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

LV 706.046 Summer Term 2019
Monday, March, 11

Introduction and Overview
From measuring Usability to measuring Causability

Assoc.Prof. Dr. Andreas Holzinger

Holzinger-Group, HCI-KDD, Institute for Medical Informatics/Statistics
Medical University Graz, Austria
&
Institute of interactive Systems & Data Science
Graz University of Technology, Austria

a.holzinger@tugraz.at
https://hci-kdd.org/intelligent-user-interfaces-2019

This is our worldwide problem …

Image by Randall Munroe https://xkcd.com

This lecture is only the overview and motivation part ☺

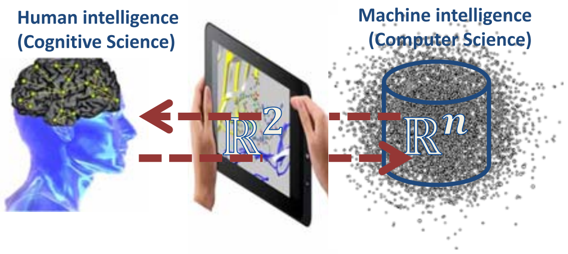
- 01 The HCI-KDD approach: integrative ML
- 02 Application Area: Health
- 03 Probabilistic Learning
- 04 aML
- 05 iML
- 06 Causality and Causability
- 07 Measuring Causality?
- Our Goal for this semester: design develop & test a System Causability Scale

01 What is the HCI-KDD approach?

- ML is a very practical field – algorithm development is at the core – however, successful ML needs a concerted effort of various topics …

Andreas Holzinger 2017. Introduction to Machine Learning and Knowledge Extraction (MAKE). *Machine Learning and Knowledge Extraction*, 1, (1), 1-20, doi:10.3390/make1010001.

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



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)

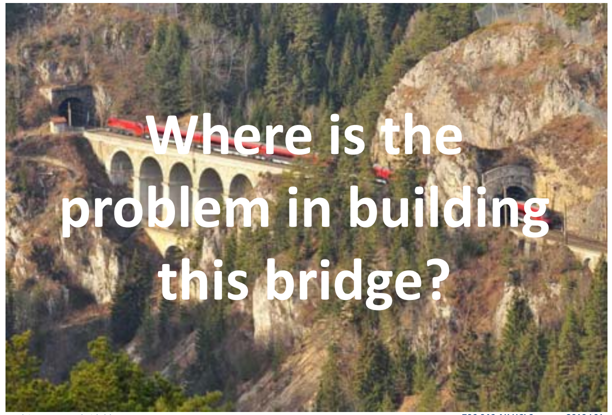
To reach a level of usable intelligence we need to …

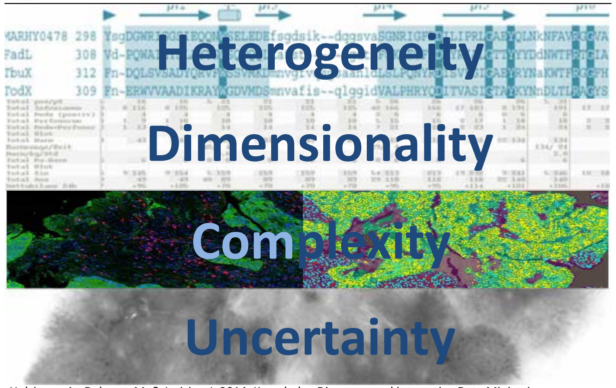
- 1) **learn** from prior data
- 2) **extract** knowledge
- 2) **generalize,** i.e. guessing where a probability mass function concentrates
- 4) fight the curse of **dimensionality**
- 5) **disentangle** underlying explanatory factors of data, i.e.
- 6) **understand** the data in the **context** of an application domain

Why is this application area complex ?

Our central hypothesis: Information may bridge this gap

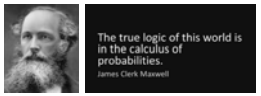
Holzinger, A. & Simonic, K.-M. (eds.) 2011. *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058*, Heidelberg, Berlin, New York: Springer.



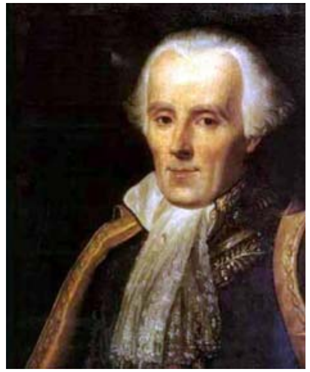


Holzinger, A., Dehmer, M. & Jurisica, I. 2014. Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. BMC Bioinformatics, 15, (S6), I1.

03 Probabilistic Learning



Probability theory is nothing but common sense reduced to calculation …



Pierre Simon de Laplace (1749-1827)

- 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
- Gauss generalized those ideas

06

What is the simplest mathematical operation for us?

$$p(x) = \sum_x (p(x,y)) \quad (1)$$

How do we call repeated adding?

$$p(x,y) = p(y|x) * p(y) \quad (2)$$

Laplace (1773) showed that we can write:

$$p(x,y) * p(y) = p(y|x) * p(x) \quad (3)$$

Now we introduce a third, more complicated operation:

$$\frac{p(x,y) * p(y)}{p(y)} = \frac{p(y|x) * p(x)}{p(y)} \quad (4)$$

We can reduce this fraction by $p(y)$ and we receive what is called Bayes rule:

$$p(x,y) = \frac{p(y|x) * p(x)}{p(y)} \qquad p(h|d) = \frac{p(d|h)p(h)}{p(d)} \quad (5)$$

07

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

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

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

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

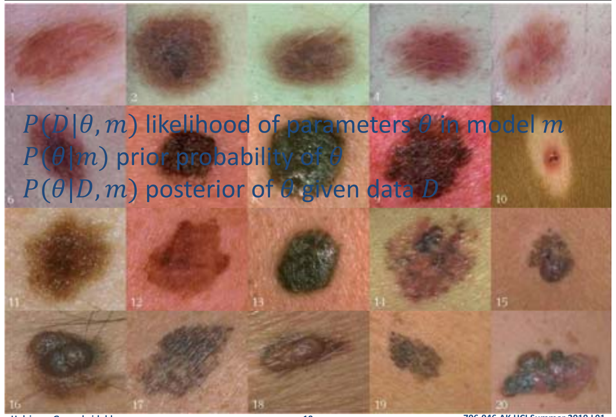
$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i,y_j)p(x_i)}$$

$$P(\text{hypothesis}|\text{data}) = \frac{P(\text{hypothesis})P(\text{data}|\text{hypothesis})}{\sum_h P(h)P(\text{data}|h)}$$

$$P(\theta|\mathcal{D},m) = \frac{P(\mathcal{D}|\theta,m)P(\theta|m)}{P(\mathcal{D}|m)}$$

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.

We need …



Bayesian Learning from data

05

$\mathcal{D} = x_{1:n} = \{x_1, x_2, \ldots, x_n\}$ $p(\mathcal{D}|\theta)$

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

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

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

Learning and Inference

04

d … data
h … hypotheses

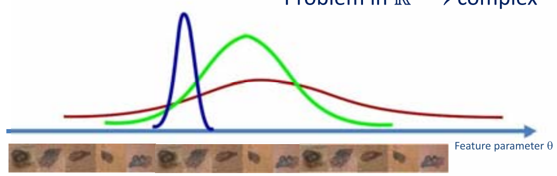
$\mathcal{H}$ … $\{H_1, H_2, \ldots, H_n\}$ $\forall\, h, d$ …

Likelihood Prior Probability

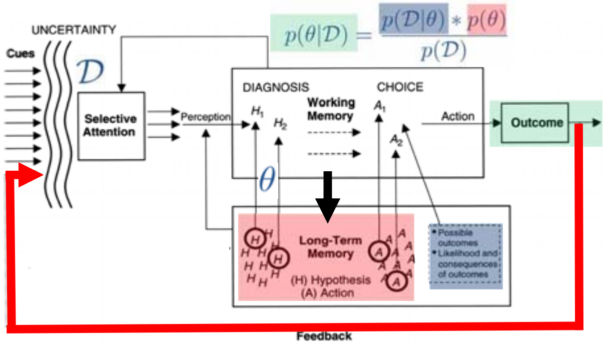
$$p(h|d) = \frac{p(d|h) * p(h)}{\sum_{h \in H} p(d|h')\, p(h')}$$

Posterior Probability

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



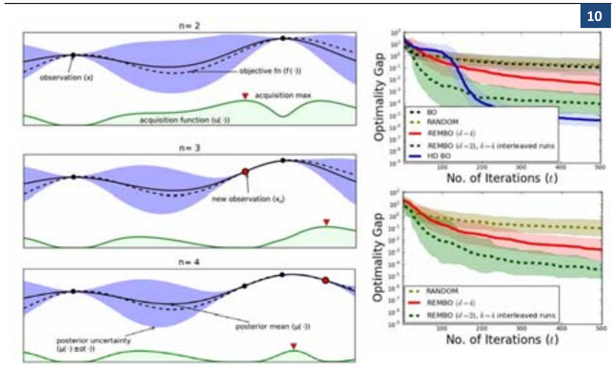
Connection to Cognitive Science: Decision Making



Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill, modified by Holzinger, A.*

Scaling to high-dimensions is the holy grail in ML

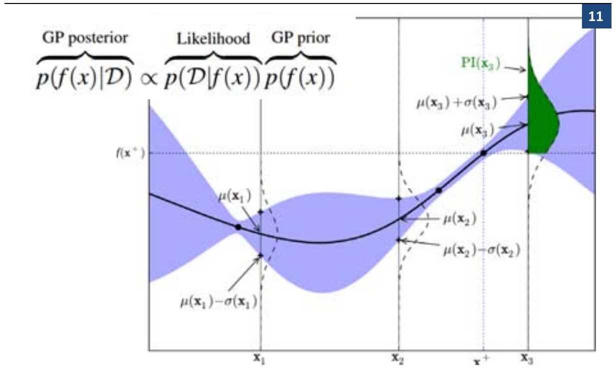
10



Wang, Z., Hutter, F., Zoghi, M., Matheson, D. & De Feitas, N. 2016. Bayesian optimization in a billion dimensions via random embeddings. Journal of Artificial Intelligence Research, 55, 361-387, doi:10.1613/jair.4806.

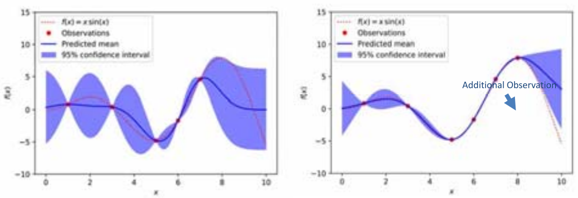
GP = distribution, observations occur in a cont. domain, e.g. t or space

11



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.

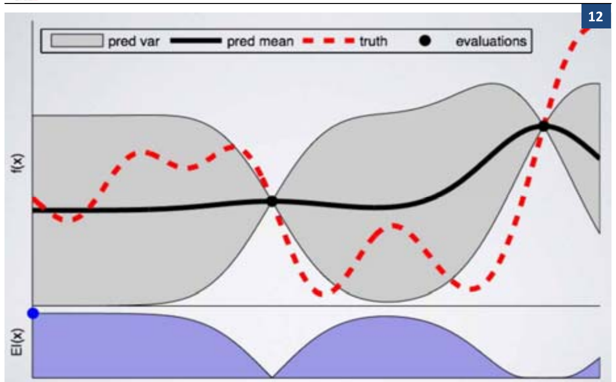
Goal: Reduce uncertainty by learning representations



$$\mathbb{E}[f] = \int p(x) f(x)\, dx \qquad \mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$

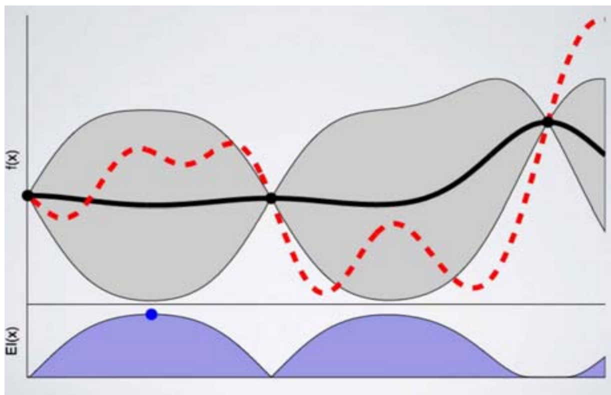
Holzinger, A. 2017. Introduction to Machine Learning and Knowledge Extraction (MAKE). Machine Learning and Knowledge Extraction, 1, (1), 1-20, doi:10.3390/make1010001.

12

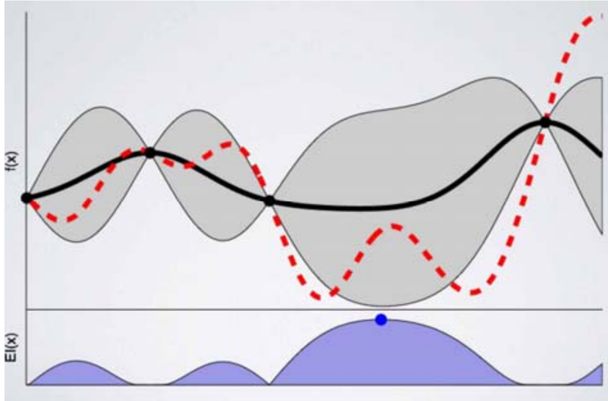


Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 2012. 2951-2959.

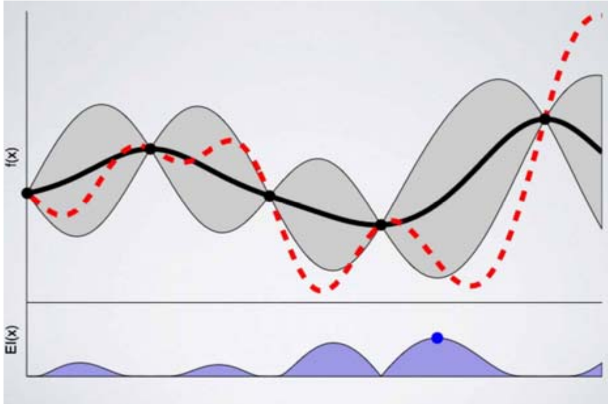
Bayesian Optimization 1



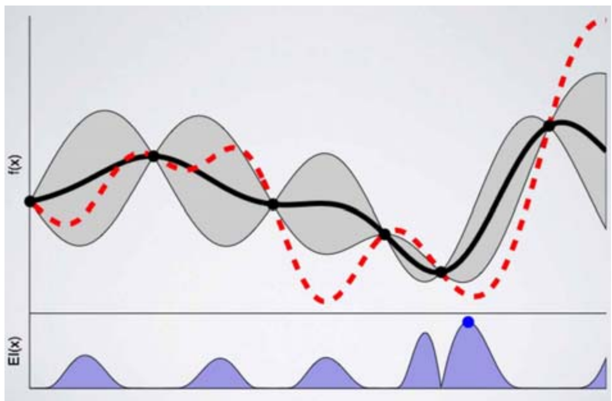
Bayesian Optimization 2



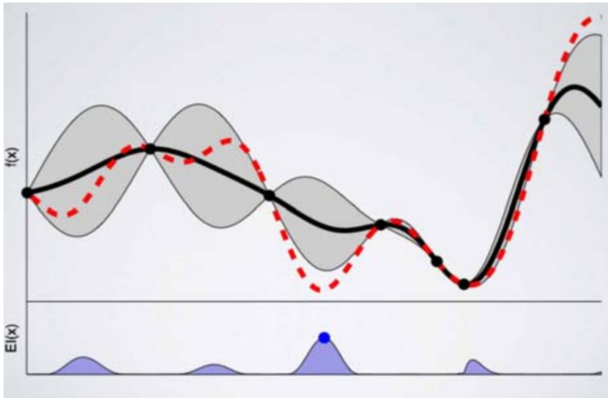
Bayesian Optimization 3



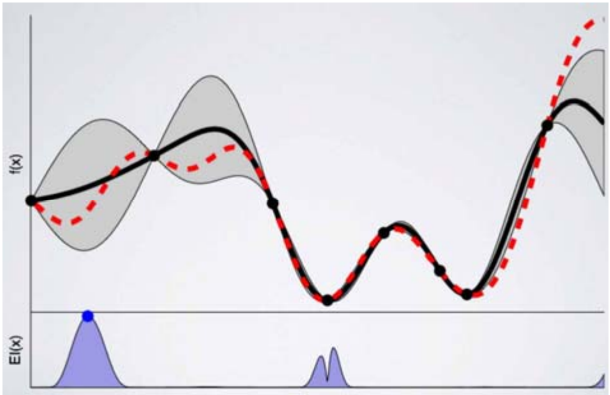
Bayesian Optimization 4



Bayesian Optimization 5

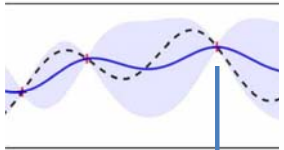


Bayesian Optimization 6



Fully automatic → Goal: Taking the human out of the loop

13

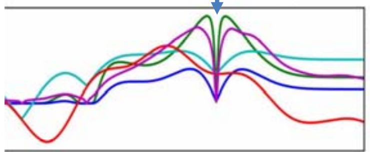


Algorithm 1 Bayesian optimization

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**



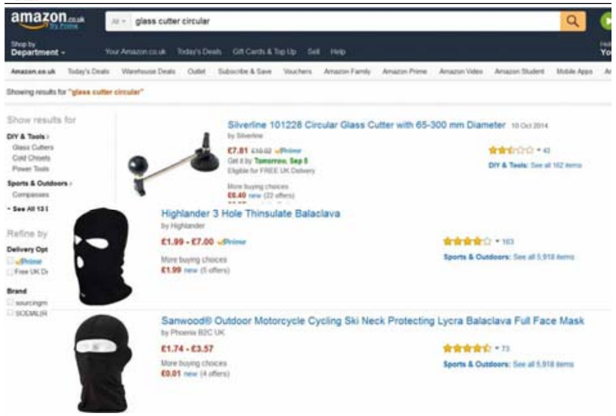
PI Probability of Improvement
EI Expected Improvement
UCB Upper Confidence Bound
TS Thompson Sampling
PES Predictive Entropy Search

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.

04 aML

Best practice examples of aML …

Recommender Systems



Fully automatic autonomous vehicles ("Google car")

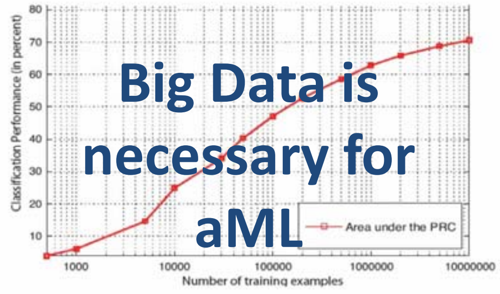
Guizzo, E. 2011. How google's self-driving car works. IEEE Spectrum Online, 10, 18.

Deep Learning Context recognition state-of-the-art

https://cs.stanford.edu/people/karpathy/deepimagesent/

Andrej Karpathy & Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 3128-3137.

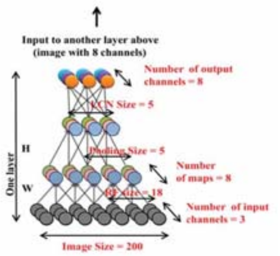
Big Data is necessary for aML !



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.

10 million 200 x 200 px images downloaded from Web



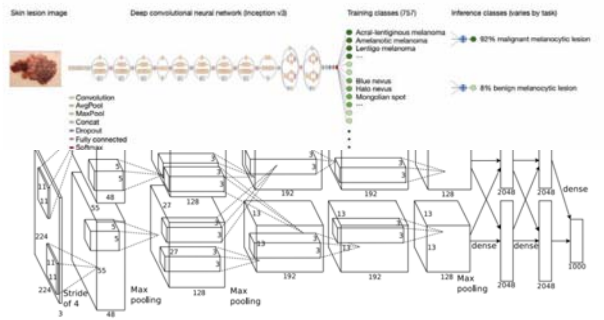
$$x^* = \arg\min_x f(x; W, H), \text{ subject to } ||x||_2 = 1.$$

Le, Q. V., Ranzato, M. A., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J. & Ng, A. Y. 2011. Building high-level features using large scale unsupervised learning. arXiv preprint arXiv:1112.6209.

Le, Q. V. 2013. Building high-level features using large scale unsupervised learning. *IEEE Intl. Conference on Acoustics, Speech and Signal Processing ICASSP.* IEEE. 8595-8598, doi:10.1109/ICASSP.2013.6639343.

Deep Convolutional Neural Network Pipeline

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. & Thrun, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542, (7639), 115-118, doi:10.1038/nature21056.

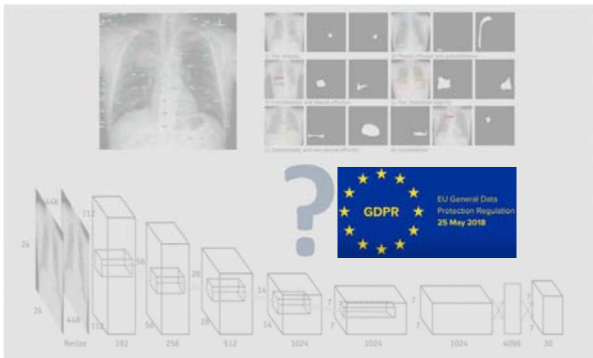


Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q., eds. Advances in neural information processing systems (NIPS 2012), 2012 Lake Tahoe. 1097-1105.

Houston, we have a problem …

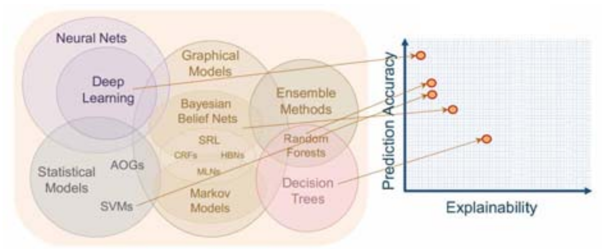
Source: NASA, Image is in the public domain

The need for explainability



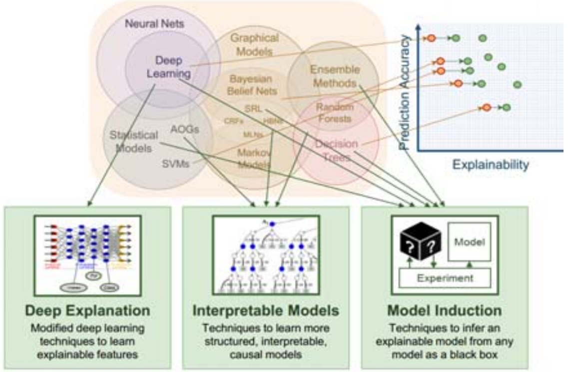
June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo & Namkug Kim 2017. Deep learning in medical imaging: general overview. Korean journal of radiology, 18, (4), 570-584, doi:10.3348/kjr.2017.18.4.570.

Interpretability vs. Performance



David Gunning 2016. Explainable artificial intelligence (XAI): Technical Report Defense Advanced Research Projects Agency DARPA-BAA-16-53, Arlington, USA, DARPA.

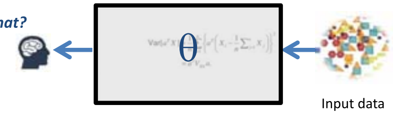
Making Models Interpretable, Explainable, Retraceable



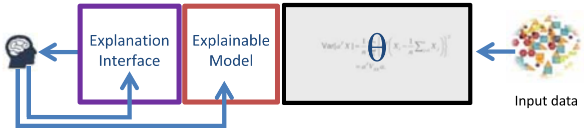
David Gunning 2016. Explainable artificial intelligence (XAI): Technical Report Defense Advanced Research Projects Agency DARPA-BAA-16-53, Arlington, USA, DARPA.

The challenge

Why did the algorithm do that?
Can I trust these results?
How can I correct an error?



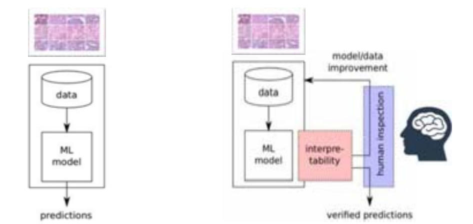
A possible solution



The domain expert can understand why …
The domain expert can learn and correct errors …
The domain expert can re-enact on demand …

05 iML

Own Work



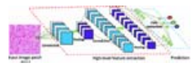
Generalization Error

Generalization Error + Human Experience

Andreas Holzinger 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Brain Informatics, 3, (2), 119-131, doi:10.1007/s40708-016-0042-6.

Why?

Verify that algorithms/classifiers work as expected
Wrong decisions can be costly and dangerous



Understanding the weaknesses and errors of the ML-Model - Detection of bias in both directions

Scientific interpretability, replicability, causality
The "why" is often more important than the prediction

Enable re-traceability, re-enactivity
Compliance to legislation "right for explanation",
retain human reliability, fosters trust and acceptance

Example: Discovery of causal relationships from data …

15

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

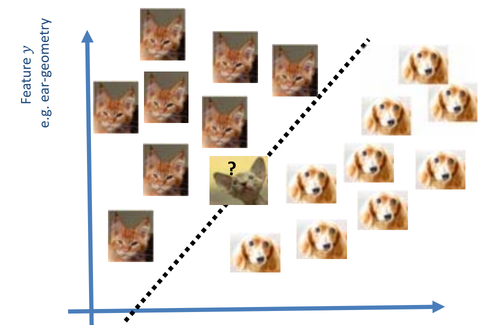
The grand question …

15b

- **"How do humans generalize from so few examples?"**
  - Learning relevant representations
  - Disentangling the explanatory factors
  - Finding the shared underlying explanatory factors, in particular between $P(x)$ and $P(Y|X)$, with a causal link between $Y \rightarrow X$

Bengio, Y., Courville, A. & Vincent, P. 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35, (8), 1798-1828, doi:10.1109/TPAMI.2013.50.

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.



Even Children can make inferences from little data …

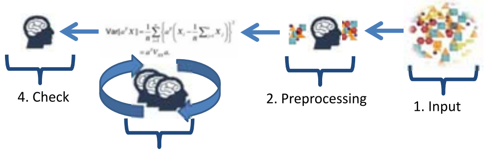
Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.1192788.

Example: Adversarial examples

See also: Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy 2014. Explaining and harnessing adversarial examples. arXiv:1412.6572, and see more examples: https://imgur.com/a/K4RWn

iML: bringing in human intuition

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 …



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.

http://hci-kdd.org/projects/iml-experiment

19

```
Input  : ProblemSize, m, β, ρ, σ, φ0
Output: Pbest
Pbest ← CreateHeuristicSolution(ProblemSize);
Pbest_cost ← Cost(Pbest);
Pheromone_init ← 1.0 / (ProblemSize × Pbest_cost);
Pheromone ← InitializePheromone(Pheromone_init);
while ¬StopCondition() do
    for i = 1 to m do
        S_i ← ConstructSolution(Pheromone, ProblemSize, β, φ0);
        Si_cost ← Cost(S_i);
        if Si_cost ≤ Pbest_cost then
            Pbest_cost ← Si_cost;
            Pbest ← S_i;
        end
        LocalUpdateAndDecayPheromone(Pheromone, S_i, Si_cost, ρ);
    end
    GlobalUpdateAndDecayPheromone(Pheromone, Pbest, Pbest_cost, ρ);
    while isUserInteraction() do
        GlobalAddAndRemovePheromone(Pheromone, Pbest, Pbest_cost, ρ);
    end
end
return P_best;
```

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. 81-95, doi:10.1007/978-3-319-45507-56.

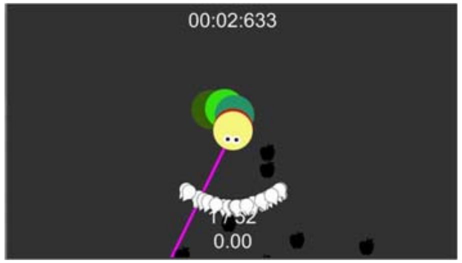
Influence the Pheromone Level by human intuition

$$p_{ij} = \frac{[\tau_{ij}]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{l \in J_i^k} [\tau(t)]^{\alpha} \cdot [\eta]^{\beta}}$$

- p_{ij} … **probability** of ants that they, at a particular node i, select the route from node $i \rightarrow j$ (**"heuristic desirability"**)
- $\alpha > 0$ and $\beta > 0$ … the **influence parameters** (α … history coefficient, β …heuristic coefficient) usually $\alpha \approx \beta \approx 2 < 5$
- τ_{ij} … the **pheromone value** for the components, i.e. the amount of pheromone on edge (i, j)
- k … the set of usable components
- J_i … the set of nodes that ant k can reach from v_i (tabu list)
- $\eta_{ij} = \frac{1}{dij}$ … attractiveness computed by a heuristic, indicating the "a-priori **desirability"** of the move

Experiment 1: The travelling Snakesman

http://hci-kdd.org/gamification-interactive-machine-learning/



LIVE DEMO

(https://iml.hci-kdd.org/imlTspSolver/)

ANDROID:

https://play.google.com/store/apps/details?id=com.hcikdd.imlacosolver

Experiment 2: The travelling tree branch

http://hci-kdd.org/gamification-interactive-machine-learning/

The explainable model

- **Ant algorithm:** Swarm algorithm driven by pheromones. Ants deposit pheromones on trails, which helps other ants decide, which trail to choose. *In our example we use the algorithm to find the shortest tour in a point set. (**T**raveling **S**alesman **P**roblem)*
  - **visualization of pheromones is a good interpretation of the ant algorithm**

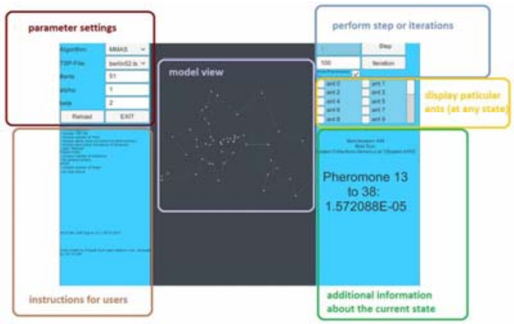
Explainable Model - the visualization

- The pheromones are showing "the state" (high or low frequented paths of ants) of the algorithm.

initial pheromone distribution | pheromones after 100 iterations | pheromones after 500 iterations

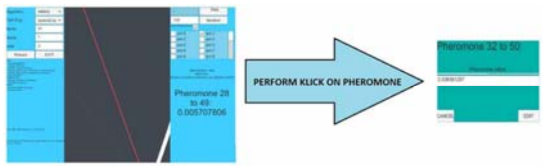
Explainable Interface - overview

http://iml.hci-kdd.org/imlTspSolver/



Explainable Interface - interactive(iML) parts

- iteration vs. step: look inside the iteration
- make the ant algorithm interactive
  - change pheromones at any time
  - *change routes of certain ants in the current iteration (future work)*



06 Causality and Causability

Causality and Explanations

David Hume (1711-1776)

Causation is a matter of perception

We remember seeing the flame, and feeling a sensation called heat; without further ceremony, we call the one cause and the other effect

Statistical ML

Forget causation! Correlation is all you should ask for.

Karl Pearson (1857-1936)

Judea Pearl (1936-)

A mathematical definition of causality

Forget empirical observations! Define causality based on a network of known, physical, causal relationships

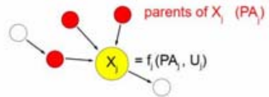
Dependence vs. Causation

Country	Area (km^2)	Storks (pairs)	Humans (10^6)	Birth rate (10^3/yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	[illegible]	[illegible]
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576

Table 1. Geographic, human and stork data for 17 European countries

Functional Causal Model *(Pearl et al.)*

- Set of observables $X_1, \ldots, X_n$
- directed acyclic graph G with vertices $X_1, \ldots, X_n$
- Semantics: parents = direct causes
- $X_i = f_i(\text{ParentsOf}_i, \text{Noise}_i)$, with independent $\text{Noise}_1, \ldots, \text{Noise}_n$.
- "Noise" means "unexplained" (or "exogenous"), we use U_i
- Can add requirement that $f_1, \ldots, f_n, \text{Noise}_1, \ldots, \text{Noise}_n$ "independent" (cf. *Lemeire & Dirkx 2006, Janzing & Schölkopf 2010* — more below)

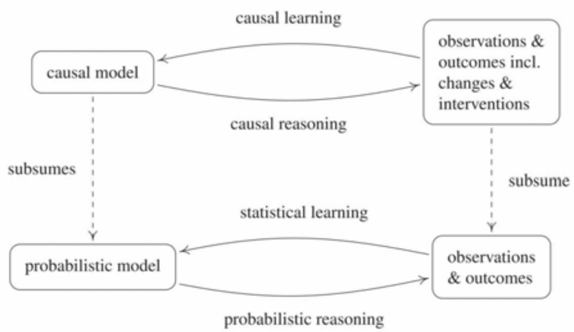


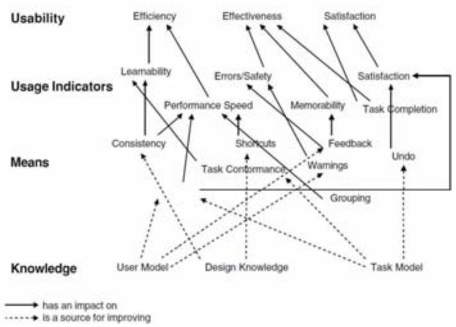
Definition:

Explainability	in a technical sense highlights decision-relevant parts of the used representations of the algorithms and active parts in the algorithmic model, that either contribute to the model accuracy on the training set, or to a specific prediction for one particular observation. It does not refer to an explicit human model.
Causability	as the extent to which an explanation of a statement to a human expert achieves a specified level of causal understanding with effectiveness, efficiency and satisfaction in a specified context of use.

- **Causability := a property of a person, while**
- **Explainability := a property of a system**

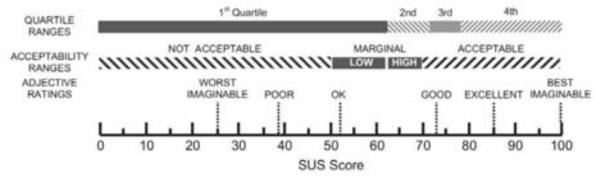
07 Measuring Causality?





Veer, G. C. v. d. & Welie, M. v. (2004) DUTCH: Designing for Users and Tasks from Concepts to Handles. In: Diaper, D. & Stanton, N. (Eds.) *The Handbook of Task Analysis for Human-Computer Interaction. Mahwah (New Jersey), Lawrence Erlbaum, 155-173.*

System Usability Scale



Bangor, A., Kortum, P. T. & Miller, J. T. (2008) An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction, 24, 6, 574-594.*

	The System Usability Scale Standard Version	Strongly Disagree 1	2	3	4	Strongly Agree 5
1	I think that I would like to use this system frequently.	O	O	O	O	O
2	I found the system unnecessarily complex.	O	O	O	O	O
3	I thought the system was easy to use.	O	O	O	O	O
4	I think that I would need the support of a technical person to be able to use this system.	O	O	O	O	O
5	I found the various functions in this system were well integrated.	O	O	O	O	O
6	I thought there was too much inconsistency in this system.	O	O	O	O	O
7	I would imagine that most people would learn to use this system very quickly.	O	O	O	O	O
8	I found the system very awkward to use.	O	O	O	O	O
9	I felt very confident using the system.	O	O	O	O	O
10	I needed to learn a lot of things before I could get going with this system.	O	O	O	O	O

Our Goal in this AK: design, develop & test a System Causability Scale