TÉCNICO+
FORMAÇÃO AVANÇADA

Introduction to Machine Learning

Descriptive *versus* predictive modeling

DASH: Data Science e Análise Não Supervisionada

Rui Henriques, rmch@tecnico.ulisboa.pt
Instituto Superior Técnico, Universidade de Lisboa

Outline

- Intelligence and learning
- Descriptive *vs* predictive learning
- Machine Learning, Data Science and AI
- Supervised *vs* unsupervised learning
- Terminology
- Machine learning tasks

Intelligence

- **Rationality**
  - ability to act in a way that maximizes some utility function
- **Curiosity**
  - ability to engage creative imaginative or inquisitive reasoning
- **Adaptability** ⇐
  - ability learn from experience
    - make abstractions (patterning)
    - deal with novelty and change

Artificial Intelligence

Qualities of intelligence?

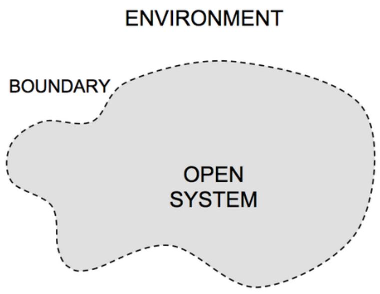
- AI with a focus on <u>learning</u> from experience => **machine/deep learning**, **reinforcement** learning …
- AI with a focus on <u>rationality</u> => **planning**, **reasoning**, **optimization** …
- AI with a focus on <u>curiosity</u> => autonomous agents, **affective computing**, …
- AI with a focus on <u>social intelligence</u> => human-agent interaction, social robots, **multi-agent systems**…

AI and Data Science

- (**machine**) **learning** is a fundamental quality of (**artificial**) **intelligence**
  - experience – ***data*** records – needed to learn intelligent agents!
- **data science** is the *principled application* of machine learning to solve data-rich problems
  - at the core of many breakthroughs
    - **chatbots**, **smart assistants**
    - personalized **recommenders**
    - **self-driving vehicles**
    - **intelligent care**
    - **cybersecurity**
    - …

Systemic world view

- **system**
  - set of elements organized with a shared purpose
  - (open) surrounded and influenced by its environment
  - described by its structure, purpose and functioning
- *open systems* evolve



- Universe → galaxy → solar system → Earth → societies → individuals → organs → cells → atoms

Systemic world view

- Everything is systemic:
  - **biological** systems
  - **ecological** systems
  - **societal** systems
  - **mechanical** systems
  - **digital** systems
  - **quantum** systems
  - **hybrid** systems
  - **astrophysical** systems

“we contain, are, interact and move within systems”

Psychoanalyst: *Know the influence of the surrounding systems in our life and be free!*

- The **behavior** of all these systems can be **monitored** (e.g. sensorization, observation)

Data everywhere!

Sensorization examples:

- **biological** systems
  - physiological signals from biosensors, molecular signals using multi-omic high-throughput profiling
  - health records (diagnostics, prescriptions, undertaken surgeries), exposomics, demographics
- **ecological** systems
  - biodiversity, plant health, crop and livestock conditions, food nutrition, forestry and fishery surveillance monitored using remote vision (satellite, drones), physical sensors, acoustic sensors, citizen notifications
- **societal** systems
  - social interactions via social networks, telecom and messaging apps
  - commerce and finance via transaction records
- **urban** systems
  - mobility from mobile phones, smart card validations, loop counters, privacy-preserving cameras
  - water and energy supply via telemetry (flowrate, pressure, smart sensors)
- … [homework: *complete the list*]

From experience to learning

- Recall: **learning** as a fundamental quality of **intelligence**
  - *"learning is any process by which a system improves performance from experience"* (Herbert Simon)
- Experience recorded as data observations acquired from:
  - **multiple systems** of the same type
    - e.g. multiple individuals (cohorts), vehicles, computers, regions, organizations
  - single system instance under **different conditions**
    - e.g. brain under different stimuli, crop under different weather, e-commerce along time
- Learn from the experience (records): multiple data observations… **statistics**!
  - discover relevant associations (patterning)

Statistical grounds

- Machine learning as…
  - the rediscovery of multivariate ***statistics*** from observable data (samples)
    - beyond descriptive and inferential statistics
  - the rediscovery of ***maths*** (linear algebra, calculus…)

- Exercise: associate naïve Bayes, kNN and deep learning (DL) with the aforementioned fields
  - *solution*: NB: statistics, kNN: algebra, DL: calculus

Sir William Petty, a 17th-century economist who used early statistical methods to analyse demographic data

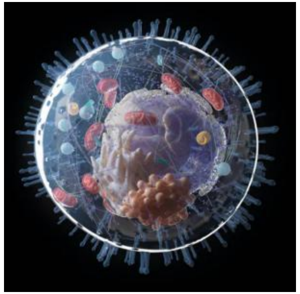
Descriptive *vs* predictive learning

- **Major ends** of pattern recognition (learning from experience):
  - understanding system behavior (**descriptive learning**)
    - data ⇒ information ⇒ knowledge
    - *unsupervised learning*: learning in the absence of outcomes of interest
    - *supervised learning*: outcomes of interest can be used to guide description
  - supporting decisions (**predictive learning**)
    - data and outcomes ⇒ decision support system
    - *supervised* or *semi-supervised learning*
      - semi-supervised when outcomes are only known for a subset of all observations

When?

- *Descriptive learning* (e.g., finding associations, categories, anomalies, summaries, informative features, representations) when:
  - human expertise does not exist (e.g. navigating on Mars, new cyberattacks)
  - humans cannot explain their expertise (e.g. speech recognition)
  - models must be customized (e.g. personalized medicine, user preferences)
  - massive amounts of data (e.g. genomics, commerce, social media, web usage)

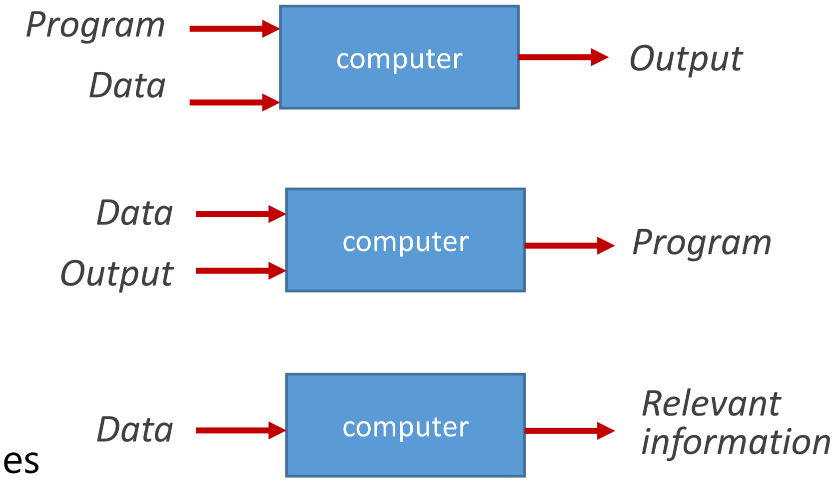


by Eric Eaton

Learning stances

- **classic AI**
  - e.g. heuristics to play chess, drive a car, recommend products, diagnose
- **supervised learning**
  - e.g. experience from good chess players, drivers, liked recommendations, clinical histories for decision support
- **unsupervised learning**
  - e.g. understand decisions, detect anomalies and behavioral patterns, summarize actions



"Machine Learning: field of study that gives computers the ability to learn without being explicitly programmed" Arthur Samuel (1959)

Data Science and Machine Learning

- Machine Learning *versus* **Data Science**
  - ML as the foundations, principles and algorithms to learn from available data
    - grounded on statistical, algebraic, mathematical and algorithmic foundations
  - Data science is the principled application of ML to solve specific data-rich problems
    - in other words, ML is a *means* to data science
    - data science has been termed as the <u>art of discovering</u> what we don’t know from data
- Machine Learning *versus* **Artificial Intelligence**
  - ML is a topic within the larger AI field
    - AI topics: optimization, search/planning, knowledge systems, autonomous agents

Coming to terms with terms

What about…

- Data-centric AI, Agentic AI
- Data Mining, Multivariate Data Analysis
- Business Intelligence

… and **terminology** choices…

- AI *vs* intelligent systems *vs* intelligent agents
- variable *vs* attribute *vs* feature
- observation *vs* instance *vs* object *vs* record *vs* data point

Structured view on ML

Learning

HOW

approaches

INPUT DATA

OUTPUT MODEL

feature data
text data
time series, event data
image, video data
heterogeneous data

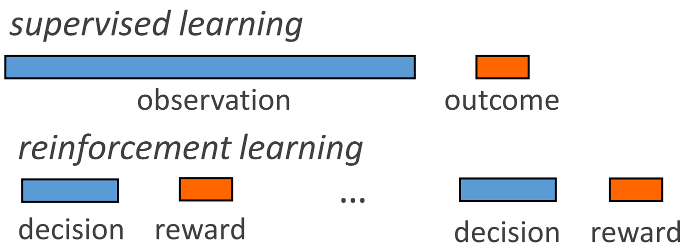
predictive models (supervised)
descriptive models (unsupervised/supervised)
prescriptive models (reinforcement)

Learning input-output functions

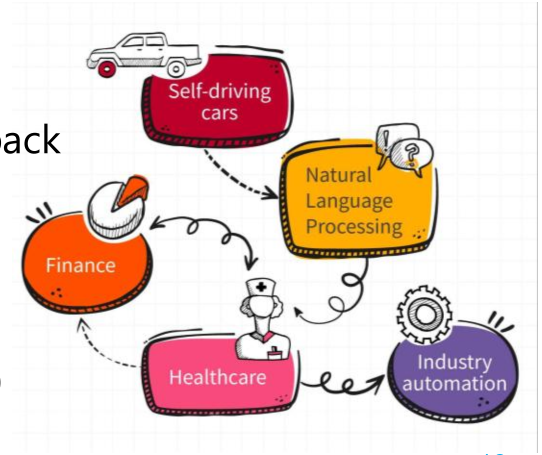
- **Supervised** learning
  - with a teacher (that tells you the ground truth)
  - learning from training data: pairs of inputs and outputs (labels, quantities, structures)
- **Unsupervised** learning
  - without a teacher
  - learning from training data without outputs (e.g. find associations, clusters/categories, anomalies)
- **Reinforcement** learning
  - with a teacher (that highlights both good and bad outputs)
  - learning rewards and penalties observed from sequence of decisions within a given environment

Reinforcement learning?

- How can we have agents making decisions with **little** or **no prior knowledge**?
  - **trial-and-error** (*reinforcement learning*) + learning from available observations



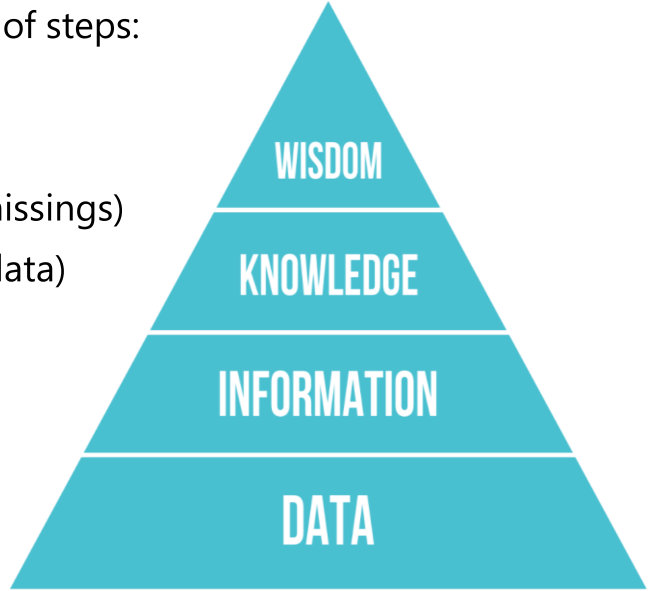
- In practice…
  - conversational AI, e-mail bots, recommendations…
    - adjusting predictors from ongoing pos or negative(!) feedback
  - self-driving car in simulated environments
    - rewards and penalties according to (un)desired risks
  - optimization in industry (e.g., automation) and healthcare (e.g., therapeutics) based on ongoing pairs (protocol, outcome)



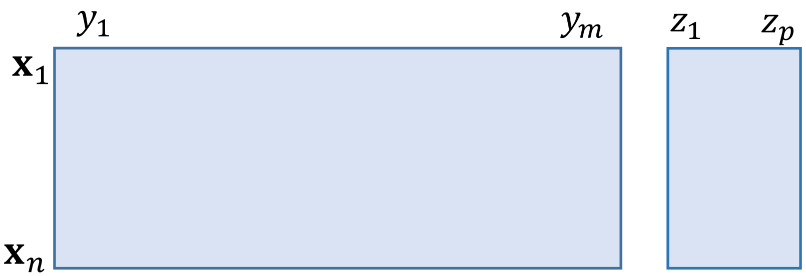
Learning in practice

Recall: knowledge discovery from data (**KDD**) is a composition of steps:

- **data acquisition** and integration
- **data preprocessing**
  - *data cleaning* (e.g. handling noise, duplicates, outliers, missings)
  - *data representation* (e.g. extract features from complex data)
  - *data transformation* (e.g. feature engineering, sampling, normalization, dimensionality reduction)
- **data mining** using *machine learning*
- **postprocessing**, explainability and **knowledge acquisition** from descriptive models or predictive models
- **validate**, consolidate and **deploy** discovered knowledge



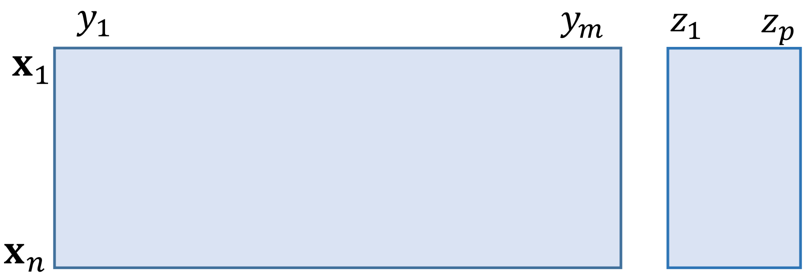
Terminology



Multivariate data:

- set of <u>observations</u>, $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ (population)
- with values/<u>features</u> along a set of <u>variables</u>, $Y = \{y_1, \dots, y_m\}$
  - <u>input variables</u> (explanatory) and optional output variables (<u>targets</u>), $Z = \{y_1, \dots, z_p\}$
- data <u>size</u> = number of observations, $|X| = n$
- data <u>dimensionality</u> = number of input variables, $|Y| = m$

Learning from data



Learning from data: retrieving relevant patterns

- relations/patterns/abstractions ≡ associations of interest on specific observations and variables
  - *unexpectedly informative*
  - *unexpectedly discriminative* (of one or more targets)
- use these relations to learn descriptors, classifiers, regressors, multi-output predictors, forecasters…

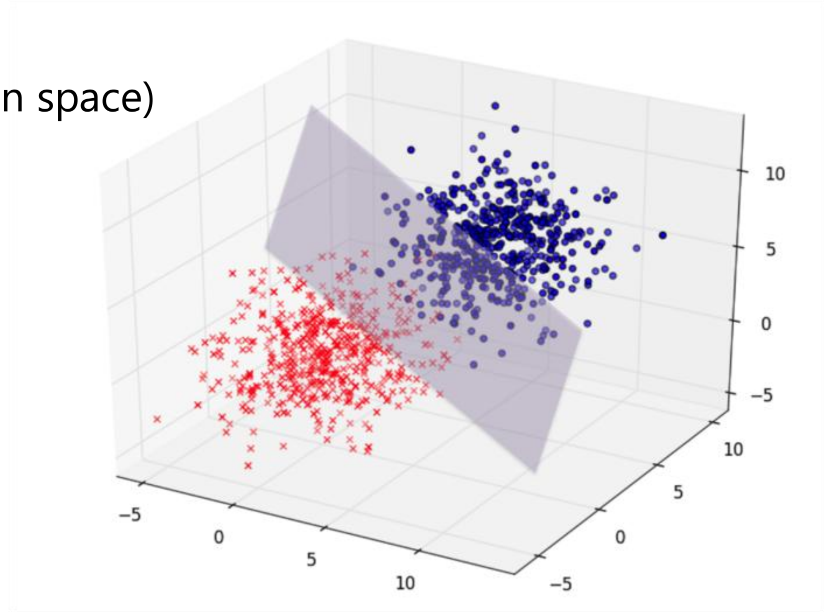
Feature space

A set of variables (dimensions) define a space

- multivariate observations are positioned in this space
- when variables are numeric:
  - feature space ≡ vector space (e.g. Euclidean space)
  - observation ≡ data point

$$\mathbf{x} = \{x_1, \ldots, x_m\} \in \mathbb{R}^m$$

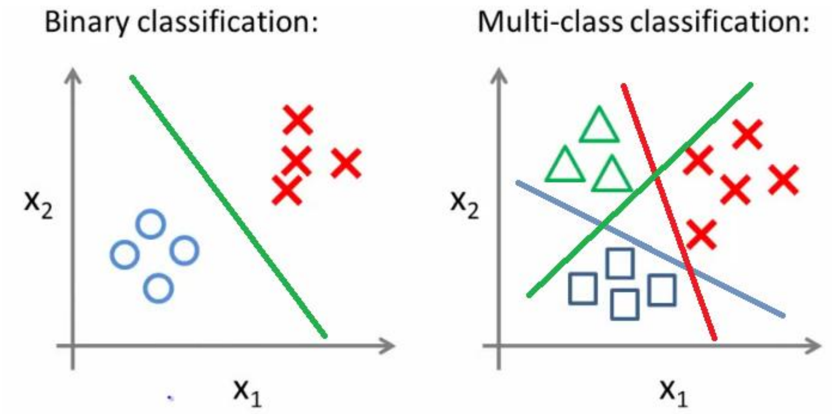
$$\|\mathbf{a} - \mathbf{b}\| = \sqrt{\sum_{i=1}^{m} (a_i - b_i)^2}$$



Classification

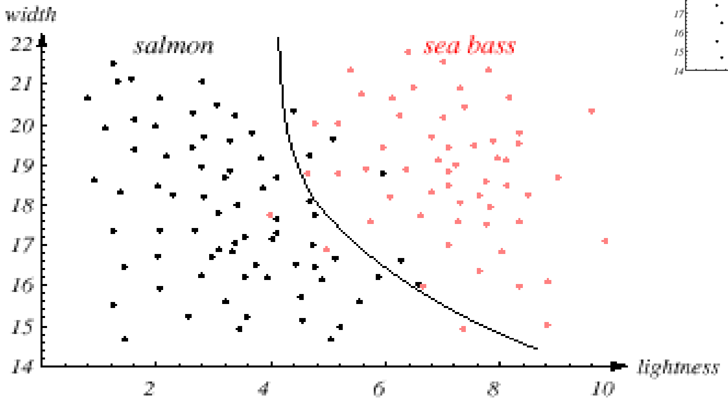
Recall: given a set of labeled observations, $\{(\mathbf{x}_1, z_1), \dots, (\mathbf{x}_n, z_n)\}$ where $z_n \in \Sigma$, a **classifier** M is a mapping function between domain variables and a categoric variable, $M : X \rightarrow Z$

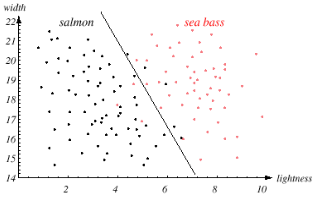
- *prediction*: given a new unlabeled observation $\mathbf{x}_{new}$, use M to classify: $\hat{z}_{new} = M(x_{new})$
- *description*: inspect M to acquire new knowledge

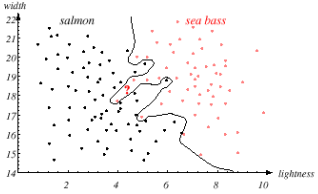


Classification: *salmon*?

- *width* and *lightness* are discriminative variables
- generalization ability linked with:
  - underfitting risks
  - overfitting risks
- *aim*: find a balanced model capacity

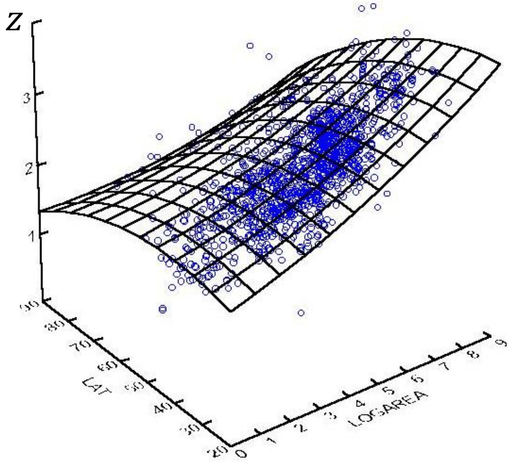






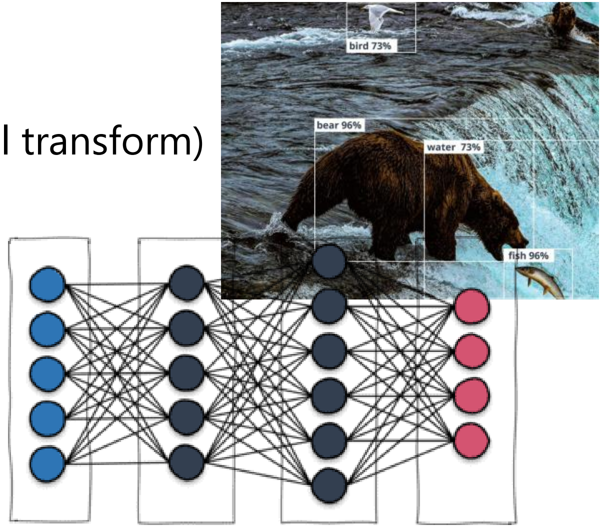
Regression

- *descriptive setting*: given a set of observations, $\{(\mathbf{x}_1, z_1), \ldots, (\mathbf{x}_n, z_n)\}$ where $z_i \in \mathbb{R}$, describe the relation between a set of (explanatory) variables and a target numeric variable
- *predictive setting*: given a set of observations, $\{(\mathbf{x}_1, z_1), \ldots, (\mathbf{x}_n, z_n)\}$ where $z_i \in \mathbb{R}$, learn a mapping, $M : X \rightarrow Z$, to estimate the outcome (quantity) of a new observation



Multi-output prediction

- Most outputs are not described by a single feature
  - generative AI (e.g. question-answer, image drawing, signal transform)
  - many others (e.g. self-driving vehicles, tagging content)



- Multi-output predictors, $M : X \rightarrow Z$
  - *predictive* setting (learn predictor M)
  - *descriptive* setting (explain predictor M)

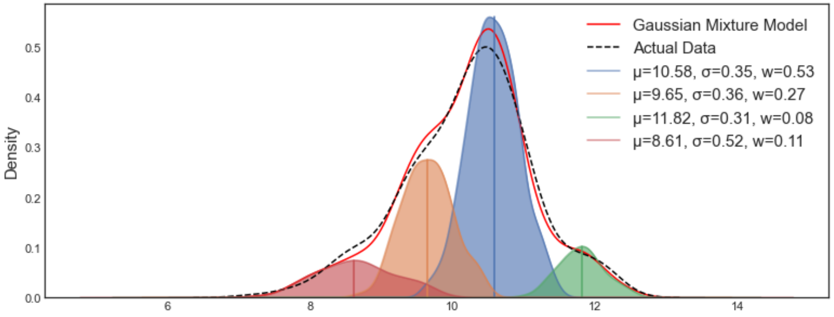
- Special cases: multi-label classification when $\mathbf{z} \in \Sigma^p$ and multi-output regression when $\mathbf{z} \in \mathbb{R}^p$

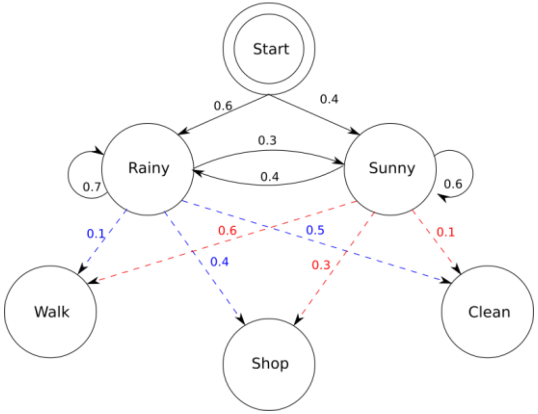
Statistical modeling

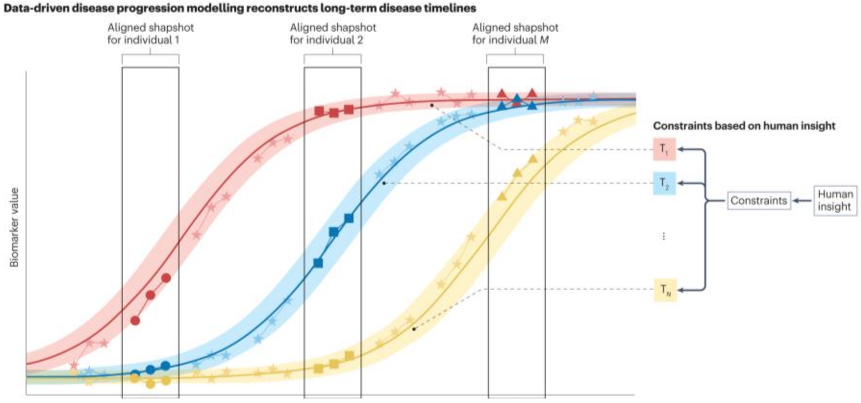
Associative analysis

Description of system dynamics

- mixture models
- generative models (such as HMMs)



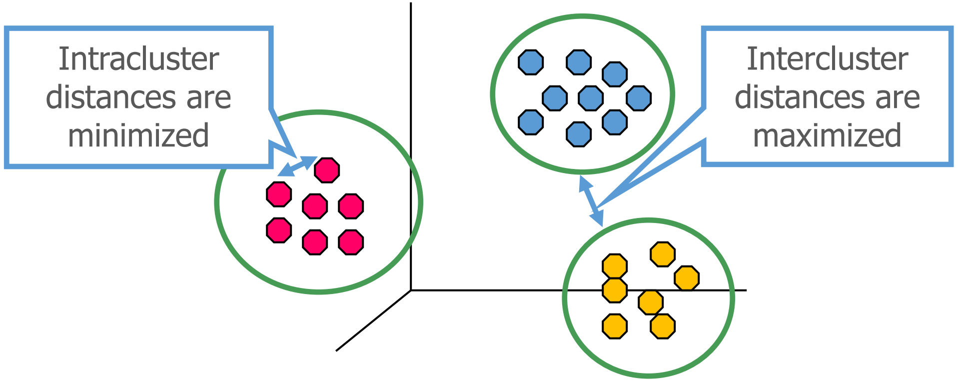


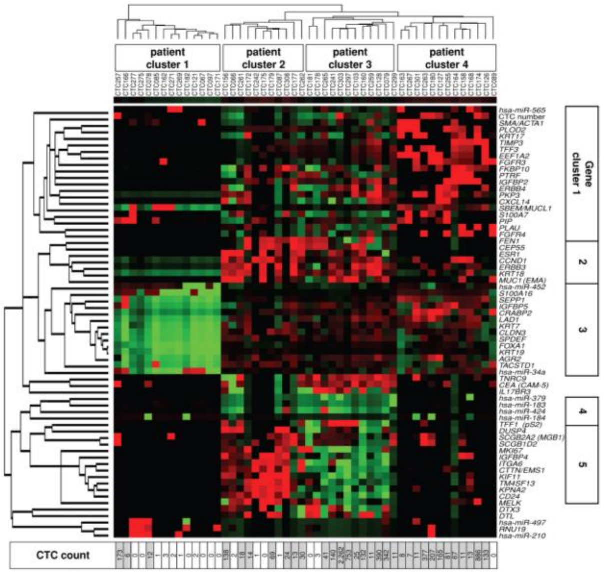


Clustering

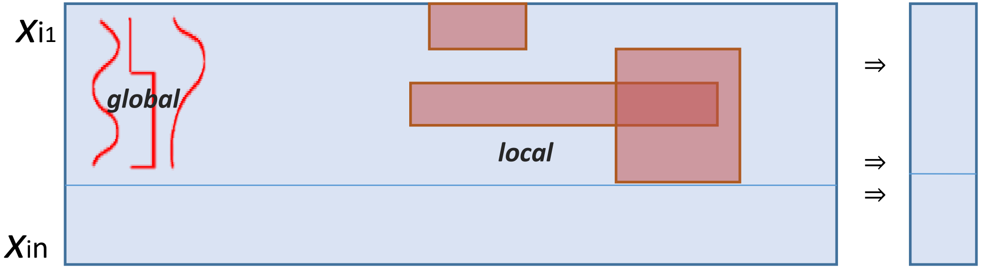
Given a set of data observations, $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, cluster analysis aims at grouping observations into clusters, $C_i \subseteq X$ with $i = 1..k$, according to their (dis)similarity:

- observations in the same cluster are more similar than those in different clusters





Pattern mining



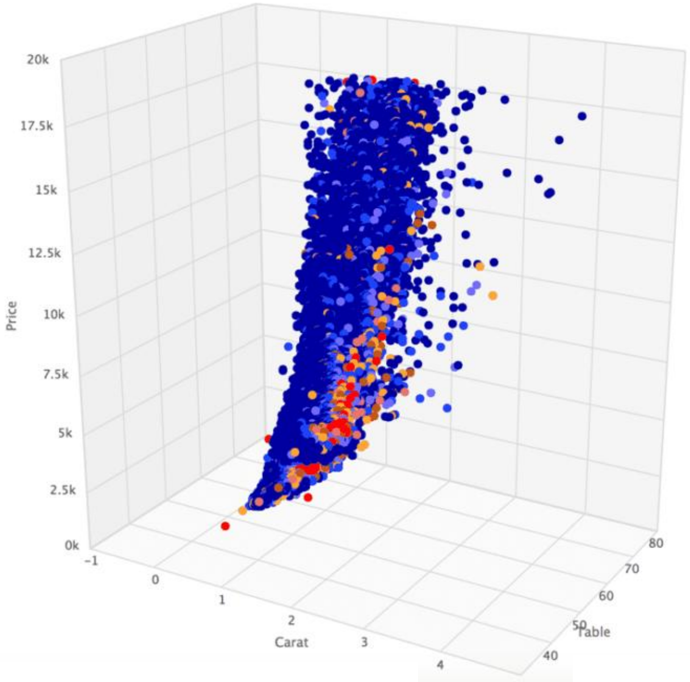
{symptomA, testB+} ⇒ condition1 [support=10%,confidence=80%, lift=1.4, p-value=1E-4]

Given a dataset, find local associations (*aka* patterns) satisfying:

- statistical significance criteria (min #observations to be unexpectedly frequent)
- discriminative power (qualitative targets) or correlation (numeric targets) criteria

Outlier analysis

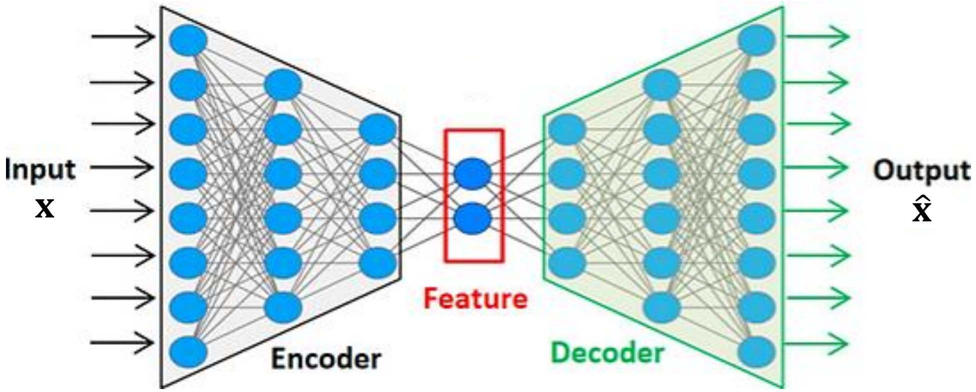
- Understand peculiar behaviors and isolate anomalous observations
  - fraud, cyberattacks, personalized health risks, adverse events, deviant social behavior, vehicle failures…



Representation learning

Describe data using a compact set of informative (unsupervised) and/or predictive (supervised) features

- dimensionality reduction: subset of features from multivariate observations with minimal info loss
- latent feature representations of complex signals (series, image, text data) using neural networks



Example: biomedicine

- clinical trials (cohort studies), e.g. case-control populations
- **observations** generally correspond to:
  - individuals
    - **input variables**: health-related features (clinical records, multi-omics, exposomics)
    - **output variable**: outcome annotations
      - qualitative conditions (diagnostics, prognostics, prescriptions, traits)
      - quantifiable phenotypes (impairments, molecular markers, risk, survivability, drug dosage)
  - hospitals, tissue samples, undertaken procedures, healthcare professionals, drugs…
- ability to **generalize** from a population to new observations
  - prevent overfitting (including non-relevant relations in the learned models)
  - prevent underfitting (excluding relevant relations from the learned models)

Example: biomedicine

- **Statistical modeling:** assess risk determinants, model health trajectories, test clinical hypotheses
- **Clustering**: group patients in accordance with biophysiological profile (e.g. stratified therapeutics)
- **Pattern mining**: discover meaningful associations to understand disease/therapeutic responses
- **Outlier analysis**: personalized care to particular needs (e.g. multimorbidity, rare diseases)
- **Representation learning**: encoders and saliency maps of medical signals/images/notes
- **Generative modeling**: comprehensive models of disease/treatment (e.g. health progression)
- **Classification**: how monitored inputs affect diagnostics/prognostics, thearapeutic choices
- **Regression**: estimate risk, drug dosage or efficacy, quantifiable phenotypes

Thank you!

Rui Henriques

rmch@tecnico.ulisboa.pt