to know that a good h will predict well?

ad 3) Occam's Razor: the most simple shall be chosen

Answers from classical Science, e.g. Newton, Rutherford

- 1) Which hypothesis space $\mathcal{H}$ to choose?
  - Deterministic h's, mathematical expressions and/or logical sentences; implicit relevance determination
- 2) How to measure the degree of fit?
  - h shall be consistent with the underlying data
- 3) How to trade-off degree of fit vs. complexity?
  - The underlying theory must be totally correct
- 4) How to find a good hypothesis h ?
  - Intuition, imagination, inspiration, inventiveness, …
- 5) How to know that a good h will predict well?
  - David Hume's problem of induction: most scientists give up

ad 2) Ernest Rutherford: If your experiment needs statistics, you ought to have done a better experiment
ad 3) excluding the experimental error
ad 4) Thomas Alva Edison: Ingenious is 1 % inspiration and 99% perspiration

Answers from Andrei Kolmogorow (1903-1987)

- 1) Which hypothesis space $\mathcal{H}$ to choose?
  - All Turing Machines or programs of a Universal TM
- 2) How to measure the degree of fit?
  - Fit must be perfect – program shall output data exactly
- 3) How to trade-off degree of fit vs. complexity?
  - Minimize the size of the program
- 4) How to find a good hypothesis h ?
  - Undecidable … unless we bound time complexity of h
- 5) How to know that a good h will predict well?
  - Theory from PAC learning

ad 5) In computational learning theory, Probably Approximately Correct learning (PAC learning) is a framework for mathematical analysis of machine learning, proposed by Leslie Valiant, 1984.

Answers from statistical ML, e.g. Hastie, Tibshirani *)

- 1) Which hypothesis space $\mathcal{H}$ to choose?
  - e.g., linear combinations of features: $h_w(x) = w^T x$
- 2) How to measure the degree of fit?
  - Loss function, e.g., squared error $\Sigma_j (y_j - w^T x)^2$
- 3) How to trade-off degree of fit vs. complexity?
  - Regularization: complexity penalty, e.g., $||w||^2$
- 4) How to find a good hypothesis h ?
  - Optimization (closed-form, numerical); discrete search
- 5) How to know that a good h will predict well?
  - Try it and see (cross-validation, bootstrap, etc.)

*) Hastie, T., Tibshirani, R. & Friedman, J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, New York, Springer, doi:10.1007/978-0-387-84858-7.

Answers from probabilistic ML, e.g. max. likelihood

- 1) Which hypothesis space $\mathcal{H}$ to choose?
  - Probability model $P(y \mid x, h)$, $e.g., Y \sim N(wTx, \sigma 2)$
- 2) How to measure the degree of fit?
  - Data likelihood $\Pi j\ P(yj \mid xj, h)$
- 3) How to trade-off degree of fit vs. complexity?
  - Regul. or prior: $argmaxh\ P(h)\ \Pi j\ P(yj \mid xj, h)$ (MAP)
- 4) How to find a good hypothesis h ?
  - Optimization (closed-form, numerical); discrete search
- 5) How to know that a good h will predict well?
  - Empirical process theory (generalizes Chebyshev, CLT, PAC…); Key assumption is (i)id

Answers from Bayesian ML, e.g. computing posterior

- 1) Which hypothesis space $\mathcal{H}$ to choose?
  - All hypotheses with nonzero a priori probability
- 2) How to measure the degree of fit?
  - Data probability, as for MLE/MAP
- 3) How to trade-off degree of fit vs. complexity?
  - Use prior, as for MAP
- 4) How to find a good hypothesis h ?
  - Don't! Bayes predictor …
    $$P(y|x,D) = \Sigma h\, P(y|x,h)\, P(D|h)\, P(h)$$
- 5) How to know that a good h will predict well?
  - Bayesian prediction is already optimal ☺

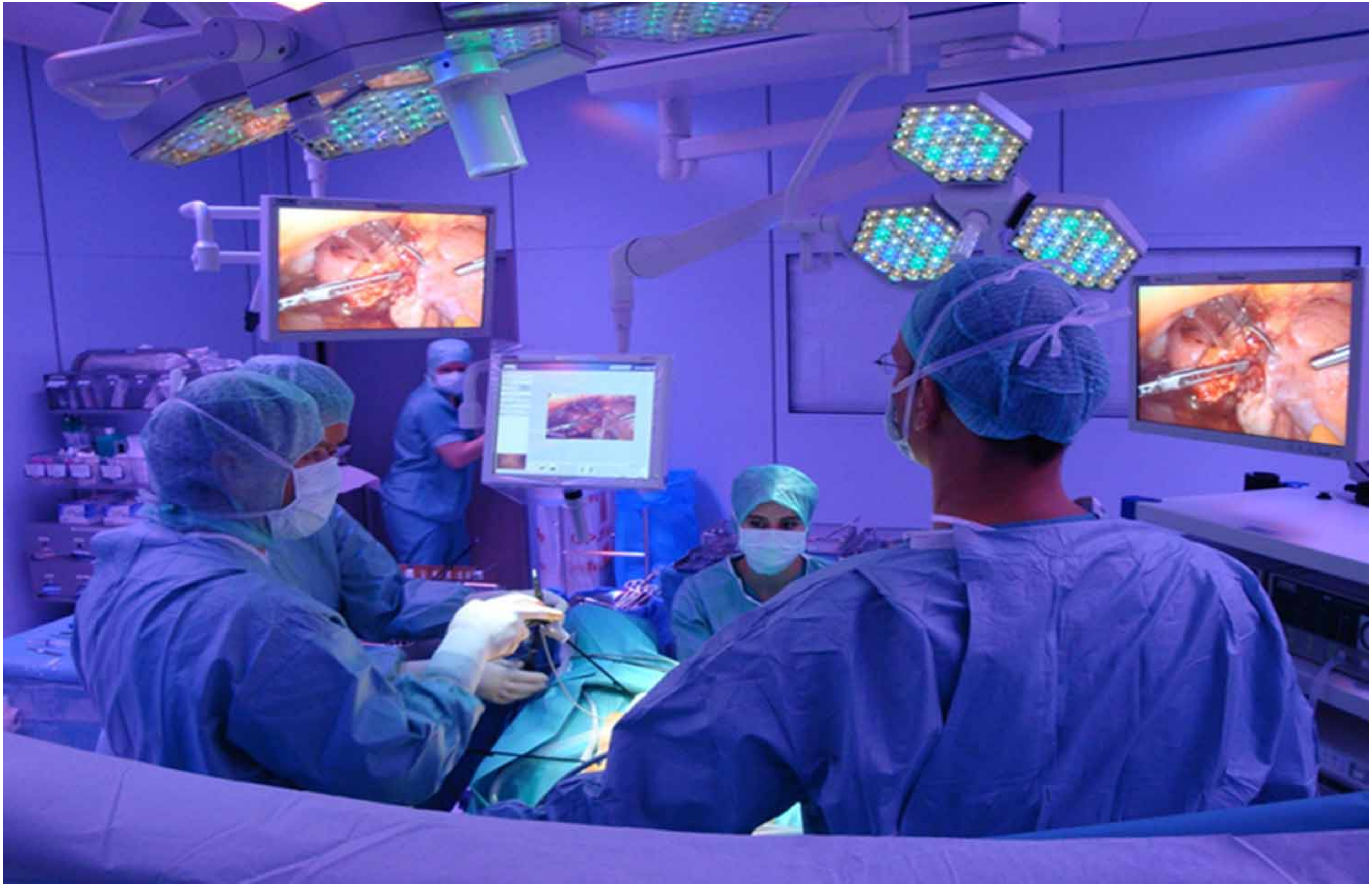
- Progress in ML is driven by the explosion in the availability of **big data** and **low-cost computation.**
- **Health is amongst the biggest challenges**

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

03 Application Area: Health Informatics

Is this our ML lab?

Does all these work here as well?



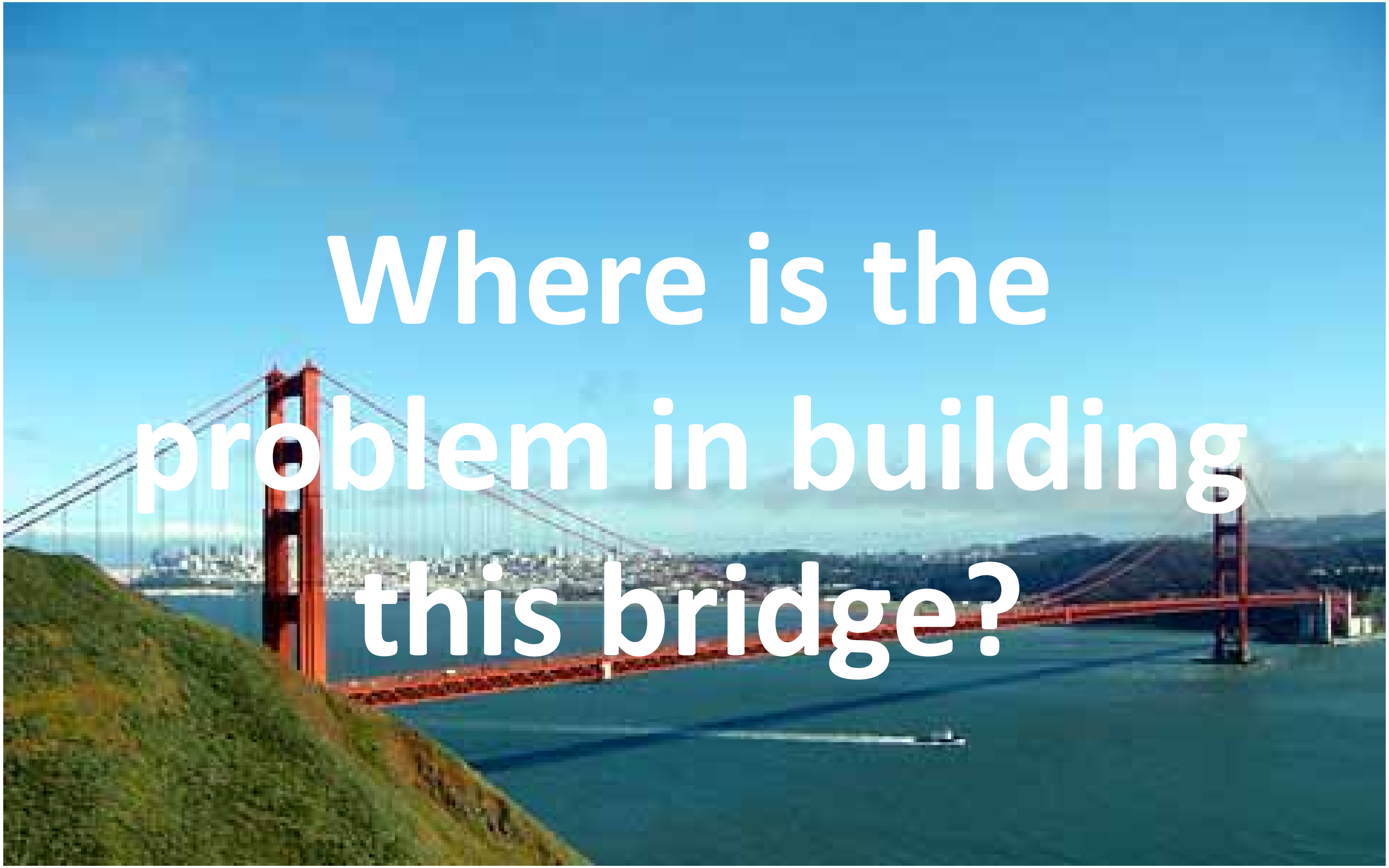
Why is this application area complex ?

In medicine we have two different worlds …

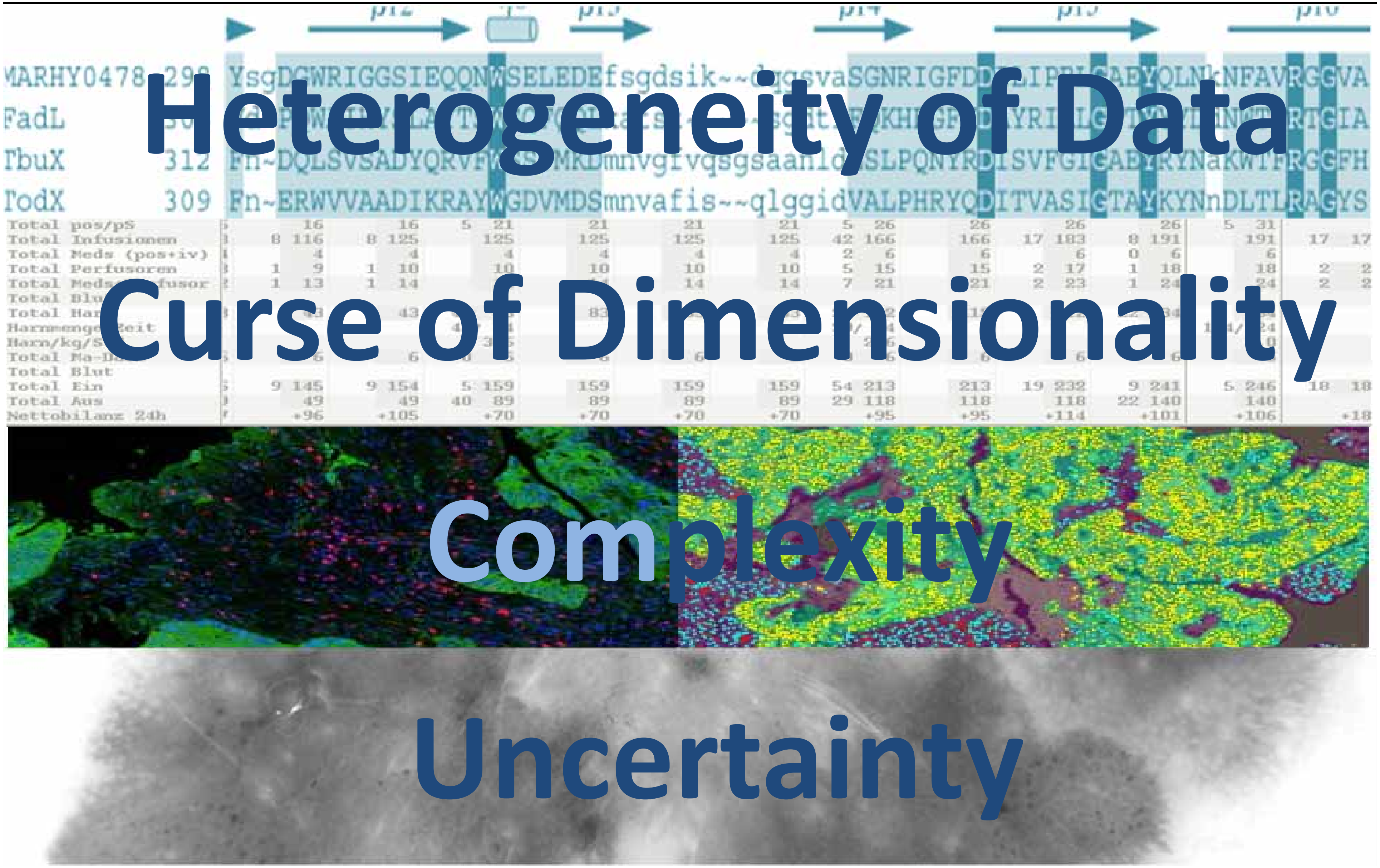
Our central hypothesis:
Information may bridge this gap

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

The bridge …



What are the (main) problems?



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.

04 Probabilistic Information p(x)

Who is this?

Probabilistic Information p(x)

Thomas Bayes
1701 - 1761

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

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

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

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.

Learning and Inference

$\boldsymbol{d}$ … $\boldsymbol{data}$ $\quad$ $\boldsymbol{h}$ … $\boldsymbol{hypothesis}$ $\quad$ $\boldsymbol{H}$ … $\{\boldsymbol{H_1}, \boldsymbol{H_2}, \ldots, \boldsymbol{H_n}\}$ $\quad$ $\forall\, \boldsymbol{h}, \boldsymbol{d}$ …

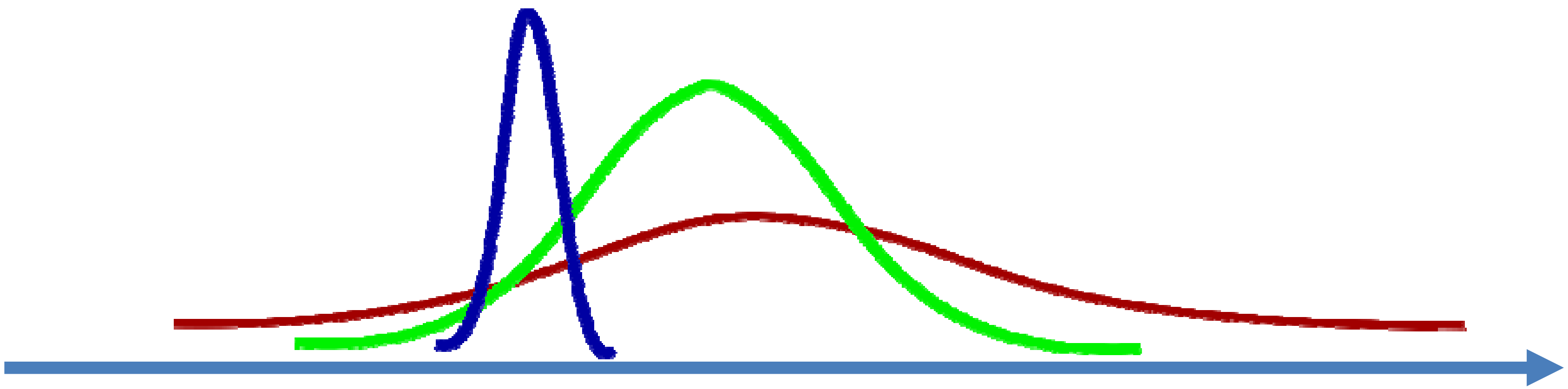
Likelihood

Prior Probability

$$p(h|d) = \frac{p(d|h) * p(h)}{\sum_{h \in H} p(d|h') \, p(h')}$$

Posterior Probability

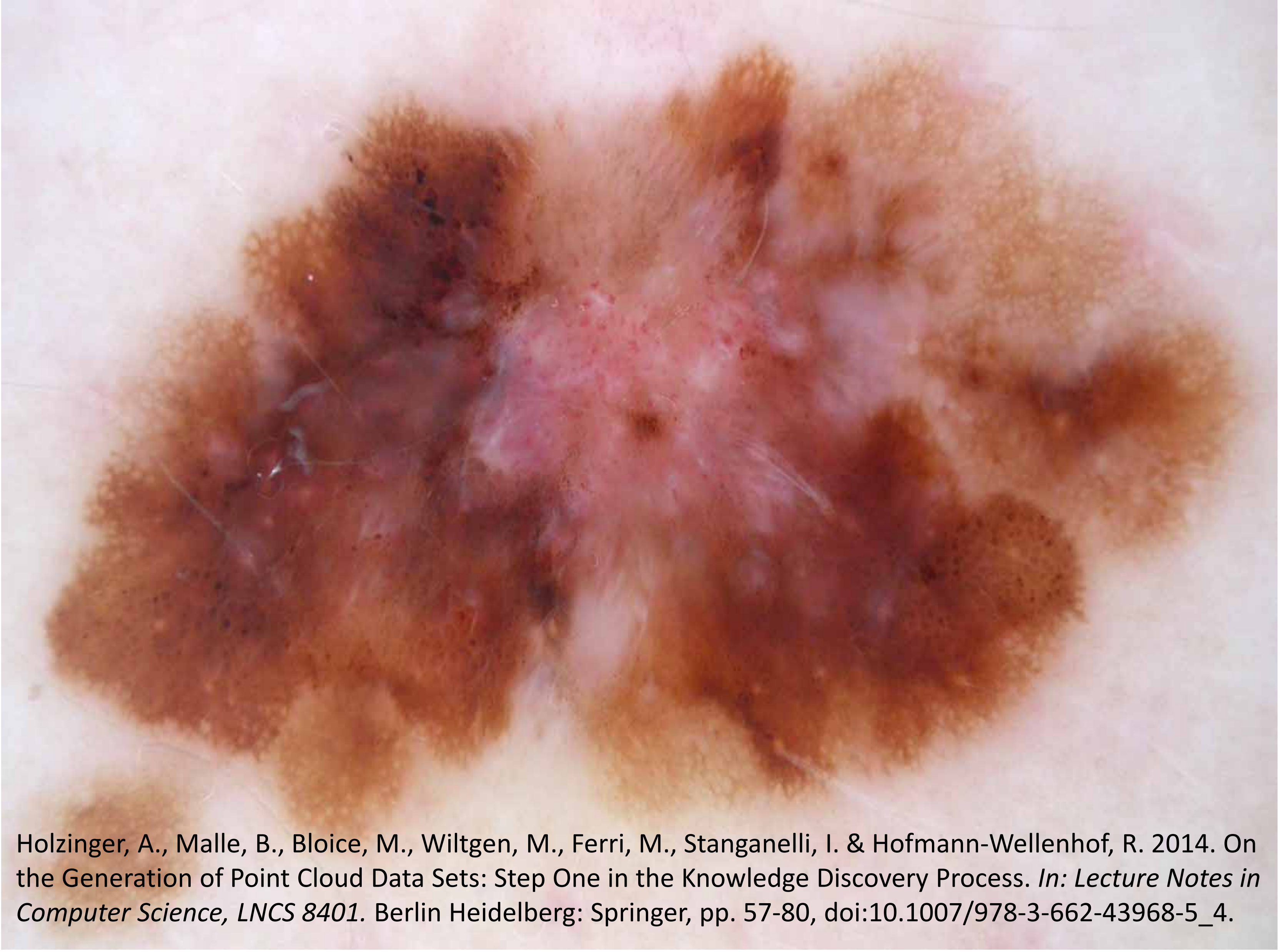
Problem in $\mathbb{R}^n \rightarrow$ complex



Feature parameter θ

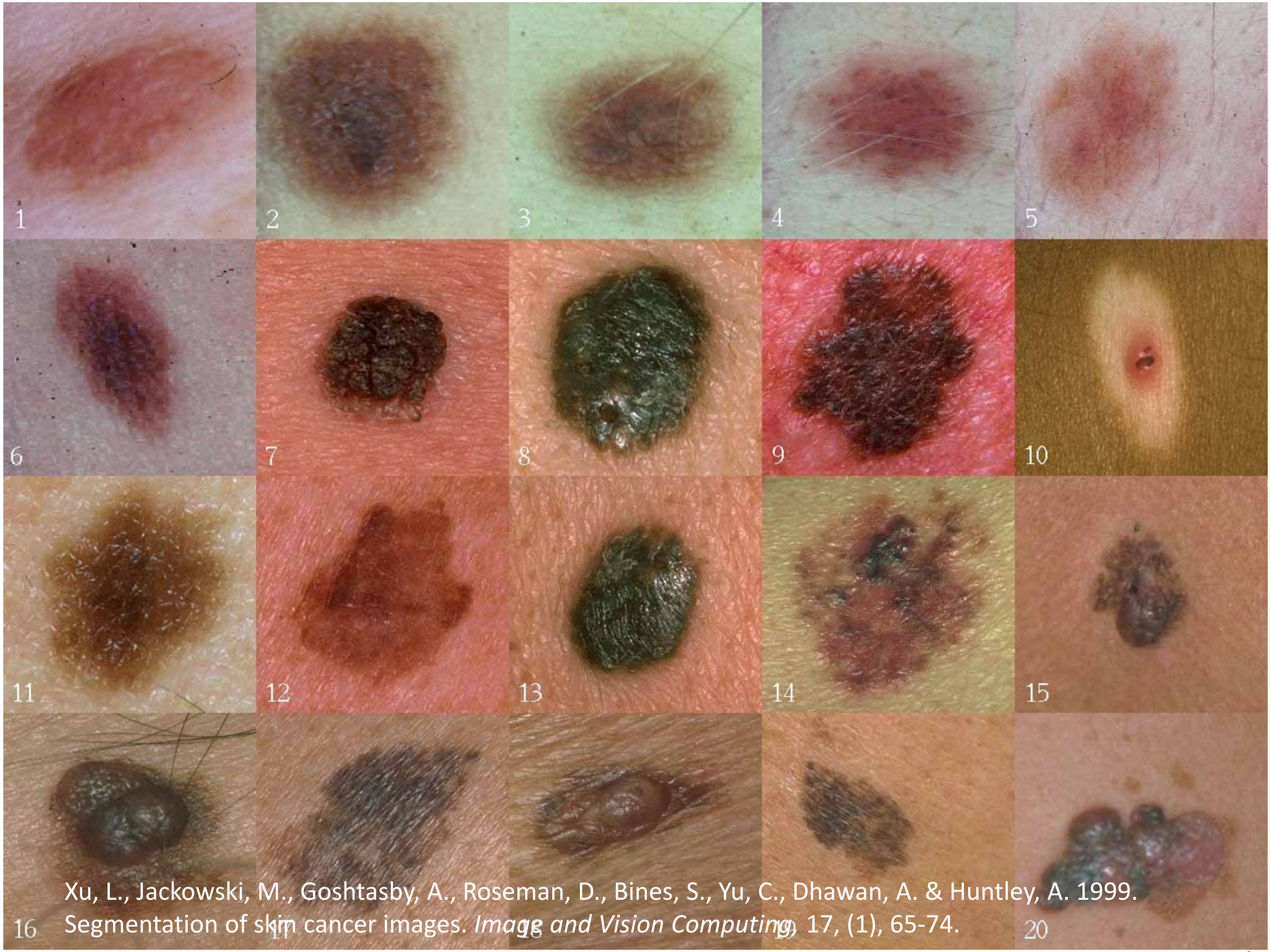
Biomedical Example

To find structural anomalies in such data …



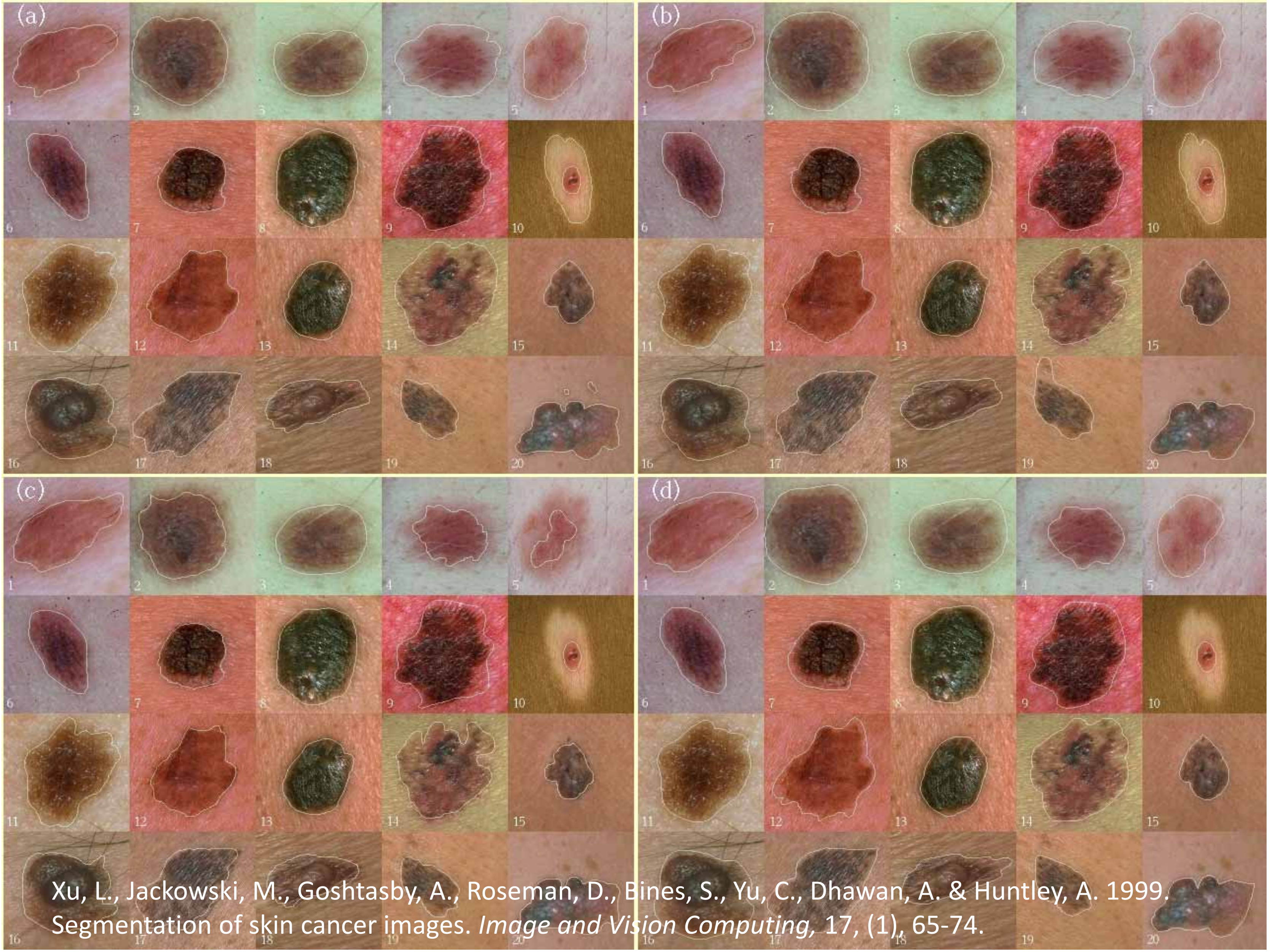
Holzinger, A., Malle, B., Bloice, M., Wiltgen, M., Ferri, M., Stanganelli, I. & Hofmann-Wellenhof, R. 2014. On the Generation of Point Cloud Data Sets: Step One in the Knowledge Discovery Process. *In: Lecture Notes in Computer Science, LNCS 8401.* Berlin Heidelberg: Springer, pp. 57-80, doi:10.1007/978-3-662-43968-5_4.

Learning from previous examples …



Xu, L., Jackowski, M., Goshtasby, A., Roseman, D., Bines, S., Yu, C., Dhawan, A. & Huntley, A. 1999. Segmentation of skin cancer images. *Image and Vision Computing*, 17, (1), 65-74.

The more examples we have the better …



Xu, L., Jackowski, M., Goshtasby, A., Roseman, D., Bines, S., Yu, C., Dhawan, A. & Huntley, A. 1999. Segmentation of skin cancer images. *Image and Vision Computing,* 17, (1), 65-74.

Five Mainstreams in Machine Learning

- Symbolic ML
  - First order logic, inverse deduction
  - Tom Mitchell, Steve Muggleton, Ross Quinlan, …
- Bayesian ML
  - Statistical learning
  - Judea Pearl, Michael Jordan, David Heckermann, …
- Cognitive ML
  - Analogisms from Psychology, Kernel machines
  - Vladimir Vapnik, Peter Hart, Douglas Hofstaedter, …
- Connectionist ML
  - Neuroscience, Backpropagation
  - Geoffrey Hinton, Yoshua Bengio, Yann LeCun, …
- Evolutionary ML
  - Nature-inspired concepts, genetic programming
  - John Holland (1929-2015), John Koza, Hod Lipson, …

Big Data is good for automatic Machine Learning

$$\mathcal{D} = x_{1:n} = \{x_1, x_2, ..., x_n\}$$

$$p(\mathcal{D}|\theta)$$

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta) * p(\theta)}{p(\mathcal{D})}$$

$$posterior = \frac{likelihood * prior}{evidence}$$

The inverse probability allows to learn from data, infer unknowns, and make predictions

$$\max_{\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^d} f(\mathbf{x})$$

$$p(h|d) \propto p(\mathcal{D}|\theta) * p(h)$$

$$p(f(x)|\mathcal{D}) \propto p(\mathcal{D}|f(x)) * p(f(x))$$

- Machine Learning is the development of algorithms which can **learn from data**
- assessment of **uncertainty,** making **predictions**
- **Automating automation - getting computers to program themselves – let the data do the work!**

Analogies



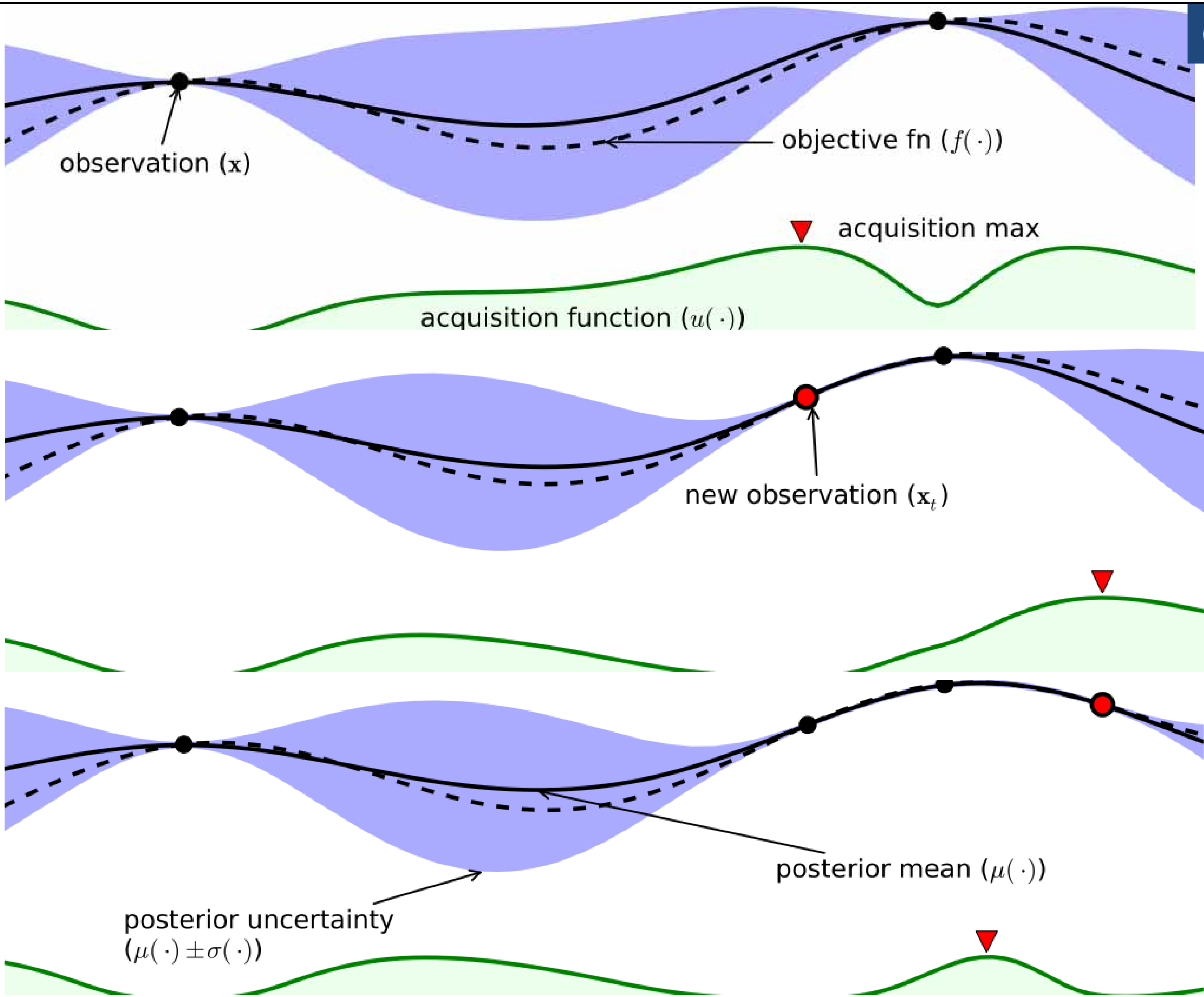

- **Newton, Leibniz, … developed calculus – mathematical language for describing and dealing with rates of change**
- **Bayes, Laplace, … developed probability theory - the mathematical language for describing and dealing with uncertainty**

Who is this?


GN4480100S8
DEUTSCHE BUNDESBANK
Banknote
10
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
1777–1855 Carl Friedr. Gauß
ZEHN DEUTSCHE MARK
Deutsche Bundesbank
Frankfurt am Main
1. September 1999
GN4480100S8

05 Gaussian Processes

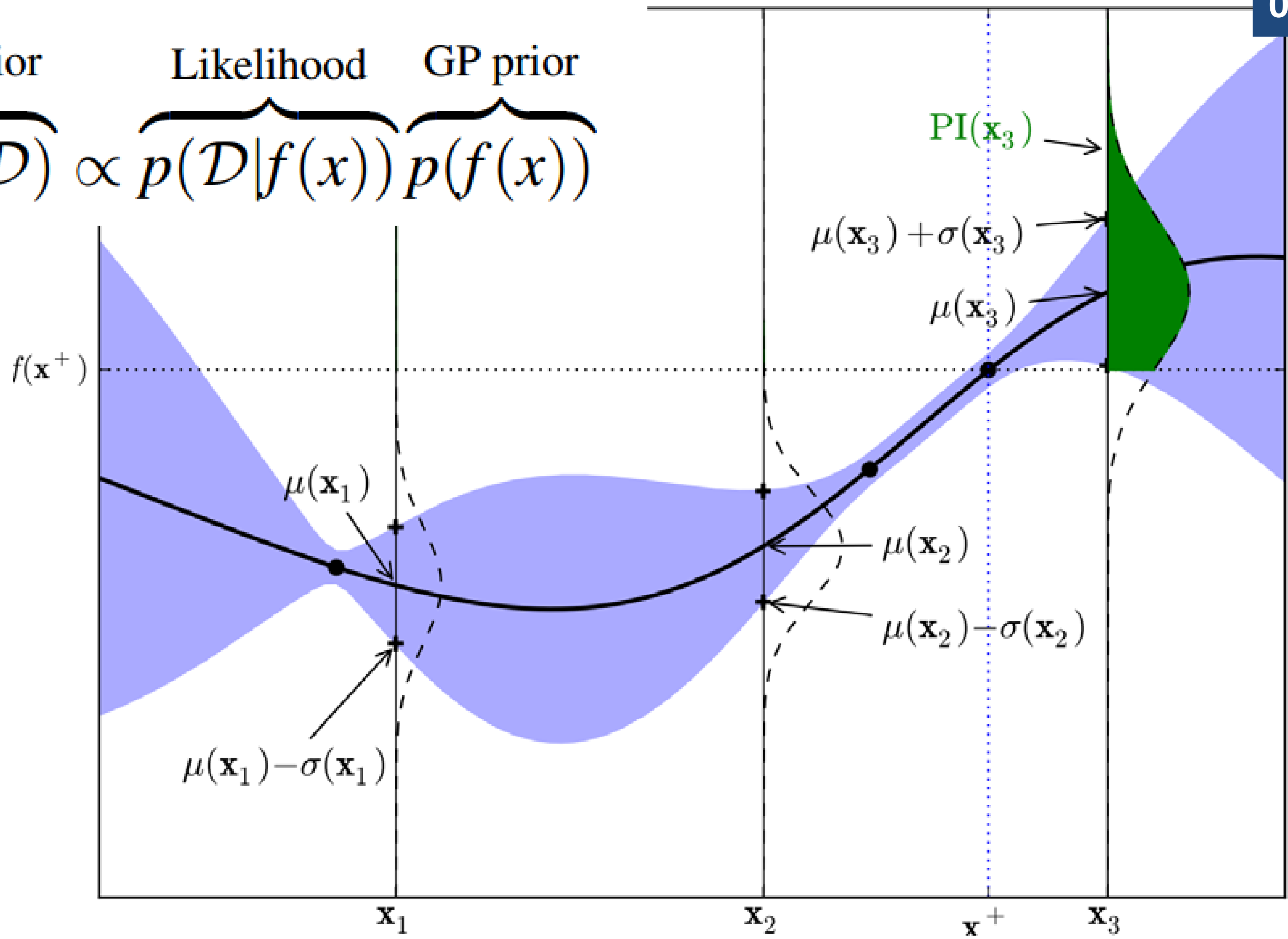
From Bayesian Optimization to Gaussian Process (GP) approximation



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.

GP = distribution, observations occur in a cont. domain, e.g. t or space

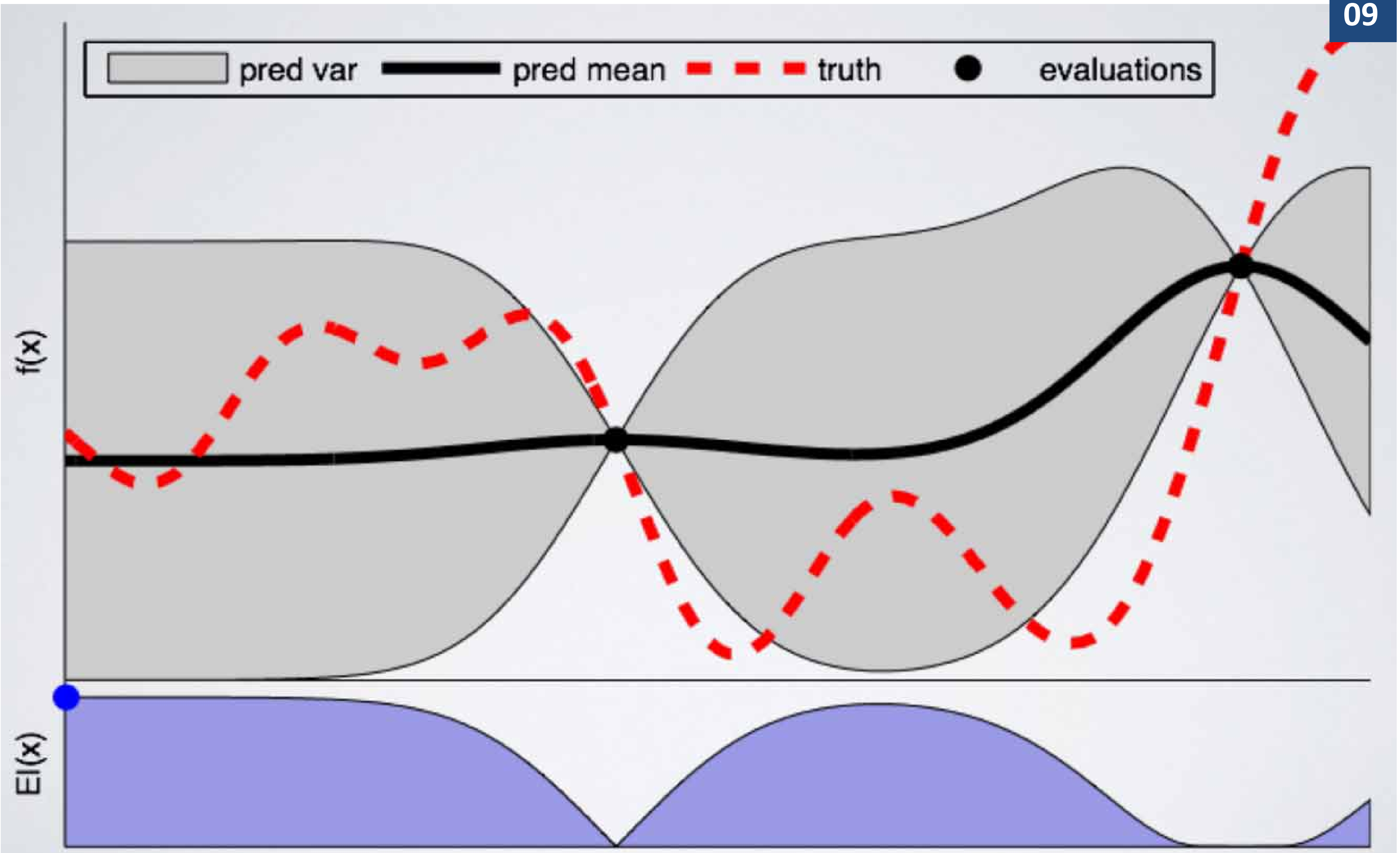
$$\overbrace{p(f(x)|\mathcal{D})}^{\text{GP posterior}} \propto \overbrace{p(\mathcal{D}|f(x))}^{\text{Likelihood}}\,\overbrace{p(f(x))}^{\text{GP prior}}$$



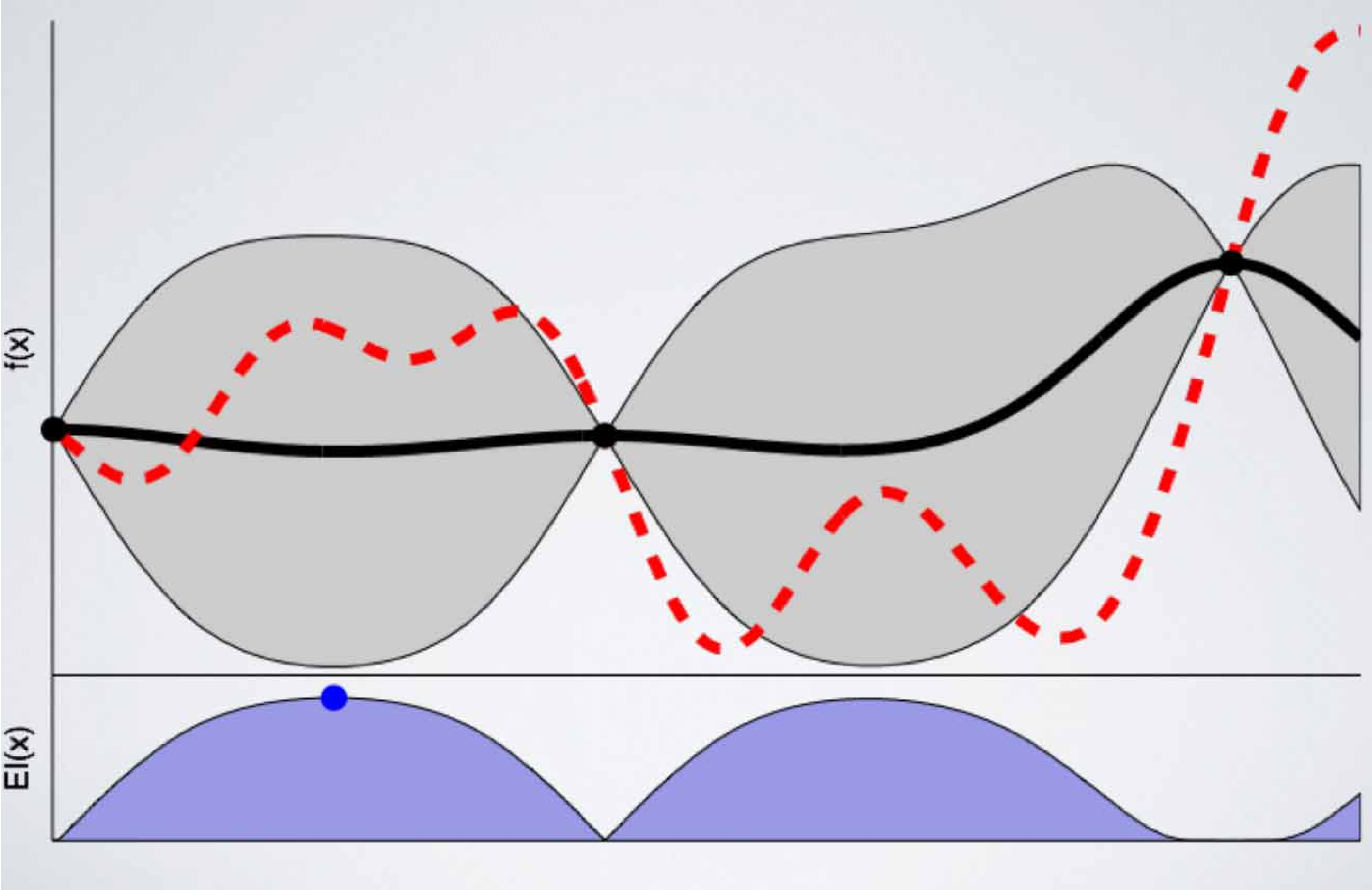
Brochu, E., Cora, V. M. & De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599.

Demo on how Bayesian Optimization works …

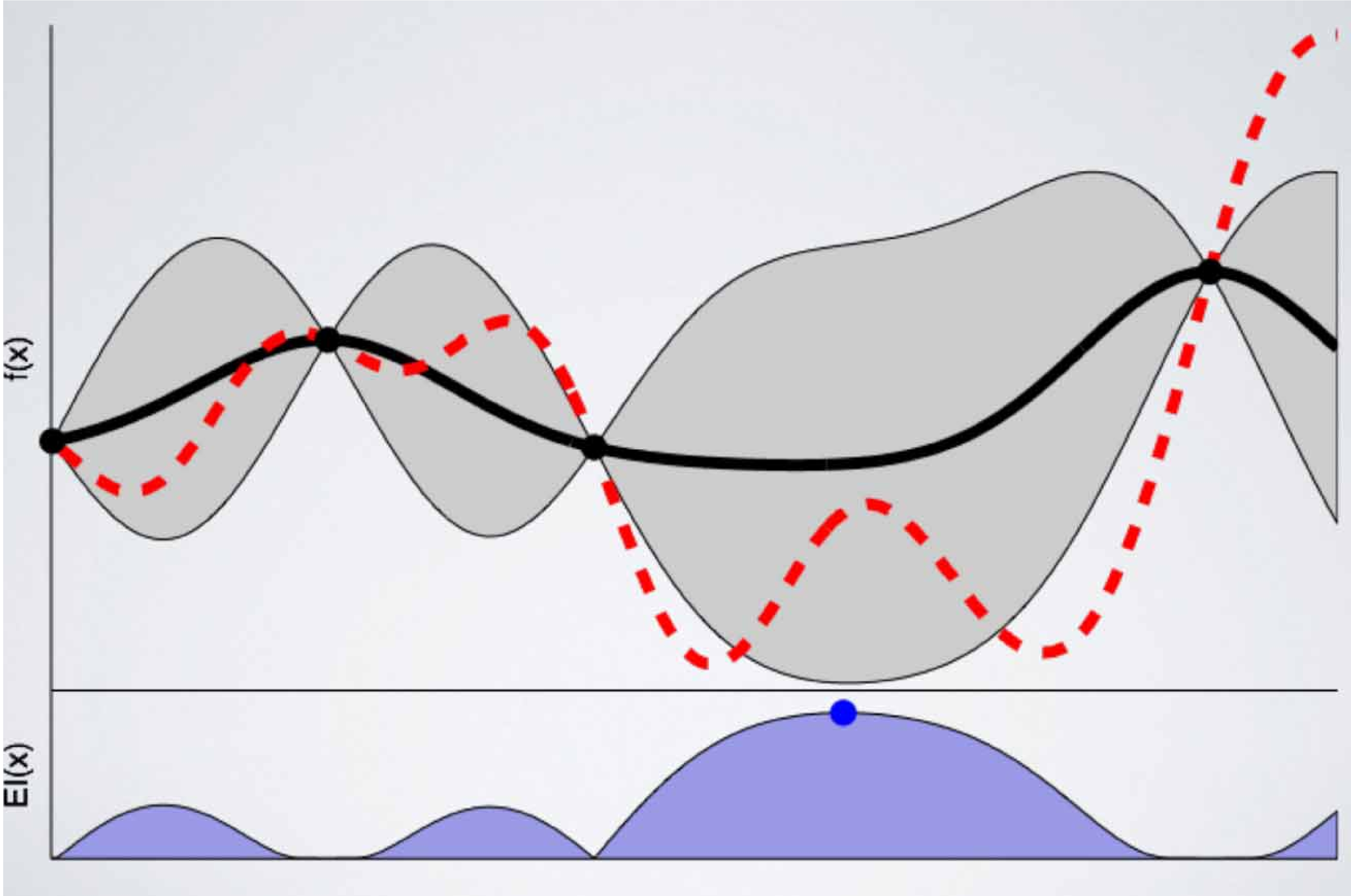
09



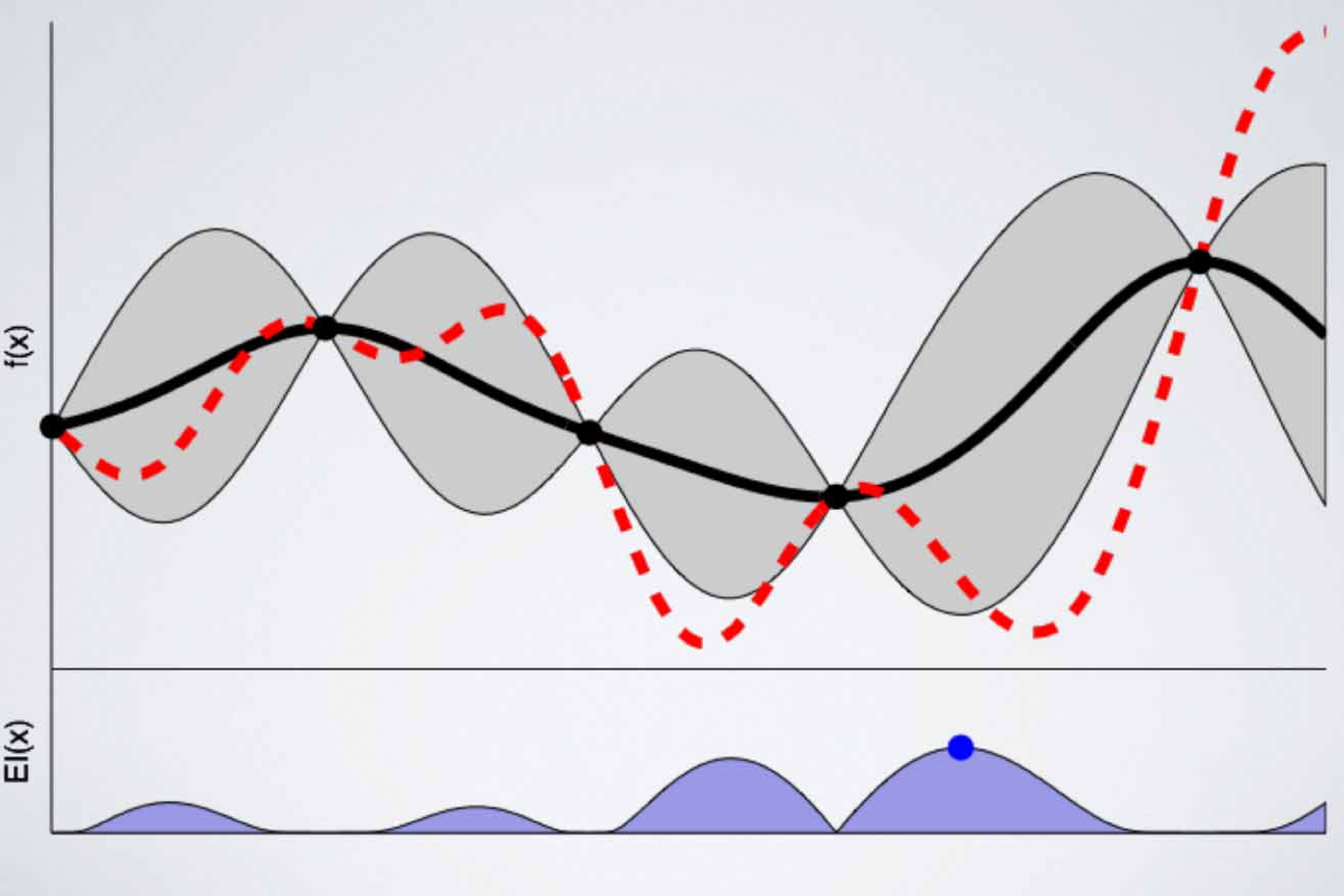
Snoek, J., Larochelle, H. & Adams, R. P. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 2012. 2951-2959.


f(x)
EI(x)

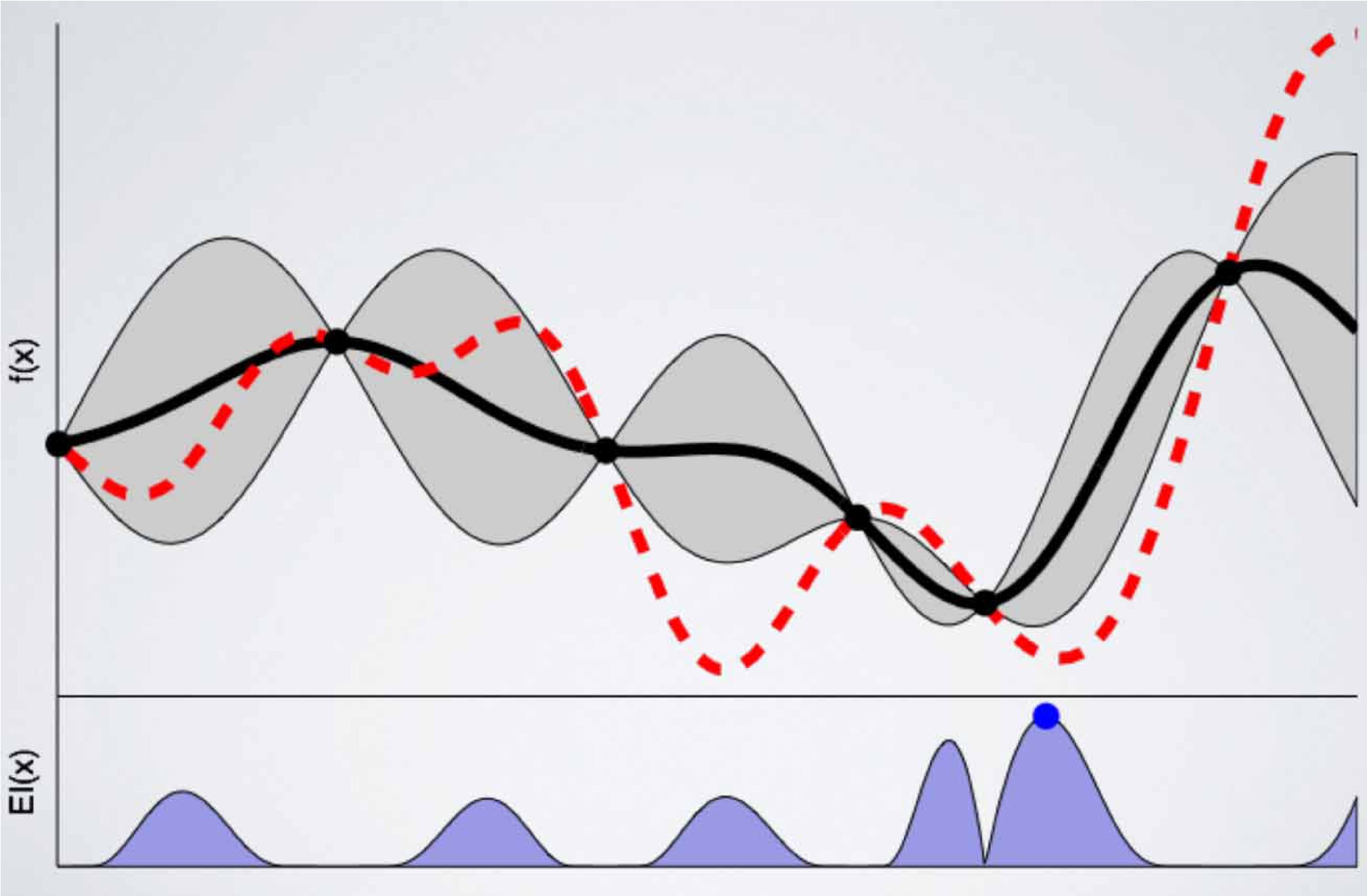
Bayesian Optimization 2



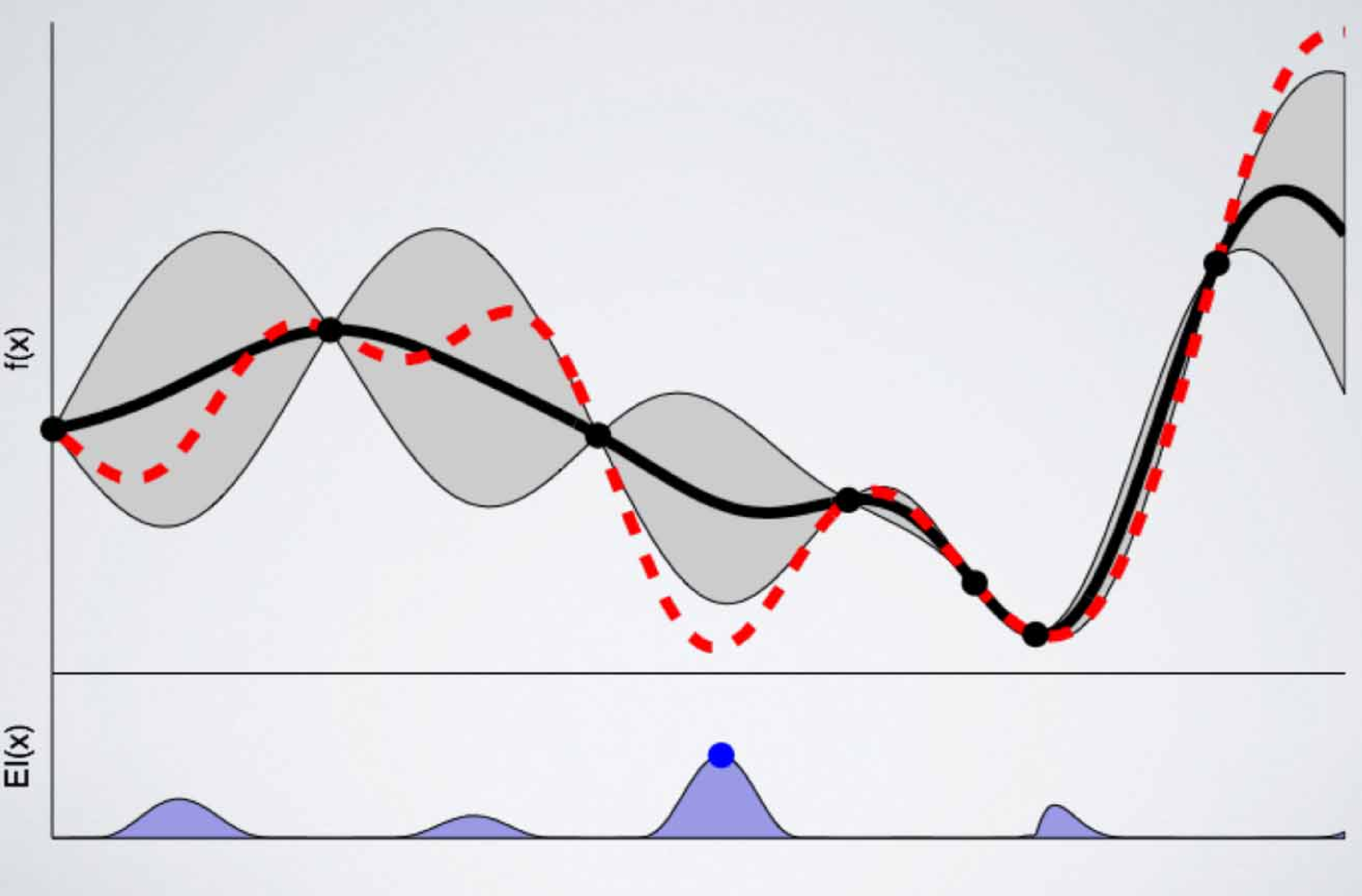
Bayesian Optimization 3



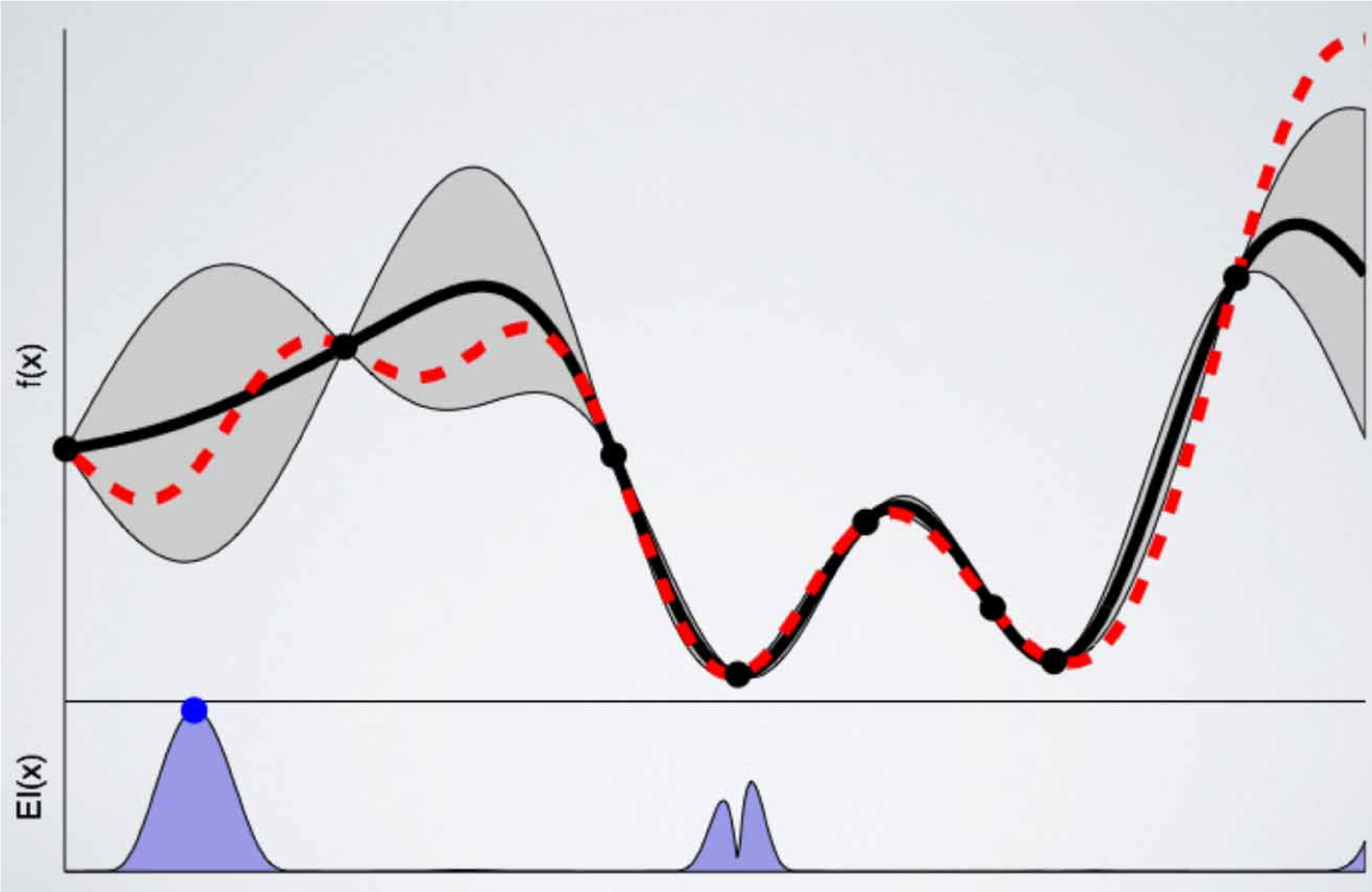
Bayesian Optimization 4



Bayesian Optimization 5



Bayesian Optimization 6



Why is this relevant for health informatics?

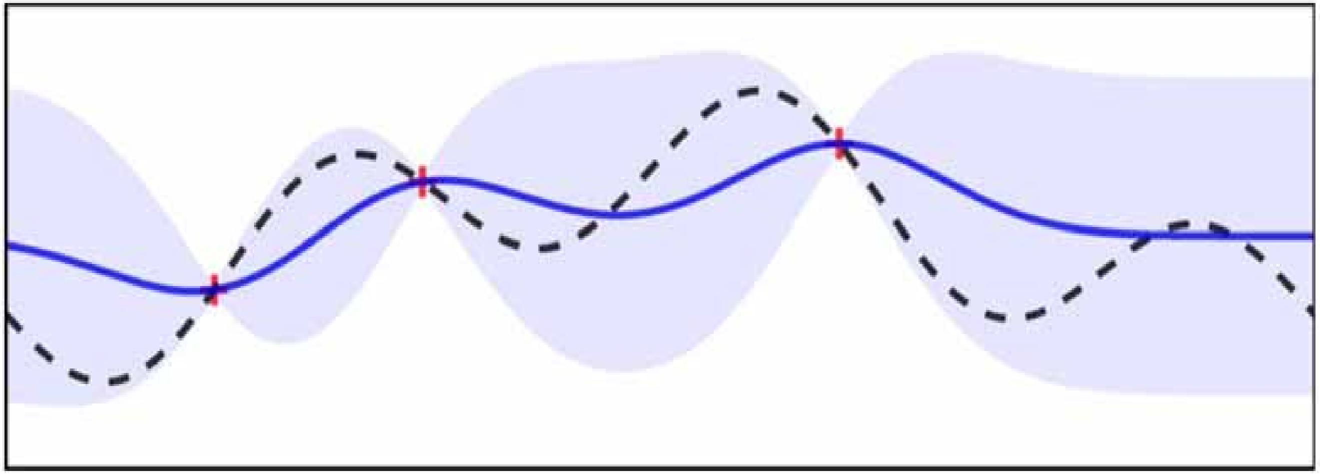
- Take patient information, e.g., observations, symptoms, test results, -omics data, etc. etc.
- Reach conclusions, and **predict** into the future, e.g. how likely will the patient be …
- Prior = belief before making a particular observation
- Posterior = belief after making the observation and is the prior for the next observation – intrinsically incremental

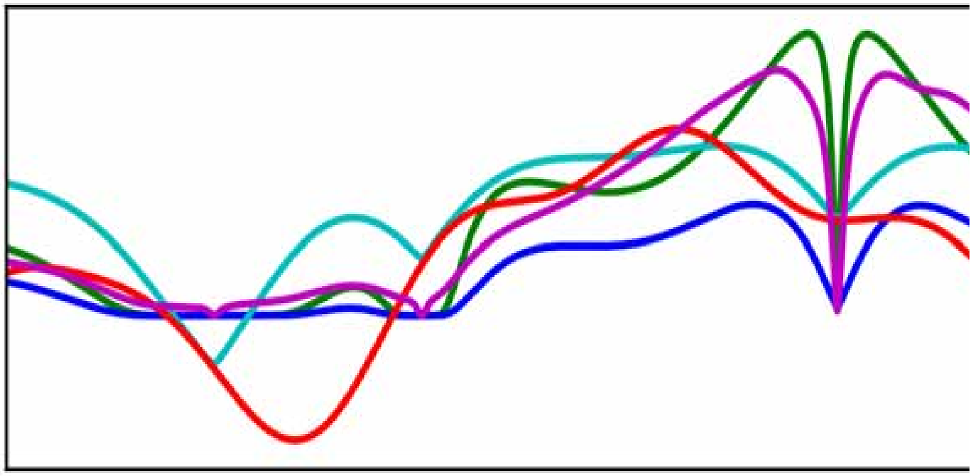
$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}$$

Grand Goal of automatic Machine Learning

…

taking the human out-of-the-loop





Algorithm 1: Bayesian optimization

1: **for** $n = 1, 2, \ldots,$ **do**
2: select new $\mathbf{x}_{n+1}$ by optimizing acquisition function α

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}} \alpha(\mathbf{x}; \mathcal{D}_n)$$

3: query objective function to obtain y_{n+1}
4: augment data $\mathcal{D}_{n+1} = \{\mathcal{D}_n, (\mathbf{x}_{n+1}, y_{n+1})\}$
5: update statistical model
6: **end for**

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. 2016. **Taking the human out of the loop:** A review of Bayesian optimization. *Proceedings of the IEEE*, 104, (1), 148-175, doi:10.1109/JPROC.2015.2494218.

06 aML

- Today most ML-applications are using automatic Machine Learning (aML) approaches

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

Best practice examples of aML

Wenger 600638 IBEX 17" Laptop Backpack with Tablet / eReader Pocket (Black / Blue)
von Wenger
EUR 66,99 Prime

Andere Angebote
EUR 60,00 neu (22 Angebote)
EUR 57,70 gebraucht (1 Angebot)

Nur Artikel von Wenger anzeigen

295

Für größere Ansicht Maus über das Bild ziehen

Lenovo Einsteiger Not
Arbeitsspeicher und W
von Lenovo
70 Kundenre
Bestseller Nr. 1 in Notebooks

Unverb. Preisempf.: ~~EUR 349,0~~
Preis: EUR 273
Sie sparen: EUR 75,01
Alle Preisan

Lieferung Mittwoch, 6. Juli: Be
Details.

Auf Lager.
Verkauf und Versand durch Ama

20 neu ab EUR 273,99 4 gebr

Größe: **500GB**

1TB **500GB**

Stil: **Intel Pentium**

Intel Core i3 **Intel Pentiu**

- Prozessor: Intel Pentium N35
- Besonderheiten: HD Glare Di
- Akku: bis zu 4 Stunden Akkul
- Herstellergarantie: 12 Monate Angaben des jeweiligen Verkä
- Lieferumfang: Lenovo ideapa

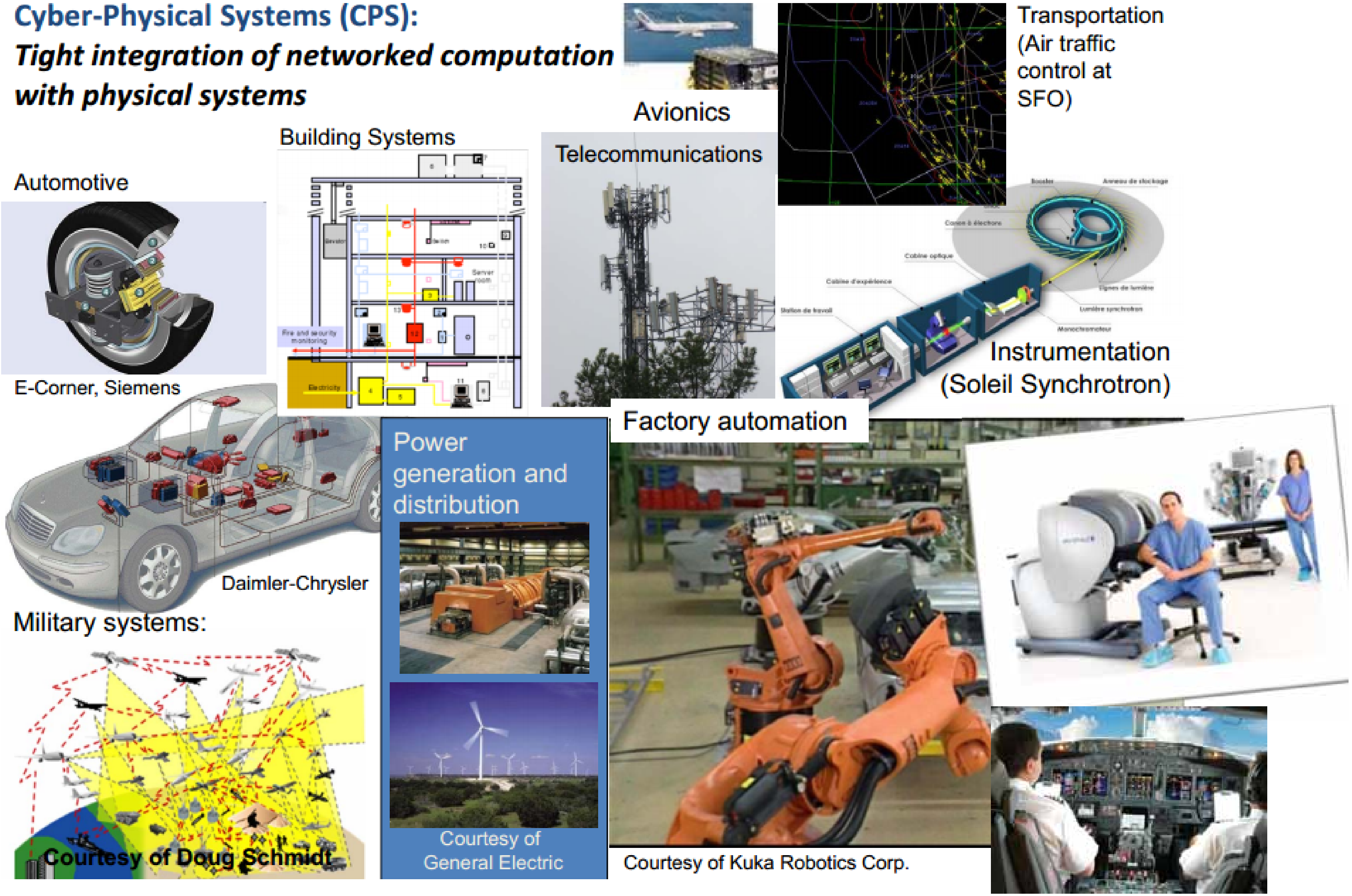
› Weitere Produktdetails

Fully automatic autonomous vehicles ("Google car")

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

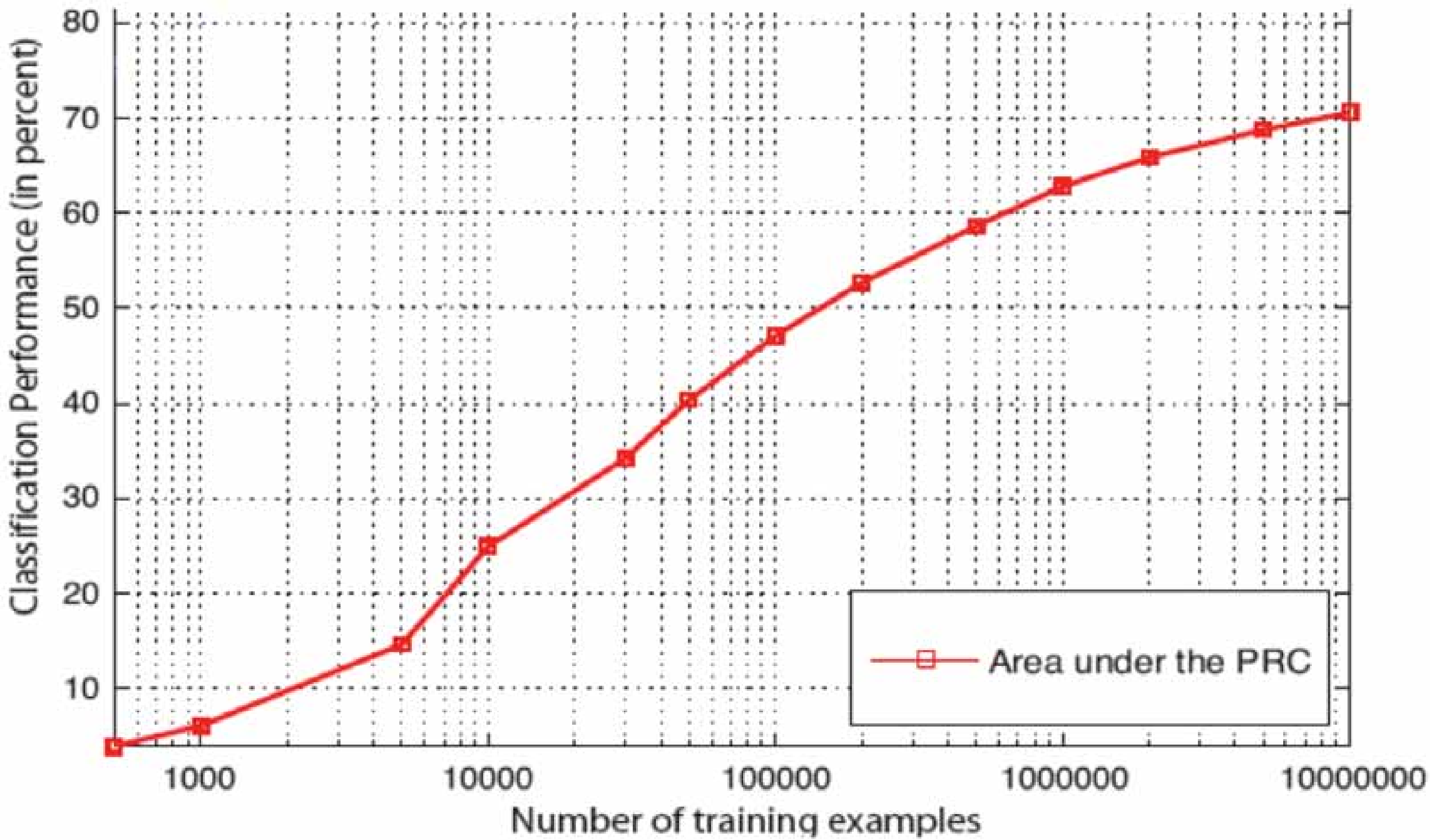
… and thousands of industrial aML applications …

Cyber-Physical Systems (CPS):
Tight integration of networked computation with physical systems



Seshia, S. A., Juniwal, G., Sadigh, D., Donze, A., Li, W., Jensen, J. C., Jin, X., Deshmukh, J., Lee, E. & Sastry, S. 2015. Verification by, for, and of Humans: Formal Methods for Cyber-Physical Systems and Beyond. Illinois ECE Colloquium.

Big Data is necessary for aML !

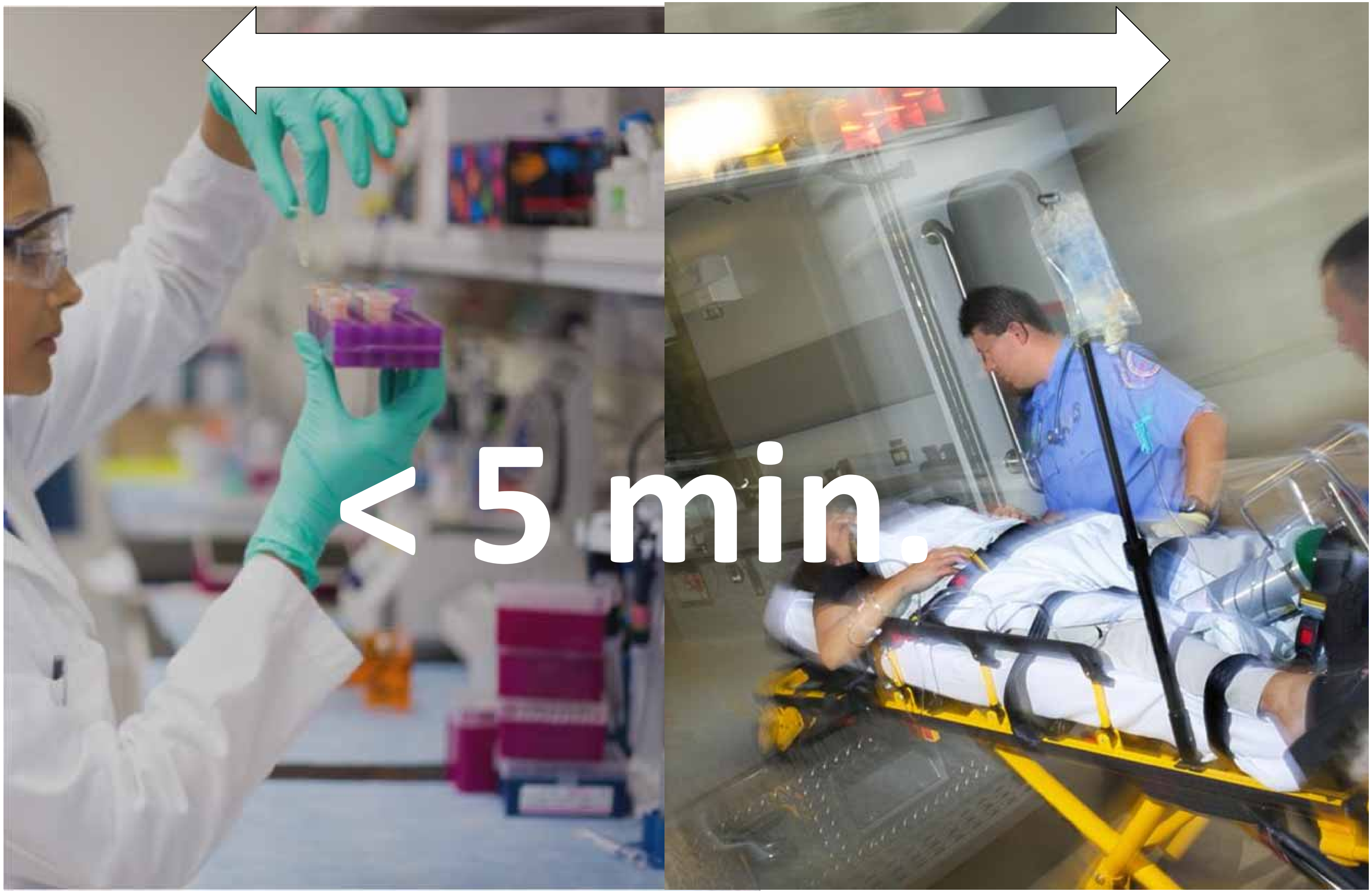


Sonnenburg, S., Rätsch, G., Schäfer, C. & Schölkopf, B. 2006. Large scale multiple kernel learning. Journal of Machine Learning Research, 7, (7), 1531-1565.

Again: Does this all work here as well?

Medical Decision Making as a Search Task in $\mathcal{H}$ Problem: Time (t)

Search in an arbitrarily high-dimensional space < 5 min.!



- Maximizing Expected Utility Theory

07 iML

When does aML fail …

- 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.…

Example: 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

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

A doctor-in-the-loop

Definition of iML

A6

- 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

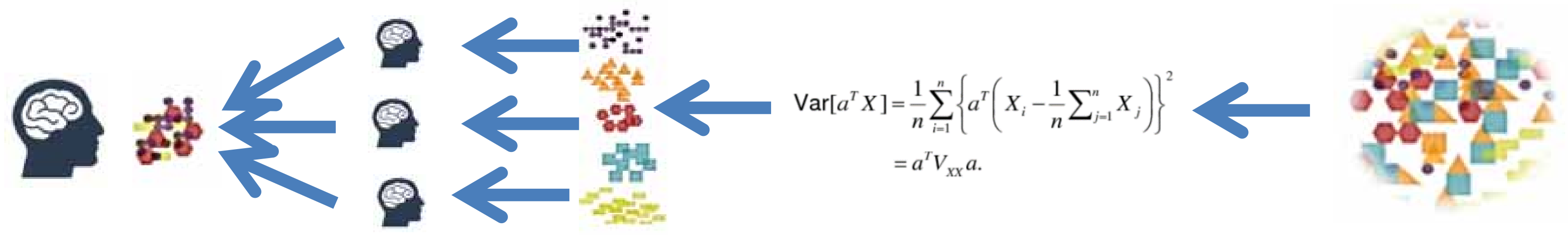
A group of experts-in-the-loop

A crowd of people-in-the-loop

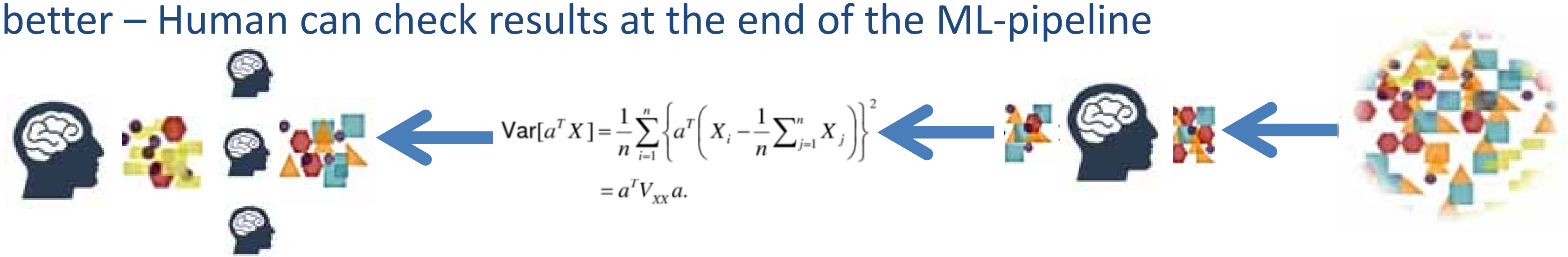
Multi-Agents-in-the-loop

Nature-inspired-agents-in-the-loop

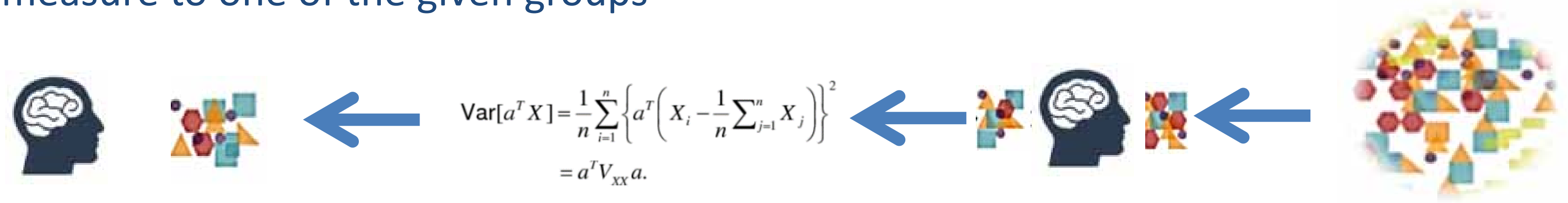
A) Unsupervised ML: Algorithm is applied on the raw data and learns fully automatic – Human can check results at the end of the ML-pipeline



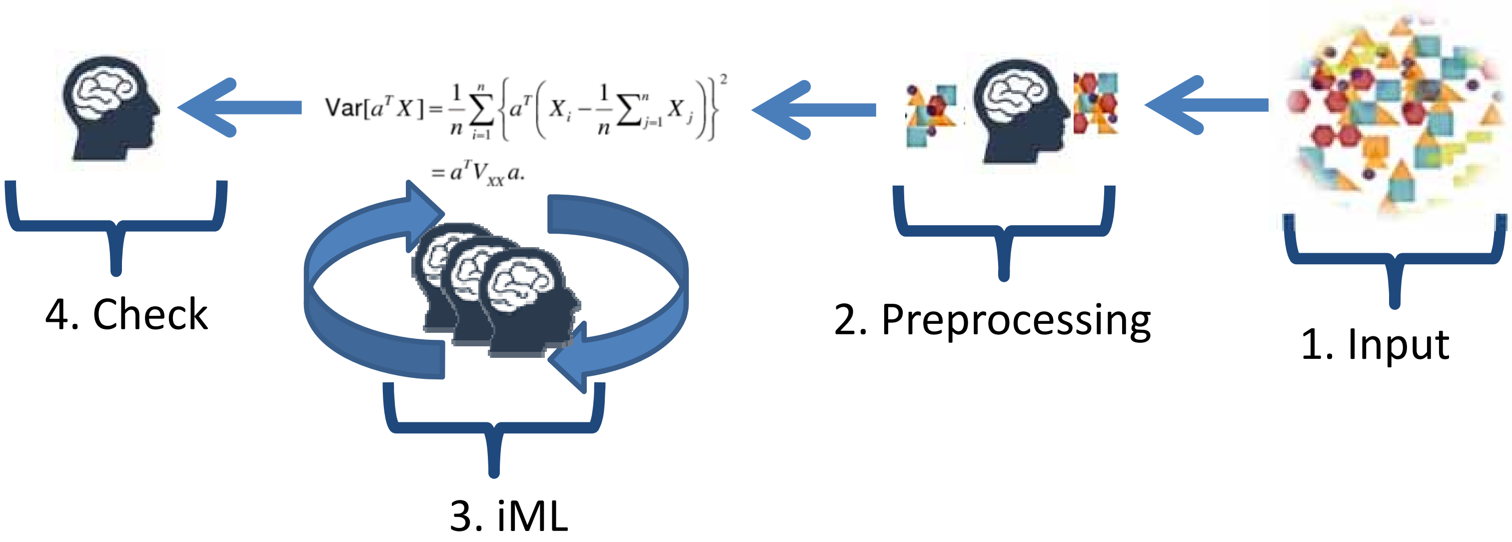
B) Supervised ML: Humans are providing the labels for the training data and/or select features to feed the algorithm to learn – the more samples the better – Human can check results at the end of the ML-pipeline



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? Brain Informatics (BRIN), 3, (2), 119-131, doi:10.1007/s40708-016-0042-6.

Three examples for the usefulness of the iML approach

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

Example 1: k-Anonymization of Medical Data

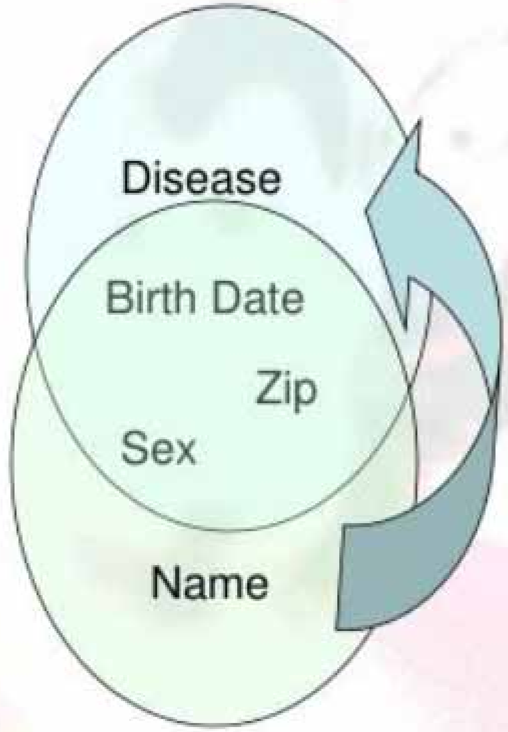
87 % of the population in the USA can be uniquely re-identified by Zip-Code, Gender and date of birth

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

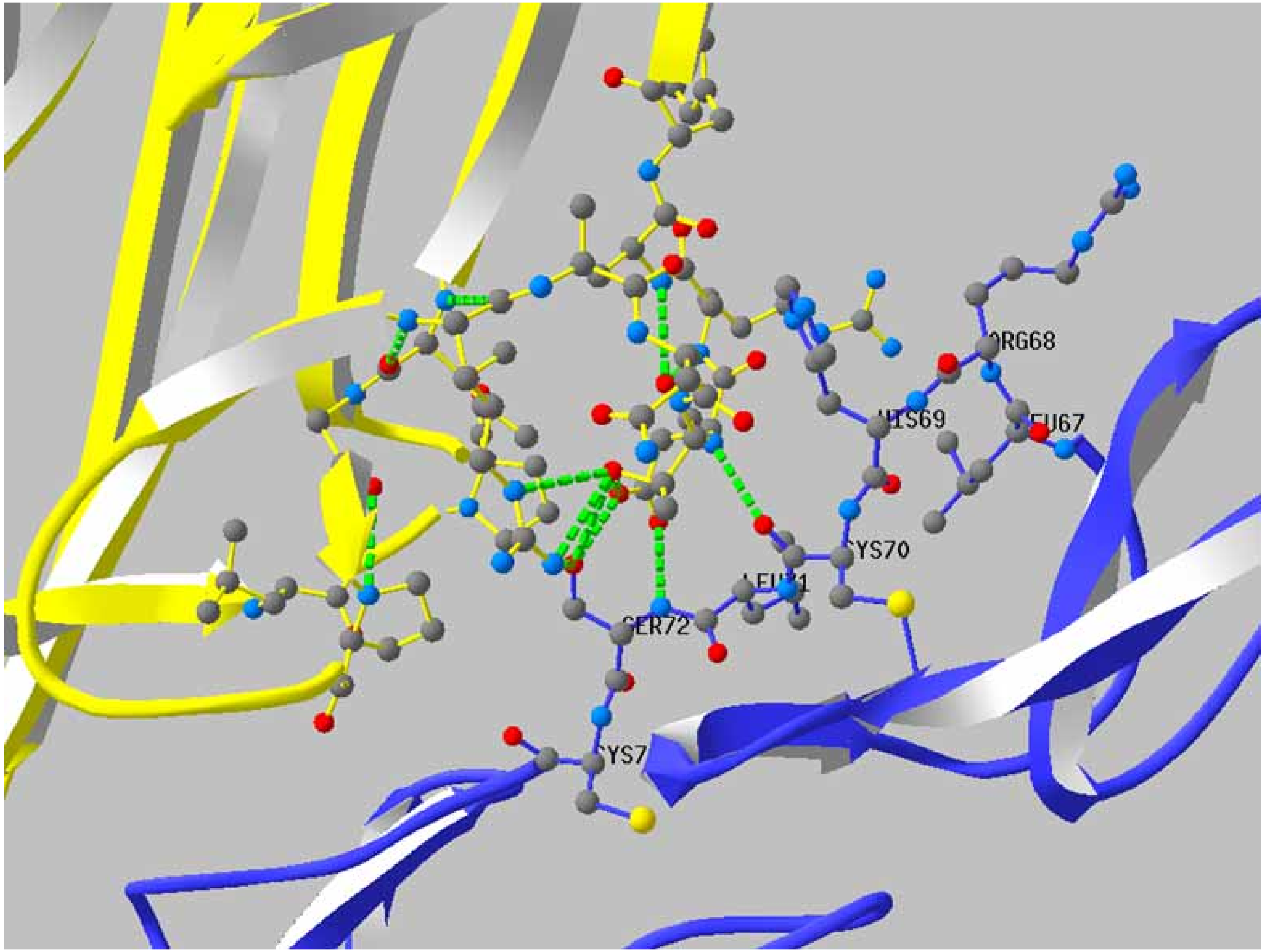
Name	Birthdate	Sex	Zipcode
Andre	1/21/76	Male	53715
Beth	1/10/81	Female	55410
Carol	10/1/44	Female	90210
Dan	2/21/84	Male	02174
Ellen	4/19/72	Female	02237



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 …

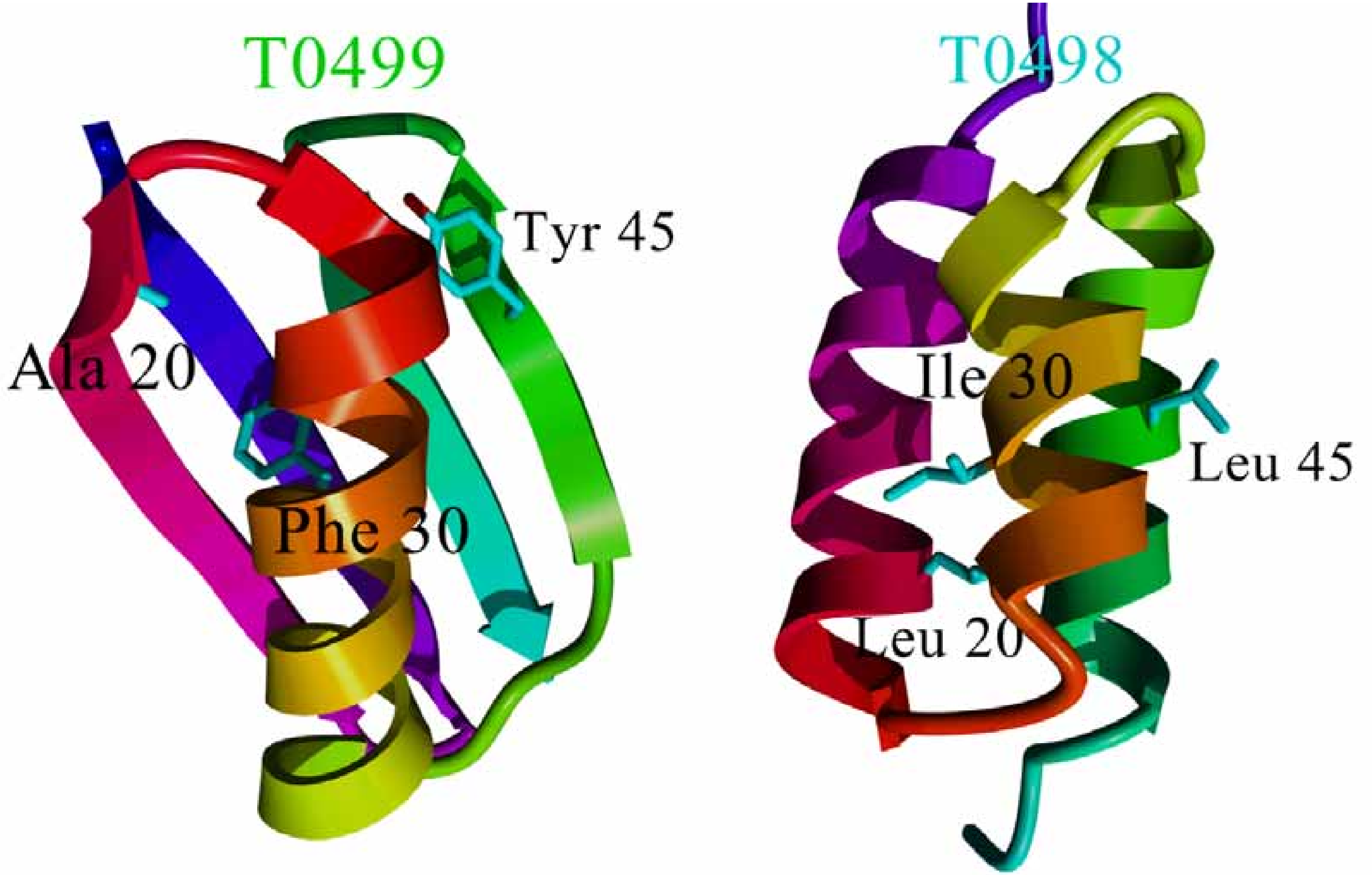
A4



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.*

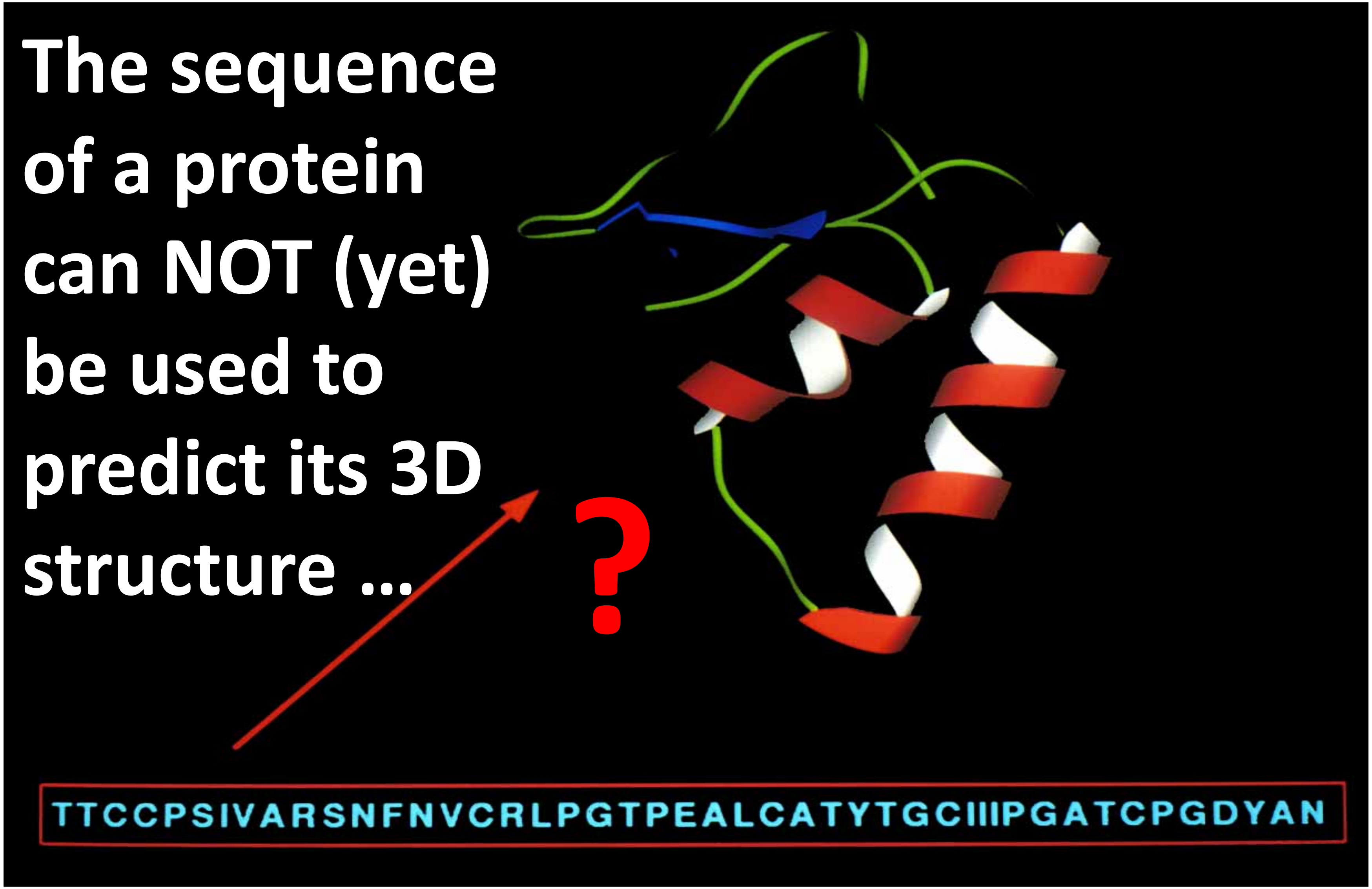
Example 2 Protein Folding: ὁμολογέω (homologeo)

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.



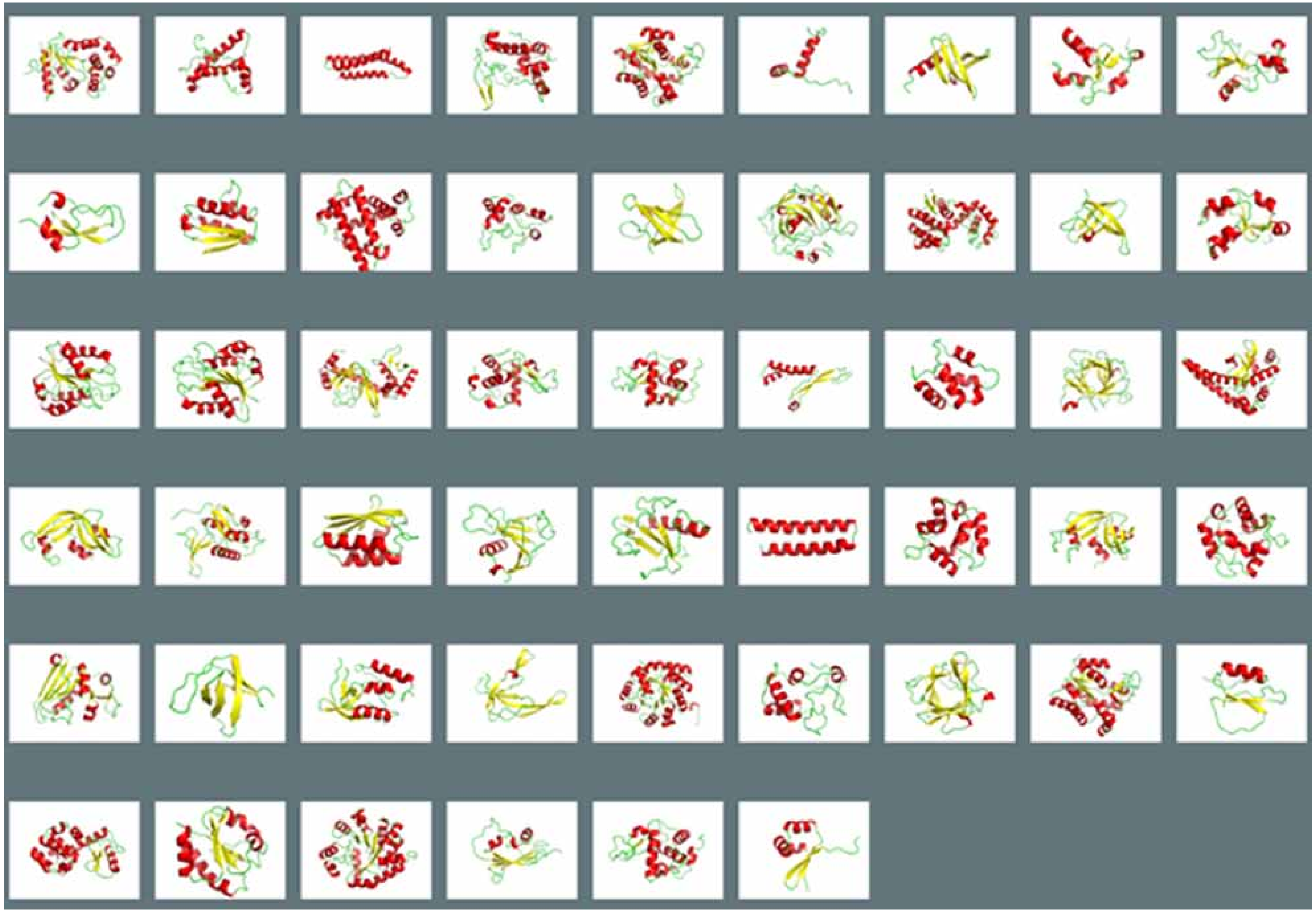

```
T0499   TTYKLILNLKQAKEEAIKEAVDAGTAEKYFKLIANAKTVEGVWTYKDEIKTFTVTE
        ||||||||||||||||||||X|||||||||X||||||||||||||X||||||||||
T0498   TTYKLILNLKQAKEEAIKELVDAGTAEKYIKLIANAKTVEGVWTLKDEIKTFTVTE
```

Protein prediction is an old – still unsolved - problem



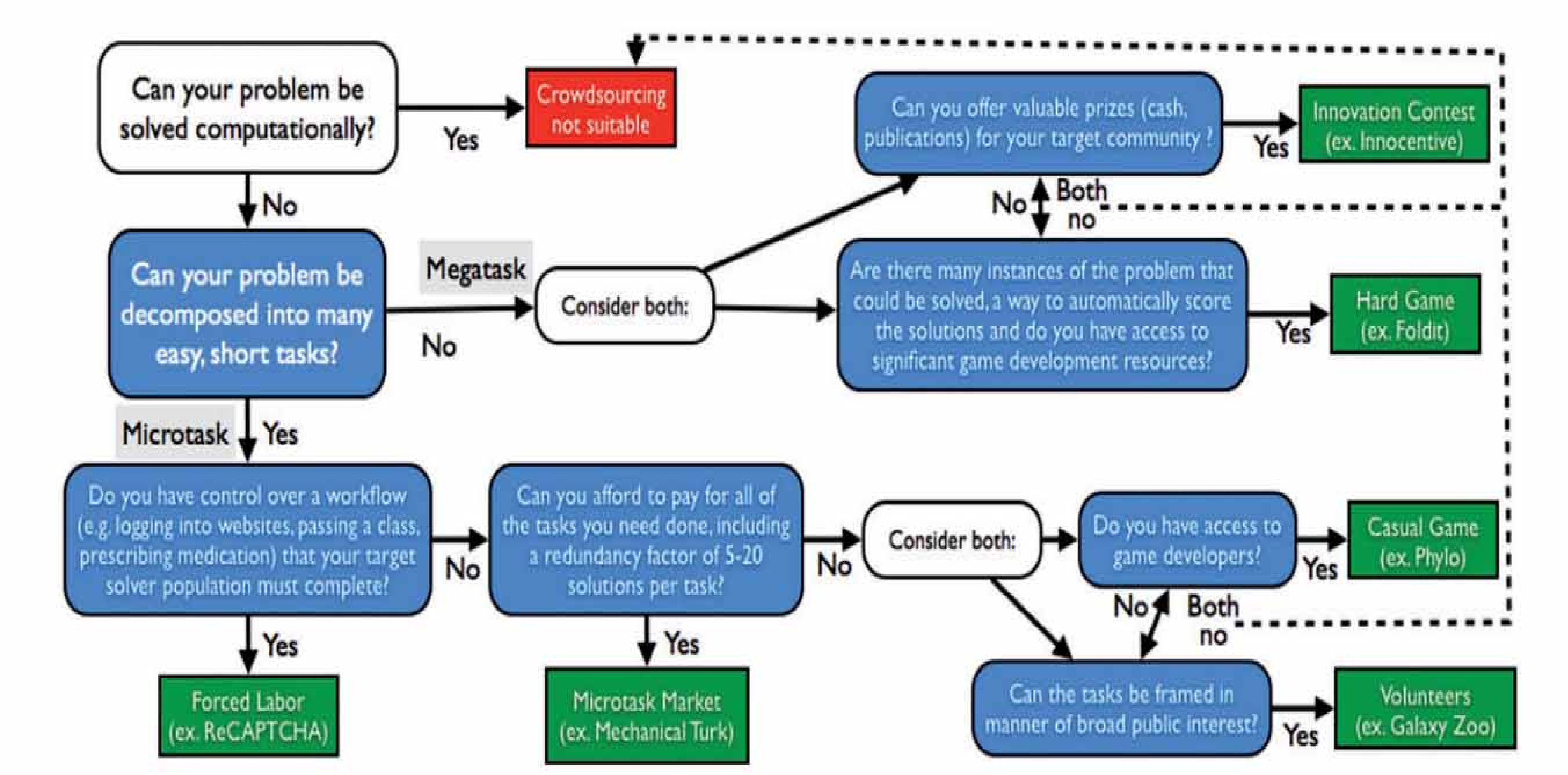
Anfinsen, C. B. **1973.** Principles that Govern the Folding of Protein Chains. Science, 181, (4096), 223-230.

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

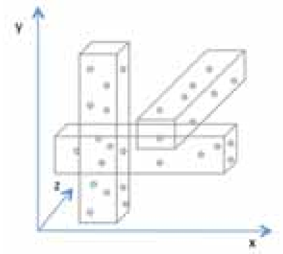


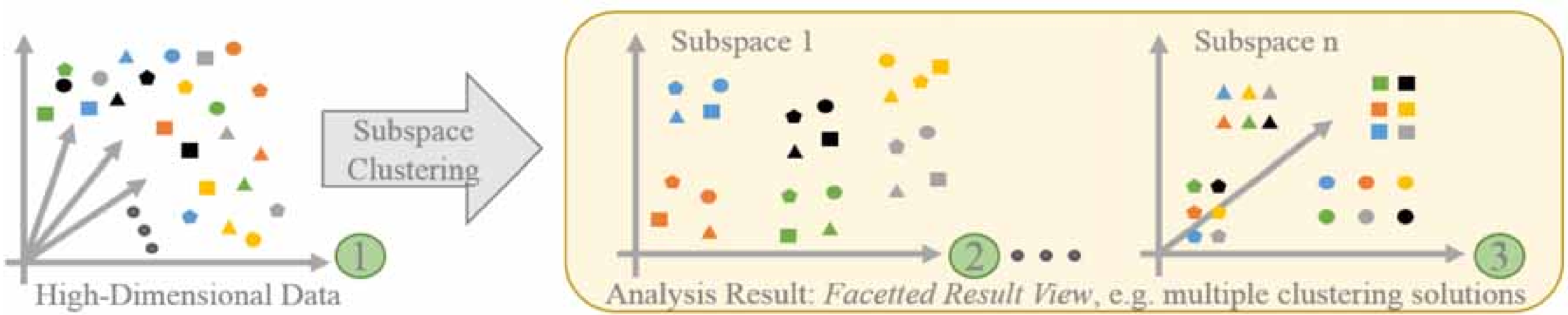
Good, B. M. & Su, A. I. 2013. Crowdsourcing for bioinformatics. *Bioinformatics,* 29, (16), 1925-1933.

Example 3: Subspace Clustering

A5

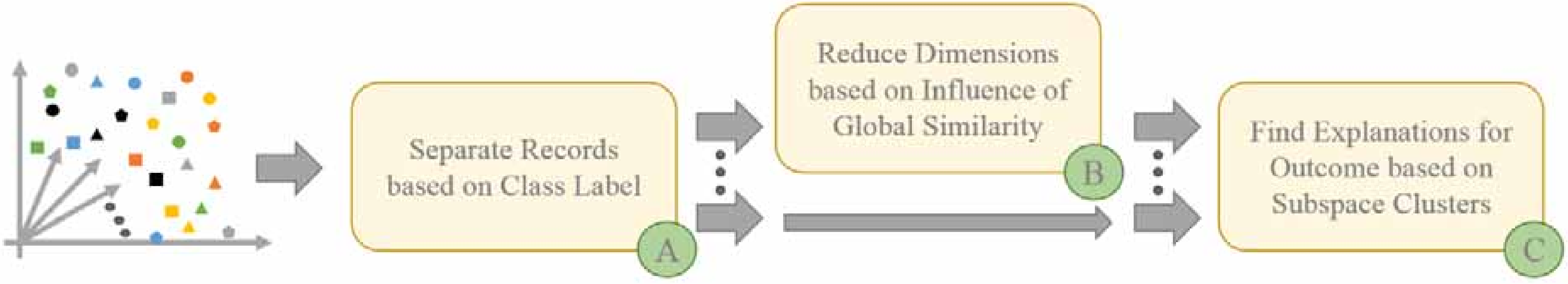
- **Patterns may be found in subspaces (dimension combinations)**
- **Clustering and subset selection: Non-convex & NP-hard**
- **Real data are often noisy and corrupted**
- **Little prior knowledge about low-dim structures**
- **Data points in different subgroups can be very close**





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.

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)

The usefulness of a Human kernel …

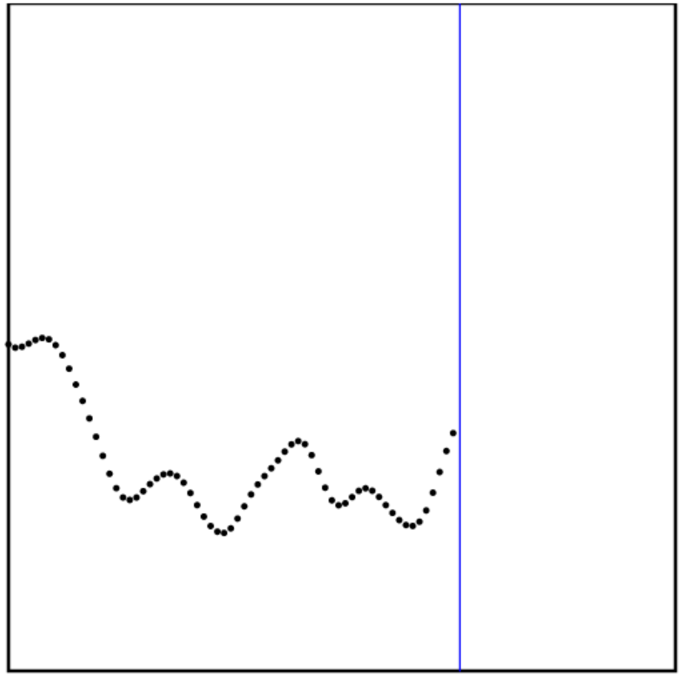
A2

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.

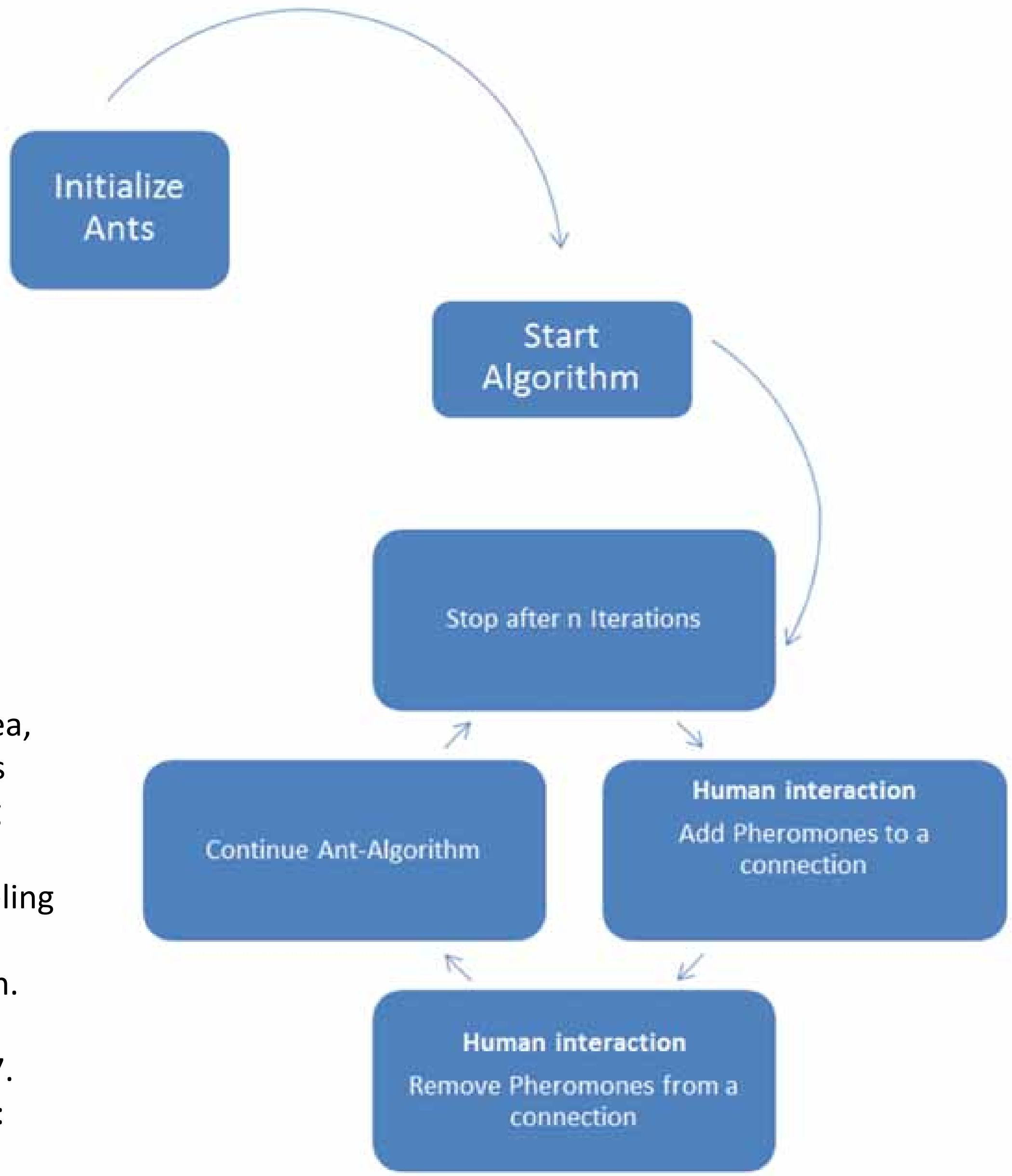
Once you have selected a position along the line, **hit the 's' key to submit the point**.



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

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

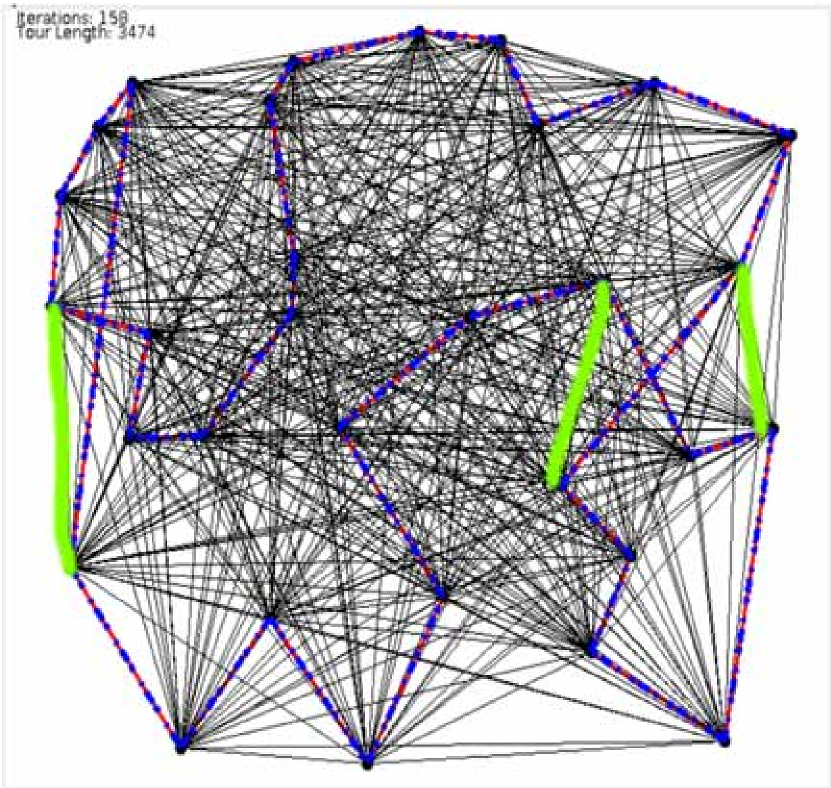
Experimental Proof for the Human in the Loop



Holzinger, A., Plass, M., Holzinger, K., Crisan, G., Pintea, C. & Palade, V. 2016. Towards interactive Machine Learning (iML): Applying Ant Colony Algorithms to solve the Traveling Salesman Problem with the Human-in-the-Loop approach. Springer Lecture Notes in Computer Science LNCS 9817. Heidelberg, Berlin, New York: Springer, pp. in print.

From Black-Box to the Glass-Box Approach

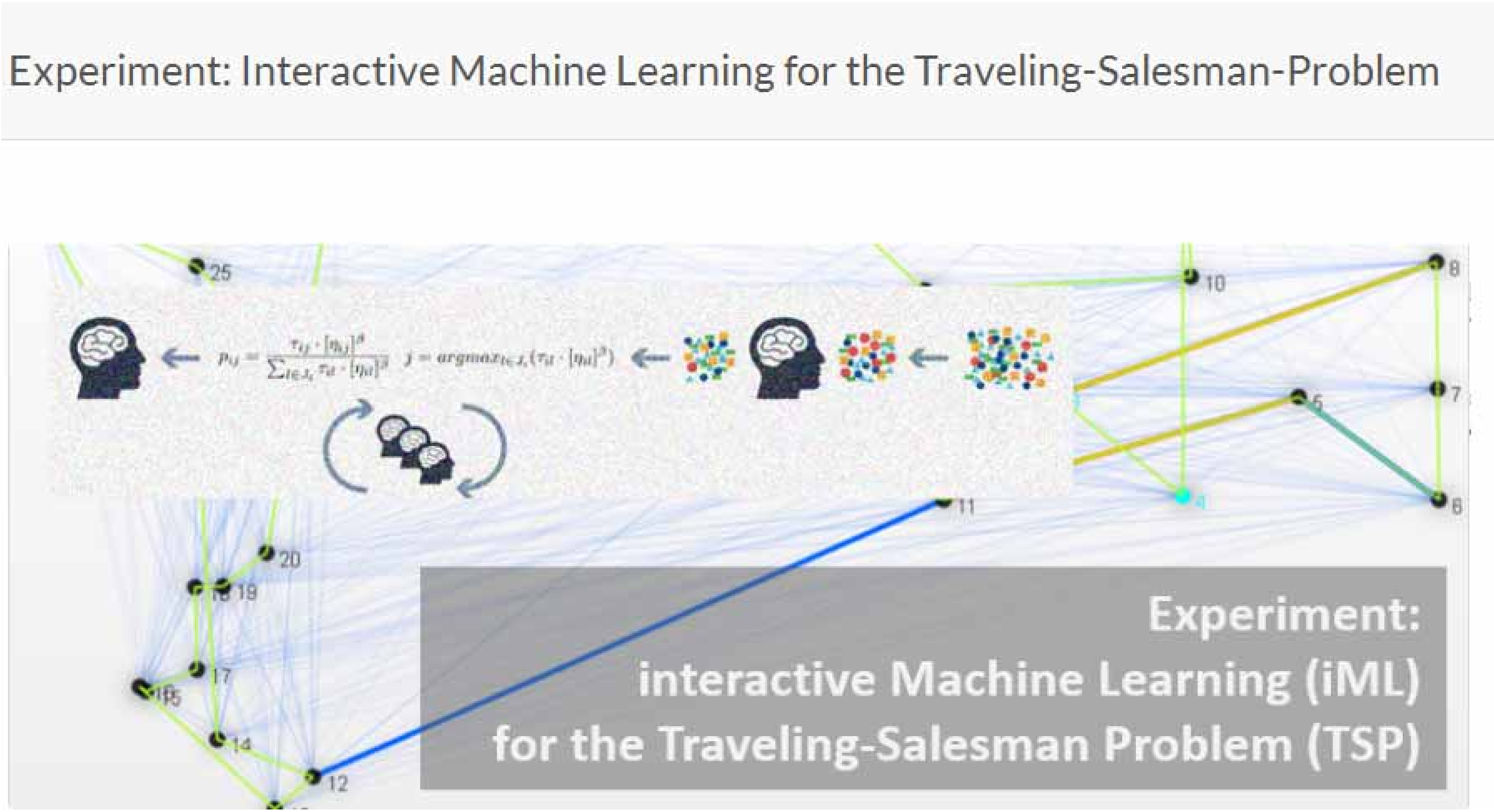
A1



Algorithm 2: Ant Colony Algorithm iML

Input : ProblemSize, $Population_{size}$, m, ρ, β, σ, $q0$
Output: P_{best}
$P_{best} \leftarrow$ CreateHeuristicSolution(ProblemSize);
$Pbest_{cost} \leftarrow \mathrm{Cost}(S_h)$;
$Pheromone_{init} \leftarrow \frac{1.0}{ProblemSize \times Pbest_{const}}$;
$Pheromone \leftarrow$ InitializePheromone($Pheromone_{init}$);
while $\neg$*StopCondition()* **do**
  for $i = 1$ to m **do**
    $S_i \leftarrow$ ConstructSolution(Pheromone, ProblemSize, β, $q0$);
    $Si_{cost} \leftarrow \mathrm{Cost}(S_i)$;
    if $Si_{cost} \leq \mathrm{Pbest}_{cost}$ **then**
      $Pbest_{cost} \leftarrow Si_{cost}$;
      $P_{best} \leftarrow S_i$;
    end
    LocalUpdateAndDecayPheromone(Pheromone, S_i, Si_{cost}, σ);
  end
  GlobalUpdateAndDecayPheromone(Pheromone, P_{best}, $Pbest_{cost}$, ρ);
  while *isUserInteraction()* **do**
    GlobalAddAndRemovePheromone(Pheromone, P_{best}, $Pbest_{cost}$, ρ);
  end
end
return P_{best};

- Please take part in this online Experiment:
- http://hci-kdd.org/projects/iml-experiment



08 Conclusion

- 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 $\mathbb{R}^N \to \mathbb{R}^2$
- 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 ML specifically.

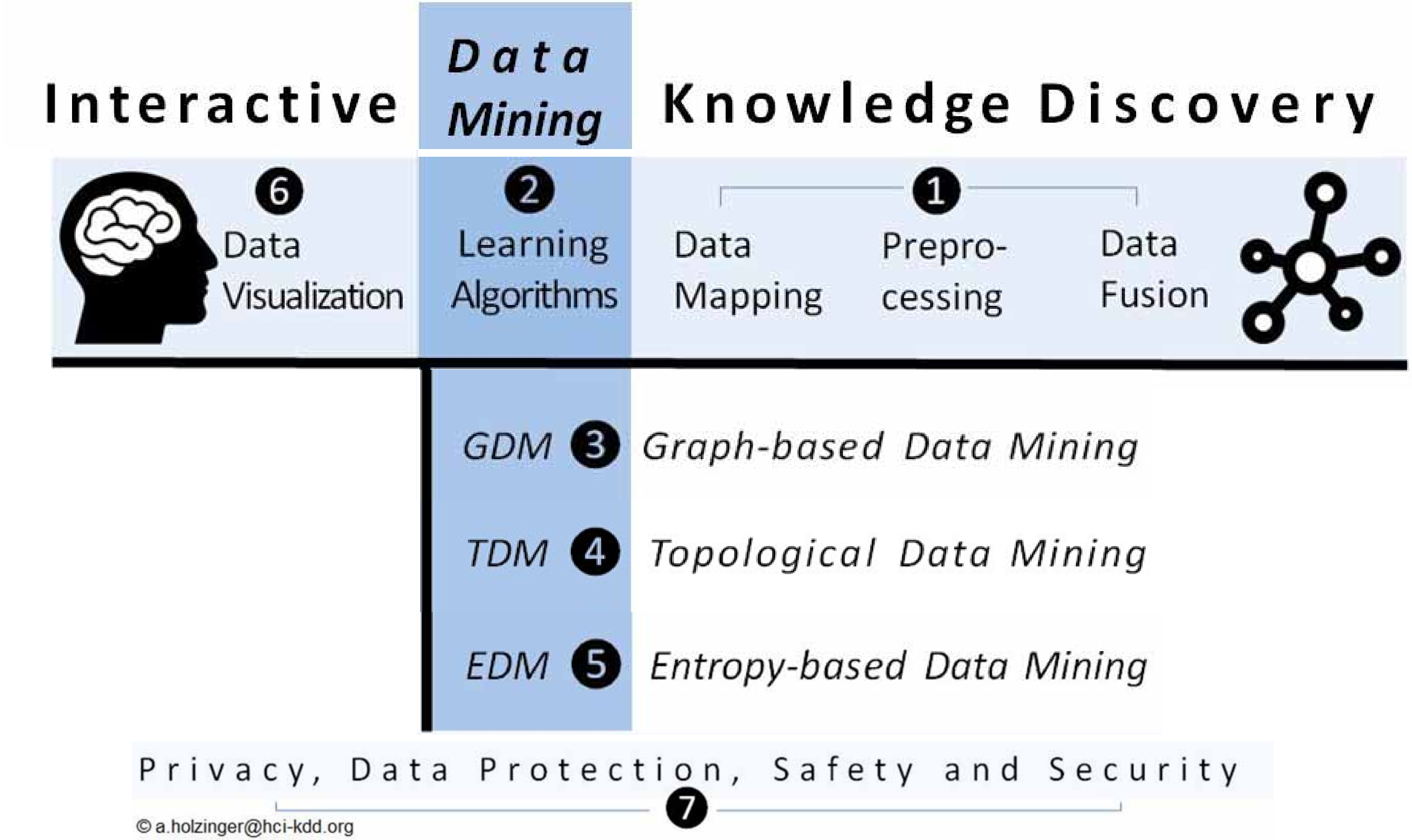
Multi-Agent-Hybrid Systems (MAHS)

To include swarm-intelligence and crowdsourcing and making use of discrete models – avoiding to seek perfect solutions – better have a good solution < 5 min.

concerted effort

international without boundaries …

ML-Algorithms are key but needs also 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.

Thank you!

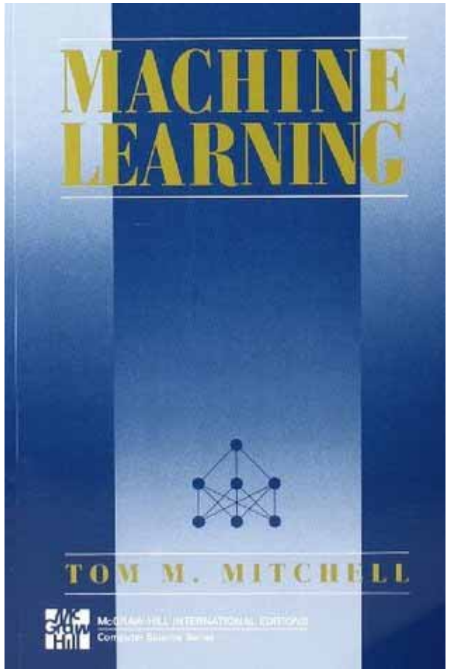
09 Questions

Sample Questions

- What is Active Learning?
- Where are the advantages of AL?
- Describe a few scenarios for AL?
- How does the robot scientist by King et al (2004) work?
- What does “Probable Approximate Correct” mean?
- What is the basic assumption of PL?
- What is the core essence of the “programming by feedback” approach?
- What could be huge disadvantages with the “human-in-the-loop”?
- What is a utility function?
- Why is multi-task learning of extreme importance for future research?
- When are humans better in TL ?
- Explain the 3 types of TL and the 4 TL approaches!
- What is the main idea of inductive TL?

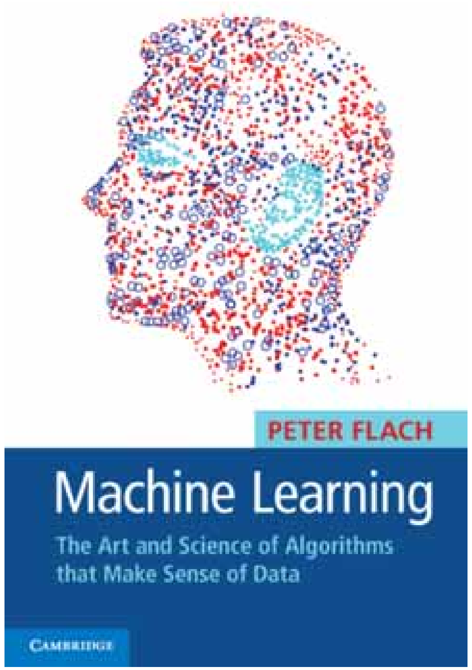
10 Appendix

Recommended Books 1 (General Introductions)



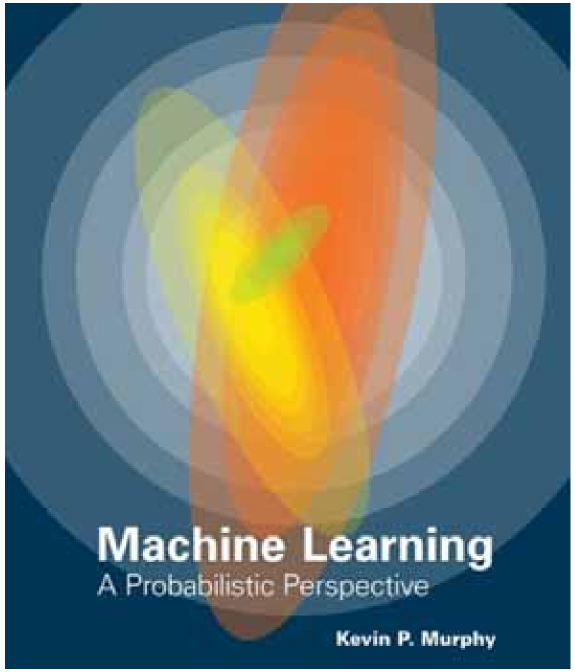
Mitchell, T. M. 1997. Machine learning, New York, McGraw Hill.

http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html



Flach, P. 2012. Machine learning: the art and science of algorithms that make sense of data, Cambridge University Press.

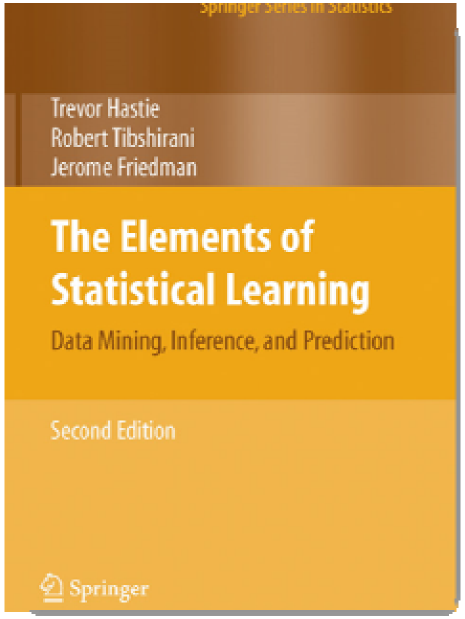
https://www.cs.bris.ac.uk/~flach/mlbook/



Murphy, K. P. 2012. Machine learning: a probabilistic perspective, Cambridge (MA), MIT press

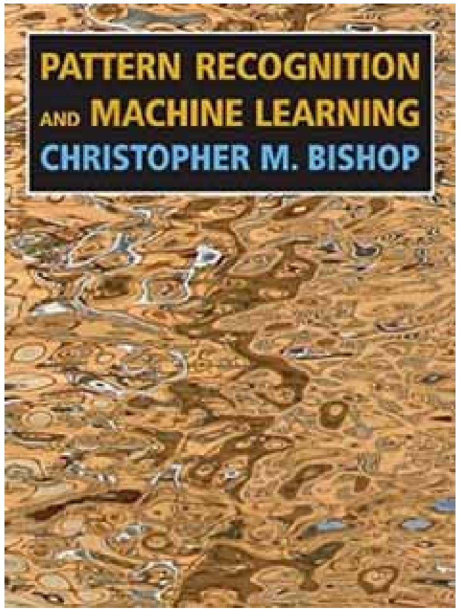
http://www.cs.ubc.ca/~murphyk/MLbook/index.html.

Recommended Books 2 (Advanced Introduction)



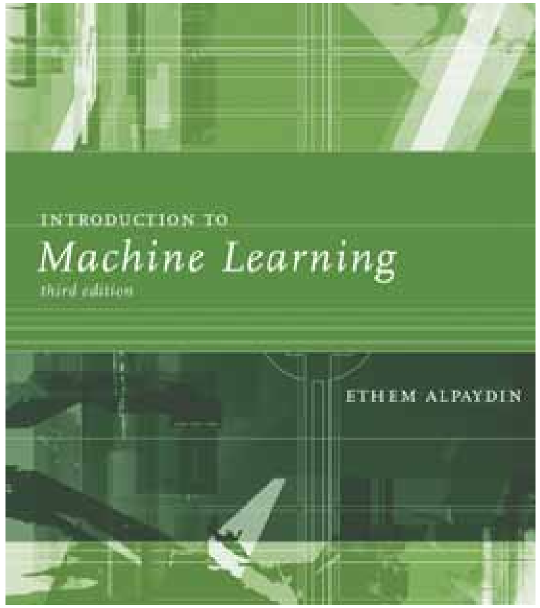
Hastie, T., Tibshirani, R. & Friedman, J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, New York, Springer

http://statweb.stanford.edu/~tibs/ElemStatLearn/



Bishop, C. M. 2007. Pattern Recognition and Machine Learning, Heidelberg, Springer.

https://www.microsoft.com/en-us/research/people/cmbishop/



Alpaydin, E. 2014. Introduction to machine learning. MIT press.

http://www.cmpe.boun.edu.tr/~ethem/i2ml3e/

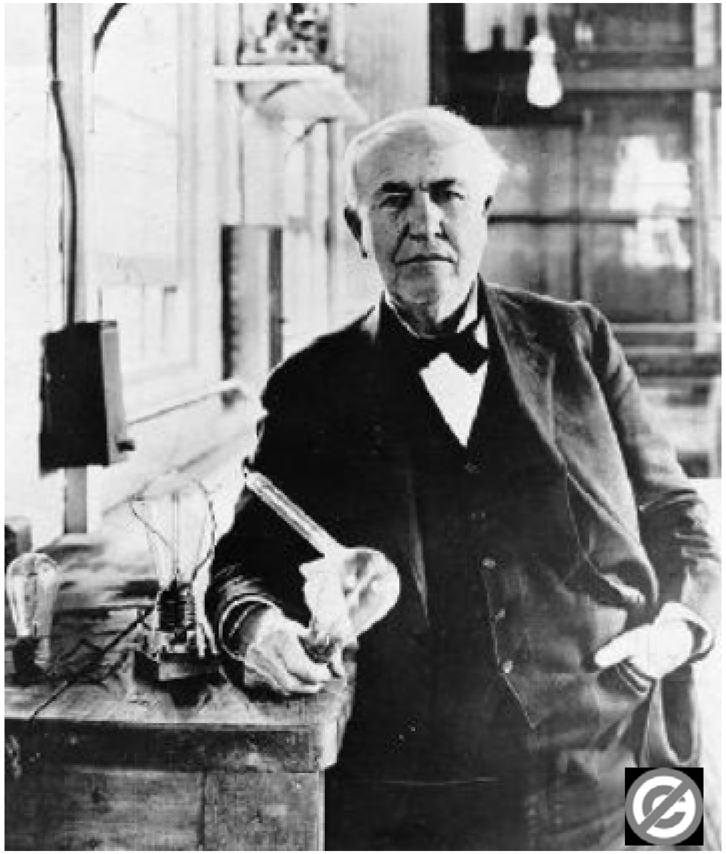
Solutions to the Quiz Questions

- ad 1) yes; e.g. Naïve Bayes classifier for spam learning; K-nearest neighbor with training data; collaborative filtering, clustering, etc. etc.
- ad 2) yes; e.g. SVD/PCA, ensemble methods, regression, clustering etc.;
- ad 3) yes; e.g. SVM, reinforcement learning, regression etc.;
- ad 4) yes; e.g. Google-Page-Rank Algorithm;
- ad 5) yes; reverse image search, e.g. feature detection algorithms for matching deformations (SIFT, PCA-SIFT and SURF) etc. Scale-invariant feature transform (SIFT) is an algorithm in computer vision to detect and describe local features in images
- ad 6) yes, recommender systems, collaborative filtering.
- ad 7) yes, fraud detection, e.g. profiling methods
- ad 8) yes, sequence motif, pattern of nucleotides in a DNA sequence or amino acids in a protein; structural motif, a pattern in a protein structure formed by the spatial arrangement of amino acids;
- ad 9) yes, structural alignment, establishing homology between polymer structures based on their shape and 3D-conformation. (protein tertiary structures);
- ad 10) yes, learning from DNA data
- Ad 11) yes, feature extraction from cancerous pathological samples
- Ad 12) yes, machine learning approach to predict protein–ligand binding affinity with applications to molecular docking in Bioinformatics

Popular ML research data sets

- University California at Irvine Machine Learning Repository: http://archive.ics.uci.edu/ml/ - some examples include:
- Wisconsin Breast Cancer data set: Given the results of a diagnostic test on breast tissue, predict whether the mass is a tumor or not
- Heart Disease data set: Given the results of various diagnostic tests on a patient, predict the risk of heart disease of the patient.
- Human activity recognition data set: From smart phone movement data predict the type of activity performed by the person holding the smart phone.
- More information see:
- http://hci-kdd.org/open-data-sets

The doctor of the future



The doctor of the future will give no medicine, but will instruct his patient in the care of the human frame, in diet and in the cause and prevention of disease.
Thomas Alva Edison (1847 – 1931)

Duerr-Specht, M., Goebel, R. & Holzinger, A. 2015. Medicine and Health Care as a Data Problem: Will Computers become better medical doctors? In: Lecture Notes in Computer Science LNCS 8700. Heidelberg, Berlin, New York: Springer, pp. 21-40, doi:10.1007/978-3-319-16226-3_2