Andreas Holzinger

VO 709.049 Medical Informatics

16.11.2016 11:15-12:45

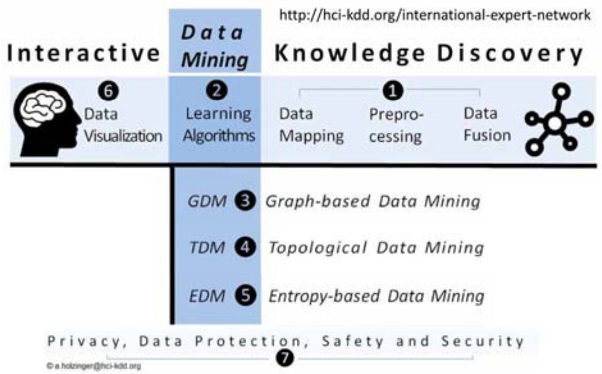
Lecture 04 Decision, Cognition, Uncertainty, Bayesian Statistics Probabilistic Modelling

a.holzinger@tugraz.at

Tutor: markus.plass@student.tugraz.at

http://hci-kdd.org/biomedical-informatics-big-data

ML needs a concerted effort fostering integrated research



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

Keywords

- Decision
- Cognition
- Intelligence
- Expected Utility Theory
- Probabilistic Inference
- Probabilistic Decision Theory
- Signal Detection Theory
- ROC curve
- Learning and Inference
- Naïve Bayes Classifier

Advance Organizer (1/2)

- **Argmax/argmin** = set of points for which f(x) attains the function's largest/smallest value.
- **Brute Force** = systematically computing all possible candidates for a solution and checking whether each candidate satisfies the problem's statement;
- **Cognition** = mental processes of gaining knowledge, comprehension, including thinking, attention, remembering, language understanding, decision making and problem-solving;
- **Cognitive Science** = interdisciplinary study of human information processing, including perception, language, memory, reasoning, and emotion;
- **Confounding Variable** = an unforeseen, unwanted variable that jeopardizes reliability and validity of a study outcome.
- **Correlation coefficient** = measures the relationship between pairs of interval variables in a sample, from r = -1.00 to 0 (no correlation) to r = +1.00
- **Decision Making** = a central cognitive process in every medical activity, resulting in the selection of a final choice of action out of alternatives; according to Shortliffe (2011) DM is still the key topic in medical informatics;
- **Diagnosis** = classification of a patient's condition into separate and distinct categories that allow medical decisions about treatment and prognostic;
- **Differential Diagnosis (DDx)** = a systematic method to identify the presence of an entity where multiple alternatives are possible, and the process of elimination, or interpretation of the probabilities of conditions to negligible levels;
- **Evidence-based medicine (EBM)** = aiming at the best available evidence gained from the scientific method to clinical decision making. It seeks to assess the strength of evidence of the risks and benefits of treatments (including lack of treatment) and diagnostic tests. Evidence quality can range from meta-analyses and systematic reviews of double-blind, placebo-controlled clinical trials at the top end, down to conventional wisdom at the bottom; NOTE: Evidence (English) is NOT Evidenz (Deutsch)!
- **Expected Utility Theory (EUT)** = states that the decision maker selects between risky or uncertain prospects by comparing their expected utility values.

Advance Organizer (2/2)

- **External Validity** = the extent to which the results of a study are generalizable or transferable;
- **Hypothetico-Deductive Model (HDM)** = formulating a hypothesis in a form that could conceivably be falsified by a test on observable data, e.g. a test which shows results contrary to the prediction of the hypothesis is the falsification, a test that could but is not contrary to the hypothesis corroborates the theory – then you need to compare the explanatory value of competing hypotheses by testing how strong they are supported by their predictions;
- **Internal Validity** = the rigor with which a study was conducted (e.g., the design, the care taken to conduct measurements, and decisions concerning what was and was not measured);
- **PDCA** = Plan-Do-Check-Act, The so called PDCA-cycle or Deming-wheel can be used to coordinate a systematic and continuous improvement. Every improvement starts with a goal and with a plan on how to achieve that goal, followed by action, measurement and comparison of the gained output.
- **Perception** = sensory experience of the world, involving the recognition of environmental stimuli and actions in response to these stimuli;
- **Qualitative Research** = empirical research exploring relationships using textual, rather than quantitative data, e.g. case study, observation, ethnography; Results are not considered generalizable, but sometimes at least transferable.
- **Quantitative Research** = empirical research exploring relationships using numeric data, e.g. surveys, quasi-experiments, experiments. Results should be generalized, although it is not always possible.
- **Reasoning** = cognitive (thought) processes involved in making medical decisions (clinical reasoning, medical problem solving, diagnostic reasoning, behind every action;
- **Receiver-operating characteristic (ROC)** = in signal detection theory this is a graphical plot of the sensitivity, or true positive rate, vs. false positive rate (1 – specificity or 1 – true negative rate), for a binary classifier system as its discrimination threshold is varied;
- **Symbolic reasoning** = logical deduction
- **Triage** = process of judging the priority of patients' treatments based on the severity of their condition;

Glossary

- CES = Central Executive System
- DDx = Differential Diagnosis
- DM = Decision Making
- DSS = Decision Support System
- EBM = Evidence-based medicine
- fMRI = functional Magnetic Resonance Image
- HDM = Hypothetico-Deductive Model
- IOM = Institute of Medicine
- LTS = Long Term Storage
- ME = Medical Error
- PDCA = Plan-Do-Check-Act
- QM = Quality Management
- ROC = Receiver Operating Characteristic
- RST = Rough Set Theory
- STS = Short Term Storage
- USTS = Ultra Short Term Storage (Sensory Register)

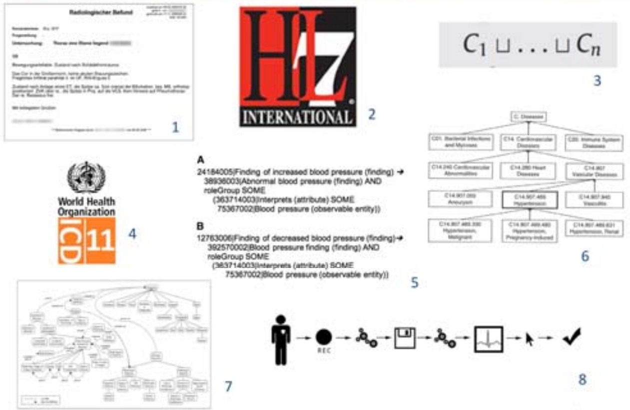
Learning Goals: At the end of this lecture you …

- … are familiar with some principles and elements of human information processing;
- … can discriminate between perception, cognition, thinking, reasoning & problem solving;
- … have got insight into some basics of human decision making processes;
- … got an overview of the Hypothetico-Deductive Method HDM versus PCDA Deming approach;
- … have acquired some basics on modeling patient health and differential diagnosis

Agenda for today

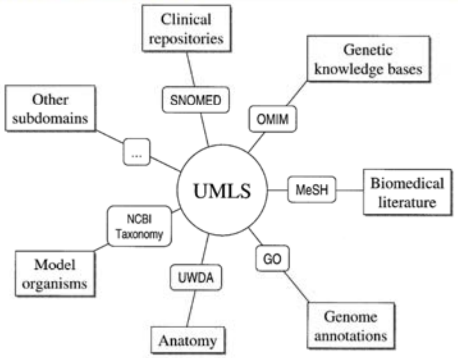
- **00 Reflection – follow-up from last lecture**
- **01 Medical Action = Decision Making**
- **02 Cognition**
- **03 Human vs. Computer**
- **04 Human Information Processing**
- **05 Probabilistic Decision Theory**
- **06 Example: Naïve Bayes Classifier**

Warm-up Quiz from topic 6 Data Mining and KDD



Patient > Data Acquisition > Preprocessing > Storage > Processing > Visualization > Exploration > Diagnosis/Decision

Why is this image important?



Bodenreider, O. (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research, 32, D267-D270.*

Key Challenges

- Medicine is an extremely complex application domain – dealing most of the time with uncertainties -> **probable information!**
- Key: Structure learning and prediction in large-scale biomedical networks with **probabilistic graphical models**
- Causality and Probabilistic Inference:
- Uncertainties are present at all levels in health related systems
- Data sets are noisy, mislabeled, atypical, dirty, wrong, etc. etc.
- Even with data of high quality from different real-world sources requires **processing uncertain information to make viable decisions.**
- In the increasingly complicated settings of modern science, model structure or **causal relationships may not be known a-priori** [1].
- Approximating probabilistic inference in Bayesian belief networks is NP-hard [2] -> here we need the "human-in-the-loop" [3]

[1] Sun, X., Janzing, D. & Schölkopf, B. Causal Inference by Choosing Graphs with Most Plausible Markov Kernels. ISAIM, 2006.
[2] Dagum, P. & Luby, M. 1993. Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artificial intelligence, 60, (1), 141-153.
[3] Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Springer Brain Informatics (BRIN), 3, 1-13, doi:10.1007/s40708-016-0042-6.

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

Slide 7-2: Decision Making is central in Biomedical Informatics

Source: Cisco (2008). Cisco Health Presence Trial at Aberdeen Royal Infirmary in Scotland

Slide 7-3: Reasoning Foundations of Medical Diagnosis

3 July 1959, Volume 130, Number 3366 SCIENCE

Reasoning Foundations of Medical Diagnosis

Symbolic logic, probability, and value theory aid our understanding of how physicians reason.

Robert S. Ledley and Lee B. Lusted

The purpose of this article is to analyze the complicated reasoning processes inherent in medical diagnosis. The importance of this problem has received recent emphasis by the increasing interest in the use of electronic computers as an aid to medical diagnostic processes

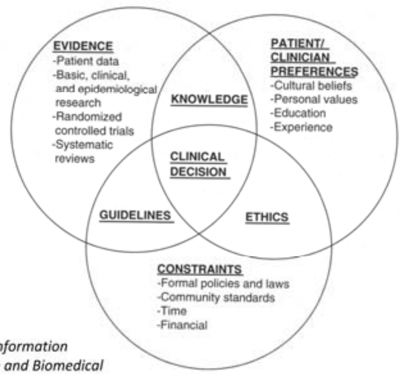
fitted into a definite disease category, or that it may be one of several possible diseases, or else that its exact nature cannot be determined." This, obviously, is a greatly simplified explanation of the process of diagnosis, for the physician might also comment that after seeing a

ance are the ones who do remember and consider the most possibilities."

Computers are especially suited to help the physician collect and process clinical information and remind him of diagnoses which he may have overlooked. In many cases computers may be as simple as a set of hand-sorted cards, whereas in other cases the use of a large-scale digital electronic computer may be indicated. There are other ways in which computers may serve the physician, and some of these are suggested in this paper. For example, medical students might find the computer an important aid in learning the methods of differential diagnosis. But to use the computer thus we must understand how the physician makes a medical diagnosis. This, then, brings us to the subject of our investigation: the reasoning foundations of medical diagnosis and treatment.

Medical diagnosis involves processes that can be systematically analyzed, as well as those characterized as "intangible." For instance, the reasoning foundations of medical diagnostic procedures

Slide 7-4 Decision Making is central in Medicine!



Hersh, W. (2010) *Information Retrieval: A Health and Biomedical Perspective. New York, Springer.*

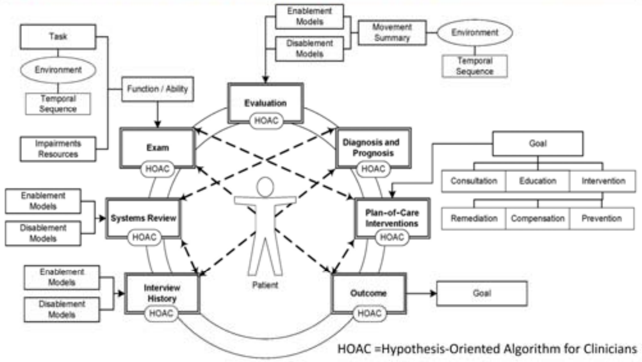
Example for Decision Support

Iserson, K. V. & Moskop, J. C. 2007. Triage in Medicine, Part I: Concept, History, and Types. Annals of Emergency Medicine, 49, (3), 275-281.

Slide 7-17 Example: Triage Tags - International Triage Tags

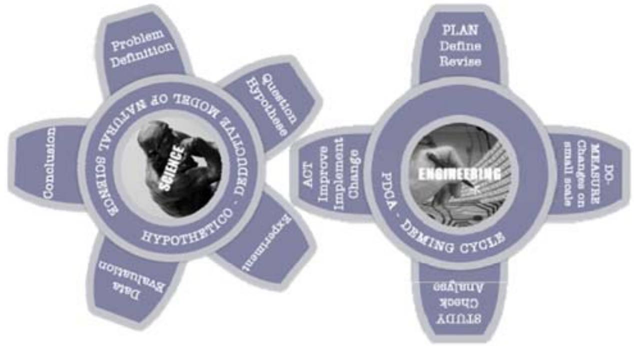
Source: http://store.gomed-tech.com

Slide 7-18 Clinical DM: Hypothesis-Oriented Algorithm



Schenkman, M., Deutsch, J. E. & Gill-Body, K. M. (2006) An Integrated Framework for Decision Making in Neurologic Physical Therapist Practice. *Physical Therapy, 86, 12, 1681-1702.*

Slide 7-19 Hypothetico-Deductive vs PDCA Deming wheel



Holzinger, A. (2010) *Process Guide for Students for Interdisciplinary Work in Computer Science/Informatics. Second Edition. Norderstedt, BoD.* http://www.hci-kdd.org

02 Cognition

02 Cognition

Cognitive Science vs. Computer Science

- **Cognitive Science → human intelligence**
  - Study the principles of ***human learning*** to understand biological intelligence
- **Human-Computer Interaction → 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 → computational intelligence**
  - Study the principles of ***machine learning*** to understand artificial intelligence

What is this: Cognitive Science?

- "By 1960 it was clear that something interdisciplinary was happening. At Harvard we called it cognitive studies, at Carnegie-Mellon they called it information-processing psychology, and at La Jolla they called it cognitive science. "

George A. Miller (1920-2012), Harvard University, well known for:

The magical number seven, plus or minus two: Some limits on our capacity for processing information.
GA Miller - Psychological review, 1956 - psycnet.apa.org
Abstract 1. A variety of researches are examined from the standpoint of information theory. It is shown that the unaided observer is severely limited in terms of the amount of information he can receive, process, and remember. However, it is shown that by the use of various ...
Zitiert von: 23560 Ähnliche Artikel Alle 70 Versionen Web of Science: 7697 In EndNote importieren

Why fitting Cognitive Science with Machine Learning?

- 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;
- ML could inspire **new directions** by novel insights about the human mind

Some definitions in Cognitive Science (very incomplete)

- Intelligence
  - Hundreds of controversial definitions – very hard to define;
  - For us: ability to solve problems, make decisions and acquire and apply knowledge and skills.
- Learning
  - Different definitions – relatively hard to define
  - basically acquisition of knowledge through previous experience
- Problem Solving
  - Process of finding solutions to complex issues
- Reasoning
  - ability of our mind to think and understand things
- Decision Making
  - Process of "de-ciding" ("ent-scheiden") between alternative options
- Sense Making
  - Process of giving meaning to experience
- Causality
  - Relationship between cause and effect

Typical Questions of Cognitive Science Research

- How does our mind work?
- How do we process information?
- How do we learn and generalize?
- How do we solve problems?
- How do we reason and make decisions?
- How do we make predictions?
- How do we behave in new situations?

$\underbrace{\hspace{10em}}_{}$

Intelligence

"Solve intelligence – then solve everything else"

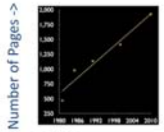
Demis Hassabis, 22 May 2015

The Royal Society,
Future Directions of Machine Learning Part 2

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

The Nobel Prize in Physiology or Medicine 2000

This book doubled in Volume every decade …

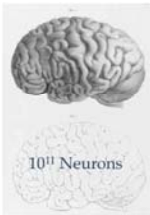


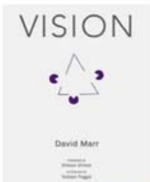
Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A. & Hudspeth, A. 2012. Principles of neural science, 5th Edition (1760 pages), New York: McGraw-Hill.

- Facts ≠ Knowledge, Descriptions ≠ Insight
- **Our goal should be the opposite: To make this book shorter!**

- Cognitive Science had its focus on specific experimental paradigms because it was embedded deeply in Psychology and Linguistics; and aimed to be cognitively/neutrally plausible …
- ML had its focus on standard learning problems and tried to optimize in the range of 1 % because it was embedded in Computer Engineering; and aimed to have working systems whether mimicking the human brain or not …

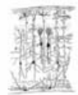
- Cerebellum: big memory to support motor learning
- Neocortex: big memory flexibly learns statistical structure from input patterns
- Hippocampus: big memory encoding memory traces via Hebbian learning
- Example Vision: process of discovering properties (what, where) of things in the real-world from 3D-images (on 2D)
- Vision = information processing task + rich internal representation
- Understanding of vision requires multiple levels of analysis: computational – algorithmic and physical (hardware)





Marr, D. 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, New York, Henry Holt.

- **Computation**
  - "What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?"
- **Representation and algorithm**
  - "What is the representation for the input and output, and the algorithm for the transformation?"
- **Implementation**
  - "How can the representation and algorithm be realized physically?"



Marr, D. 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, New York, Henry Holt.

- Human learning
- Categorization
- Causal learning
- Function learning
- Representations
- Language
- Experiment design

- Machine learning
- Density estimation
- Graphical models
- Regression
- Nonparametric Bayes
- Probabilistic grammars
- Inference algorithms

"People who are interested in machine learning should be cognitive scientists and vice versa" Joshua Tenenbaum, MIT

http://web.mit.edu/cocosci/josh.html

- 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

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

Problem Solving: Humans vs. Computers

When is the human *) better?

*) human intelligence/natural intelligence/human mind/human brain/ learning

- **Natural Language Translation/Curation**
  Computers cannot understand the context of sentences [3]
- **Unstructured problem solving**
  Without a pre-set of rules, a machine has trouble solving the problem, because it lacks the creativity required for it [1]
- **NP-hard Problems**
  Processing times are often exponential and makes it almost impossible to use machines for it, but human make heuristic decisions which are often not perfect but sufficiently good [4]

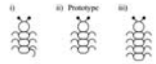
When is the computer **) better?

**) Computational intelligence, Artificial Intelligence/soft computing/ML

- **High-dimensional data processing**
  Humans are very good at dimensions less or equal than 3, but computers can process data in arbitrarily high dimensions
- **Rule-Based environments**
  Difficulties for humans in rule-based environments often come from not recognizing the correct goal in order to select the correct procedure or set of rules [2]
- **Image optimization**
  Machine can look at each pixel and apply changes without human personal biases, and with more speed [1]

[1] Kipp, M. 2006. Creativity Meets Automation: Combining Nonverbal Action Authoring with Rules and Machine Learning. In: LNCS 4133, pp. 230-242, doi:10.1007/11821830_19.
[2] Cummings, M. M. 2014. Man versus Machine or Man + Machine? IEEE Intelligent Systems, 29, (5), 62-69, doi:10.1109/MIS.2014.87.
[3 Pizlo, Z., Joshi, A. & Graham, S. M. 1994. Problem Solving in Human Beings and Computers. Purdue TR 94-075.
[4] Griffiths, T. L. Connecting human and machine learning via probabilistic models of cognition. Interspeech, 2009, ISCA, 9-12..

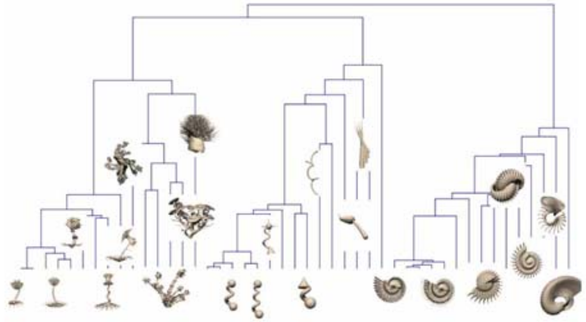
Human Cognitive capacities of Inference and Prediction



- Similarity [1]
- Representativeness and evidential support
- Causal judgment [2]
- Coincidences and causal discovery
- Clinical diagnostic inference [3]
- Predicting the future

[1] Kemp, C., Bernstein, A. & Tenenbaum, J. B. A generative theory of similarity. Proceedings of the 27th Annual Conference of the Cognitive Science Society, 2005. 1132-1137.
[2] Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J. & Blum, B. 2003. Inferring causal networks from observations and interventions. Cognitive science, 27, (3), 453-489.
[3] Krynski, T. R. & Tenenbaum, J. B. 2007. The role of causality in judgment under uncertainty. Journal of Experimental Psychology: General, 136, (3), 430.

Humans learn from very few examples …



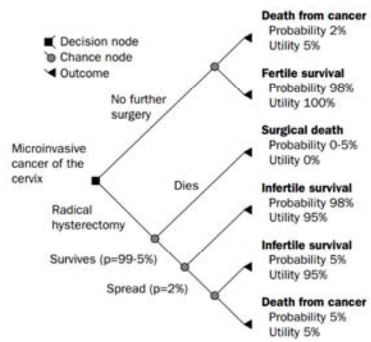
Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.1192788.

Important: Statistics meet Knowledge

- 1. How does abstract knowledge guide learning and inference from sparse data?
  - (Approximate) Bayesian inference in probabilistic models.
- 2. What are the forms and contents of that knowledge?
  - Probabilities defined over a range of structured representations: graphs, grammars, predicate logic, schemas… programs.
- 3. How is that knowledge itself acquired?
  - Hierarchical Bayesian models, with inference at multiple levels of abstraction ("learning to learn"). Learning as (hierarchical Bayesian) program induction.
- **Central Question:**
  How does our mind get so much out of so little?

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.

Decision trees are coming from Clinical Practice



Physician treating a patient approx. 480 B.C. Beazley (1963), Attic Red-figured Vase-Painters, 813, 96. Department of Greek, Etruscan and Roman Antiquities, Sully, 1st floor, Campana Gallery, room 43 Louvre, Paris

Elwyn, G., Edwards, A., Eccles, M. & Rovner, D. 2001. Decision analysis in patient care. The Lancet, 358, (9281), 571-574.

Expected Utility Theory $E\ (U|d)$

For a single decision variable an agent can select $D = d$ for any $d \in dom(D)$.
The expected utility of decision $D = d$ is

http://www.eoht.info/page/Oskar+Morgenstern

$$E(U \mid d) = \sum_{x_1,\ldots,x_n} P(x_1,\ldots,x_n \mid d) U(x_1,\ldots,x_n,d)$$

An optimal single decision is the decision $D = dmax$ whose expected utility is maximal:

$$d_{\max} = \arg \max_{d \in \mathrm{dom}(D)} E(U \mid d)$$

Von Neumann, J. & Morgenstern, O. 1947. Theory of games and economic behavior, Princeton university press.

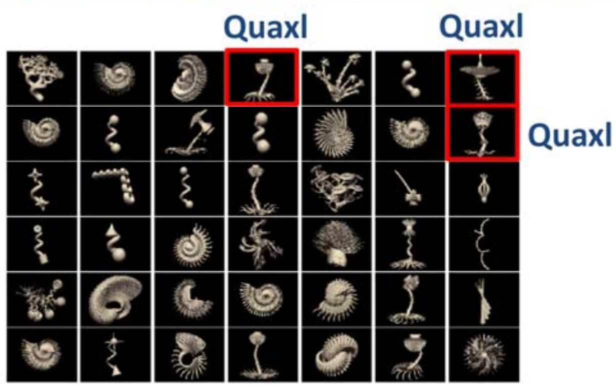
04 Human Information Processing

04 Human Information Processing

How does our mind get so much out of so little …

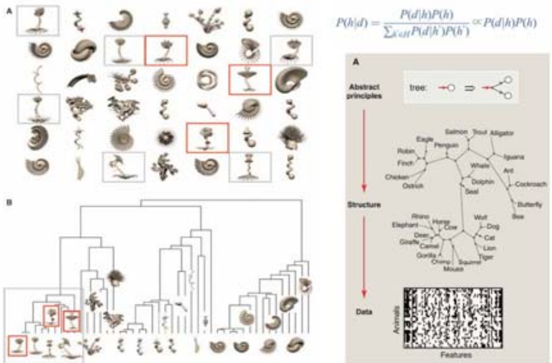
Salakhutdinov, R., Tenenbaum, J. & Torralba, A. 2012. One-shot learning with a hierarchical nonparametric Bayesian model. Journal of Machine Learning Research, 27, 195-207.

Learning words for objects – concepts from examples



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

How do we understand our world ...



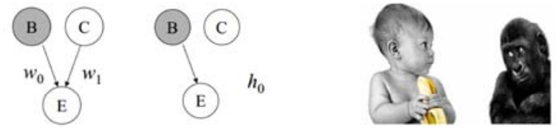
Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285.

One of the unsolved problems in human concept learning

- which is highly relevant for ML research, concerns the factors that determine the subjective difficulty of concepts:
- Why are some concepts psychologically extremely simple and easy to learn,
- while others seem to be extremely difficult, complex, or even incoherent?
- These questions have been studied since the 1960s but are still unanswered ...

Feldman, J. 2000. Minimization of Boolean complexity in human concept learning. Nature, 407, (6804), 630-633, doi:10.1038/35036586.

A few certainties

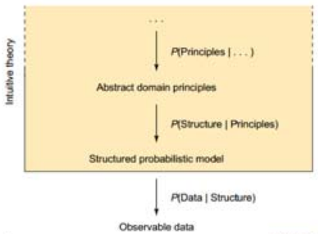


- Cognition as probabilistic inference
  - Visual perception, language acquisition, motor learning, associative learning, memory, attention, categorization, reasoning, causal inference, decision making, theory of mind
- Learning concepts from examples
- Learning and applying intuitive theories (balancing complexity vs. fit)

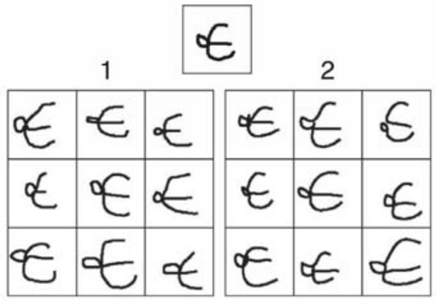
Modeling basic cognitive capacities as intuitive Bayes

- Similarity
- Representativeness and evidential support
- Causal judgement
- Coincidences and causal discovery
- Diagnostic inference
- Predicting the future

Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. 2006. Theory-based Bayesian models of inductive learning and reasoning. Trends in cognitive sciences, 10, (7), 309-318.



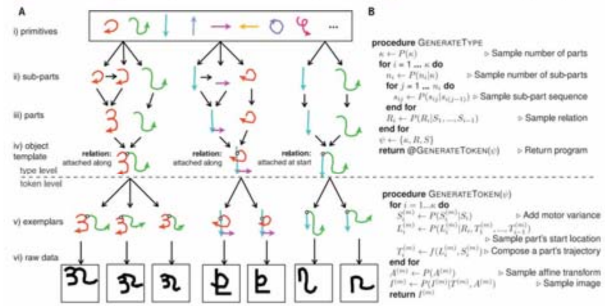
Drawn by Human or Machine Learning Algorithm?



Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. 2015. Human-level concept learning through probabilistic program induction. Science, 350, (6266), 1332-1338, doi:10.1126/science.aab3050.

Human-Level concept learning – probabilistic induction

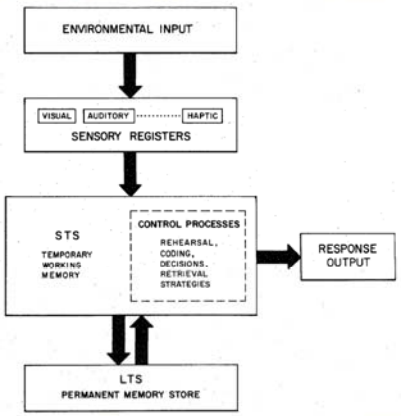
A Bayesian program learning (BPL) framework, capable of learning a large class of visual concepts from just a single example and generalizing in ways that are mostly indistinguishable from people



Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. 2015. Human-level concept learning through probabilistic program induction. Science, 350, (6266), 1332-1338, doi:10.1126/science.aab3050.

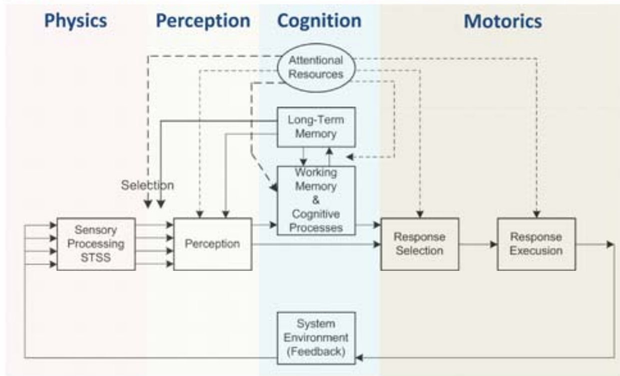
How does our mind get so much out of so little?

Human Information Processing Model (A&S)

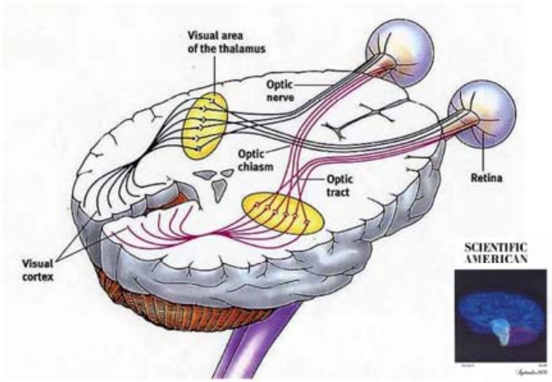


Atkinson, R. C. & Shiffrin, R. M. (1971) *The control processes of short-term memory (Technical Report 173, April 19, 1971). Stanford, Institute for Mathematical Studies in the Social Sciences, Stanford University.*

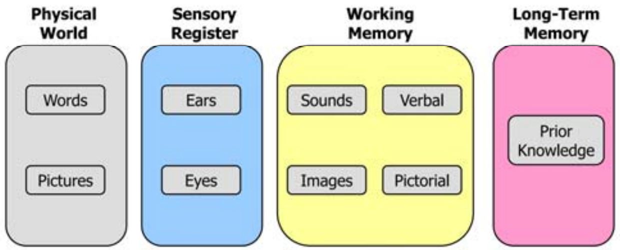
General Model of Human Information Processing



Wickens, C., Lee, J., Liu, Y. & Gordon-Becker, S. (2004) *Introduction to Human Factors Engineering: Second Edition. Upper Saddle River (NJ), Prentice-Hall.*

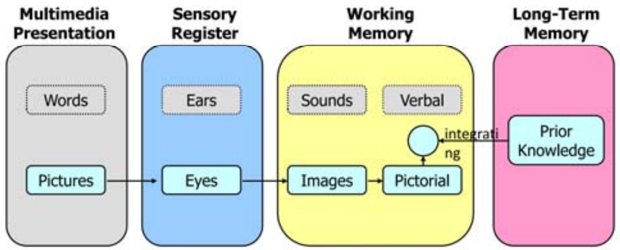


Source: Department of Neuroscience, The Mount Sinai School of Medicine (2004)



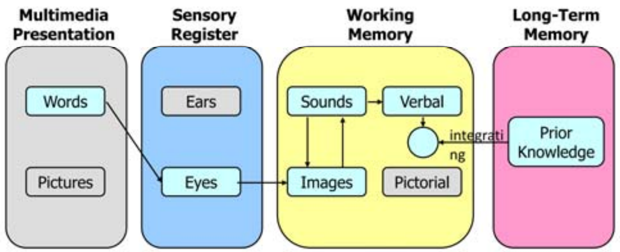
cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

a) Processing of visual information (PICTURES)



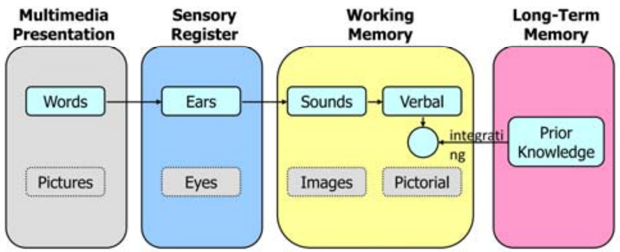
cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

b) Processing of visual information (PRINTED WORDS)

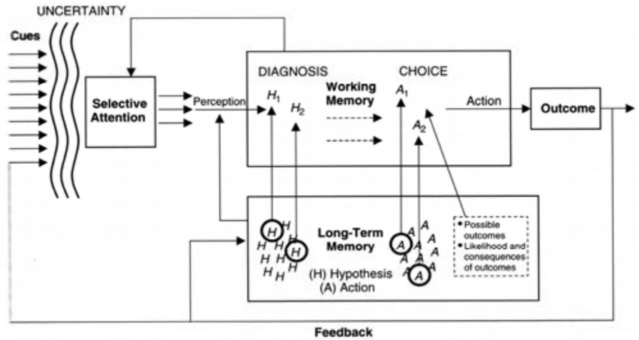


cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

c) Processing of audio information (SPOKEN WORDS)



cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

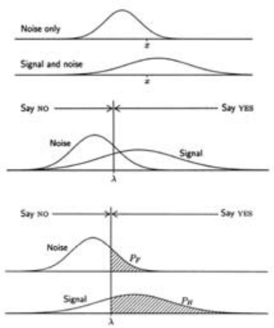


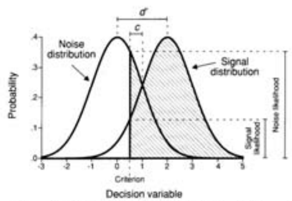
Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill.*

05 Probabilistic Decision Theory

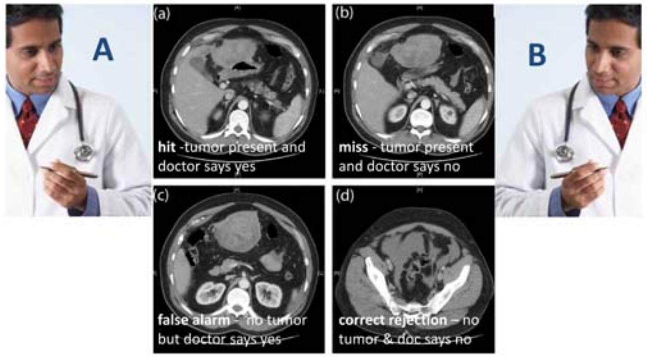
"It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge"
Pierre Simon de Laplace, 1812

Image source: Staffordshire University Computing Futures Museum http://www.fcet.staffs.ac.uk/jdw1/sucfm/malvern.htm





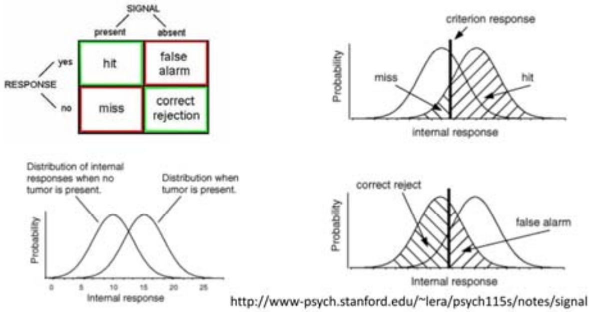
Stanislaw, H. & Todorov, N. 1999. Calculation of signal detection theory measures. Behavior research methods, instruments, & computers, 31, (1), 137-149.



Two doctors, with equally good training, looking at the same CT scan, will have the same information … but they may have a different bias/criteria!

Slide 7-25 Decision Making Process - Signal Detection

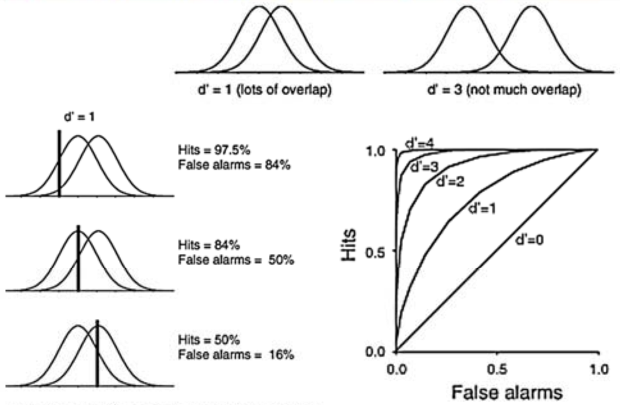
Remember: Two doctors, with equally good training, looking at the same CT scan data, will have the same information … but they may gain different knowledge due to *bias/criteria*.



http://www-psych.stanford.edu/~lera/psych115s/notes/signal

For an example see: Braga & Oliveira (2003) Diagnostic analysis based on ROC curves: theory and applications in medicine. *Int. Journal of Health Care Quality Assurance, 16, 4, 191-198.*

Slide 7-26 Receiver Operating Characteristics (ROC curve)



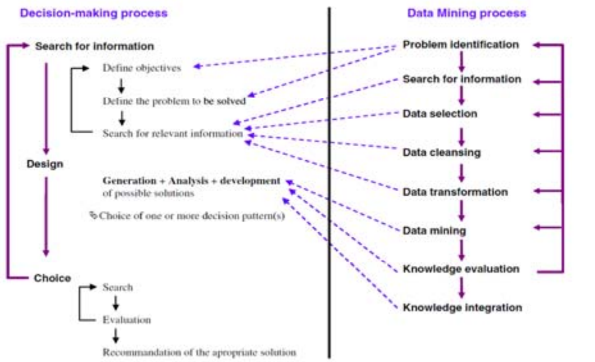
http://gru.stanford.edu/doku.php/tutorials/sdt

Slide 7-23 Information Acquisition and criteria - bias

- **Information acquisition:** in the CT data, e.g. healthy lungs have a characteristic shape; the presence of a tumor might distort that shape (= anomaly).
- Tumors have different image characteristics: brighter or darker, different texture, etc.
- With proper training a doctor learns what kinds of things to look for, so with more practice/training they will be able to acquire more (and more reliable) information.
- Running another test (e.g., MRI) can be used to acquire more (relevant!) information.
- The effect of information is to increase the likelihood of getting either a hit or a correct rejection, while reducing the likelihood of an outcome in the two error boxes (slide 33).
- **Criterion:** Additionally to relying on technology/testing, the medical profession allows doctors to use their own judgment.
- Different doctors may feel that the different types of errors are not equal.
- For example, a doctor may feel that missing an opportunity for early diagnosis may mean the difference between life and death.
- A false alarm, on the other hand, may result only in a routine biopsy operation. They may chose to err toward ``yes'' (tumor present) decisions.
- Other doctors, however, may feel that unnecessary surgeries (even routine ones) are very bad (expensive, stress, etc.).
- They may chose to be more conservative and say ``no'' (no turmor) more often. They will miss more tumors, but they will be doing their part to reduce unnecessary surgeries. And they may feel that a tumor, if there really is one, will be picked up at the next check-up.

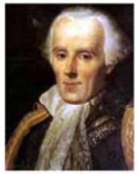
Mohamed, A. et al. (2010) Traumatic rupture of a gastrointestinal stromal tumour with intraperitoneal bleeding and haematoma formation. *BMJ Case Reports, 2010.*

Slide 7-24: Decision Making Process vs. Data Mining process



Ayed, B. M., Ltifi, H., Kolski, C. & Alimi, A. (2010) A user-centered approach for the design & implementation of KDD-based DSS: A case study in the healthcare domain. *Decision Support Systems, 50, 64-78.*

Repetition Bayes Foundations



What is the simplest mathematical operation for us?

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

How do we call repeated adding?

$$p(x, y) = p(y|x) * p(y)$$

Laplace (1773) showed that we can write:

$$p(x, y) * p(y) = p(y|x) * p(x)$$

Now we introduce a third, more complicated operation:

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

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

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

Learning and Inference

04

d … *data*

h … *hypotheses*

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

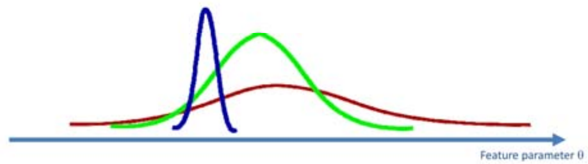
Likelihood

Prior Probability

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

Posterior Probability

Evidence $p(d)$ = marginal likelihood



Slide 7-27b Bayes Law of Total Probability = data modelling

d … data; h … hypothesis

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

$P(h)$: prior belief (probability of hypothesis h before seeing any data)

$P(d\,|\,h)$: likelihood (probability of the data if the hypothesis h is true)

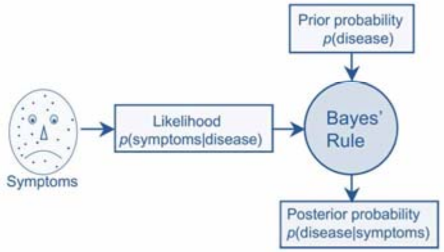
$P(d) = \sum_h P(d\,|\,h)P(h)$: data evidence (marginal probability of the data)

$P(h\,|\,d)$: posterior (probability of hypothesis h after having seen the data d)

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

- **evidence = marginal likelihood = "normalization"**
- **Remember: The inverse probability allows to infer unknowns, learn from data and make predictions … machine learning!**

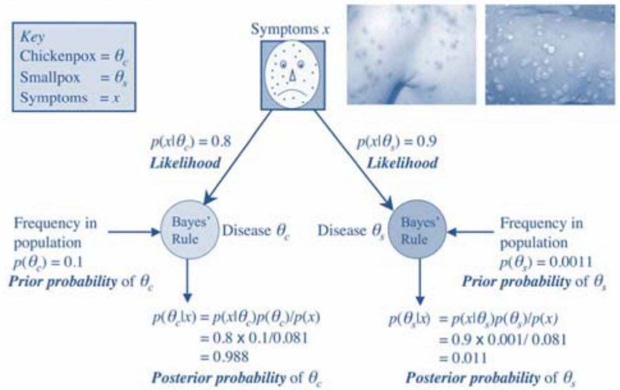
Bayes Rule for Medical Diagnosis



$$p(\text{disease}|\text{symptoms}) = \frac{p(\text{symptoms}|\text{disease})p(\text{disease})}{p(\text{symptoms})}$$

Stone, J. V. 2013. Bayes' rule: a tutorial introduction to Bayesian analysis. Sebtel Press.

Bayesian Inference



Python Code Samples

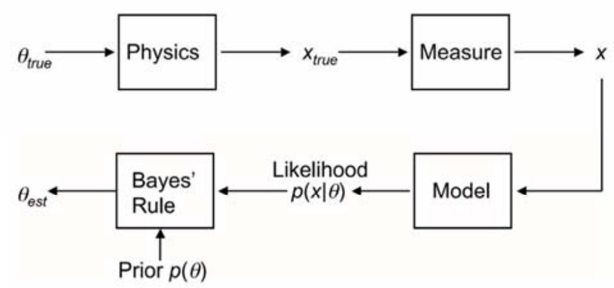
Code Example 1.1: Smallpox
File: Ch1Eq06.py

```
# likelihood = prob of spots given smallpox
pSpotsGSmallpox = 0.9
# prior = prob of smallpox
pSmallpox = 0.001
# marginal likelihood = prob of spots
pSpots = 0.081
# find posterior = prob of smallpox given spots
pSmallpoxGSpots = pSpotsGSmallpox * pSmallpox / pSpots

print('Posterior,␣pSmallpoxGSpots␣=␣%.3f.' % pSmallpoxGSpots)
# Output: Posterior, pSmallpoxGSpots = 0.011.
```

Code Example 1.2: Chickenpox
File: Ch1Eq09.py

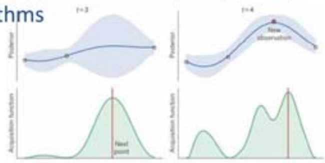
```
#  likelihood = prob of spots given chickenpox
pSpotsGChickenpox = 0.8
#  prior = prob of chickenpox
pChickenpox = 0.1
#  marginal likelihood = prob of spots
pSpots = 0.081
#  find posterior = prob of chickenpox given spots
pChickenpoxGSpots = pSpotsGChickenpox * pChickenpox / pSpots
print('Posterior,␣pChickenpoxGSpots␣=␣%.3f.' % pChickenpoxGSpots)

# Output:   Posterior, pChickenpoxGSpots = 0.988.
```

http://jim-stone.staff.shef.ac.uk/BookBayes2012/BayesRulePythonCode.html



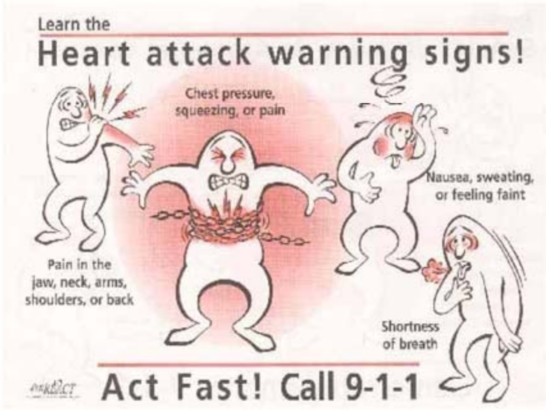
Slide 7-27c Representation of uncertainty

- Many aspects of intelligence and learning depend on **probabilistic representation of uncertainty:**
- Forecasting
- Decision support
- Learning from noisy, missing, uncertain data …
- Knowledge discovery
- Probabilistic programming (e.g. Stochastic Python, Julia)
- Universal inference algorithms
- Global optimization



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

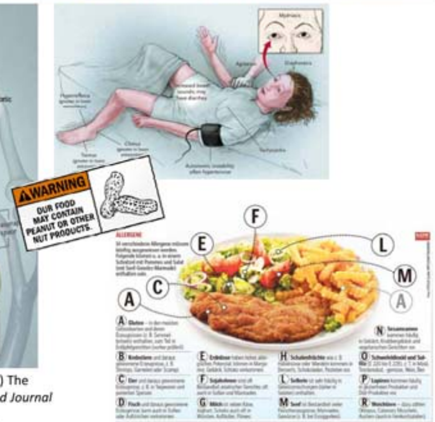
A practical example



Slide 7-27d Bayes' in clinical practice

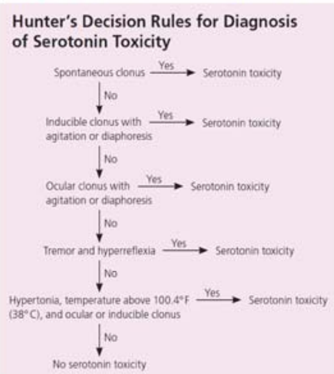
- Clinical Example:
- D … acute heart attack
- U_+… instable chest pain
- $p(D)$ … 37 of 1000 = 0,037 (heart attack)
- $p(\overline{D})$ … 963 of 1000 = 0,963 (no heart attack)
- 40% of patients report on instable chest pain
- $p(U_+|D) = 0{,}4$
- Unfortunately this symptoms also occur in 5 % of the healthy population
- $p(U_+|\overline{D}) = 0{,}05$
- We find the probability for a heart attack during this symptoms therefore by using Bayes' Rule:
- $p(D|U_+) = \frac{p(U_+|D)*p(D)}{p(U_+|D)*p(D)+p(U_+|\overline{D})*p(\overline{D})} = 0{,}235$

Slide 7-28: Example Clinical Case: Serotonin Syndrome



Boyer, E. W. & Shannon, M. (2005) The Serotonin Syndrome. *New England Journal of Medicine, 352, 11, 1112-1120.*

Slide 7-29 Differential Diagnosis on Serotonin Syndrome 1/2



Signs & Symptoms of Serotonin Syndrome

Agitation (restlessness)*	Multi-organ failure†
Diaphoresis*	Myoclonus*
Diarrhea*	Ocular clonus
Disseminated intravascular coagulation†	Rhabdomyolysis†
Fever above 100.4°F (38°C)	Shivering*
Hyperreflexia*	Tonic-clonic seizures†
Incoordination (ataxia)*	Tremor*
Mental status changes	
Confusion*	
Hypomania*	

*—Sternbach's diagnostic criteria require three of 10 signs and symptoms.
†—Extremely severe cases.

Ables, A. Z. & Nagubilli, R. (2010) Prevention, recognition, and management of serotonin syndrome. *American family physician, 81, 9, 1139.*

Slide 7-30 Differential Diagnosis on Serotonin Syndrome 2/2

Clinical condition	History	Vital signs	Clinical features
Anticholinergic syndrome	Use of tricyclic antidepressants or other anticholinergic drugs	Tachycardia, tachypnea, hyperthermia (usually 102.2°F [39°C] or below)	Dry mouth, blurred vision, mydriasis, flushed skin, agitation/delirium, decreased bowel sounds
Malignant hyperthermia	Administration of halogenated inhalational anesthetics or depolarizing muscle relaxants	Hypertension, tachycardia, tachypnea, hyperthermia (up to 114.8°F [46°C])	Diaphoresis, mottled skin, agitation, decreased bowel sounds, muscular rigidity, hyporeflexia
Neuroleptic malignant syndrome	Ingestion of antipsychotic medications	Hypertension, tachycardia, tachypnea, hyperthermia (above 105.8°F [41°C])	Sialorrhea, diaphoresis, pallor, stupor, mutism, coma, normal or decreased bowel sounds, lead-pipe rigidity, bradyreflexia

Ables, A. Z. & Nagubilli, R. (2010) Prevention, recognition, and management of serotonin syndrome. *American family physician, 81, **9, 1139.***

06 Naïve Bayes

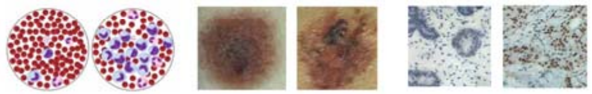
06 Practical Example: Naïve Bayes Classifier

- What can we do if the data sets d have different attributes?
- Naïve (simple - independent) Bayes assumption:
  - Attributes that describe data instances are conditionally independent given the classification hypothesis

$$P(\mathbf{d} \mid h) = P(a_1, \ldots, a_T \mid h) = \prod_t P(a_t \mid h)$$

What is classification?

- Predict labels y (classes C) for inputs x
  - Spamfilter (input: document, classes: spam / ham)
  - OCR (input: images, classes: characters)
  - Fraud detect (in: account activity, cl: fraud / no fraud)
  - Medical diagnosis (input: symptoms, classes: cancer / no cancer)



https://www.cancer.gov/about-cancer/understanding/what-is-cancer

MAP

- We can compute the *Maximum A Posterior* (MAP) hypothesis h_{MAP} for the data D
- We are interested in the best hypothesis for some hypothesis space $\mathcal{H}$ given observed training data D

$$h_{MAP} \equiv \underset{h \in H}{\operatorname{argmax}} \; P(h \mid D) = \underset{h \in H}{\operatorname{argmax}} \frac{P(D \mid h) P(h)}{P(D)} = \underset{h \in H}{\operatorname{argmax}} \; P(D \mid h) P(h)$$

Maximum Likelihood

- Now assume that all hypotheses are equally probable a priori, i.e., $P(hi) = P(h_j)$ for all h_i, h_j belong to H.
- This is called assuming a *uniform prior*.
- It simplifies computing the posterior:

$$h_{ML} = \underset{h \in H}{\arg\max} \; P(D \mid h)$$

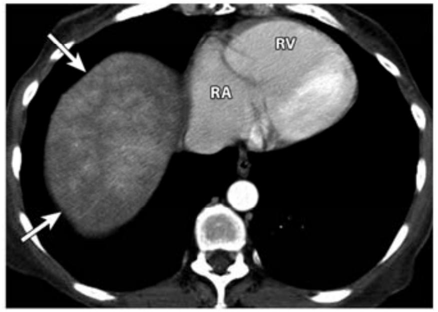
- This hypothesis is called the *maximum likelihood hypothesis.*

Why is this so interesting

- *Incrementality:* with each training example, the prior and the likelihood can be updated dynamically: flexible and robust to errors!
- *Combination of prior knowledge and observed data:* prior probability of a hypothesis multiplied with probability of the hypothesis given the training data
- *Probabilistic hypothesis:* outputs are not only a classification, but a probability distribution over all classes!

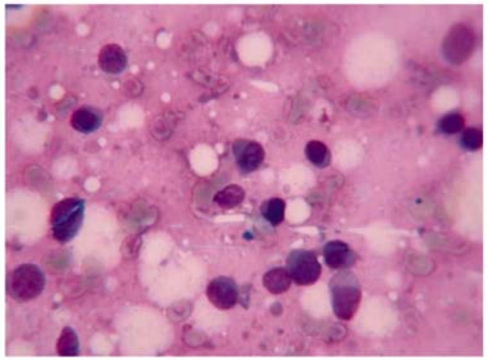
Example 1/4: carcinoid heart disease (chd)

Hepatic venous congestion and carcinoid heart disease secondary to an ovarian carcinoid tumor in a 56-year-old woman with elevated liver enzyme levels and right upper quadrant pain.



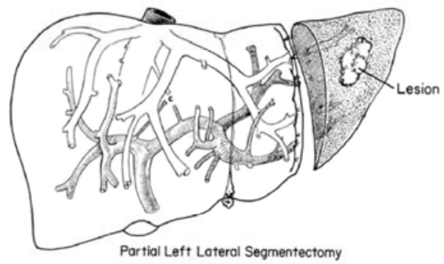
Shanbhogue, A. K. P., Shanbhogue, D. K. P., Prasad, S. R., Surabhi, V. R., Fasih, N. & Menias, C. O. (2010) Clinical Syndromes Associated with Ovarian Neoplasms: A Review. *Radiographics, 30, 4, 903-919.*

Example 2/4: bone-marrow depression (bmd)



Prasad, M., Maitra, A., Sethiya, N., Bharadwaj, V. K., Chowdhury, V., Valecha, J. & Biswas, R. (2009) Acute renal failure followed by low back ache. *BMJ Case Reports, 2009.*

Example 3/4: partial liver resection (plr)

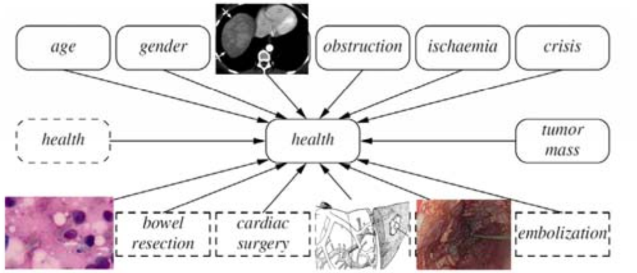


Zollinger, R. M. & Ellison, C. (2010) *Zollinger's Atlas of Surgical Operations (9th Edition). New York, McGraw Hill.*

Example 4/4: radiofrequency ablation (rfa)

van Vilsteren, F. G. I. et al. (2011) Stepwise radical endoscopic resection versus radiofrequency ablation for Barrett's oesophagus with high-grade dysplasia or early cancer: a multicentre randomised trial. *GUT.*

Slide 7-20 Modeling Patient health (1/2)



chd = carcinoid heart disease; bmd = bone-marrow depression; plr = partial liver resection;
rfa = radiofrequency ablation; dashed … past states; square objects … treatments

van Gerven, M. A. J., Taal, B. G. & Lucas, P. J. F. (2008) Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics, 41, 4, 515-529.*

Slide 7-21 Modeling Patient Health (2/2)

Let $\mathbf{U} \subseteq \mathbf{X}$ denote this risk factors and
Let $\mathbf{V} = \mathbf{X} \setminus \mathbf{U}$ denote the complement.

The risk of immediate death $p(health(t) = death|X)$ can be expressed by calculation of the following product:

$$\prod_{U \in \mathbf{U}} p(U)$$

Further, we obtain

$$p(health(t) = h|\boldsymbol{X}) = p(h|\boldsymbol{V}) \prod_{U \in \mathbf{U}} p(health(t) \neq death|U, health(t-1))$$

for $h \neq death$

van Gerven, M. A. J., Taal, B. G. & Lucas, P. J. F. (2008) Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics, 41, 4, 515-529.*

What can we do if we have not only probabilistic, but also I c pl e d t …

Rough Set Theory

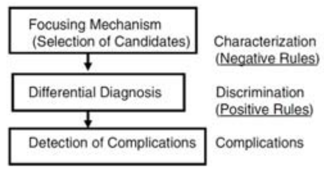
https://www.calvin.edu/~pribeiro/othrlnks/Fuzzy/fuzzyeng.htm

Slide 7-31 Rough Set Theory for dealing with incomplete data

- … is an extension of the Classical Set Theory, for use when representing incomplete knowledge.
- RS are sets *with fuzzy boundaries* – sets that cannot be precisely characterized using the available set of attributes, exactly like it is in medical decision making; based on 2 ideas:
- 1) a given concept can be approximated by partition-based knowledge as upper and lower approximation – which corresponds to the focusing mechanism of differential medical diagnosis: upper approximation as selection of candidates and lower approximation as concluding a final diagnosis.
- 2) a concept or observation can be represented as partitions in a given data set, where rough sets provides a rule induction method from given data. Thus, this model can be used to extract rule-based knowledge from medical databases.

Slide 7-32 Diagnostic Procedure (Differential Diagnostic)

Example Symptom: Headache



Tsumoto, S. (2006) Pawlak Rough Set Model, Medical Reasoning and Rule Mining. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H. & Slowinski, R. (Eds.) *Rough Sets and Current Trends in Computing. Berlin, Heidelberg, Springer, 53-70.*

Conclusion

- Medical action is continuing decision making
- Decision making is central in health informatics!
- Decision is inherently connected with cognition
- Studying algorithms that can learn from data needs understanding the human learning
- Bayes, Laplace, Gauß, … provided powerful tools for medical reasoning and decision support
- Evidence based medicine is following these approaches
- Implementation Examples in the following lectures

Thank you!

Questions

Quiz (1/2)

- Explain the Hypothetico-Deductive Method!
- What is the PDCA Deming wheel?
- Why is understanding intelligence a grand goal?
- Give an example for causality!
- When is the human better than a computer?
- When is the computer better than a human?
- Explain how humans learn from very few examples!
- What describes the Expected Utility Theory (EUT)?
- How do humans make a decision?
- What can we learn from Signal Detection Theory?
- How can an algorithm learn from data?
- Where is Bayes used in clinical practice?
- What is the problem with incomplete data?

Quiz (2/2)

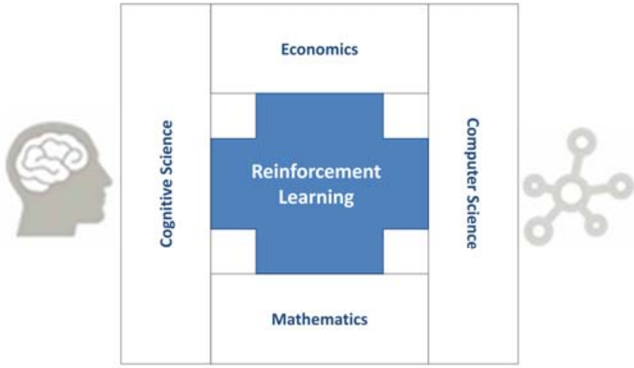
- Give three examples of where iML is beneficial in health informatics?
- What is the difference between Medical/Biomedical/Health Informatics?
- What are the key problems in health informatics?
- Why is both time and structure so important?
- What is life (in the sense of Erwin Schrödinger)?
- What are the building blocks of life?
- Please define BMI according to the AMIA!
- What are open problems in health informatics?
- What is personalized medicine?
- What is a biomarker? Why are biomarkers important?
- What is the famous time limit to reach a medical decision?

Appendix

Additional Reading

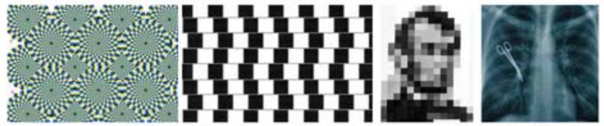
- **History of Probability Theory**
- Franklin, J. *The Science of Conjecture: Evidence and Probability Before Pascal.* John Hopkins University Press, 2001.
- Jaynes, E. T. *Probability Theory: The Logic of Science.* Cambridge University Press, 2003.
- **Probabilistic Reasoning**
- Gigerenzer, G., and D. J. Murray. *Cognition as Intuitive Statistics.* Hillsdale, NJ: Erlbaum, 1987.
- Gilovich, T., D. Griffin, and D. Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment.* Cambridge University Press, 2002.
- Kahneman, D., P. Slovic, and A. Tversky, eds. *Judgment under Uncertainty: Heuristics and Biases.* Cambridge University Press, 1982.
- **Bayesian Networks**
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman, San Mateo, CA, 1988.
- Breese, J. S. "Construction of Belief and Decision Networks." *Computational Intelligence* 8, 4 (1992): 624–647.
- F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. "Statistical Foundations for Default Reasoning." *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI).* Chambery, France, August 1993, pp. 563-569.
- **Multiple-Instance Bayesian Networks**
- Pasula, H., and S. Russell. "Approximate Inference for First-order Probabilistic Languages." *IJCAI-01.* Seattle, WA, 2001, pp. 741–748.
- Halpern, J. Y. "An Analysis of First-order Logics of Probability." *Artificial Intelligence* 46, 3 (1990): 311–350.
- D. Koller, and A. Pfeffer. "Object-Oriented Bayesian Networks." *Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI).* Providence, Rhode Island, 1997, pp. 302-313.

RL is multi-disciplinary and a bridge within ML



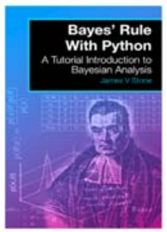
Slide 7-1 Key Challenges

- Time to make a decision = "5 Minutes" [1], [2]
- Limited perceptual, attentive and cognitive human resources [3], and Human error
- Noisy, missing, probabilistic, uncertain data



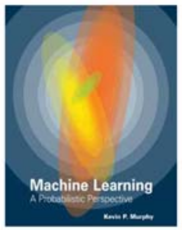
[1] Gigerenzer, G. 2008. *Gut Feelings: Short Cuts to Better Decision Making* London, Penguin.
[2] Gigerenzer, G. & Gaissmaier, W. 2011. Heuristic Decision Making. In: Fiske, S. T., Schacter, D. L. & Taylor, S. E. (eds.) Annual Review of Psychology, Vol 62. pp. 451-482.
[3] Bialek, W. 1987. Physical Limits to Sensation and Perception. Annual Review of Biophysics and Biophysical Chemistry, 16, 455-478.

Recommended Books



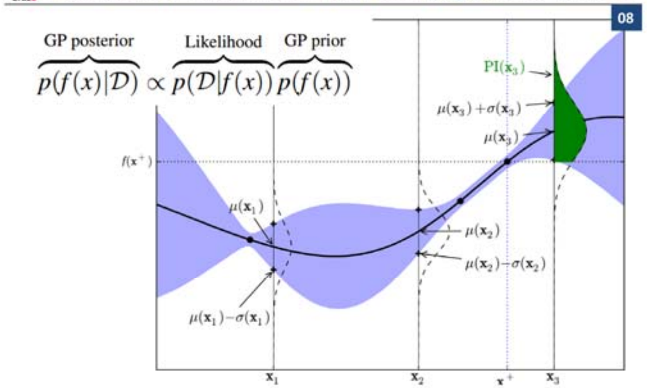
Stone, J. V. 2013. Bayes' rule: a tutorial introduction to Bayesian analysis. Sebtel Press. http://jim-stone.staff.shef.ac.uk

Barber, D. 2012. Bayesian reasoning and machine learning, Cambridge, Cambridge University Press.



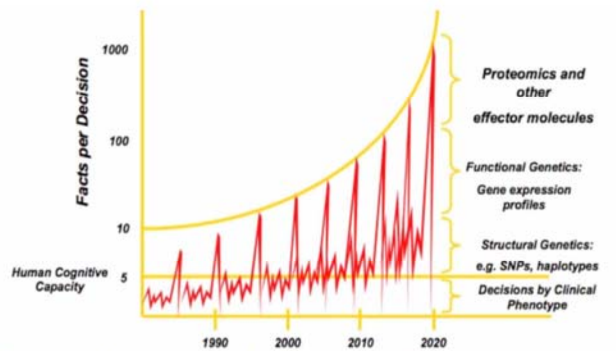
Murphy, K. P. 2012. Machine learning: a probabilistic perspective, Cambridge (MA), MIT press.

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



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.

Slide 7-47 Future Outlook



William Stead, IOM Meeting, 8 October 2007. Growth in facts affecting provider decisions versus human cognitive capacity.

More Questions (1)

- What is still considered the main and central topic in medical informatics?
- Please explain the information flow within the memory system according to Atkinson & Shiffrin!
- Explain the general model of human information processing following the model of Wickens!
- Explain the processing of visual (image, pictorial) information!
- Why is Attention of importance for medical informatics?
- Please explain the process of human decision making according to the model of Wickens (1984)!
- What is Triage?
- Please explain the hypothesis-oriented algorithm for Clinicians!

Sample Questions (2)

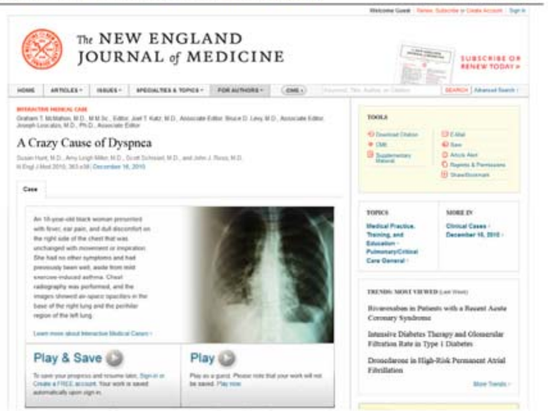
- What is the big difference between the Hypothetico-Deductive Method and the Plan-Do-Check-Act Deming Model?
- How can we model patient health – please provide an example!
- Please contrast the decision making process with the data mining process!
- Why is Signal Detection Theory important for us?
- Please provide an Example for the application of Bayes' Theorem!
- How does Differential Diagnosis work?
- How can we apply Rough Set Theory for differential diagnostics?

Some Useful Links

- http://www.anaesthetist.com/mnm/stats/roc
- http://sbml.org
- http://www.lcb.uu.se/tools/rosetta
- http://wise.cgu.edu/sdtmod/overview.asp (excellent Tutorial on SDT)
- http://www.iom.edu (Institute of Medicine)
- http://www.ahrq.gov/qual/patientsafetyix.htm (Agency for Health Care Research and Quality)
- http://www.fda.gov/drugs/drugsafety/medicationerrors/default.htm (Food and Drug Administration, FDA, medication errors)

Appendix: NEJM Interactive Multimedia Cases

Hunt, S., Miller, A. L., Schissel, S. & Ross, J. J. (2010) A Crazy Cause of Dyspnea. Interactive Multimedia Case *New England Journal of Medicine, 363, 25, e38.*



http://www.nejm.org/doi/full/10.1056/NEJMimc1008281

Slide 7-34 Rough Set Theory Example Symptom: Headache 2

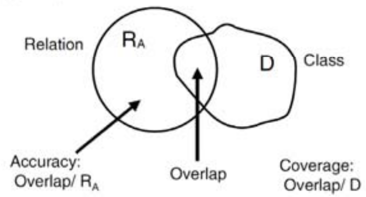
- The atomic formula over
- B ⊆ A ∪ {d} and V are expressions of the form [a = v]
- called descriptors over B, where a ∈ B and v ∈ Va.
- The set F(B, V) of formulas over B is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction and negation. For example, [location = occular] is a descriptor of B.
- For each f ∈ F(B, V), fA denote the meaning of f in A, i.e., the set of all objects in U with property f, defined inductively as follows.
- 1. If f is of the form [a = v] then, fA = {s ∈ U|a(s) = v}
- 2. (f ∧ g)A = fA ∩ gA; (f ∨ g)A = fA ∨ gA; (¬f)A = U − fa
- For example, f = [location = whole] and fA = {2, 4, 5, 6}. As an example of a conjunctive formula, g = [location = whole] ∧ [nausea = no] is a descriptor of U and fA is equal to glocation, nausea = {2, 5}.

Slide 7-35 Classification Accuracy and Coverage

Definition 1. *Let R and D denote a formula in $F(B, V)$ and a set of objects which belong to a decision d. Classification accuracy and coverage(true positive rate) for $R \to d$ is defined as:*

$$\alpha_R(D) = \frac{|R_A \cap D|}{|R_A|}(= P(D|R)), \text{ and}$$
$$\kappa_R(D) = \frac{|R_A \cap D|}{|D|}(= P(R|D)),$$

where $|S|$, $\alpha_R(D)$, $\kappa_R(D)$ and $P(S)$ denote the cardinality of a set S, a classification accuracy of R as to classification of D and coverage (a true positive rate of R to D), and probability of S, respectively.

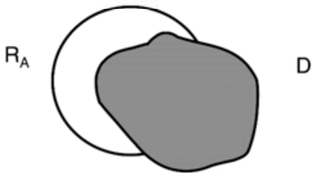


Tsumoto (2006)

Slide 7-36 Probabilistic Rules – modus ponens

By the use of accuracy and coverage, a probabilistic rule is defined as:

$$R \xrightarrow{\alpha,\kappa} d \quad s.t.\ R = \wedge_j [a_j = v_k], \alpha_R(D) \geq \delta_\alpha \text{ and } \kappa_R(D) \geq \delta_\kappa.$$



$$\boxed{R \to D \quad s.t. \quad \alpha_R(D) > \delta_\alpha,\ \kappa_R(D) > \delta_\kappa}$$

Tsumoto (2006)

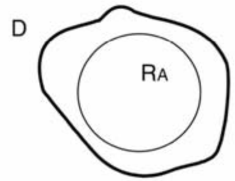
Slide 7-37 Positive Rules

A positive rule is defined as a rule supported by only positive examples, the classification accuracy of which is equal to 1.0. It is notable that the set supporting this rule corresponds to a subset of the lower approximation of a target concept, which is introduced in rough sets [1]. Thus, a positive rule is represented as:

$$R \to d \quad s.t. \quad R = \wedge_j [a_j = v_k], \quad \alpha_R(D) = 1.0$$

Figure 4 shows the Venn diagram of a positive rule. As shown in this figure, the meaning of R is a subset of that of D. This diagram is exactly equivalent to the classic proposition $R \to d$. In the above example, one positive rule of "m.c.h." (muscle contraction headache) is:

$$[nausea = no] \to m.c.h. \quad \alpha = 3/3 = 1.0.$$



Tsumoto (2006)

Slide 7-38 Exclusive Rules

Before defining a negative rule, let us first introduce an exclusive rule, the contrapositive of a negative rule [7]. An exclusive rule is defined as a rule supported by all the positive examples, the coverage of which is equal to 1.0. That is, an exclusive rule represents the necessity condition of a decision. It is notable that the set supporting an exclusive rule corresponds to the upper approximation of a target concept, which is introduced in rough sets [1]. Thus, an exclusive rule is represented as:

$$R \to d \quad s.t. \quad R = \vee_j [a_j = v_k], \quad \kappa_R(D) = 1.0.$$

Figure 5 shows the Venn diagram of a exclusive rule. As shown in this figure, the meaning of R is a superset of that of D. This diagram is exactly equivalent to the classic proposition $d \to R$. In the above example, the exclusive rule of "m.c.h." is:

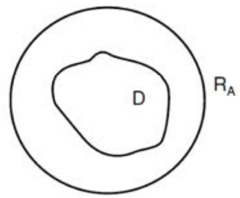
$$[M1 = yes] \vee [nau = no] \to m.c.h$$

From the viewpoint of propositional logic, an exclusi as:

$$d \to \vee_j [a_j = v_k],$$

because the condition of an exclusive rule correspo of conclusion d. Thus, it is easy to see that a ne contrapositive of an exclusive rule:

$$\wedge_j \neg [a_j = v_k] \to \neg d,$$



Tsumoto (2006)

Slide 7-39 Negative Rule

of conclusion d. Thus, it is easy to see that a negative rule is defined as the contrapositive of an exclusive rule:

$$\wedge_j \neg[a_j = v_k] \rightarrow \neg d,$$

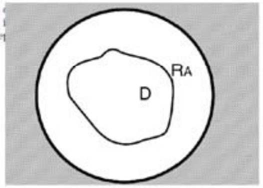
which means that if a case does not satisfy any attribute value pairs in the condition of a negative rules, then we can exclude a decision d from candidates. For example, the negative rule of m.c.h. is:

$$\neg[M1 = yes] \wedge \neg[nausea = no] \rightarrow \neg m.c.h.$$

In summary, a negative rule is defined as:

$$\wedge_j \neg[a_j = v_k] \rightarrow \neg d \quad s.t. \quad \forall[a_j = v_k]\ \kappa_{[a_j = v_k]}(D) = 1.0,$$

where D denotes a set of samples which belong to a class
Venn diagram of a negative rule. As shown in this figure,
negative region is the "positive region" of "negative concep

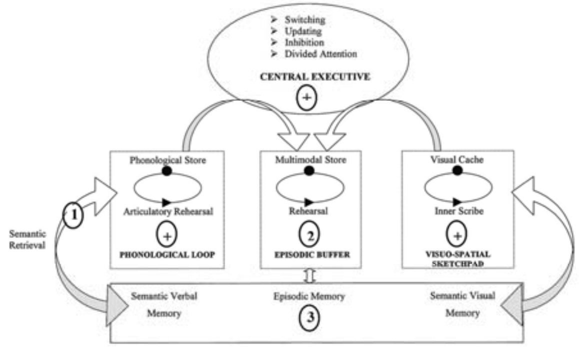


Tsumoto (2006)

Slide 7-40 Example: Algorithms for Rule Induction

```
procedure Exclusive and Negative Rules;
  var
    L : List;
      /* A list of elementary attribute-value pairs */
  begin
    L := P_0;
    /* P_0: A list of elementary attribute-value pairs given in a database */
    while (L ≠ {}) do
      begin
        Select one pair [a_i = v_j] from L;
        if ([a_i = v_j]_A ∩ D ≠ φ) then do /* D: positive examples of a target class d */
          begin
          L_ir := L_ir + [a_i = v_j];    /* Candidates for Positive Rules */
          if (κ_[a_i=v_j](D) = 1.0)
          then R_er := R_er ∧ [a_i = v_j];
            /* Include [a_i = v_j] into the formula of Exclusive Rule */
          end
      L := L − [a_i = v_j];
      end
    Construct Negative Rules:
        Take the contrapositive of R_er.
  end {Exclusive and Negative Rules};
```

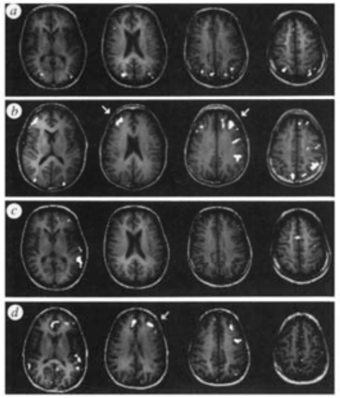

Alternative Model: Baddeley - Central Executive



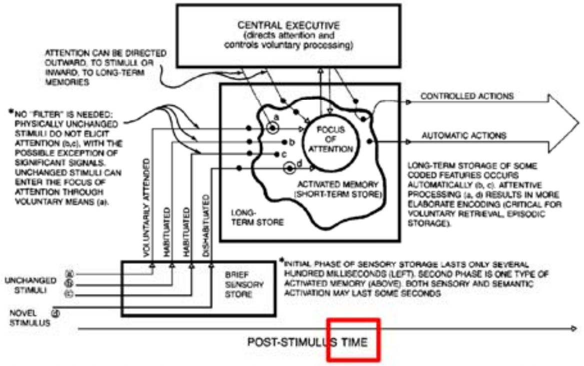
Quinette, P., Guillery, B., Desgranges, B., de la Sayette, V., Viader, F. & Eustache, F. (2003) Working memory and executive functions in transient global amnesia. *Brain, 126, 9, 1917-1934.*

Neural Basis for the "Central Executive System"

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S. & Grossman, M. (1995) The neural basis of the central executive system of working memory. *Nature, 378, 6554, 279-281.*



Slide 7-14 Central Executive – Selected Attention



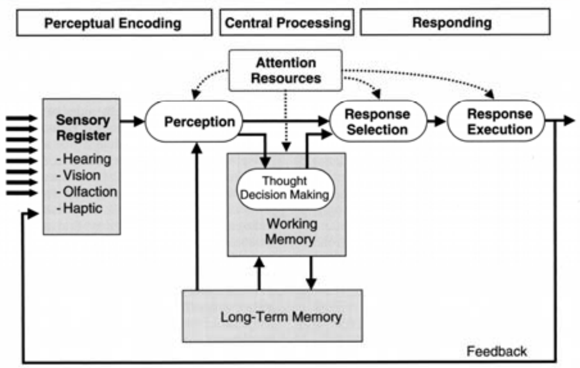
Cowan, N. (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104, 2, 163.*

Selective Attention Test

Note: The Test does NOT properly work if you know it in advance or if you do not concentrate on counting

Simons, D. J. & Chabris, C. F. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception, 28, (9), 1059-1074.

Human Attention is central for decision making



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