Machine Learning II: SS 19

Nils Bertschinger / Visvanathan Ramesh

Lecture Bridging Foundations/Applications

***With contributions from numerous collaborators in over 25 years.**

*Slide source credits: U Washington, Stanford U. (1994/1995), European Conference on Computer Vision 2010 presentation from Siemens AG (publically released industrial perspective), Jian-Bin-Huang and Joerg Bornschein, Patel et al (2015) (deep learning), BFNT-Frankfurt team (2011-2015).

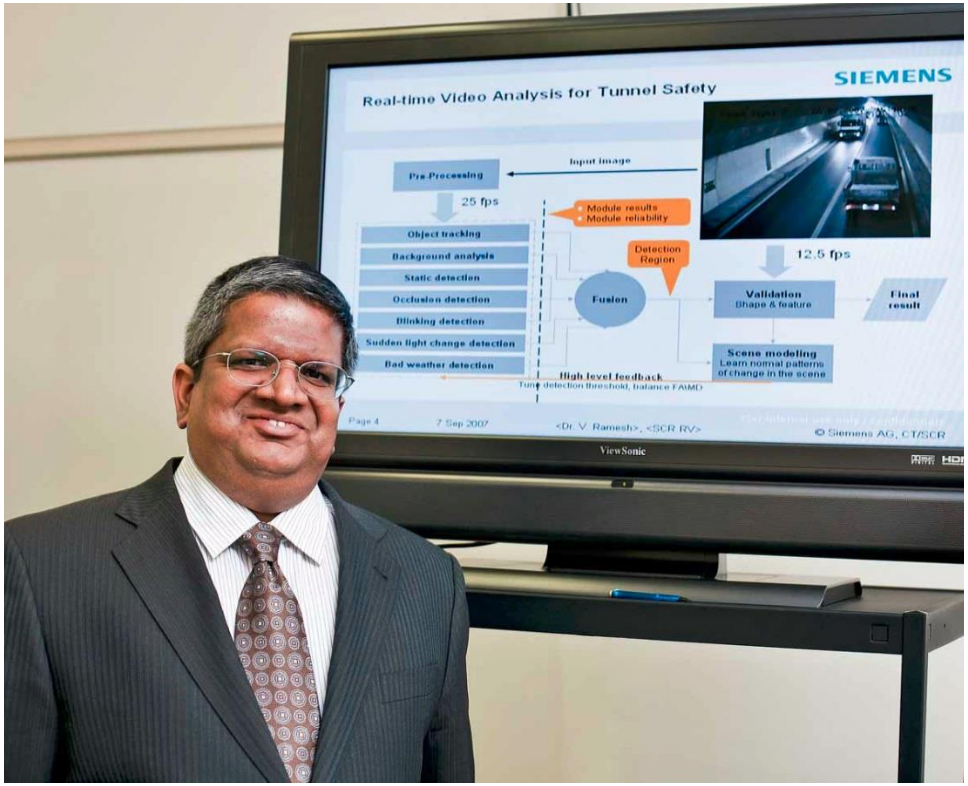
Software Engineering Chair (Intro)

My Background

- Bachelors in Engineering, India
- M.S. Electrical & Computer Engineering, Virginia Tech
- Ph.D. Electrical Engineering, U of Washington, Seattle.
- Education in Modeling, Analysis, and Validation of Complex Systems

- Industrial Career:
  - Led large international research organization
  - Numerous patents, publications, real-world systems
  - IEEE Best paper award (2000)
  - IEEE Longuet Higgins award (2010)

- Academic Career (2011-Present)
  - Professor of Software Engineering focusing on systems science & engineering for Intelligence



‘Siemens Inventor the Year’ – 2008 for contributions in ‘Real-time Vision & Modeling’

‘Holistic Systems Thinking in Science, Technology, Management and Education’

Learning over the years

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Present

Systems Engineering for Intelligence
(Goethe University, FIAS)

2011

Systems Engineering Applied in Practice (Real world Products in Industry, Energy, Mobility, Security, and Healthcare)
(Siemens Corporate Research)

1994

Systems Engineering for Vision
(Methodology for Model-based design and analysis of vision systems)
(U of Washington)

1987

Computer Engineering (Breadth, Virginia Tech)

1985

Electronics & Communication Engineering (India)

Competence

Focus of Software Engineering Chair Activity

Mobility

Industry, Energy

Healthcare

(Neben Taetigkeit)

Applications

Tools & SW platforms

Systems and Software Engineering Tools addressing entire lifecycle of Intelligent Systems -

Funding Sources:

Conti
Siemens
Others:
BMBF,
EU

GU/
FIAS

Systems Science & Engineering Research

Research in: Design patterns for Intelligent systems, Formal models of tasks, performance requirements, and contexts and mapping to approximate inference engines

BMBF
DFG
EU
Others

Teaching/Education Synergistic with Research

Courses in:
Bayesian Statistics, Machine Learning, Computer Vision, Software & Systems Engineering

GU

Recap from Prof: Bertschinger's past lectures

Approaches to Machine Learning

Two approaches to Machine Learning

- Data-driven
  - Very large data sets ... "Big Data"
  - Non-parametric models, e.g. k-NN
- Model-driven
  - Can be used for small data sets
  - Parametric models

Note: As models become more complex any data set is "small"

$\Longrightarrow$ Recent rise of model based machine learning

Model based Machine Learning

General setup of model based ML:

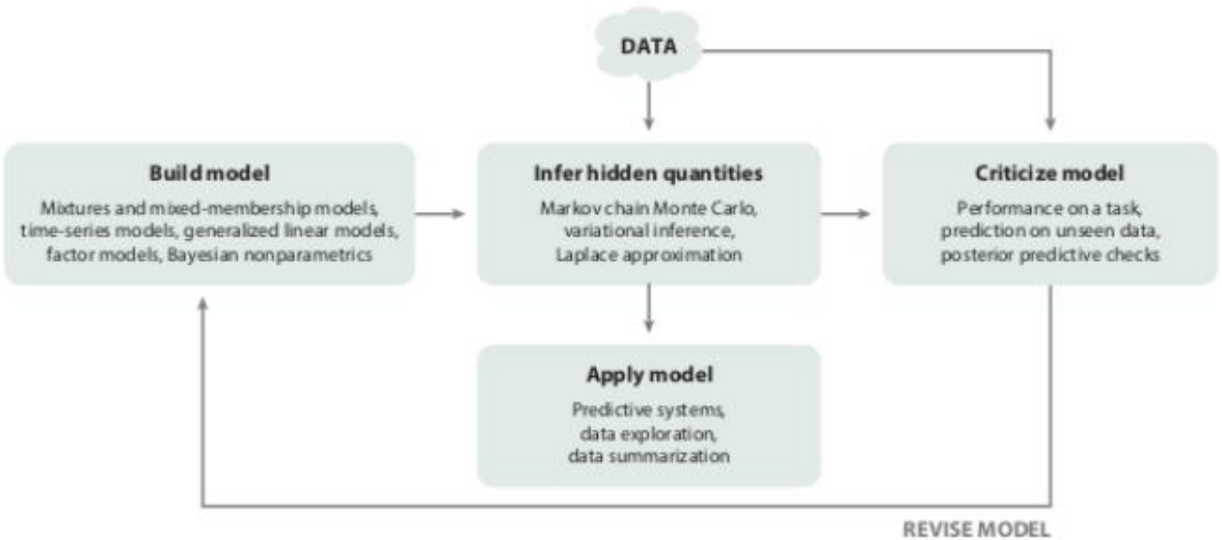


Fig. from: David M. Blei, *Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models*, Annu. Rev. Stat. Appl. 2014. 1:20332

Basic Terminology

- **Supervised:** Patterns whose class/output is known a-priori are used for training (*labelled training data*)
  - *Regression:* Real-valued output
    Typical examples: Interpolation, (Time-series) Prediction
  - *Classification:* Categorical output
    Typical examples: Face recognition, Identity authentification, Speech recognition
- **Unsupervised:** Number of classes is (in general) unknown and no labelled data are available
  Typical examples: Cluster analysis, Recommendation systems

Machine Learning 2: Topics So far

Machine Learning II course ... Focus on Bayesian methods

- Motivation: Bayesian vs frequentist statistics
- Probability theory: Conjugate priors
- Model selection: Marginal likelihood, sparsity priors
- Modeling:
  - Latent variable models
  - Bayesian non-parametrics: Gaussian processes
  - Deep neural networks
- Algorithms: Sampling methods, variational Bayes

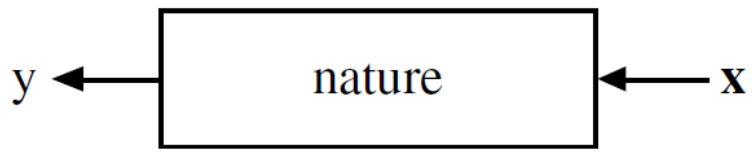
Potential applications

- Social data: Voting results, network models
- Economic data: GDP forecasting, volatility modeling
- Computer vision: Detection, tracking, recognition, segmentation
- ...

Philosophy - Modeling vs Algorithmic Cultures (Leo Breiman, 2001)

Statistics (Breiman, 2001)

Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables **x** (independent variables) go in one side, and on the other side the response variables **y** come out. Inside the black box, nature functions to associate the predictor variables with the response variables, so the picture is like this:

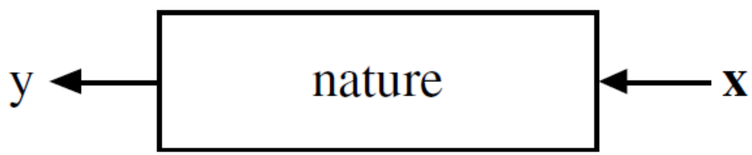


There are two goals in analyzing the data:

Prediction. To be able to predict what the responses are going to be to future input variables;
Information. To extract some information about how nature is associating the response variables to the input variables.

Statistics (Breiman, 2001)

Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables **x** (independent variables) go in one side, and on the other side the response variables **y** come out. Inside the black box, nature functions to associate the predictor variables with the response variables, so the picture is like this:



There are two goals in analyzing the data:

Prediction. To be able to predict what the responses are going to be to future input variables;
Information. To extract some information about how nature is associating the response variables to the input variables.

Data Modeling Culture

The analysis in this culture starts with assuming a stochastic data model for the inside of the black box. For example, a common data model is that data are generated by independent draws from

$$\text{response variables} = f(\text{predictor variables}, \text{random noise}, \text{parameters})$$

The values of the parameters are estimated from the data and the model then used for information and/or prediction. Thus the black box is filled in like this:

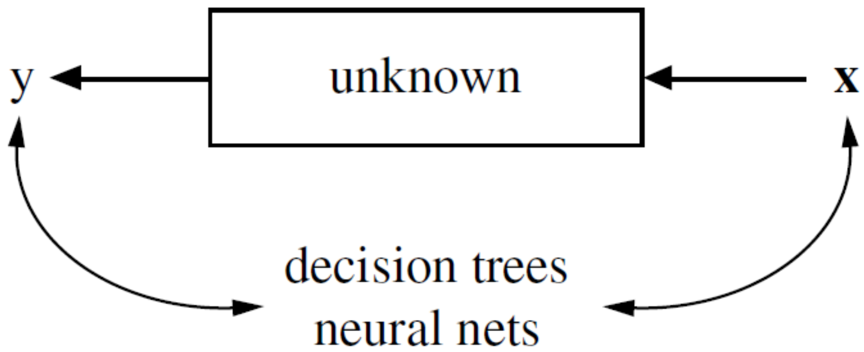
$$y \longleftarrow \boxed{\begin{array}{l}\text{linear regression}\\ \text{logistic regression}\\ \text{Cox model}\end{array}} \longleftarrow x$$

Model validation. Yes–no using goodness-of-fit tests and residual examination.
Estimated culture population. 98% of all statisticians.

ML Culture

The Algorithmic Modeling Culture

The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function $f(\mathbf{x})$—an algorithm that operates on $\mathbf{x}$ to predict the responses $\mathbf{y}$. Their black box looks like this:



Model validation. Measured by predictive accuracy.
Estimated culture population. 2% of statisticians, many in other fields.

Breiman on Statistical Analysis

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

3.3 Perceptions on Statistical Analysis

As I left consulting to go back to the university, these were the perceptions I had about working with data to find answers to problems:

(a) Focus on finding a good solution—that's what consultants get paid for.

(b) Live with the data before you plunge into modeling.

(c) Search for a model that gives a good solution, either algorithmic or data.

(d) Predictive accuracy on test sets is the criterion for how good the model is.

(e) Computers are an indispensable partner.

Breiman – Key Comments (continued)

- If the model is a poor emulation of nature, the conclusions may be wrong.

Rashomon: the multiplicity of good models;
Occam: the conflict between simplicity and accuracy;
Bellman: dimensionality—curse or blessing.

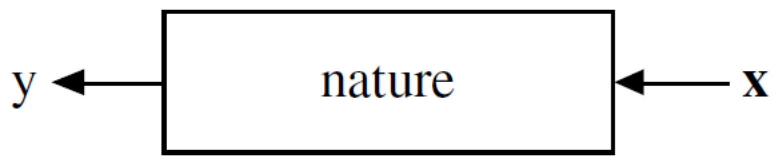
* Leading to Bagging, Boosting(VR)

- Accuracy generally requires more complex prediction methods. Simple and interpretable functions do not make the most accurate predictors.

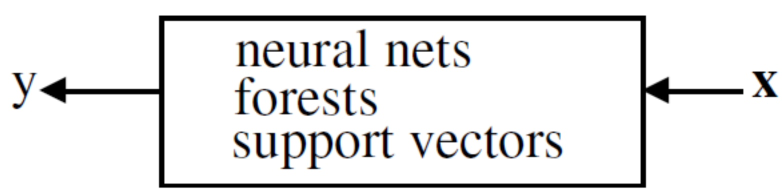
* Unless simple models reflect nature closely (VR)

Information from Black box (Breiman)

The dilemma posed in the last section is that the models that best emulate nature in terms of predictive accuracy are also the most complex and inscrutable. But this dilemma can be resolved by realizing the wrong question is being asked. Nature forms the outputs **y** from the inputs **x** by means of a black box with complex and unknown interior.



Current accurate prediction methods are also complex black boxes.



Breiman comments - continued

The point of a model is to get useful information about the relation between the response and predictor variables. Interpretability is a way of getting information. But a model does not have to be simple to provide reliable information about the relation between predictor and response variables; neither does it have to be a data model.

- The goal is not interpretability, but accurate information.

- Depends on the question – if the goal is systematic extensibility of the model space and validation for new problems then careful Analysis is needed.(VR)

Breiman’s arguments on Algorithmic approaches - continued

- Higher predictive accuracy is associated with more reliable information about the underlying data mechanism. Weak predictive accuracy can lead to questionable conclusions.
- Algorithmic models can give better predictive accuracy than data models, and provide better information about the underlying mechanism.

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Computer Vision lectures (V. Ramesh)

- Vision as Inverse Graphics
  - Vision as Bayesian Estimation
  - History & Examples
    - MRF's for Image Segmentation (Geman & Geman)
    - Bayesian methods for various vision sub-tasks – detection, tracking, recognition, motion analysis, etc. (various authors)
    - Conditional Random Fields (Kumar et al)
    - Pattern Grammars for Vision (Zhu, Mumford)
  - Probabilistic Programming for Vision (Kulkarni et al)
- Modern Practice in ML for Vision
  - Deep CNN's, Variational Auto-encoders, Generative Adversarial Networks
  - Link between modern ML and Bayesian viewpoints

Thank you!

Backup