Medical Information Science for Decision Support

EXECUTIVE ACADEMY

Assoc. Prof. Dr. Andreas HOLZINGER (Med. Uni Graz)

Day 1 –Part 1 -17.4.2018

Information Sciences meets Life Sciences

Andreas Holzinger: Background

- PhD in Cognitive Science 1998
- Habilitation Computer Science 2003
- Lead Holzinger Group HCI-KDD www.hci-kdd.org
- Visiting Professor for Machine Learning in Health Informatics: TU Vienna, Univ. Verona, UCL London, RWTH Aachen
- Research Statement see: Holzinger, A. (2016) Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Springer Brain Informatics, 3, 1-13, doi:10.1007/s40708-016-0042-6
- Most recent: Holzinger, A. 2018. Explainable AI (ex-AI). Informatik-Spektrum, doi:10.1007/s00287-018-1102-5

Mini-Course Syllabus

- At the end of this course you will …
- … be fascinated to see our world in **data sets;**
- … understand the differences between **data, information and knowledge**
- … be aware of some problems and challenges in **biomedical informatics**
- … understand the importance of the concept of **probabilistic information p(x)**
- … know what **AI/Machine Learning** can (not) do
- … have some fundamental insight into medical information science for **decision making**

Reading on Paper or on any electronic device

Holzinger, A. 2014. Biomedical Informatics: Discovering Knowledge in Big Data, New York, Springer, doi:10.1007/978-3-319-04528-3.

Overview

Day 1 - Fundamentals

01 Information Sciences meets Life Sciences

02 Data, Information and Knowledge

03 Decision Making and Decision Support

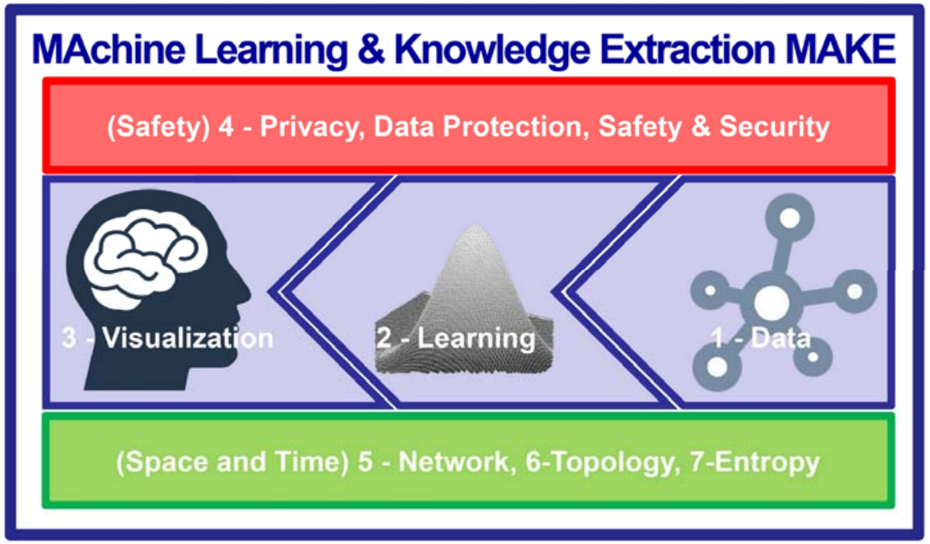
04 From Expert Systems to Explainable AI

Agenda

- 01 What is the HCI-KDD approach?
- 02 Application Area: Health Informatics
- 03 Probabilistic Information
- 04 Automatic Machine Learning
- 05 Interactive Machine Learning
- 06 Key Problems in Biomedical Informatics

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.

Cognitive Science AND Computer Science

- Cognitive Science → human intelligence
- Computer Science → computational intelligence
- Human-Computer Interaction → the bridge

To reach a level of usable intelligence we need to …

04

- 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

Understanding Intelligence

"Solve intelligence – then solve everything else"

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

Demis Hassabis, 22 May 2015

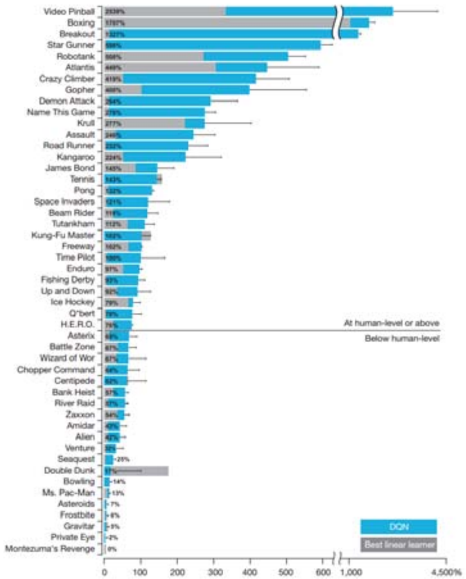
The Royal Society,
Future Directions of Machine Learning Part 2

Google DeepMind

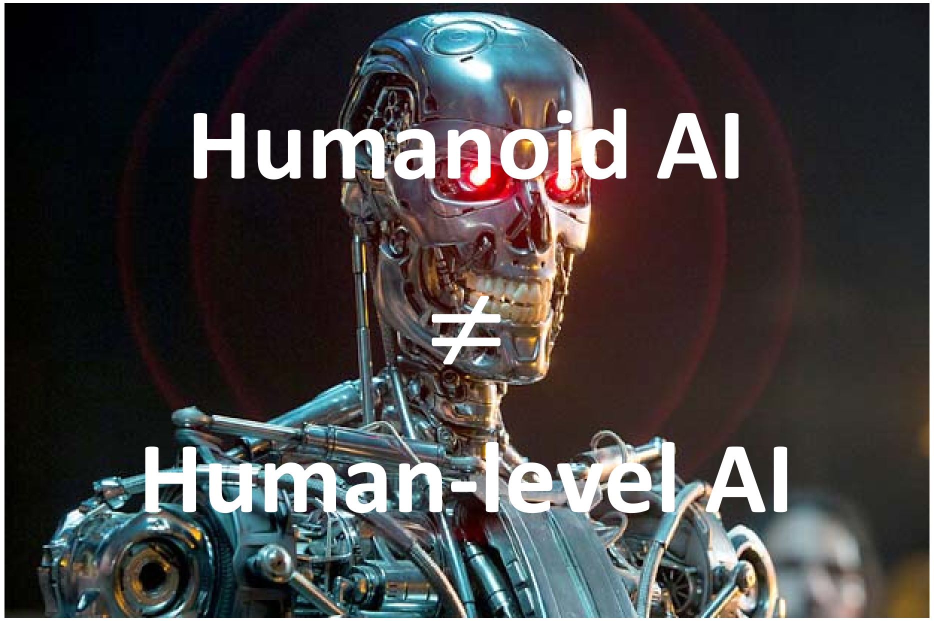
How far are we already ?

Compare your best ML algorithm with a seven year old child …

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, doi:10.1038/nature14236



Not our Goal: Humanoid AI



Health is a complex area

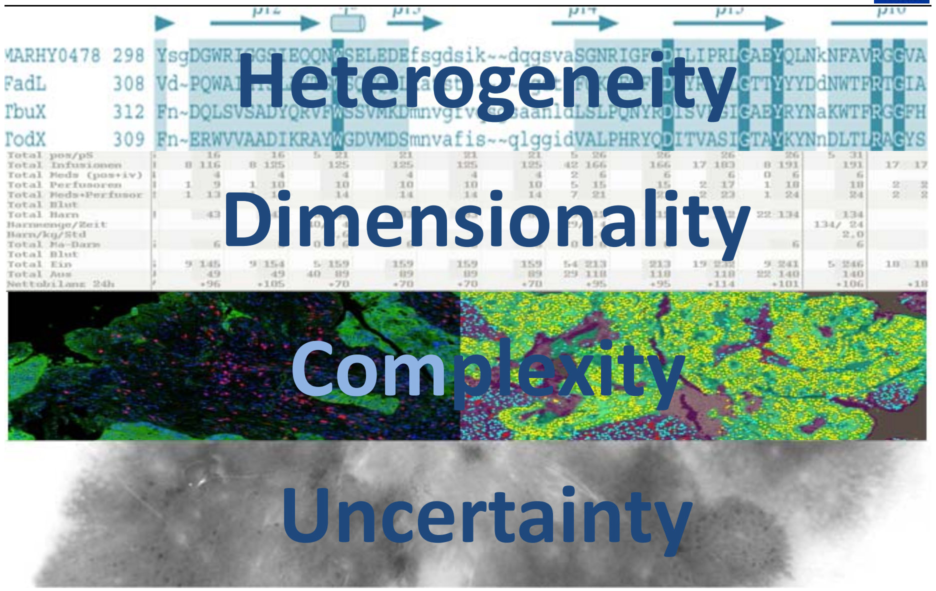
Why is this application area complex ?

In medicine we have two different worlds …

Our central hypothesis:
Information may bridge this gap

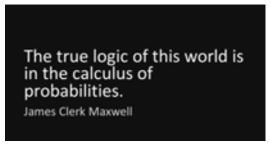
Holzinger, A. & Simonic, K.-M. (eds.) 2011. *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058, Heidelberg, Berlin, New York: Springer.*

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.

03 Probabilistic Information p(x)

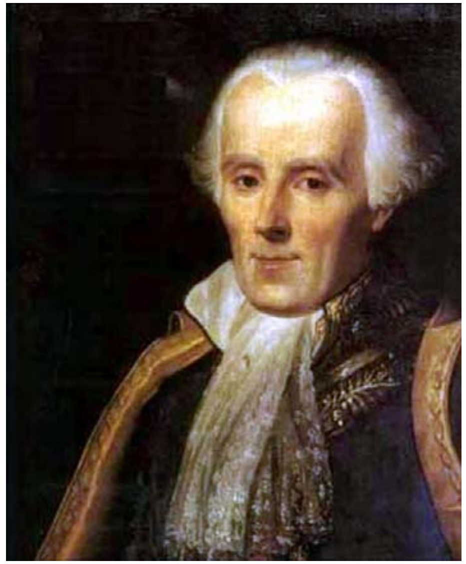


Analogies

- 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
- Gauss generalized those ideas

The foundation for modern machine learning …

Probability theory is nothing but common sense reduced to calculation …



Pierre Simon de Laplace (1749-1827)

Bayesian Learning from data

05

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

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

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

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

Learning and Inference

04

d … $data$

h … $hypotheses$

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

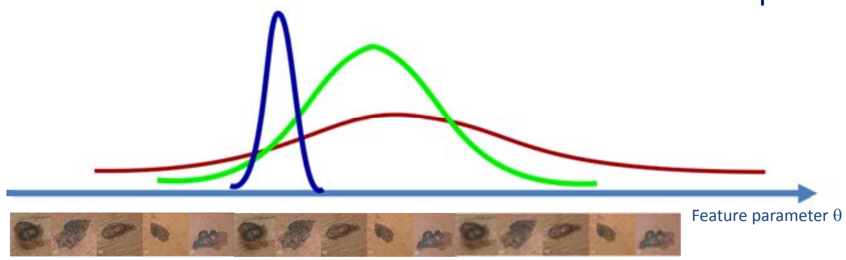
Likelihood

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



Why is this relevant for health informatics?

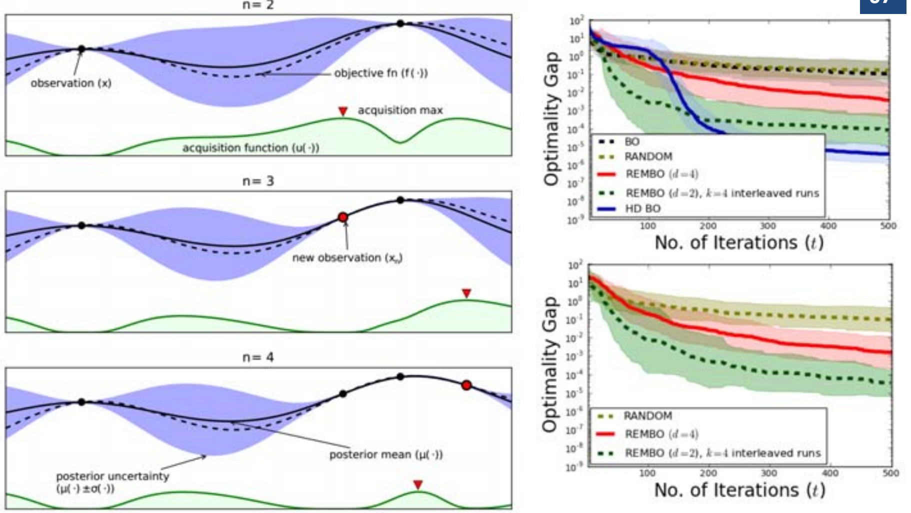
Reasoning under uncertainty: Decision Making

- 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

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

Scaling to high-dimensions is the holy grail in ML

07

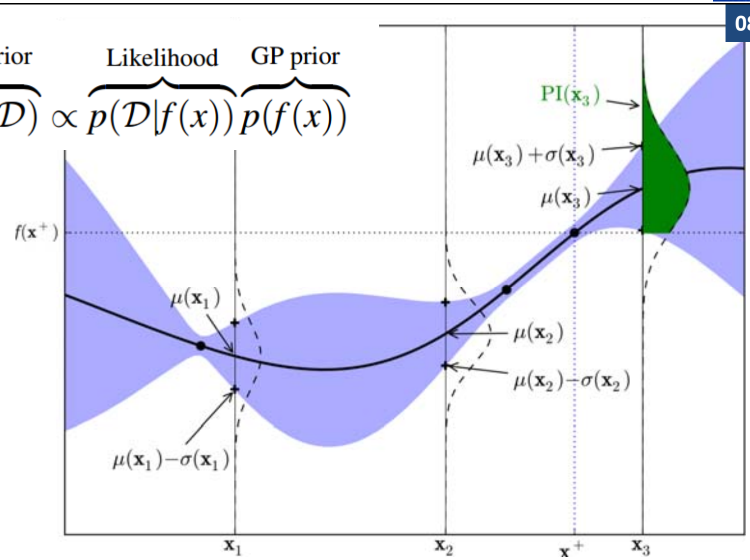


Wang, Z., Hutter, F., Zoghi, M., Matheson, D. & De Feitas, N. 2016. Bayesian optimization in a billion dimensions via random embeddings. Journal of Artificial Intelligence Research, 55, 361-387, doi:10.1613/jair.4806.

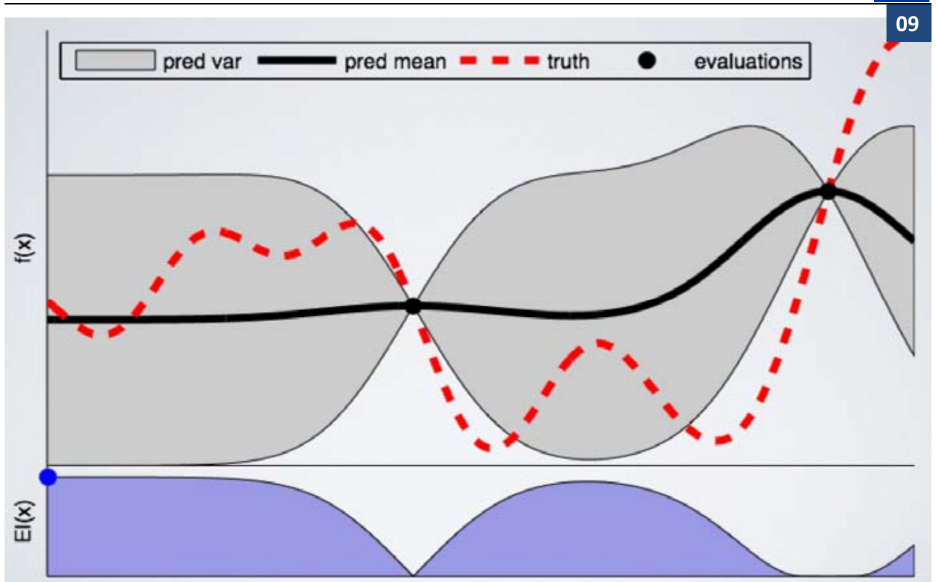
GP = distribution, observations occur in a cont. domain, e.g. t or space

08

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

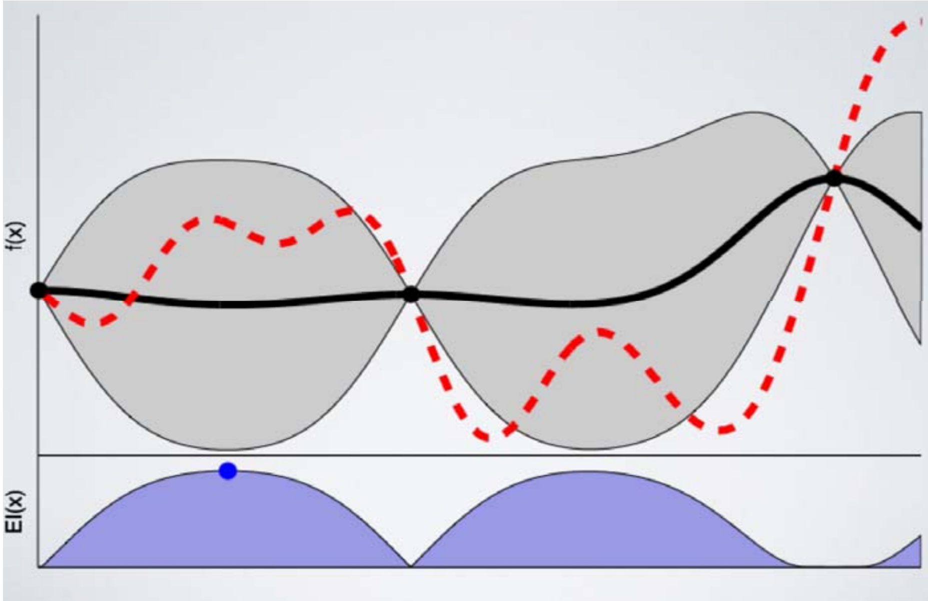


Brochu, E., Cora, V. M. & De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599.

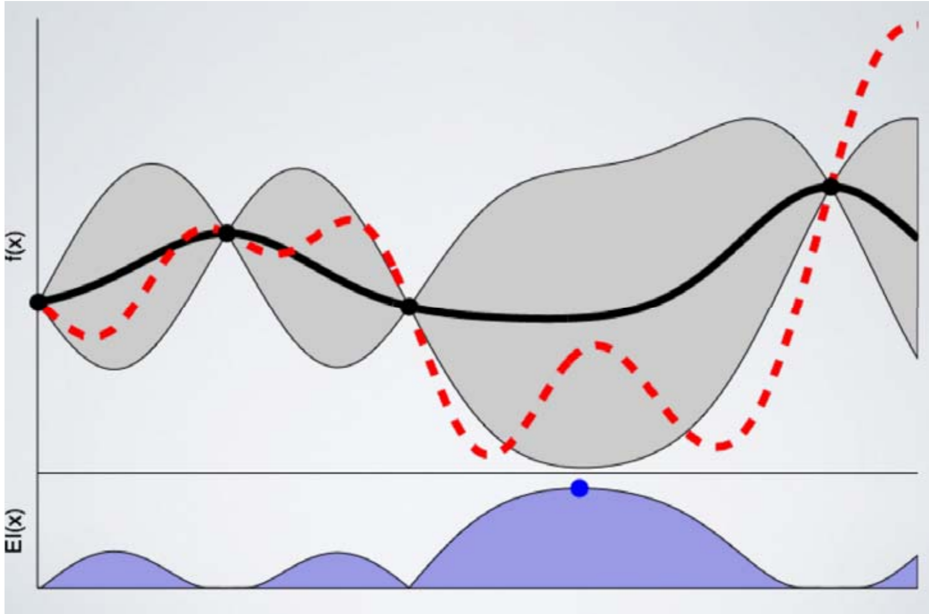


Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 2012. 2951-2959.

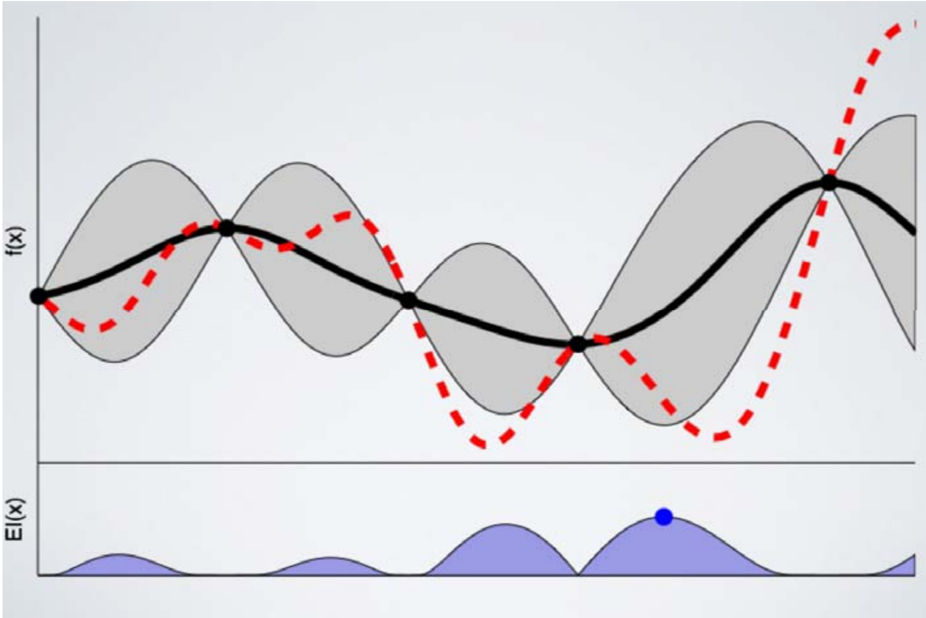
Bayesian Optimization 1



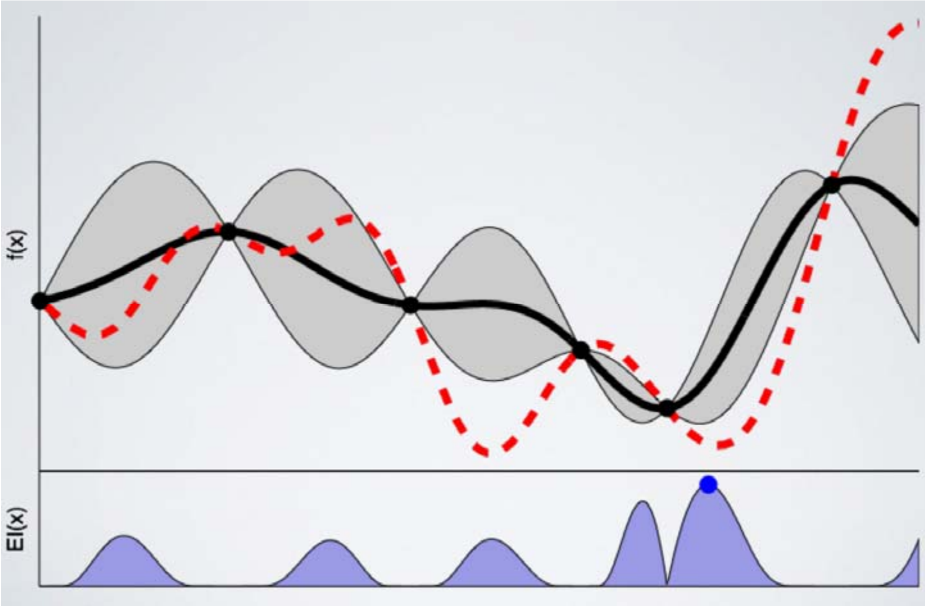
Bayesian Optimization 2



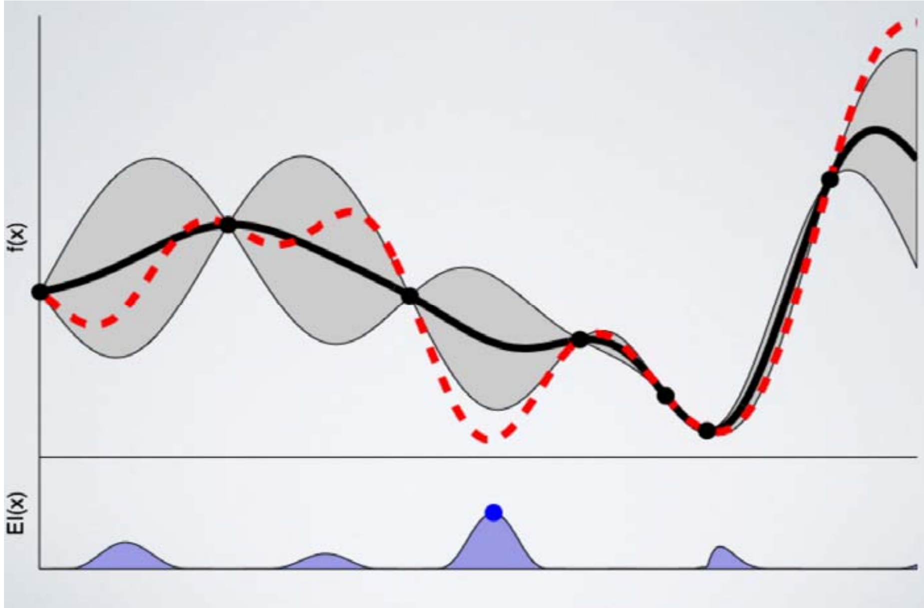
Bayesian Optimization 3



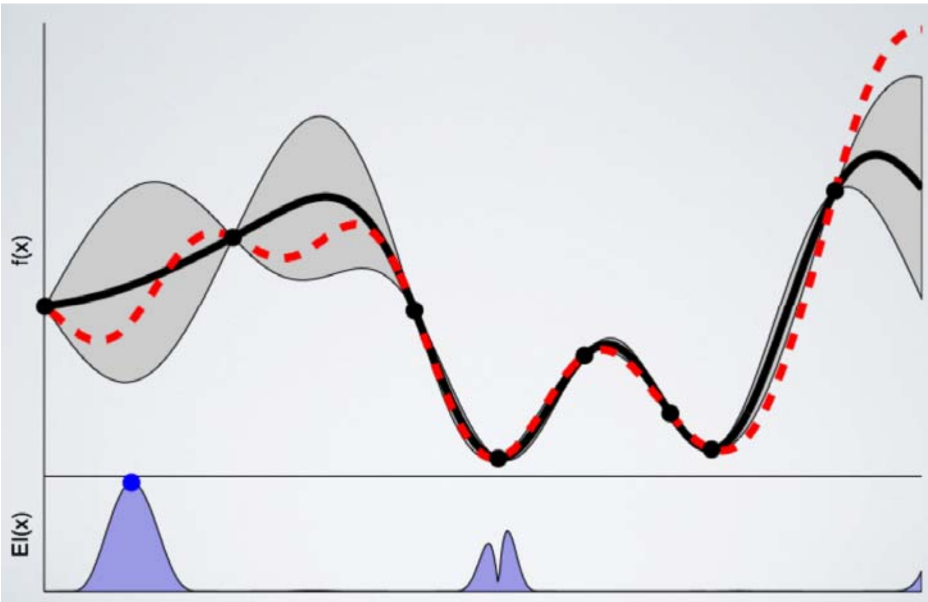
Bayesian Optimization 4



Bayesian Optimization 5

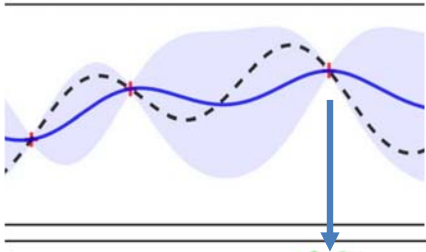


Bayesian Optimization 6



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

07

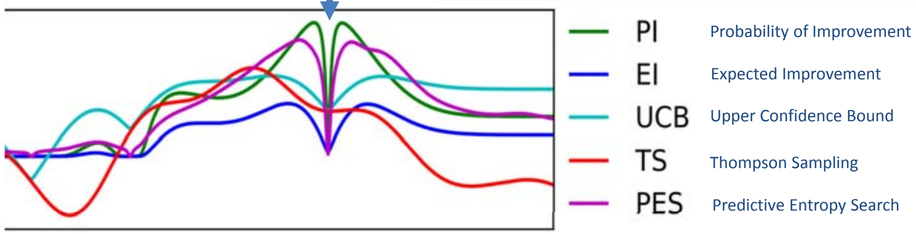


Algorithm 1 Bayesian optimization

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

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

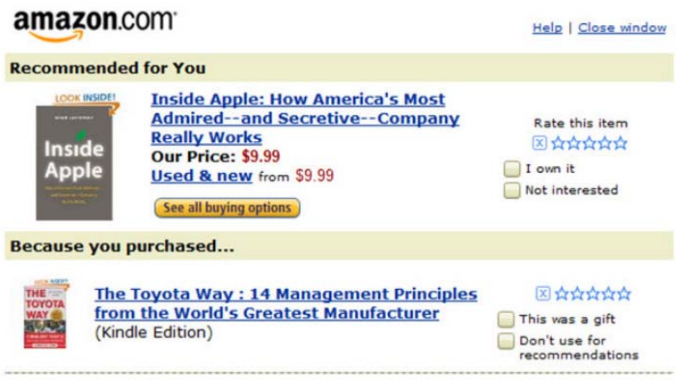
3: query objective function to obtain y_{n+1}
4: augment data $\mathcal{D}_{n+1} = \{\mathcal{D}_n, (\mathbf{x}_{n+1}, y_{n+1})\}$
5: update statistical model
6: **end for**



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.

04 aML

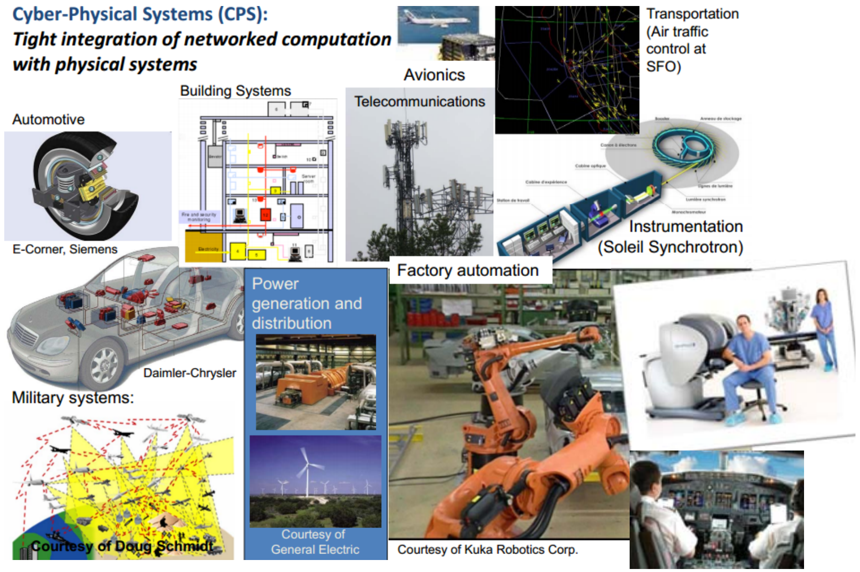
Example for aML: Recommender Systems



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.

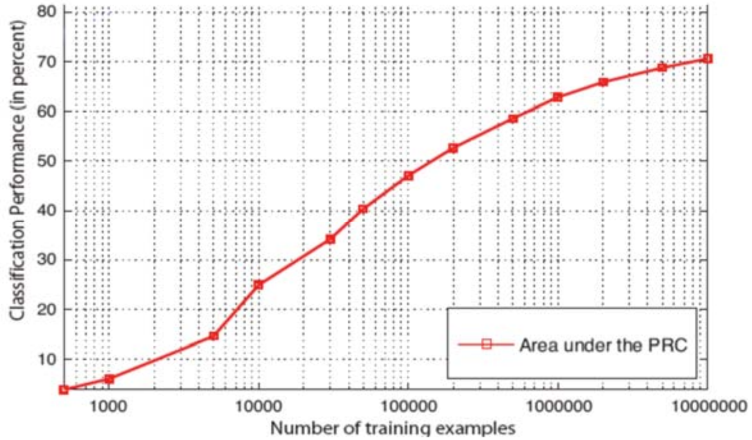
… and thousands of industrial aML applications …



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.

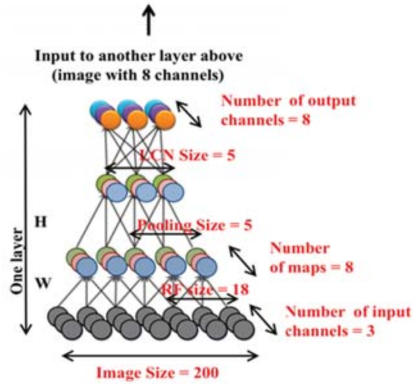
Big Data is necessary for aML !

14



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



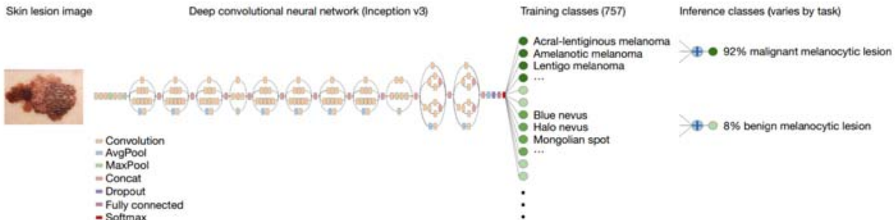
$$x^* = \arg\min_x f(x; W, H), \text{ subject to } ||x||_2 = 1.$$

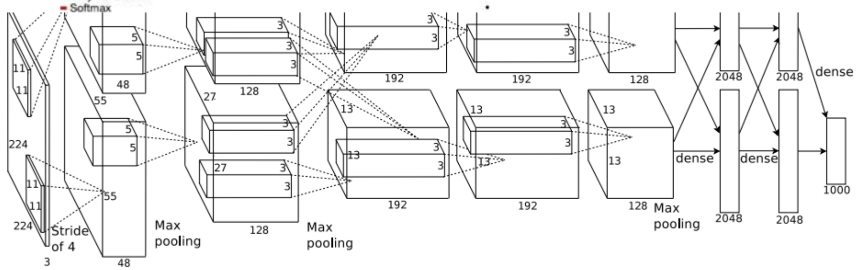
Le, Q. V., Ranzato, M. A., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J. & Ng, A. Y. 2011. Building high-level features using large scale unsupervised learning. arXiv preprint arXiv:1112.6209.

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.

Limitations of Deep Learning approaches

- Computational resource intensive (supercomps, cloud CPUs, **federated learning,** …)
- Black-Box approaches – lack **transparency,** do not foster trust and acceptance among end-user, legal aspects make “black box” difficult!
- **Non-convex:** difficult to set up, to train, to optimize, needs a lot of expertise, error prone
- Very bad in dealing with **uncertainty**
- **Data intensive, needs often millions of training samples …**

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,
    - k-Anonymization,
    - Protein-Folding, …

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

Consequently …

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

05 iML

Definition of iML (Holzinger – 2016)

A6

- iML := algorithms which interact with agents***)** and can optimize their learning behaviour through this interaction

***) where the agents can be human**

Holzinger, A. 2016. Interactive Machine Learning (iML). Informatik Spektrum, 39, (1), 64-68, doi:10.1007/s00287-015-0941-6.

Sometimes we need a doctor-in-the-loop

A group of experts-in-the-loop

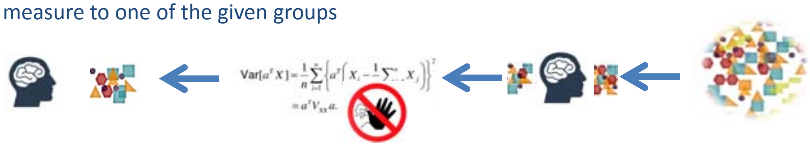
A crowd of people-in-the-loop

aML: taking the human-out-of-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

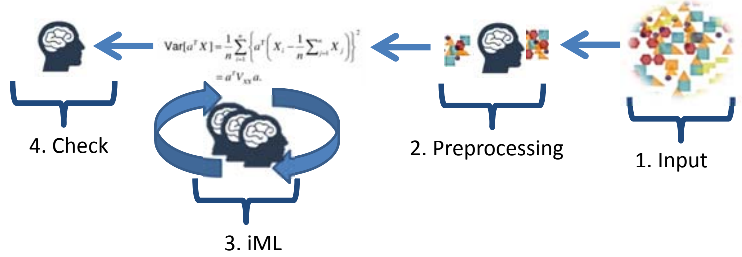
B) Supervised ML: Humans are providing the labels for the training data and/or select features to feed the algorithm to learn – the more samples the better – Human can check results at the end of the ML-pipeline

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



iML: bringing the human-in-the-loop

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: Subspace Clustering**
- **Example 2: k-Anonymization**
- **Example 3: Protein Design**

Hund, M., Böhm, D., Sturm, W., Sedlmair, M., Schreck, T., Ullrich, T., Keim, D. A., Majnaric, L. & Holzinger, A. 2016. Visual analytics for concept exploration in subspaces of patient groups: Making sense of complex datasets with the Doctor-in-the-loop. Brain Informatics, 1-15, doi:10.1007/s40708-016-0043-5.

Kieseberg, P., Frühwirt, P., Weippl, E. & Holzinger, A. 2015. Witnesses for the Doctor in the Loop. In: Guo, Y., Friston, K., Aldo, F., Hill, S. & Peng, H. (eds.) Lecture Notes in Artificial Intelligence LNAI 9250. Springer, pp. 369-378, doi:10.1007/978-3-319-23344-4_36.

Lee, S. & Holzinger, A. 2016. Knowledge Discovery from Complex High Dimensional Data. In: Michaelis, S., Piatkowski, N. & Stolpe, M. (eds.) Solving Large Scale Learning Tasks. Challenges and Algorithms, Lecture Notes in Artificial Intelligence LNAI 9580. Springer, pp. 148-167, doi:10.1007/978-3-319-41706-6_7.

06 Key Problems in health informatics

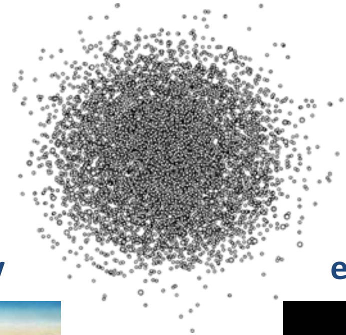
Key Problems

- **Zillions** of different biological species (humans, animals, bacteria, virus, plants, …);
- Enormous **complexity** of the medical domain [1];
- **Complex,** heterogeneous, high-dimensional, big data in the life sciences [2];
- Limited **time,** e.g. a medical doctor in a public hospital has only 5 min. to make a decision [3];
- Limited **computational power** in comparison to the complexity of life (and the natural limitations of the Von-Neumann architecture, …);

1. Patel VL, Kahol K, & Buchman T (2011) Biomedical Complexity and Error. *J. Biomed. Inform.* 44(3):387-389.
2. 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.
3. Gigerenzer G (2008) Gut Feelings: Short Cuts to Better Decision Making (Penguin, London).

Our World in Data - Macroscopic Structures

Two thematic mainstreams in dealing with data …



Time

e.g. Entropy

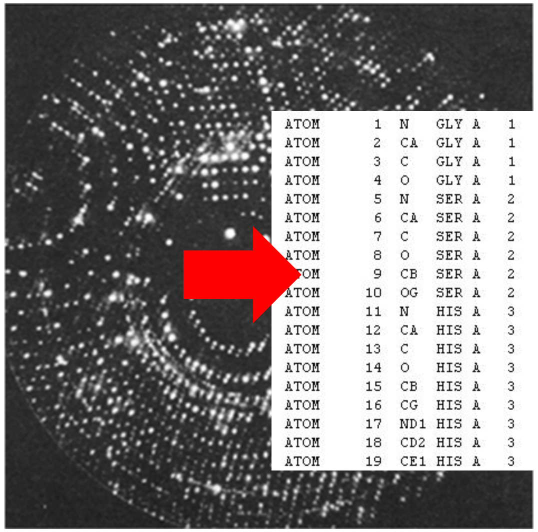
Dali, S. (1931) The persistence of memory

Space

e.g. Topology

Bagula & Bourke (2012) Klein-Bottle

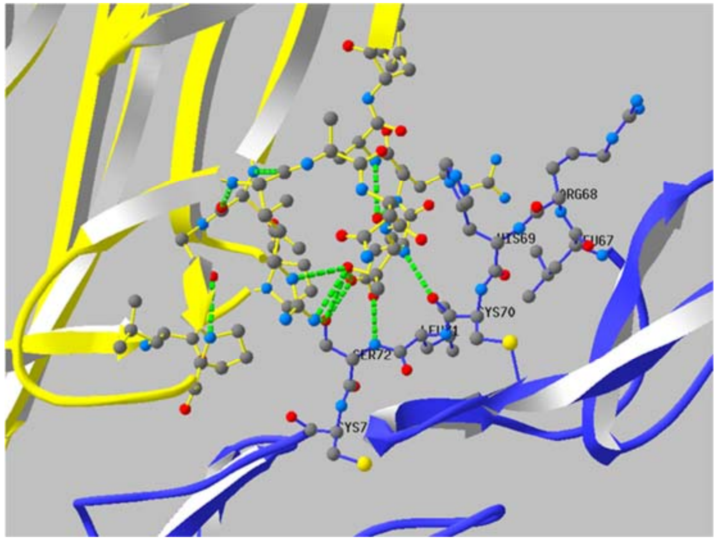
Our World in Data – Microscopic Structures



ATOM	1	N	GLY	A	1	44.842	51.034	101.284	0.01	27.20
ATOM	2	CA	GLY	A	1	45.640	50.230	100.389	0.01	26.99
ATOM	3	C	GLY	A	1	46.692	49.648	101.308	0.01	26.80
ATOM	4	O	GLY	A	1	46.895	50.222	102.381	0.01	26.91
ATOM	5	N	SER	A	2	47.283	48.516	100.951	1.00	26.26
ATOM	6	CA	SER	A	2	48.277	47.866	101.761	1.00	26.17
ATOM	7	C	SER	A	2	49.212	47.031	100.845	1.00	24.21
ATOM	8	O	SER	A	2	49.060	47.195	99.630	1.00	19.77
ATOM	9	CB	SER	A	2	47.438	47.091	102.800	1.00	26.31
ATOM	10	OG	SER	A	2	46.276	46.356	102.404	1.00	27.99
ATOM	11	N	HIS	A	3	50.147	46.186	101.370	1.00	23.93
ATOM	12	CA	HIS	A	3	51.129	45.389	100.609	1.00	21.44
ATOM	13	C	HIS	A	3	50.953	43.905	100.849	1.00	20.32
ATOM	14	O	HIS	A	3	50.530	43.595	101.950	1.00	22.00
ATOM	15	CB	HIS	A	3	52.555	45.674	100.990	1.00	19.69
ATOM	16	CG	HIS	A	3	52.940	47.090	100.611	1.00	21.44
ATOM	17	ND1	HIS	A	3	53.371	47.470	99.422	1.00	20.87
ATOM	18	CD2	HIS	A	3	52.956	48.175	101.433	1.00	21.69
ATOM	19	CE1	HIS	A	3	53.676	48.730	99.476	1.00	20.57

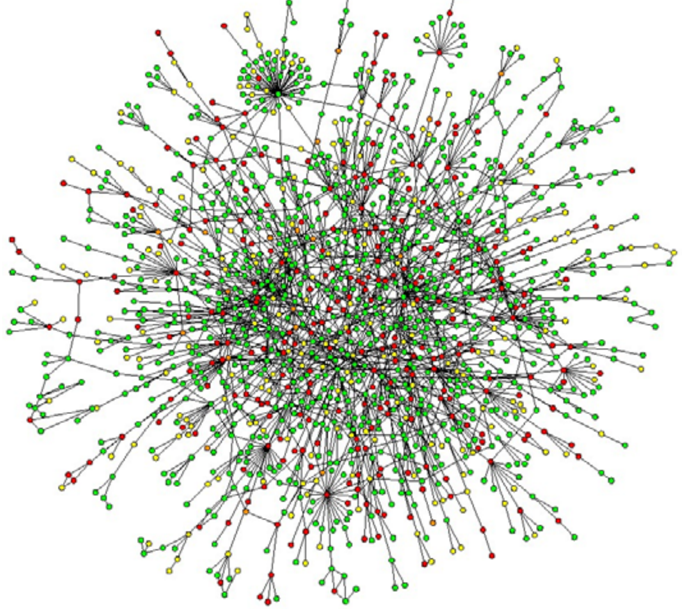
Wiltgen, M. & Holzinger, A. (2005) Visualization in Bioinformatics: Protein Structures with Physicochemical and Biological Annotations. In: *Central European Multimedia and Virtual Reality Conference. Prague, Czech Technical University (CTU), 69-74*

Knowledge Discovery from Data



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.*

Slide 1-4: First yeast protein-protein interaction network

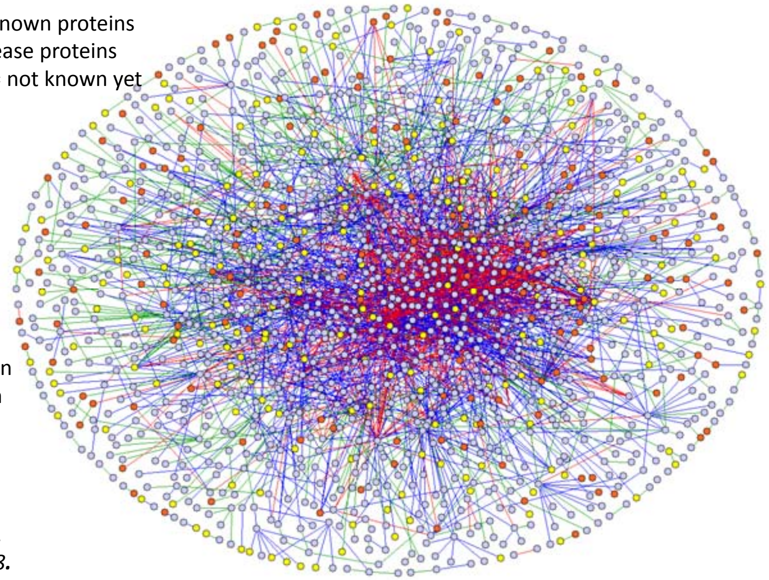


Nodes = proteins
Links = physical interactions (bindings)
Red Nodes = lethal
Green Nodes = non-lethal
Orange = slow growth
Yellow = not known

Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. (2001) Lethality and centrality in protein networks. *Nature, 411, 6833, 41-42.*

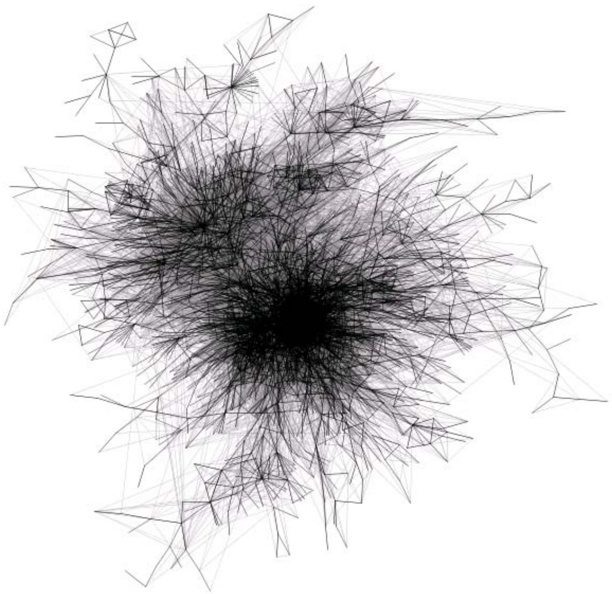
First human protein-protein interaction network

Light blue = known proteins
Orange = disease proteins
Yellow ones = not known yet



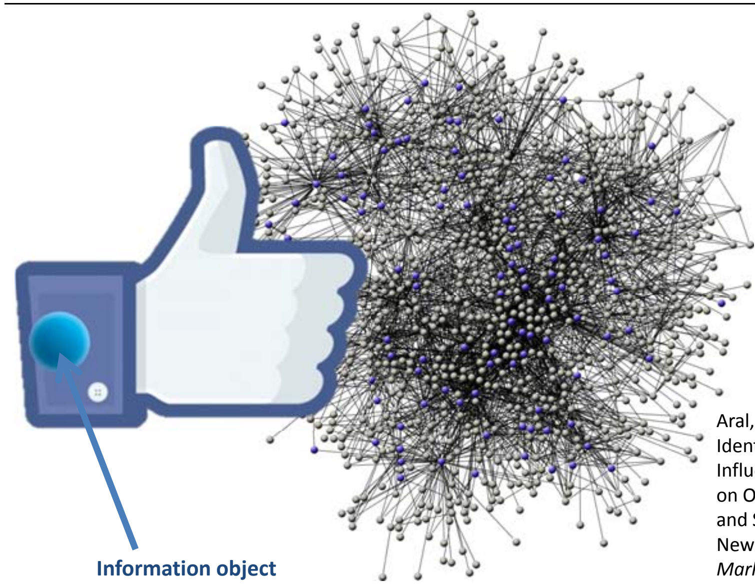
Stelzl, U. et al. (2005) A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell, 122, 6, 957-968.*

Non-Natural Network Example: Blogosphere



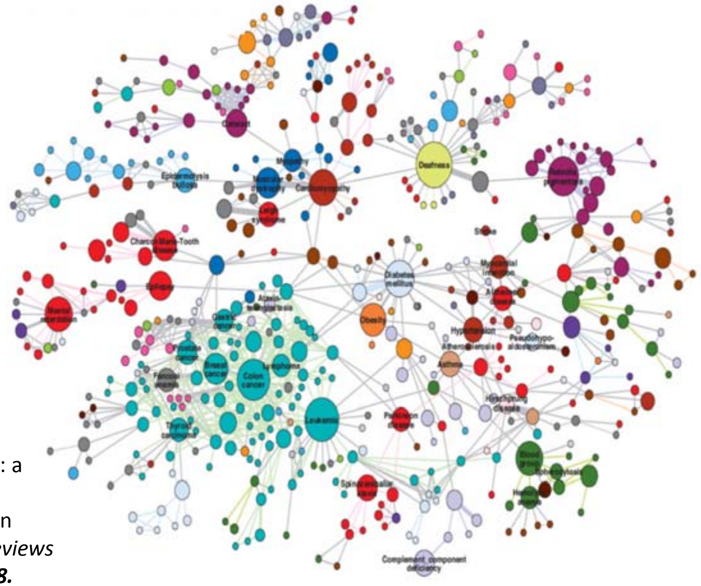
Hurst, M. (2007), Data Mining: Text Mining, Visualization and Social Media. Online available: http://datamining.typepad.com/data_mining/2007/01/the_blogosphere.html, last access: 2011-09-24

Social Behavior Contagion Network



Aral, S. (2011) Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science, 30, 2, 217-223.*

Human Disease Network -> Network Medicine



Barabási, A. L., Gulbahce, N. & Loscalzo, J. 2011. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics, 12*, ***56-68.***

Five decades of Health Informatics

- 1960+ Medical Informatics (Early "AI")
  - Focus on data acquisition, storage, accounting (typ. "EDV"), Expert Systems
  - The term was first used in 1968 and the first course was set up 1978 !
- 1985+ Health Telematics (AI winter)
  - Health care networks, Telemedicine, CPOE-Systems, …
- 1995+ Web Era (AI is forgotten)
  - Web based applications, Services, EPR, distributed systems, …
- 2005+ Ambient Era (AI renaissance)
  - Pervasive, ubiquitous Computing, Internet of things, …
- 2010+ Data Era – Big Data (good for AI)
  - Massive increase of data – data integration, mapping, …
- 2020+ Information Era – (towards Explainable AI)
  - Sensemaking, disentangling the underlying concepts, causality, …

Thank you!

Appendix

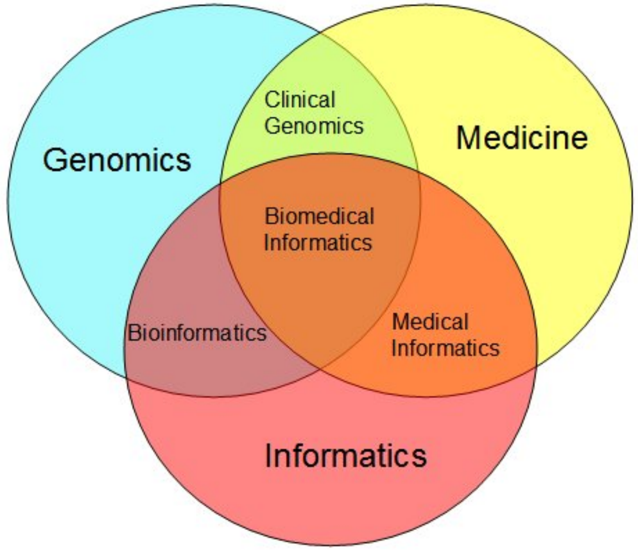
What is biomedical informatics?

INFORMATICS PROFESSIONALS. LEADING THE WAY.

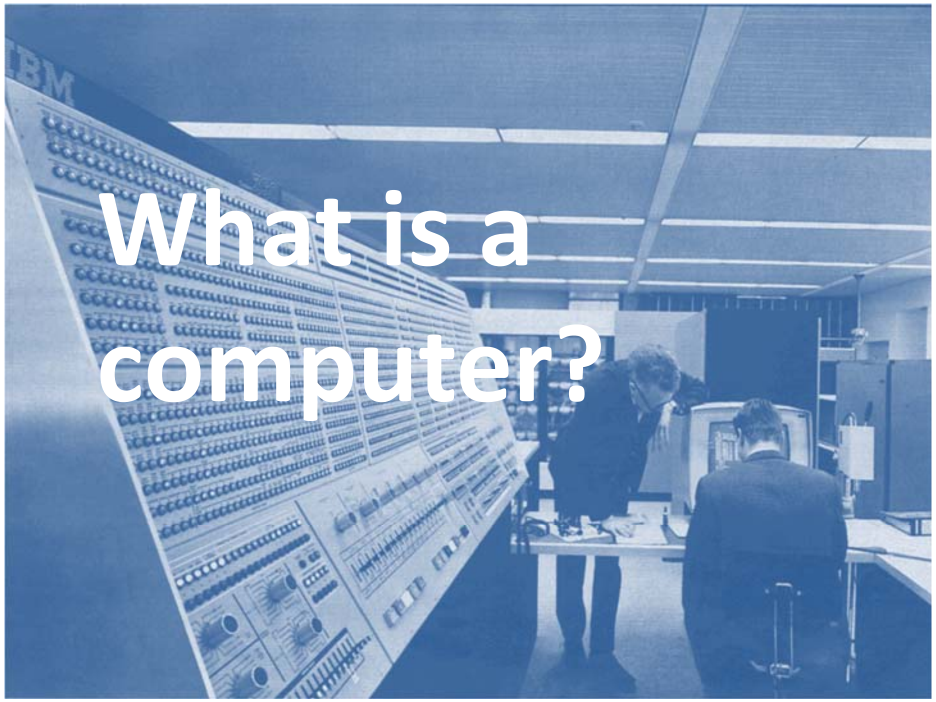
- *Biomedical informatics* (BMI) is the interdisciplinary field that studies and pursues the effective use of biomedical data, information, and knowledge for scientific problem solving, and decision making, motivated by efforts to improve human health

Shortliffe, E. H. (2011). Biomedical Informatics: Defining the Science and its Role in Health Professional Education. In A. Holzinger & K.-M. Simonic (Eds.), *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058 (pp. 711-714). Heidelberg, New York: Springer.*

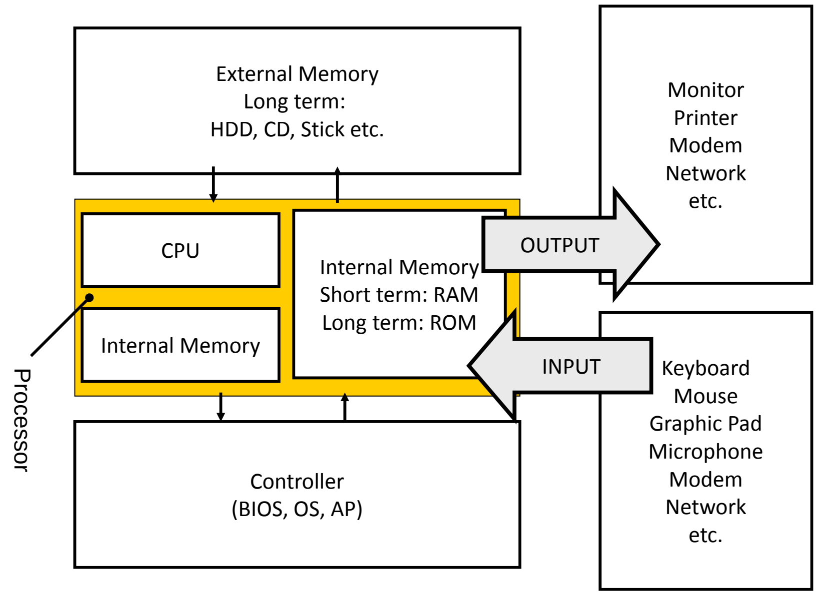
Computational Sciences meet Life Sciences



http://www.bioinformaticslaboratory.nl/twiki/bin/view/BioLab/EducationMIK1-2



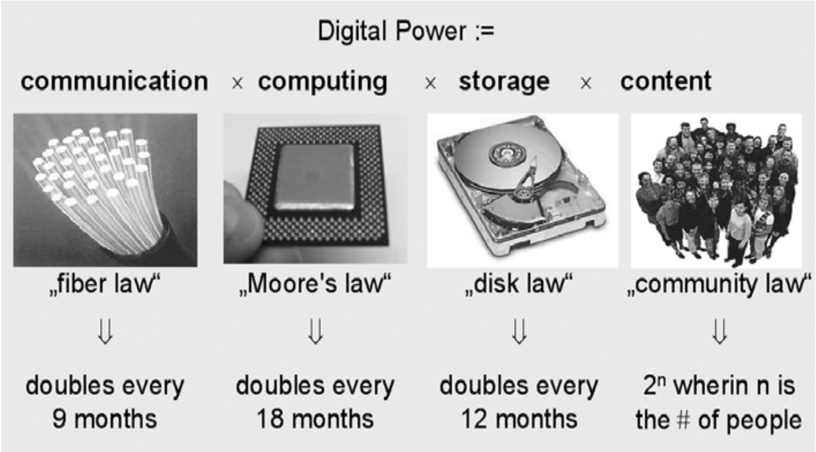
Computer: Von-Neumann Architecture



Holzinger (2002), 90-134

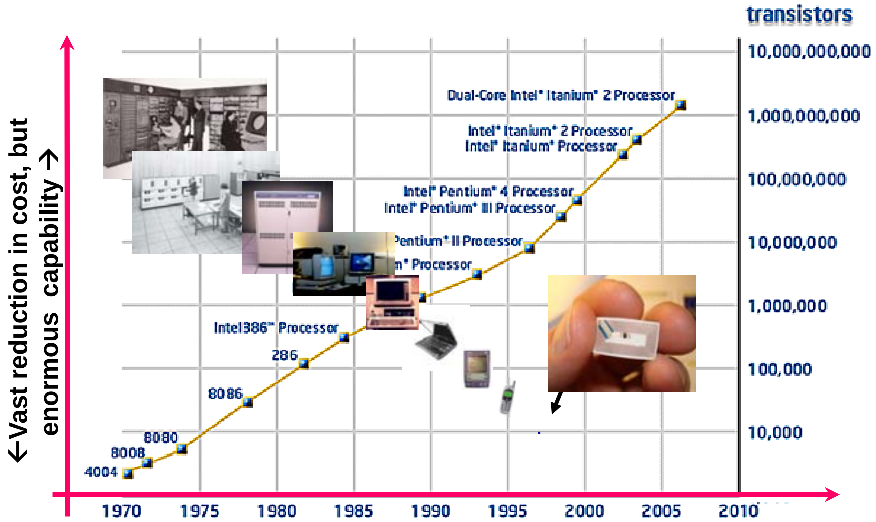
Technological Performance / Digital Power

Gordon E. Moore (1965, 1989, 1997)



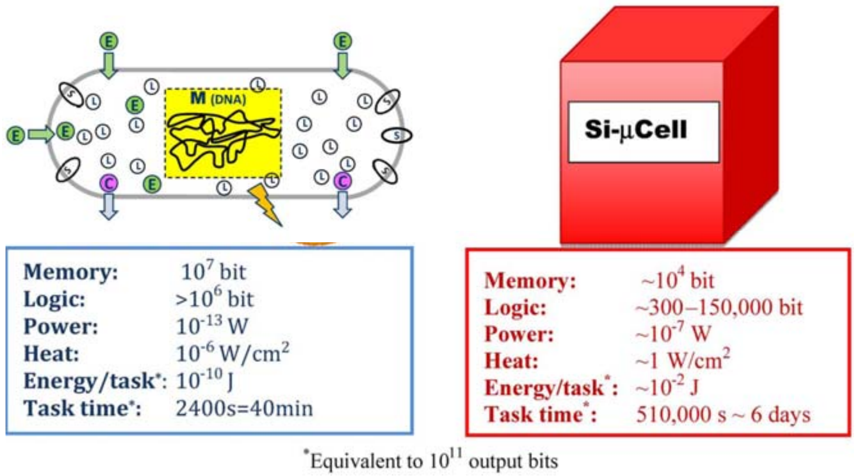
Holzinger, A. 2002. *Basiswissen IT/Informatik Band 1: Informationstechnik. Das Basiswissen für die Informationsgesellschaft des 21. Jahrhunderts, Wuerzburg, Vogel Buchverlag.*

Computer cost/size versus Performance



Cf. with Moore (1965), Holzinger (2002), Scholtz & Consolvo (2004), Intel (2007)

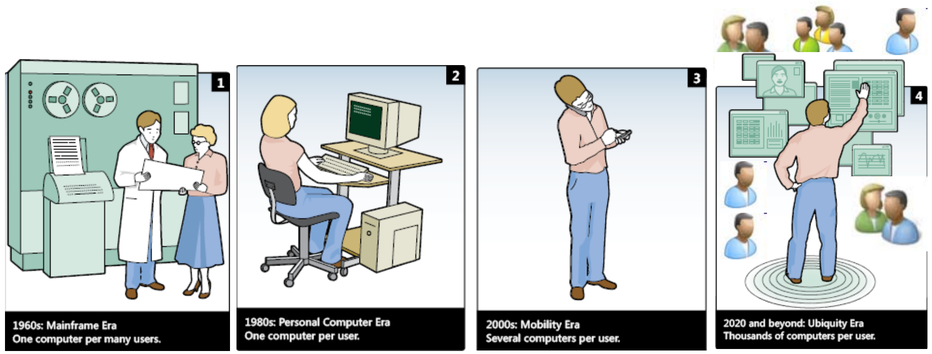
Beyond Moore’s Law -> biological computing



Cavin, R., Lugli, P. & Zhirnov, V. 2012. Science and Engineering Beyond Moore's Law. *Proc. of the IEEE, 100, 1720-49* (L=Logic-Protein; S=Sensor-Protein; C=Signaling-Molecule, E=Glucose-Energy)

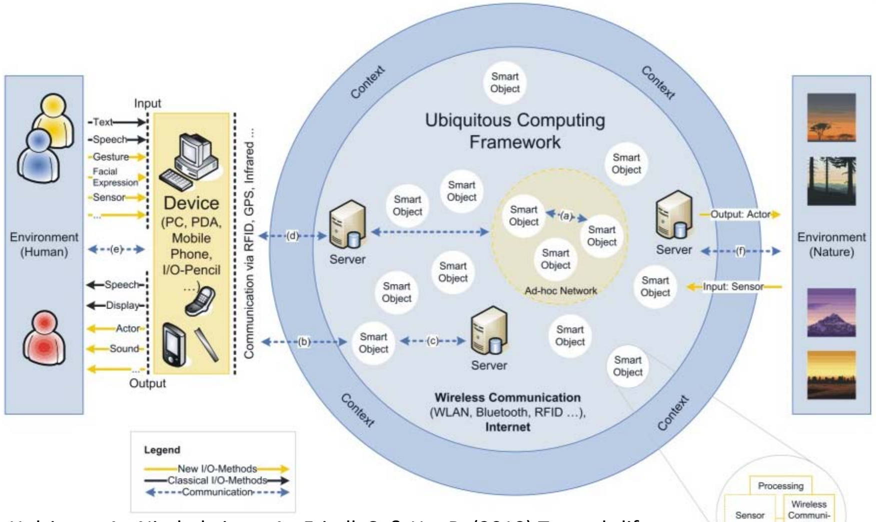
From mainframe to Ubiquitous Computing

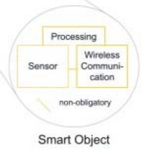
- … using technology to augment human capabilities for structuring, retrieving and managing information



Harper, R., Rodden, T., Rogers, Y. & Sellen, A. (2008) *Being Human: Human-Computer Interaction in the Year 2020. Cambridge, Microsoft Research.*

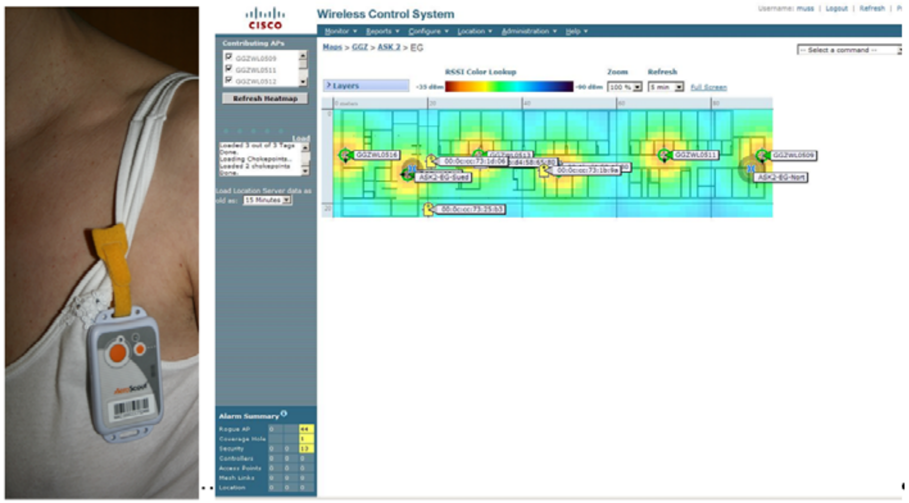
Ubiquitous Computing – Smart Objects





Holzinger, A., Nischelwitzer, A., Friedl, S. & Hu, B. (2010) Towards life long learning: three models for ubiquitous applications. *Wireless Communications and Mobile Computing, 10, 10, 1350-1365.*

Slide 1-34 Example: Pervasive Health Computing

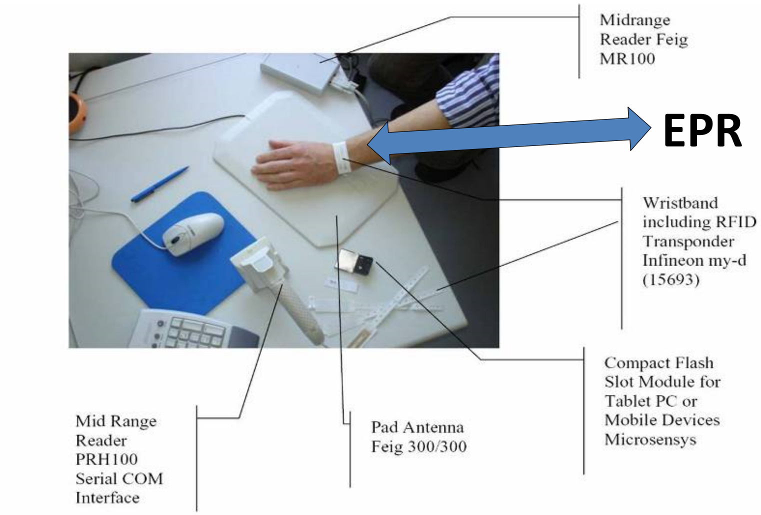


Holzinger, A., Schaupp, K. & Eder-Halbedl, W. (2008) An Investigation on Acceptance of Ubiquitous Devices for the Elderly in an Geriatric Hospital Environment: using the Example of Person Tracking In: *Lecture Notes in Computer Science (LNCS 5105). Heidelberg, Springer, 22-29.*

Ambient Assisted Living - pHealth

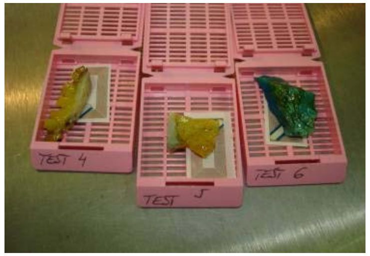
Alagoez, F., Valdez, A. C., Wilkowska, W., Ziefle, M., Dorner, S. & Holzinger, A. (2010) From cloud computing to mobile Internet, from user focus to culture and hedonism: The crucible of mobile health care and Wellness applications. *5th International Conference on Pervasive Computing and Applications (ICPCA). IEEE, 38-45.*

Example Pervasive Computing in the Hospital



Holzinger, A., Schwaberger, K. & Weitlaner, M. (2005) Ubiquitous Computing for Hospital Applications: RFID-Applications to enable research in Real-Life environments *29th Annual IEEE International Computer Software & Applications Conference (IEEE COMPSAC), 19-20.*

Smart Objects in the pathology



Holzinger et al. (2005)

The medical world is mobile (Mocomed)

Holzinger, A., Kosec, P., Schwantzer, G., Debevc, M., Hofmann-Wellenhof, R. & Frühauf, J. 2011. Design and Development of a Mobile Computer Application to Reengineer Workflows in the Hospital and the Methodology to evaluate its Effectiveness. *Journal of Biomedical Informatics, 44****, 968-977.***

1970 Turning Knowledge into Data

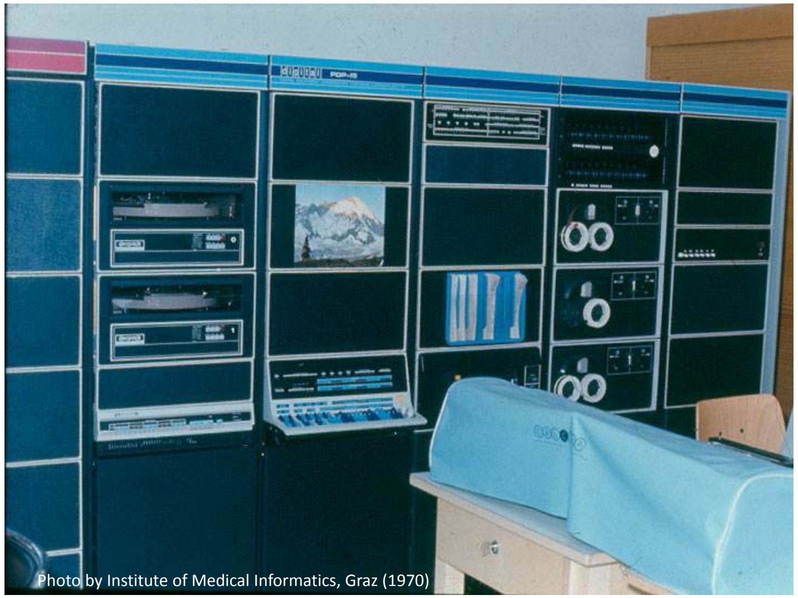


Photo by Institute of Medical Informatics, Graz (1970)