OPPA European Social Fund
Prague & EU: We invest in your future.

Machine Learning and Data Analysis – overview

Jiří Kléma

Department of Cybernetics,
Czech Technical University in Prague

http://ida.felk.cvut.cz

Syllabus

Lecture	Lecturer	Content
1.	J. Kléma	Introduction, (un)supervised learning. Cluster analysis, formalization.
2.	J. Kléma	Cluster analysis, EM algorithm, k-means, hierarchical clustering.
3.	J. Kléma	Spectral, conceptual, fuzzy clustering. Biclustering.
4.	J. Kléma	Frequent itemsets, Apriori algorithm, ascociation rules.
5.	J. Kléma	Frequent sequences, epizodal rules, sequence models.
6.	J. Kléma	Frequent subtrees/subgraphs.
7.	J. Kléma	Learning from texts and web, applications.
8.	F. Železný	Computational learning theory, concept space, PAC learning.
9.	F. Železný	PAC-learning logic forms, learning in predicate logic.
10.	F. Železný	Infinite concept spaces.
11.	F. Železný	Risk estimates, empirical validation of hypotheses.
12.	F. Železný	Inductive logic programming, least generalization, inverse entailment.
13.	F. Železný	Learning from logic interpretations, relational decision trees, relational features.
14.	F. Železný	Statistical relational learning, probablistic relational models, Markov logic.

Unsupervised learning. Descriptive models.

Symbolic learning – concepts.

Inductive a statistical learning of logic forms.

Unsupervised learning

:: Assumptions:

- exists an instance (observation) space X
  - real vectors, graphs, sequences, relational structures, . . . ,
- exists probability density P_X on X.

:: Learner receives:

- finite *sample* $(m \in N)$

$$S = \{x_1, x_2, \ldots, x_m\}$$

  drawn i.i.d. from P_X
  - S is a multiset, elements called *examples*.

:: Goals:

- general: learn P_X: density estimation task, or
- special: learn something about P_X: *manifold learning task*.

Density estimation

:: Non-parametric

- No prior knowledge about P_X
- Unfeasible in general
  - unless P_X very simple and/or m very large

:: Parametric, e.g.

- Mixture of multivariate Gaussian distributions
  - $X = R^n$
  - Number of mixed Gaussians known
  - Learned parameters: means $\vec{\mu}$ and covariance matrices Σ
- Bayesian networks
  - Usually $X = \{0,1\}^n$ (i.e., random events)
  - Independence structure (graph) known
  - Learned parameters: probabilities at vertices (CPT's)
- etc.

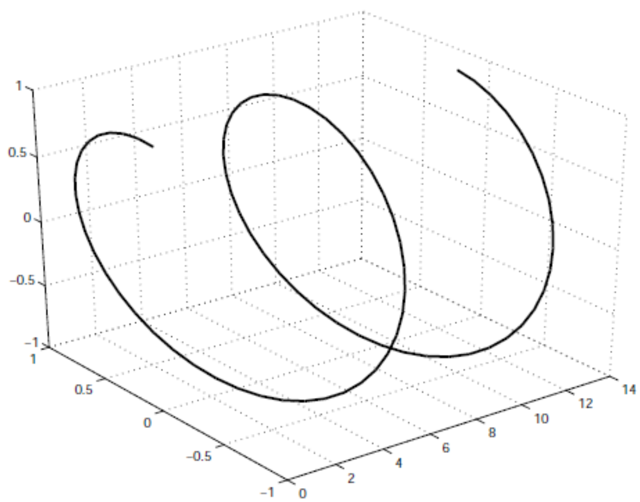
Manifold learning

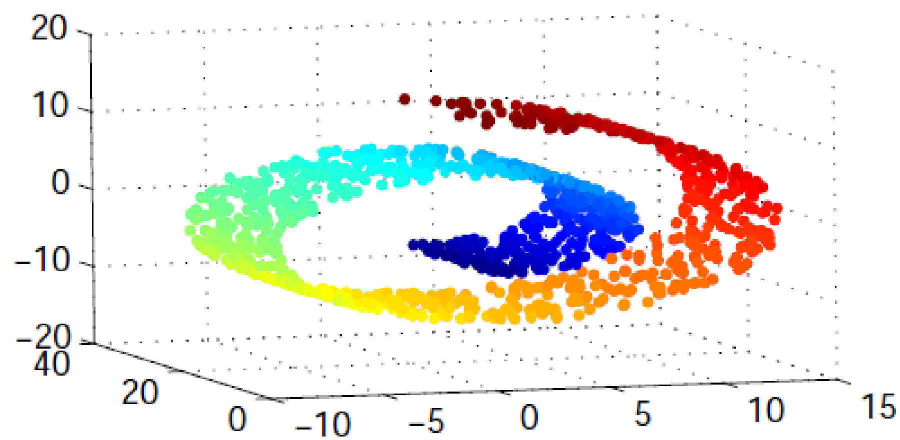
:: Manifold

- a topological space that on a small enough scale resembles the Euclidean space,
- globally typically nonlinear,

:: Learning

- identify a manifold dimension (it is embedded in a space of a higher dimension),
- project the problem (objects) into the low dimensional space – nonlinear dimension reduction,
- linear analogy: PCA or factor analysis.





Cayton: Algorithms for Manifold Learning.

Manifold learning – examples

:: Dimension reduction

- Linear – PCA, factor analysis
- Nonlinear – kernel PCA, locally linear embedding
- Learning in? Problem simplification, the transformation shows the manifold structure.

:: Clustering

- Learns partitions with high P_X
- Represented explicitly (examples assigned to partitions)

:: Pattern learning

- Patterns define manifolds of X with unexpectedly high P_X
- Frequent itemsets, subgraphs, subsequences, ...
- *How* do patterns define manifolds?

Supervised learning

:: Assumed:

- Instance (observation) space X
  - real vectors, graphs, sequences, relational structures, ...
- State space Y
  - also various kinds, but usually subsets of R
- Probability density P_{XY} on $X \times Y$

Learner receives

- Finite *sample* $(m \in N)$

$$S = \{(x_1, y_1), (x_2, y_2) \ldots, (x_m, y_m)\}$$

drawn i.i.d. from P_{XY}. S is a multiset, elements called *examples*.

Goal?

Supervised learning: goals

:: The most general goal, answer any question

- learn P_{XY}
  - principally same methods applicable as for learning P_X

:: The most often goal, to estimate the states y from observations x

- learn $P_{Y|X}$
  - this is more special than learning P_{XY}, why?

:: Estimates are single guesses, not distributions, so we need to learn only

- $f : X \to Y$ such that

$$f(x) = \arg\max_{y \in Y} P_{Y|X}(y|x)$$

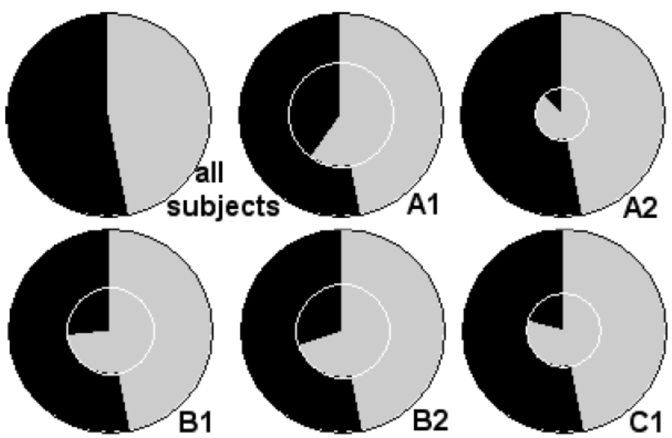
  - this is more special than learning $P_{Y|X}$, why?

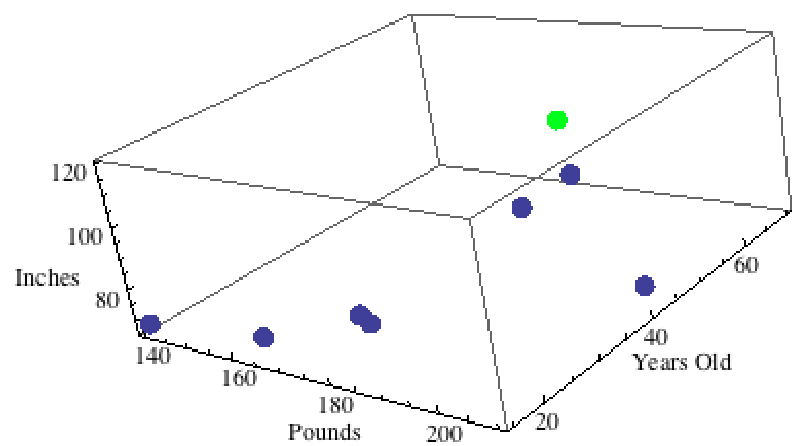
Data mining

- What is it?
  - "The great challenge in (biological) research today is how to turn **data** into **knowledge**. I have met people who think data is knowledge but these people are then striving for a means of turning knowledge into understanding." (Sydney Brenner – The Scientist, 2002)
  - Data mining is application of algorithms that extract **meaningful patterns**.
- Relation with learning
  - similar methods used,
  - DM emphasizes comprehensibility, originality and usability in practice,
  - rather technology than science.
- Unified theory
  - $T = \{\phi \in \mathcal{L} \mid \mathrm{q}(D, \phi) \text{ is true}\}$
  - $\mathcal{L}$ . . . a formal language (a countable formula set),
  - predicate q gives quality of the formula $\phi \in \mathcal{L}$ wrt the input data $D \subseteq X$,
  - T represents knowledge extracted from D, the formulae $\phi \in T$ are called patterns.

Descriptive models

- aim at **compressed data description**
  - focus on the intrinsic structure, relations, interconnectedness of the data,
- descriptive model taxonomy
  - focus – generate a global data model?
    - ∗ search for dominant structures
      - · subgroup discovery, segmentation, clustering, associations,
    - ∗ search for nuggets, deviation detection
      - · frauds, network attacks, harmful www pages,



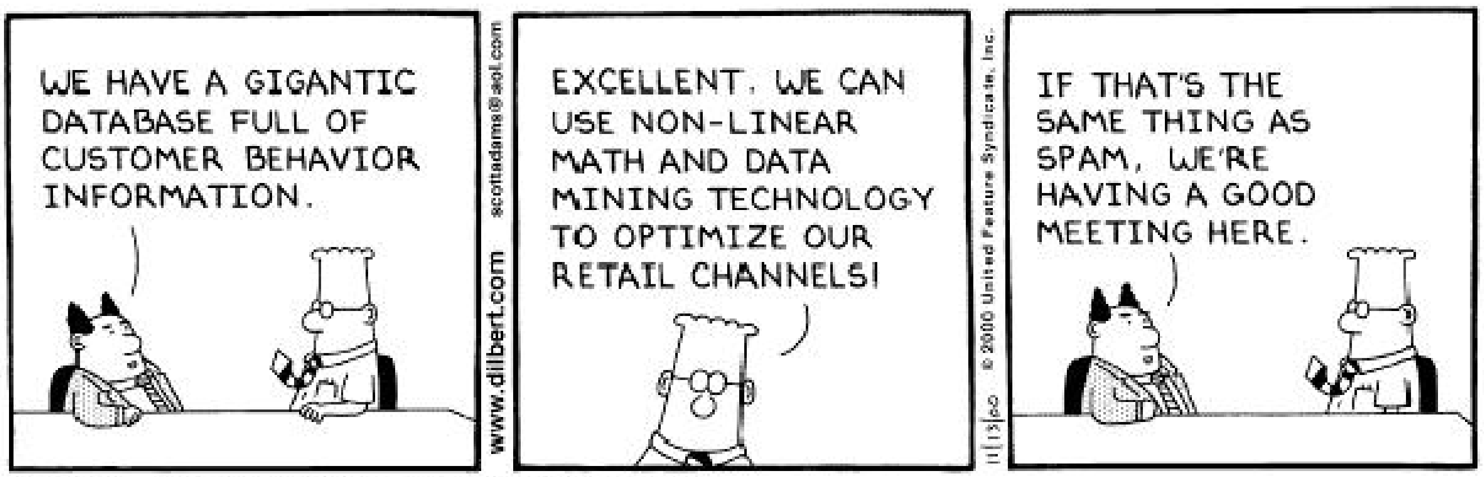


Descriptive models

- descriptive model taxonomy (continuation)
  - what type of models they use?
    - probabilistic models – data described by a probabilistic distribution
      - parametric, non-parametric, mixtures,
    - **symbolic** models – interpret data conceptually (with concepts and their relations),
      - graphs, rules, taxonomies, logic formulae,
      - characteristic: clearly and comprehensibly express knowledge,
    - combined models
      - eg. graphical prob models – bayesian networks, markov models,
  - input data formats?
    - numeric data, symbolic data, texts,
    - attribute representation, relational databases,
    - time series and sequential data.

Descriptive models – applications

- private companies
  - banks, insurance companies, business,
  - cost cutting, sale promotion, marketing, fraud detection,
- public sector
  - public administration, medicine, intelligent service,
  - efficiency, loss and fraud prevention.



Interactions with other courses

- direct prerequisite
  - A4B33RPZ – Rozpoznávání a strojové učení,
    * (non)bayesian decision making, loss minimization when deciding under uncertainty,
    * statistical learning – linear, kNN, SVM, neural nets,
    * parameter estimation using data – likelihood, EM algorithm,
- connections
  - A4B33ZUI – Základy umělé inteligence,
  - A0B01LGR – Logika a grafy,
  - A4B33FLP – Funkcionální a logické programování,

  - AE4M33GMM – Boris Flach, Graphical Markov Models.

OPPA European Social Fund
Prague & EU: We invest in your future.