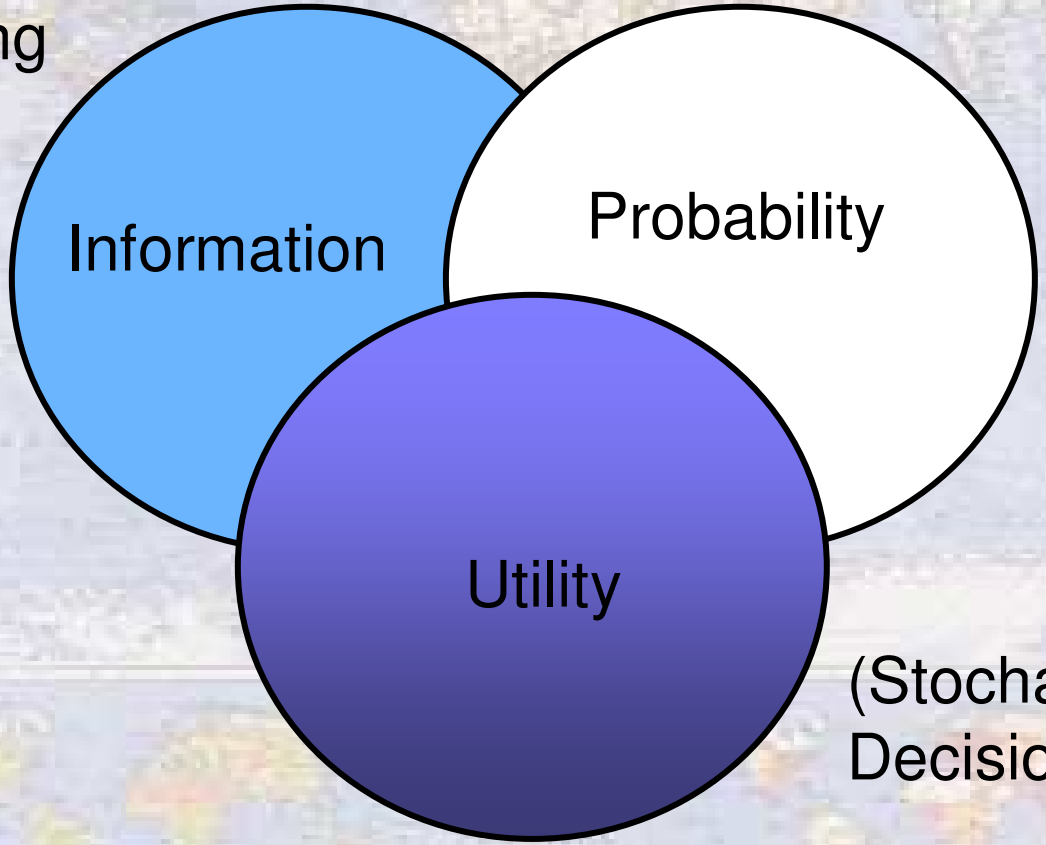
581287
Three concepts: Probability
Spring 2006

Petri Myllymäki
Complex Systems Computation Group
Department of Computer Science
University of Helsinki, Finland

http://www.cs.helsinki.fi/group/cosco/Teaching/Probability/

Three concepts

Compression,
coding, modeling



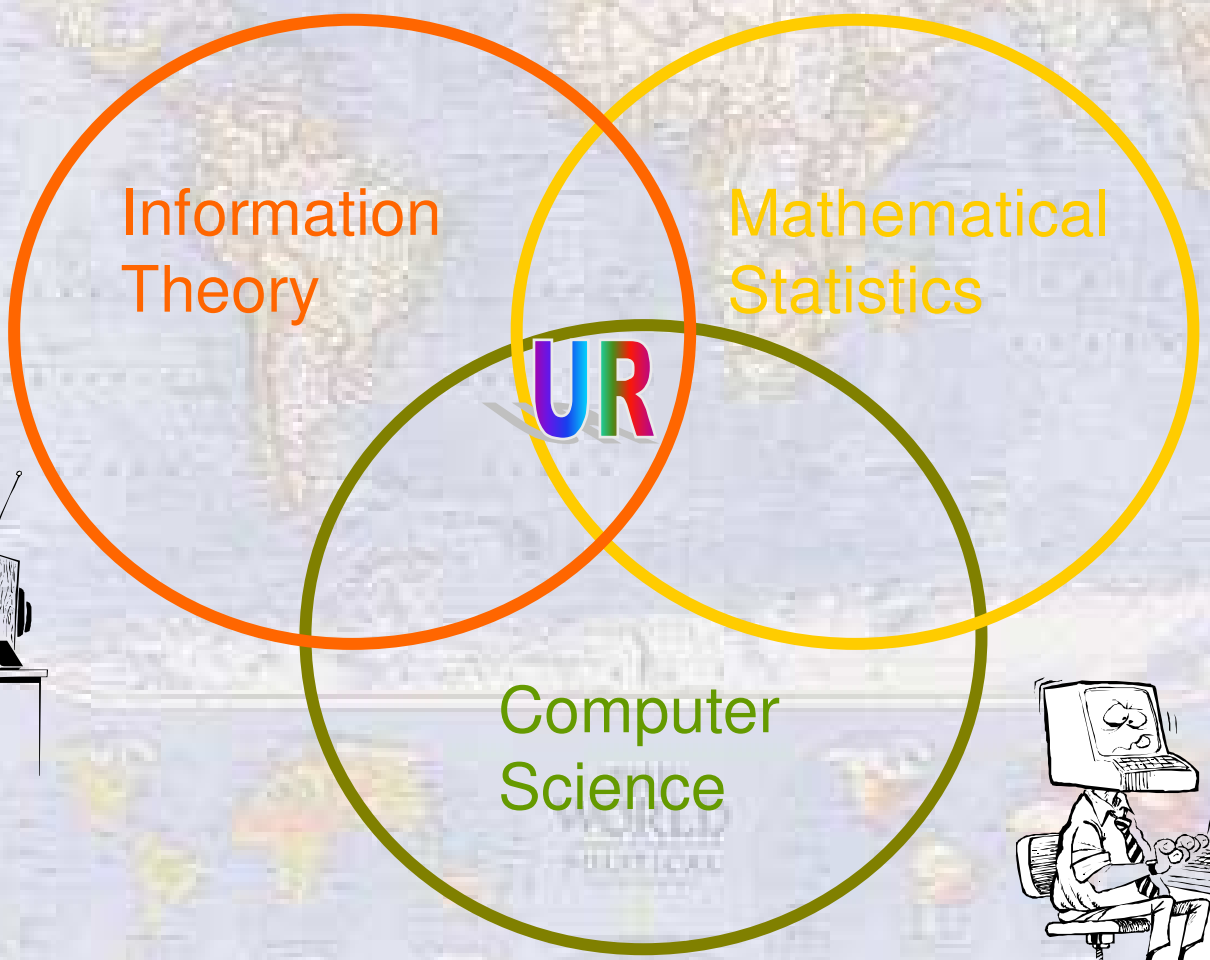
Uncertain
reasoning

(Stochastic) search
Decision making

On modeling

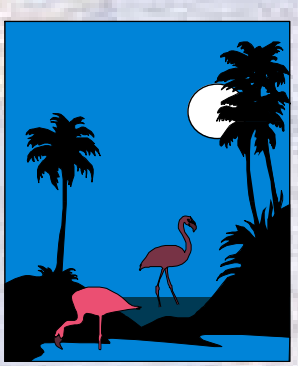
In building intelligent systems, statistics and in the rest of the world ...

Our motivation: Uncertain reasoning



Do I really need this stuff?

- "Machine learning"
- "Data mining"
- "Intelligent Systems"
- "Neural networks"
- "Pattern recognition"
- "Computational Intelligence"

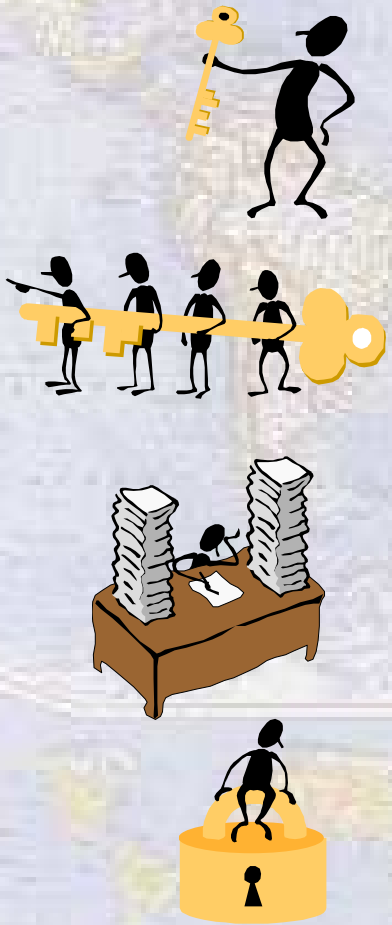


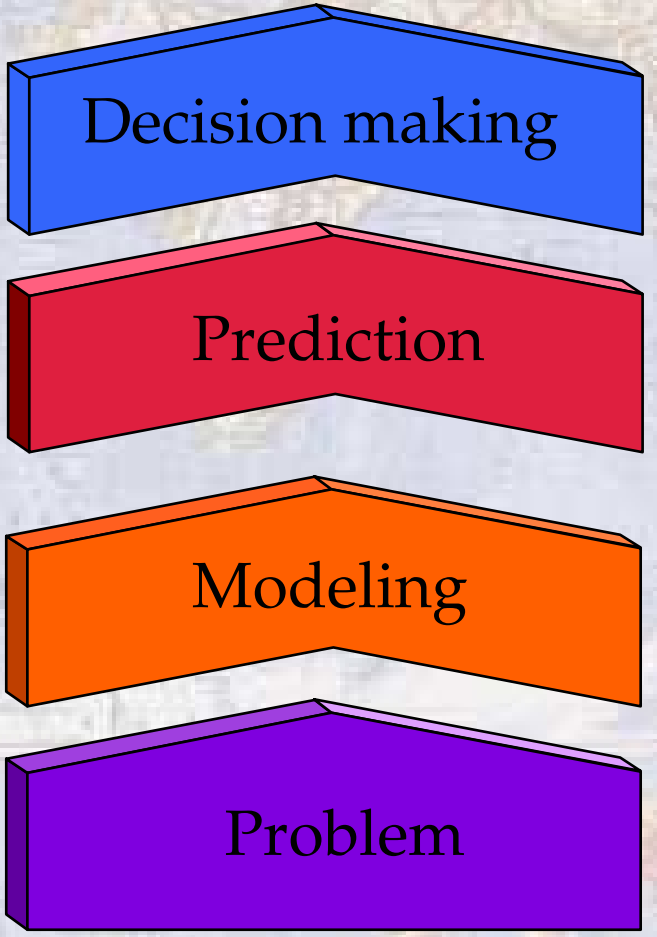
or

…and more?

- "Statistics"
- "Robotics"
- "Expert system design"
- "Medical informatics"
- "User interface design"
- "Artificial life" (Evolutionary computation)
- ….

Modeling framework



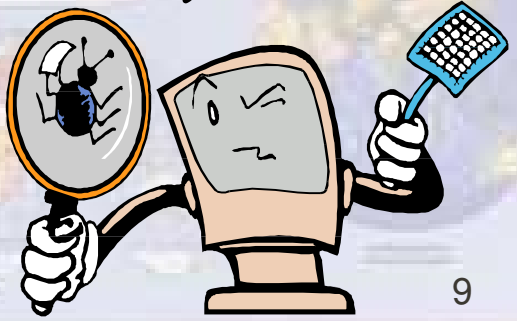


What does this mean?

- Problem: there is a need to model some part of the universe and make decisions based on the model
- Modeling: build the best model possible from a priori knowledge and data available
- Prediction: use the model to predict properties of interest
- Decision making: decide actions based on the predictions

For example

- Problem: online troubleshooting of software/hardware
- Modeling: build a latent variable (Bayes) model of the problems user encounters based on knowledge about the software and symptom data
- Prediction: use the model to predict the underlying problem given symptoms
- Decision making: propose actions to remove the problem (or to find more symptoms)



Microsoft Technical support

Bayesian email spam filters

- SpamBayes, OPFile, Outclass, bayespam, bogofilter, ifile, PASP, spamoracle, Spam Assassin Annoyance Filter, BSpam, Spam Bully, Death2Spam, InBoxer, …
- Software:
  - http://spambayes.sourceforge.net/related.html
- Background:
  - http://spambayes.sourceforge.net/background.html

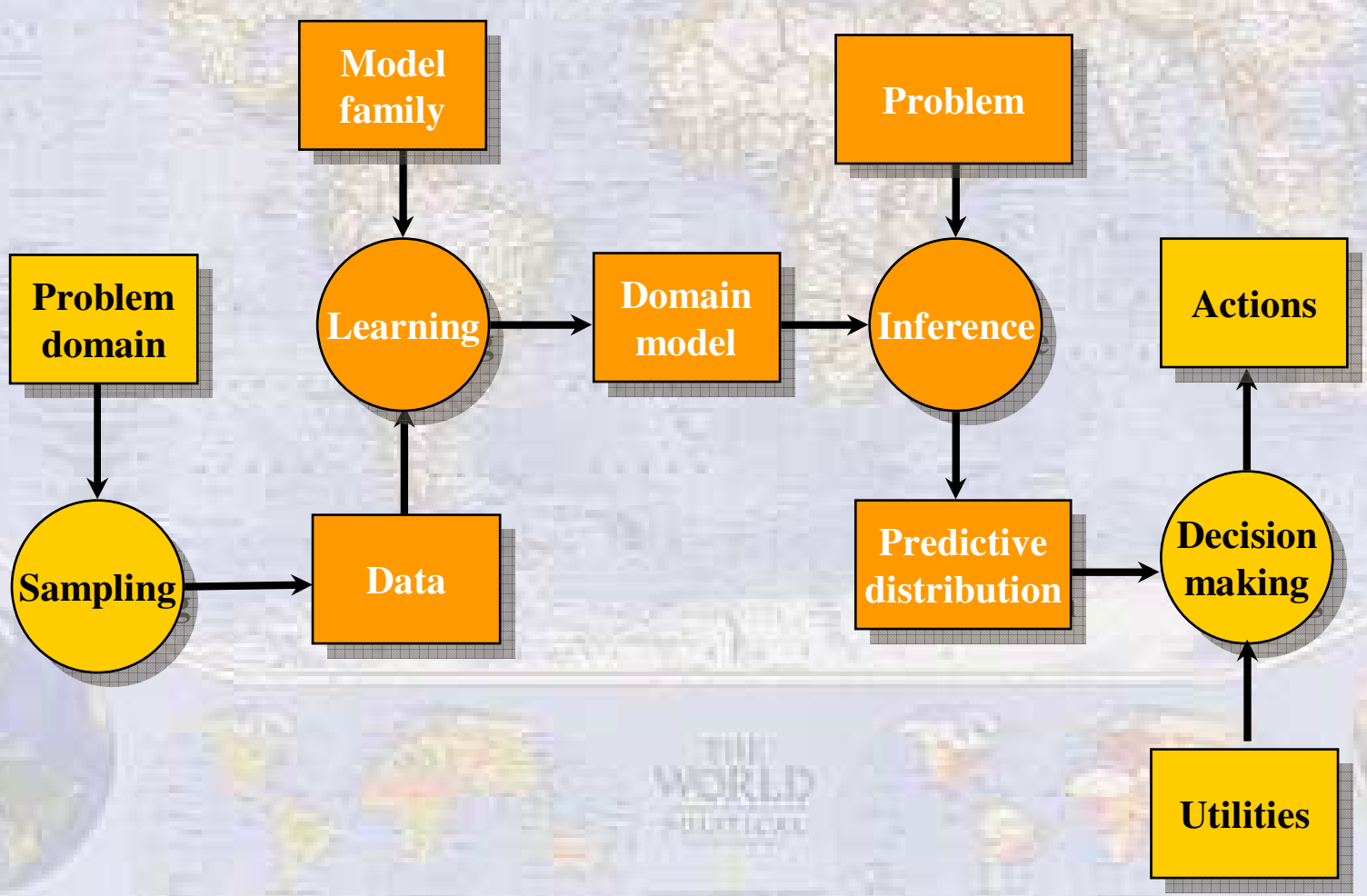
Real questions are ...

- Infinite number of models - what models do we consider?
  - Model is always chosen from a set of possible models!
- How do we compare models (i.e., measure that one model is better than another one) given some data?
- How do we find good models?

…and more

- How do we use the models to predict unobserved quantities of interest?
- What actions do we choose given the predictions?

General "rational agent" framework



Choice of models

- Simple models vs. complex models
- Linear models vs. non-linear models
- Parametric models vs. non-parametric models
- Flat models vs. structural models

Paradise lost

- What is simple and what is complex?*
  - The Catastrophic
    - ✓ Intuition: small, gradual changes in causes give rise to small, gradual changes in effects
  - The Chaotic
    - ✓ Intuition: Deterministic rules of behavior give rise to completely predictable events
  - The Lawless
    - ✓ Intuition: All real-world truths are logical outcome of following a set of rules
  - The Irreducible
    - ✓ Intuition: Complicated systems can always be understood by breaking them down into simpler parts
  - The Emergent
    - ✓ Intuition: Surprising behavior results only from complicated, hard-to-understand interactions among a system's component parts

* John. L. Casti, Complexification

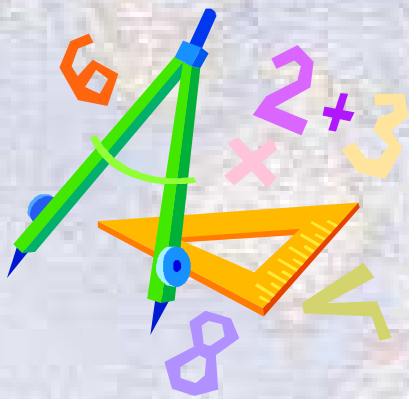
What is complex?

- what is complex is a totally nontrivial question
- one intuition: a complex model has more effective parameters
  - example: 100 factor model of IQ is more complex than 10 factor model!

$$P(D|M) = \frac{(l-1)!}{(N+l-1)!}\prod_{k=1}^{l}|c_k|!\prod_{j=1}^{m}\prod_{k=1}^{l}\left(\frac{n_j-1\ !}{(|c_k|+n_j-1)!}\prod_{i=1}^{n_i}F_k(a_{ji})!\right)$$

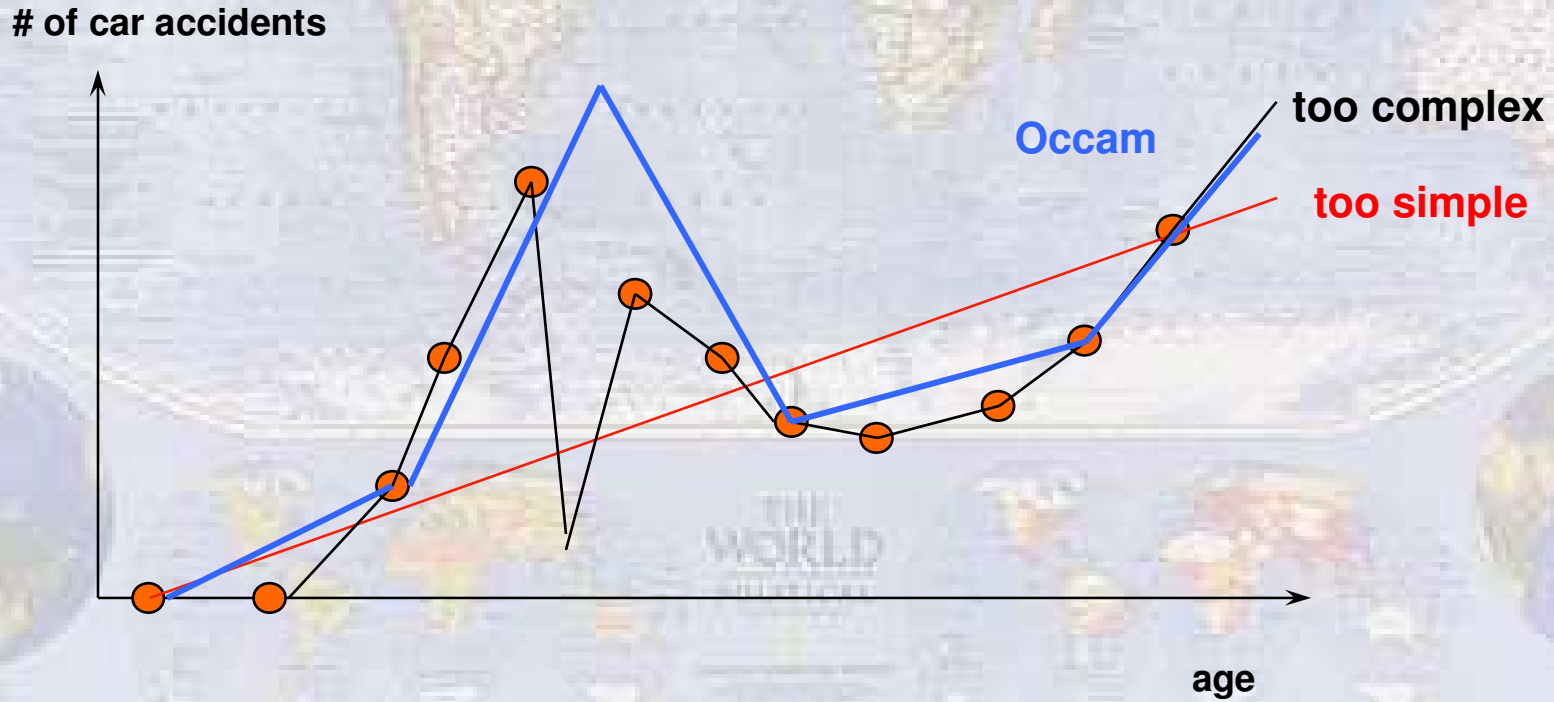
The Occam's razor principle

- The problem:
  - You are given the following sequence: -1, 3, 7, 11
  - Question: What are the two next numbers?
- Solution 1:
  - Answer: 15 and 19
  - Explanation: add 4 to the previous number
- Solution 2:
  - Answer: -19.9 and 1043.8
  - Explanation: if the previous number is x, the next one is $-x^3/11 + 9/11x^2 + 23/11$
- "Of two competing hypotheses both conforming to our observations, choose the simpler one."



Occam's Razor in Modeling

- there is a trade-off between the model complexity and fit to the data



Simpler models are better than complex models

- interpretation: they are easier to understand
- computation: predictions are typically easier to compute (not necessarily!)
- universality: they can be applied in more domains (more accurate predictions)
- "models should be only as complex as the data justifies"
- Bayesian model selection: automatic Occam's razor for model complexity regularization

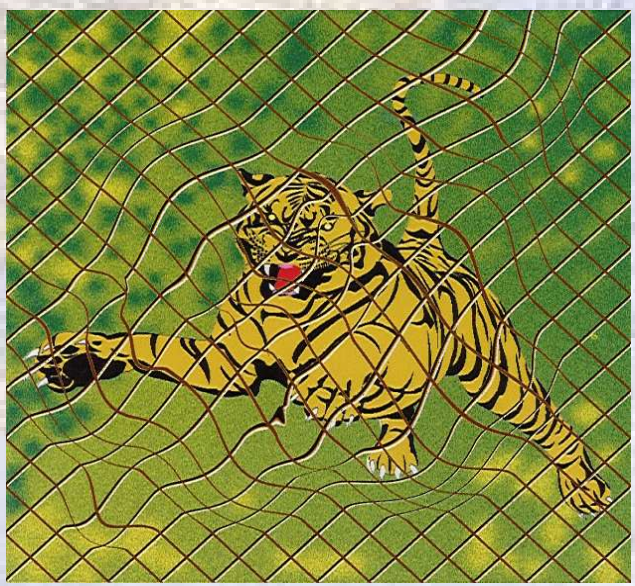
The source of great confusion

- Descriptive models ("Statistical modeling")
  - describe objects (e.g., data) as they are
  - typically exploratory structures
- Predictive models ("Predictive inference")
  - models that are able to predict unknown objects (e.g., future data)
  - models of the underlying process

Some viewpoints

- "prediction is our business"
- why the best fit to data is not the best predictor
  - data can be erroneous - perfect fit is too "specialized" and models the errors also!
  - a sample can only "identify" up to a certain level of complexity
- intuitive goal: minimize model complexity + prediction error - it keeps you honest!

Alternatives



- Probabilistic inference
  - Statistical inference
  - Bayesian inference
- Fuzzy inference
- Dempster-Shafer inference
- Non-monotonic logic

All models have probabilistic counterparts

Probability distributions

$$P(x) = K\,2^{-L(x)}$$

$$P(y \mid H, x) = \frac{1}{z} e^{-\beta ER(y|H,x)}$$

Kraft inequality

(Shannon-Fano) codes

$$-\log P(x)$$

Entropification

Non-probabilistic models with error function ER(y|H,x)