1. Introduction

1.1 Learning and Statistical Estimation

- Problem of learning from data
- Goal of learning
  - predictive accuracy (generalization)
  - interpretation (explanation)
- first-principle model

  = basic scientific model + building applications

  -> data : verify model + estimate model parameters

- Paradigm shift

  classical modeling based on first-principles

  -> developing models from data

- Learning capabilities of biological system is done in a data-driven fashion
  - 1980's neural network
  - 1990's fuzzy rules

  -> neurofuzzy systems

- Statistical framework

  describe methods for learning from data

- Statistical estimation
  - predictive learning from data
  - known samples -> properties of statistical distribution
- Operation of learning system
  - learning/estimation
  - operation/prediction

- **Supervised learning**

  estimate unknown mapping from known samples

- **Unsupervised learning**
  - only input is given, no notion of output
  - estimate probability distribution of input
  - discover natural structure in the input data

General experimental procedures

- **Statement of the Problem**
  - domain-specific knowledge/experience

- **Hypothesis Formulation**
  - hypothesis specifies unknown dependency and is estimated from experimental data
  - close interaction between a modeler and application experts

- **Data Generation/Experiment Design**
  - designed experiment setting
  - observational setting (=random data generation)
  - sampling distribute is unknown and implicit in the data collection procedure
  - past data and future data come from the same sampling distribution

- **Data Collection and Preprocessing**
  - Outlier processing

    = detection/removal + robust modeling methods
  - variable scaling/different types of encoding techniques
  - selection of informative features from high-dimensional data

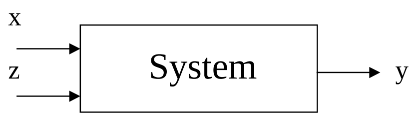
    = feature selection

    -> making the task of estimating dependency much simpler

- Model Estimation
  - main goal : construct models for accurate prediction of future outputs from (known) input vales
  - Goal of predictive accuracy = generalization
  - fixed parametric functions = linear in parameters
  - estimating nonlinear dependencies of an arbitrary form
- Interpretation of the Model and Drawing Conclusions
  - decision making
  - simple <-> complex dilemma
  - highly interpretable parametric models
  - high prediction accuracy + interpretation -> separate tasks

1.2 Statistical Dependency and Causality

- Statistical inference/learning system
  estimating unknown dependencies hidden in the data
- Statistical dependency <- unobserved factors



z : an unobserved/uncontrolled data

- Causality cannot be inferred from data analysis alone
  + argument outside the statistical analysis
- Common instances of learning problem
  - Manufacturing process control
  - Person’s height/weight
  - Life expectancy : place, marriage
  - Medical diagnosis

1.3 Characterization of Variables

- Numeric : order relation, distance relation
- Categorical : equal/unequal
- Periodic : numeric variable with distance relation
- Ordinal : categorical variable with order relation
  - closely related to linguistic or fuzzy variables
  - subjectively defined in a particular context
  - no crisp boundary
  - denote overlapping sets
  - a single (numeric) input value can belong (simultaneously) to several values of an ordinal/fuzzy variable

1.4 Characterization of Uncertainty

- Describing uncertainty is based on the notion of probability and statistical distribution

- Frequentist interpretation
  - probability = relative frequency of a random experiment
  - learning = estimating parameters/structures of the unknown input-output dependency from data and a priori knowledge about the problem

- Bayesian interpretation of probabilities
  - see probability as a subjective degree of belief
  - specifying a priori knowledge (encoded as a priori probability distribution)
  - combining this knowledge with data via Bayes theorem
  - Bayes formula provides a rule for updating priori probabilities after the data are known

  = Bayes inference = Bayesian inductive principle

- Fuzzy membership function
  - quantify the degree of subjective belief
  - specify the degree of partial membership

- Probability
  - describe randomness (uncertainty of event occurrence)

- Fuzziness
  - describe uncertainty related to event ambiguity (subject degree to which an event occurs)

- Bayesian/Fuzzy are useful for specification of a priori knowledge about unknown system

- Both provide subjective characterization of uncertainty