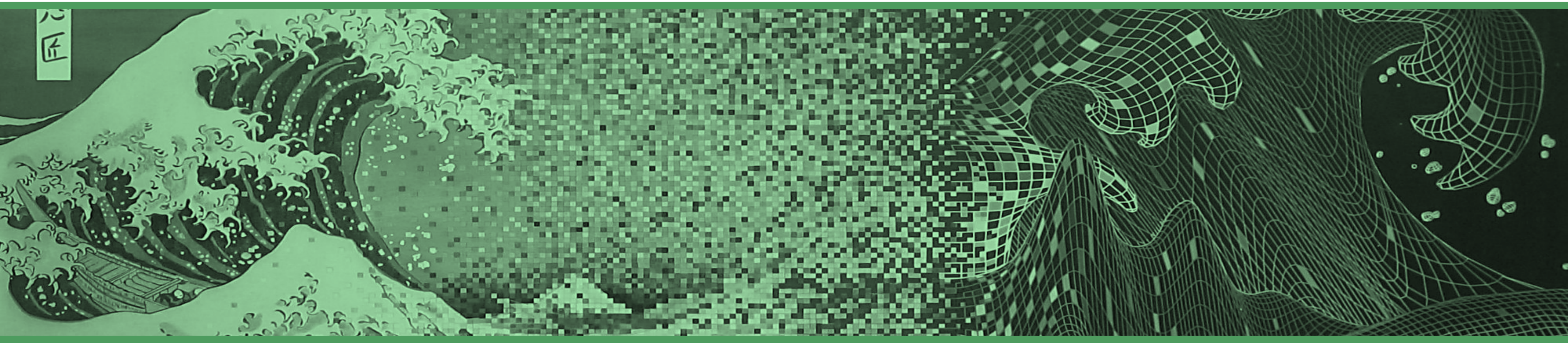
TÉCNICO LISBOA

Introduction to Machine Learning

Learning and Univariate Data Analysis



Aprendizagem 2024

Miguel Couceiro miguel.j.couceiro@tecnico.ulisboa.pt
Andreas Wischert andreas.wichert@tecnico.ulisboa.pt

Outline

- **Machine learning**
  - intelligence and learning
  - data science and AI
  - symbolic learning
  - terminology
  - descriptive and predictive tasks
- **Univariate data analysis**
  - numeric and categoric variables
  - empirical and theoretical distributions
  - summary statistics
  - outlier removal
  - discriminant analysis
  - correlation

Outline

- **Machine learning**
  - **intelligence and learning**
  - **data science and AI**
  - **symbolic learning**
  - **terminology**
  - **descriptive and predictive tasks**
- Univariate data analysis
  - numeric and categoric variables
  - empirical and theoretical distributions
  - summary statistics
  - outlier removal
  - discriminant analysis
  - correlation

Intelligence

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

What is machine learning?

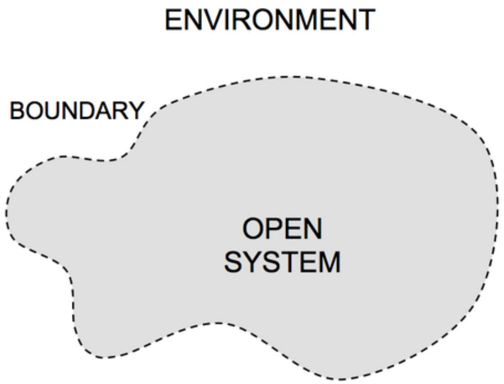
- **Artificial Intelligence (AI)** emulates **qualities of human intelligence** to answer real-world problems
  - parallels between human and artificial intelligence?
  - many AI techniques inspired from human psychology, biology, behavior

- **Learning** is a fundamental quality intelligence
  - "*learning is any process by which a system improves performance from experience*"
    (by Herbert Simon)

- **Machine learning (ML) as a subfield of AI**
  - AI with a focus on <u>rationality</u> => optimization, planning, reasoning, …
  - AI with a focus on <u>curiosity</u> => autonomous agents, affective computing, …
  - AI with a focus on <u>adaptability</u> from experience (data records) => ML

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

By monitoring systems (e.g. sensorization, observation)…
- **data** ⇒ information (descriptive learning) ⇒ **knowledge**
- **data** ⇒ **decision support** system (predictive learning)

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

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

Data everywhere!

Sensorization examples:

- **biological** systems
  - physiological signals from biosensors, molecular signals using multi-omic high-throughput technologies
  - health records (diagnostics, prescriptions, undertaken surgeries), exposomics, demographics
- **knowledge** systems
  - corpora from digital libraries and the Web
- **ecological** systems
  - biodiversity, plant health, crop and livestock conditions, water quality, food nutrition, forestry and fishery surveillance from 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
  - traffic records from mobile phones, smart card validations, inductive loop counters, privacy-preserving
  - water and energy supply via telemetry (flowrate, pressure, smart sensors)
- … [*complete the list*]

From experience to learning

- Data records acquired from:
  - multiple systems of the same type
    - e.g. different individuals, vehicles, computers, organizations
  - single system under different conditions
    - e.g. brain under different stimuli, crop under different weather conditions, e-commerce along time

- Multiple data records… statistics!
  - discover of relevant relations/associations (patterning)

- Pattern recognition aids us in:
  - understanding systems' behavior (descriptive learning)
  - supporting decisions (predictive learning)

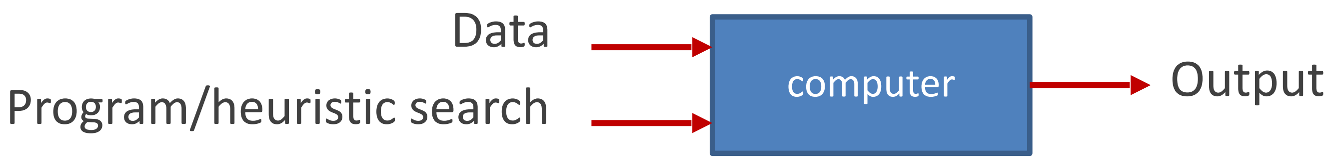
Machine Learning

- Machine Learning *versus* **Artificial Intelligence**
  - recall: ML is a subfield of the larger AI field

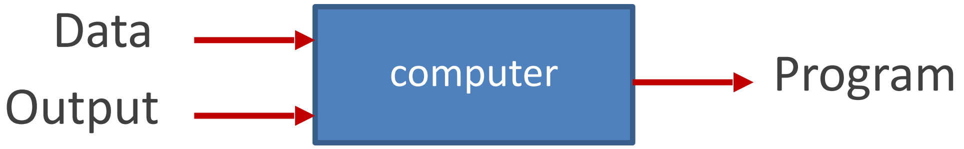
- Machine Learning *versus* **Data Science**
  - ML as a set of concepts, principles and computational methods
    to aid decisions and accomplish other digital tasks from available data
    - grounded on statistical, algebraic, mathematical and algorithmic foundations
  - Data Science has been termed the art of discovering what we don't know from data
    - the non-trivial extraction of implicit, previously unknown, and potentially useful knowledge from data
    - ML provides the foundational concepts and algorithmic means for Data Science

The ML stance

- Traditional programming and classic AI



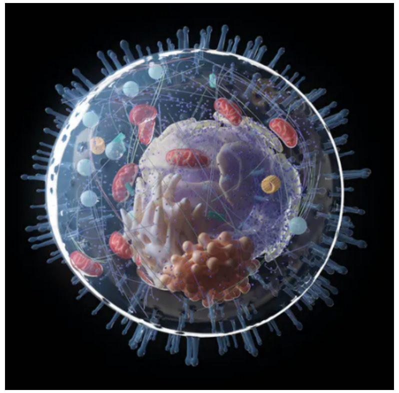
- Machine learning



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

When?

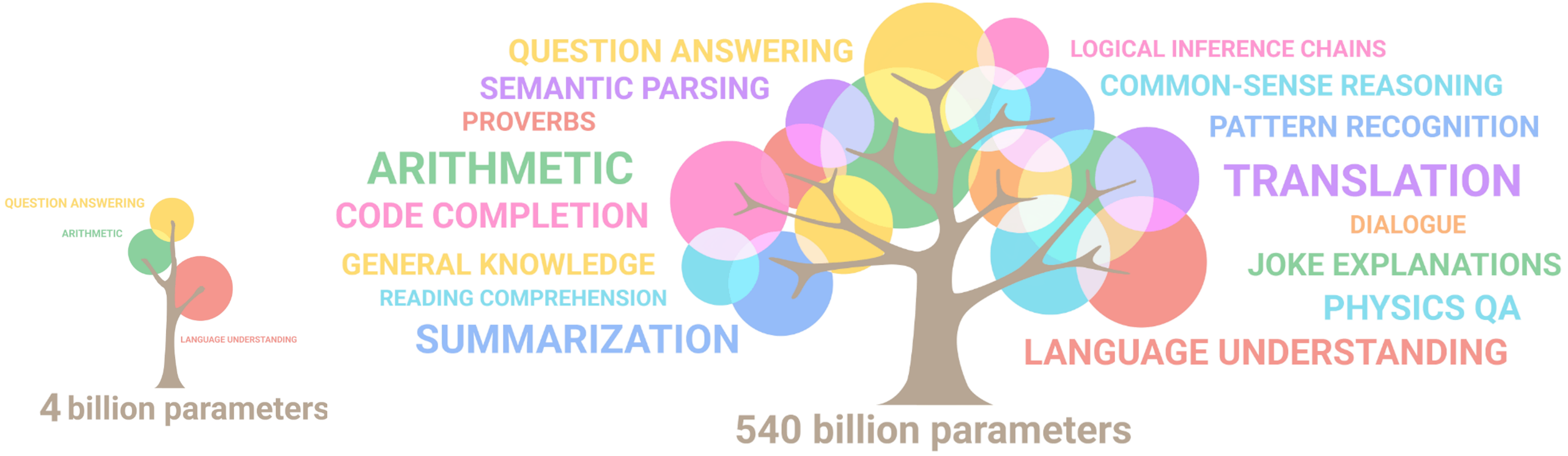
- Human expertise does not exist (e.g. navigating on Mars)
- Humans cannot explain their expertise (e.g. speech recognition)
- Models must be customized (e.g. personalized medicine)
- Models are based on huge amounts of data (e.g. genomics)

- Learning isn't always useful: there is no need to learn to calculate payroll!

by Eric Eaton

When?

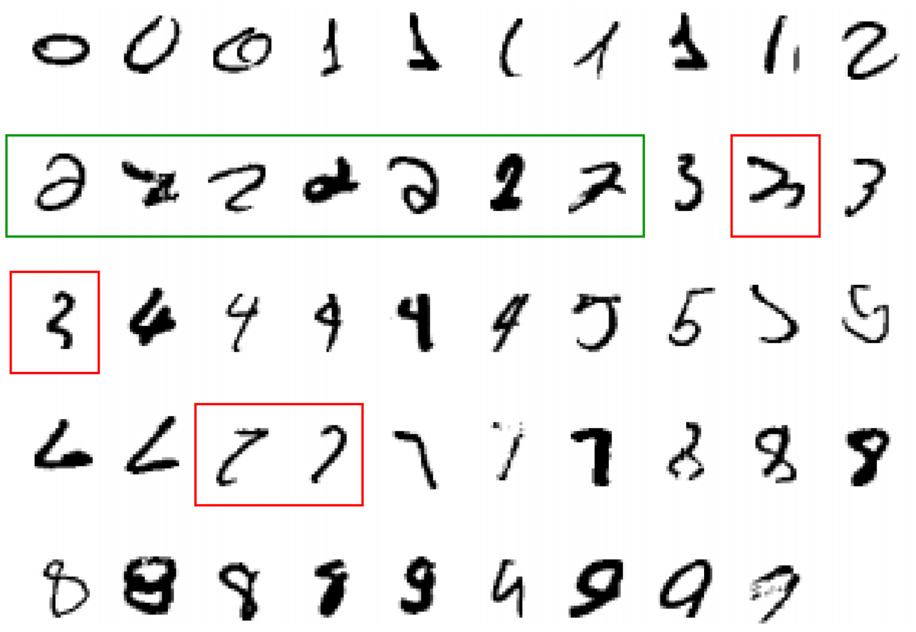


- Task-specific *versus* multi-task/purpose learning – the era of Large Language-and-Vision Models

https://github.com/Hannibal046/Awesome-LLM

When?

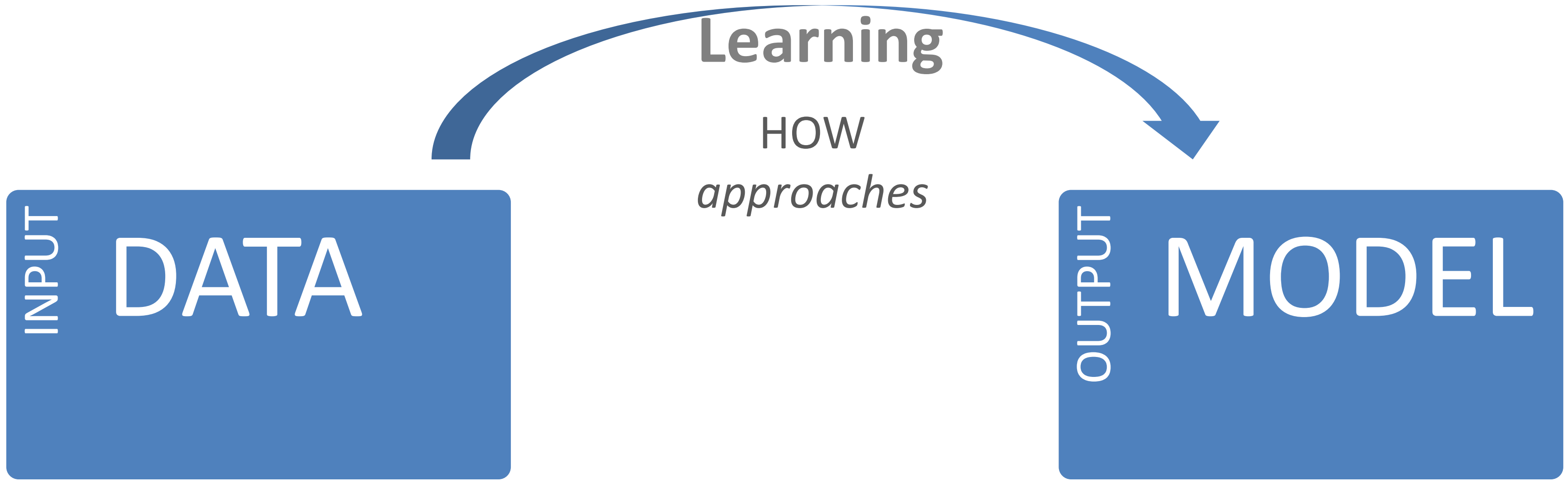
- Classic example of a task that requires machine learning:
  - Hard to say what makes a 2!



  - What about clinical diagnostics? Product recommendations?

by Geoffrey Hinton

Machine Learning



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

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

Learning input-output functions

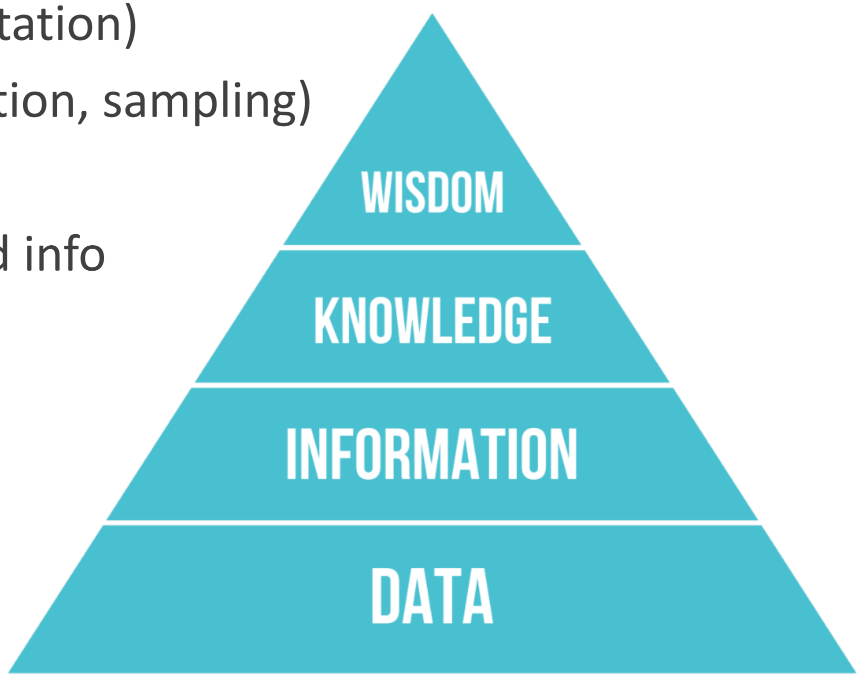
- **Supervised** learning
  - with a teacher
  - learning from training data and desired outputs (labels, quantities, structures)

- **Unsupervised** learning
  - without a teacher
  - learning from training data without desired outputs

- **Reinforcement** learning
  - absence of a designated teacher to give positive and negative examples
  - learning rewards and penalties observed from sequence of actions within a given environment

Machine learning in practice

- The process of knowledge discovery is a composition of steps:
  - **data preprocessing**
    - data acquisition and integration
    - data cleaning (e.g. duplicate and outlier removal, missing imputation)
    - data transformations (e.g. normalization, dimensionality reduction, sampling)
  - **data mining** recurring to *machine learning*
  - **postprocessing** needs and knowledge retrieval from the extracted info (descriptive stance) or learned models (predictive stance)
    - interpret and validate results
    - consolidate and deploy discovered knowledge



Data Science

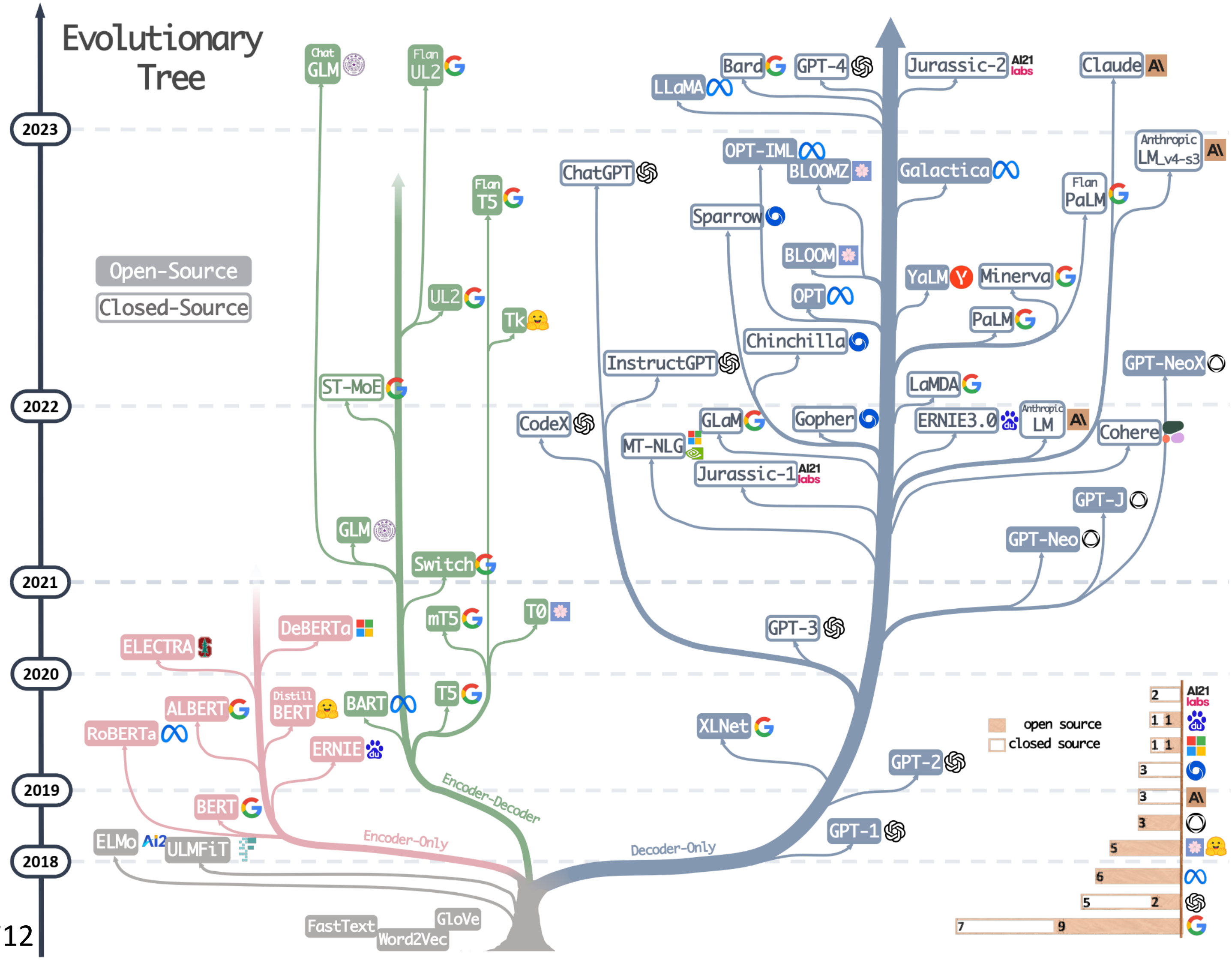
- Data science
  - the rediscovery of “statistics” ...
    - descriptive statistics
    - inferential statistics
  - the rediscovery of “maths” …
    - linear algebra
    - calculus

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

History of Machine Learning

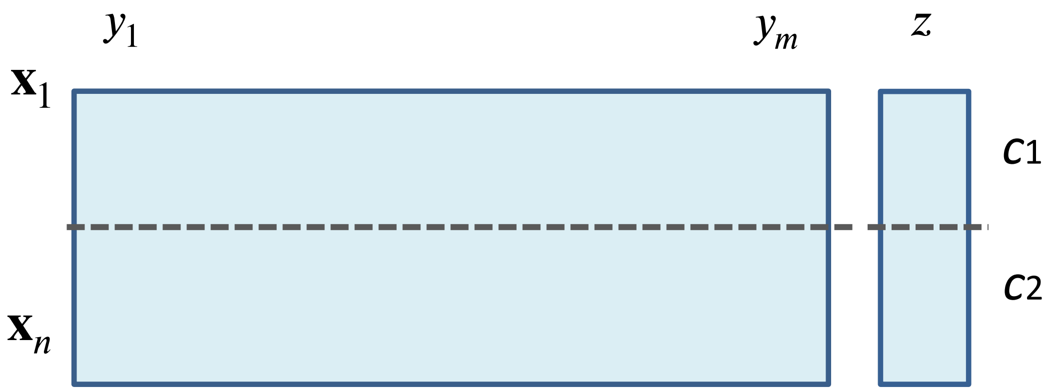
- **1950s**
  - Samuel’s checker player
  - Selfridge’s Pandemonium
- **1960s**
  - Perceptron and its limitations
- **1970s**
  - Symbolic learning
  - Expert systems
  - Decision trees
- **1980s**
  - Resurgence of neural networks: backpropagation
  - Learning and planning
  - Explanation-based learning
  - Inductive logic programming
  - Utility problem, analogy
  - Cognitive architectures
  - PAC Learning Theory
- **1990s**
  - Data mining, text mining
  - Adaptive software agents
  - Reinforcement learning (RL)
  - Ensembles: bagging, boosting, stacking
  - Bayes network learning
- **2000s**
  - Support vector machines, kernel methods
  - Learning in robotics and vision
  - Graphical models
  - Relational learning
- **2010s**
  - Deep learning
  - Big data
  - Uncertainty
  - Multi-task learning
  - Large language models

Deep Learning and Large Language Models



https://arxiv.org/abs/2304.13712

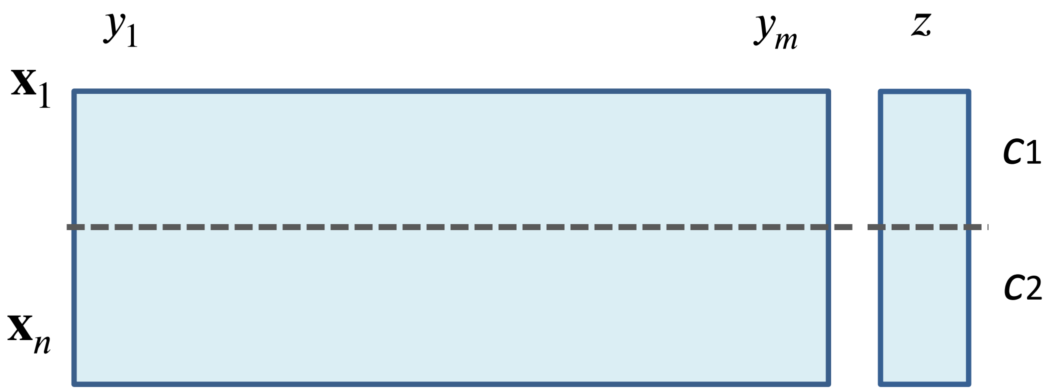
Terminology



Dataset:

- set of observations/instances/records, $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ (population)
- with values/features along a set of variables/attributes, $Y = \{y_1, \ldots, y_m\}$
  - input variables (explanatory)
  - optional output variables (targets)
- data size = number of observations, $|X| = n$
- data dimensionality = number of variables, $|Y| = m$

Learning



Learning from a dataset: retrieving relevant **data relations**

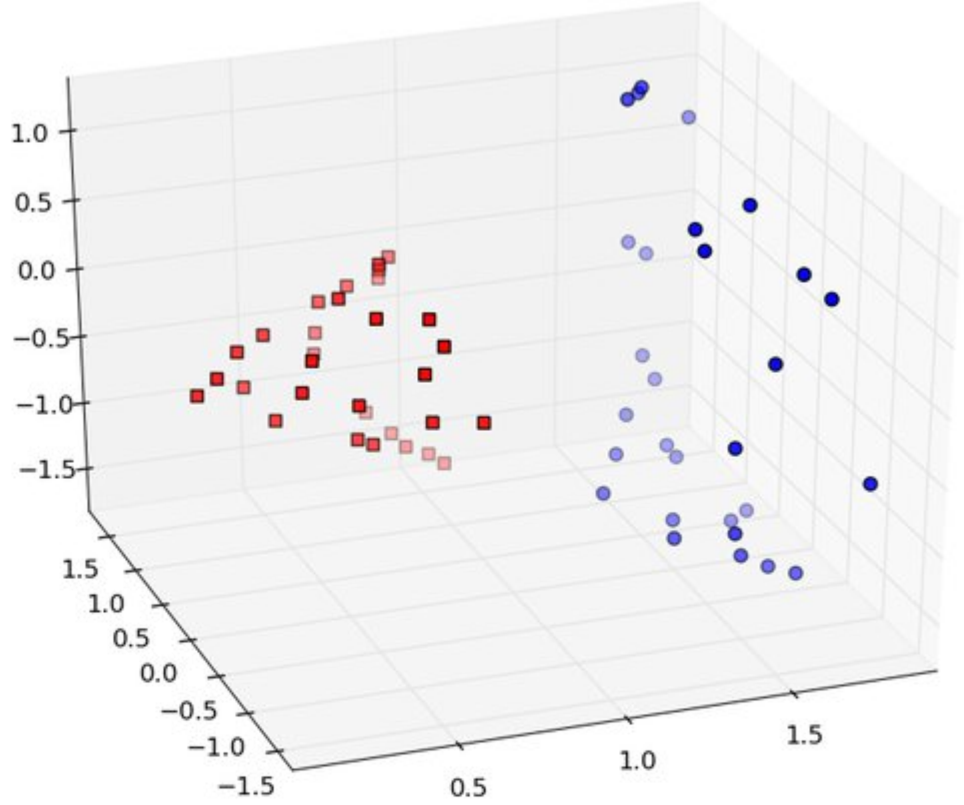
- relations/patterns/abstractions ≡ distributions of interest on specific observations and attributes
  - *unexpectedly informative*
  - *unexpectedly discriminative* (different distribution between populations)
- learn classifiers, regressors, descriptors, forecasters, autoencoders from these relations

Feature space

- When variables are numeric:
  - feature space ≡ vector space (e.g. Euclidean space)
  - observation ≡ data point

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

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



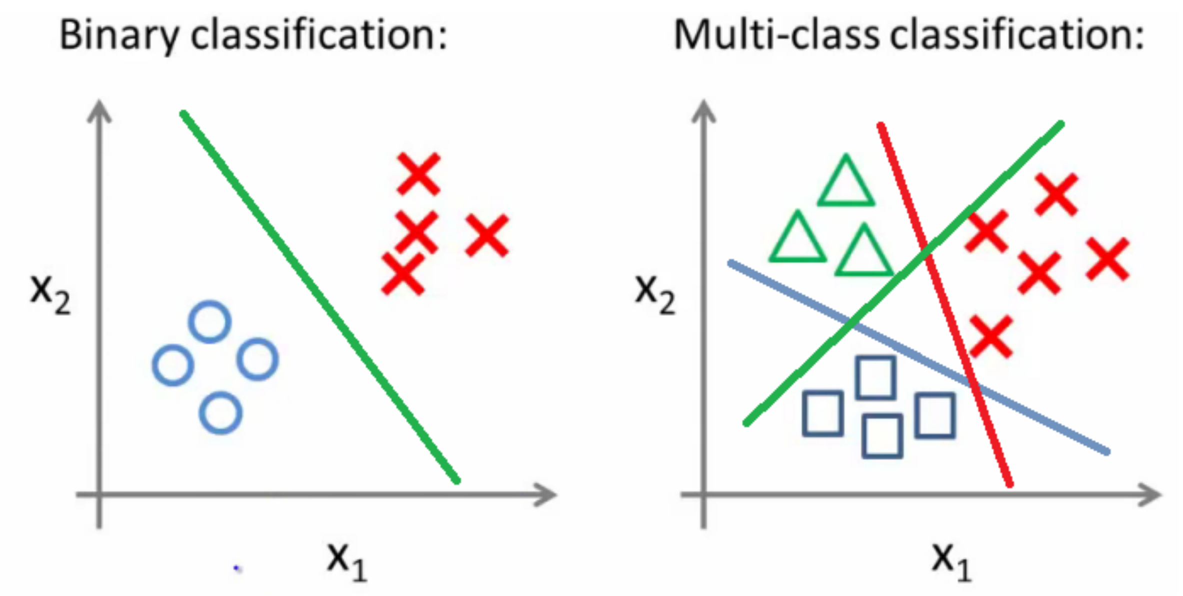
NB: Encyclopedia of Distances (M. & E. Deza), Springer

https://link.springer.com/book/10.1007/978-3-662-52844-0

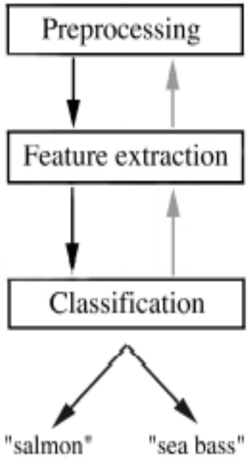
Classification

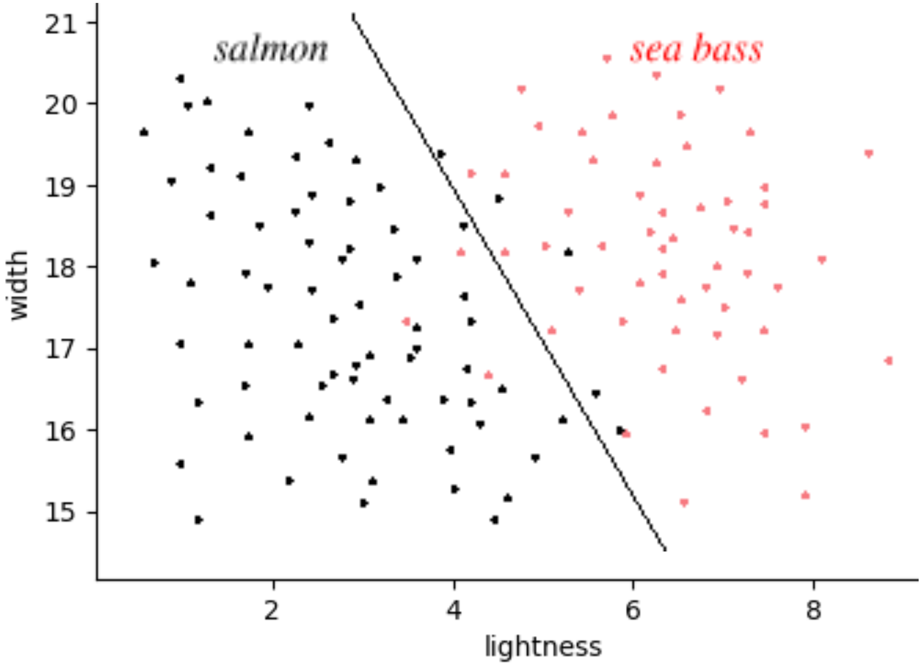
Given a set of labeled observations, $\{(\mathbf{x}_1, z_1), \ldots, (\mathbf{x}_n, z_n)\}$ where $z_n \in \Sigma$, a **classifier** M is a mapping function between input variables and a categoric variable: $M : X \rightarrow Z$

– given a new unlabeled observation $\mathbf{x}_{new}$, use M to classify: $\hat{z}_{new} = M(\mathbf{x}_{new})$



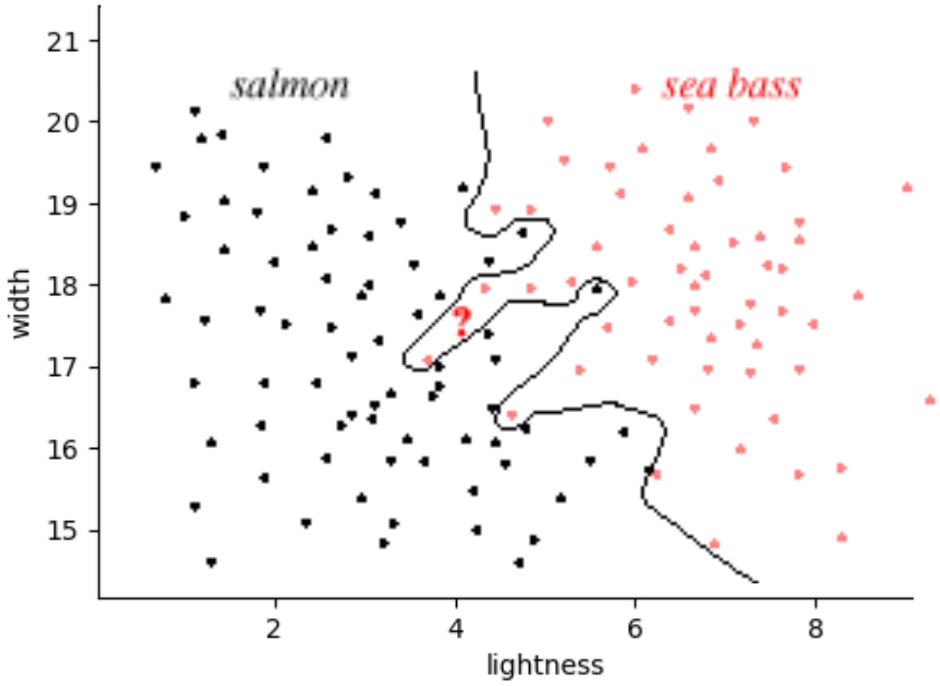
Classification: *salmon?*





Classification: *salmon?*

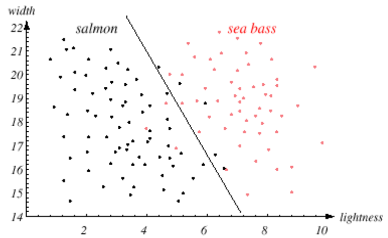
- we might add other variables that are not correlated with the ones we already have
  – caution should be taken not to reduce the performance by adding such “noisy features”
- the best decision boundary should be the one which provides an optimal performance

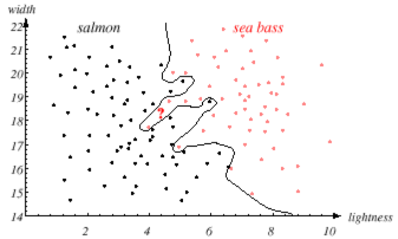


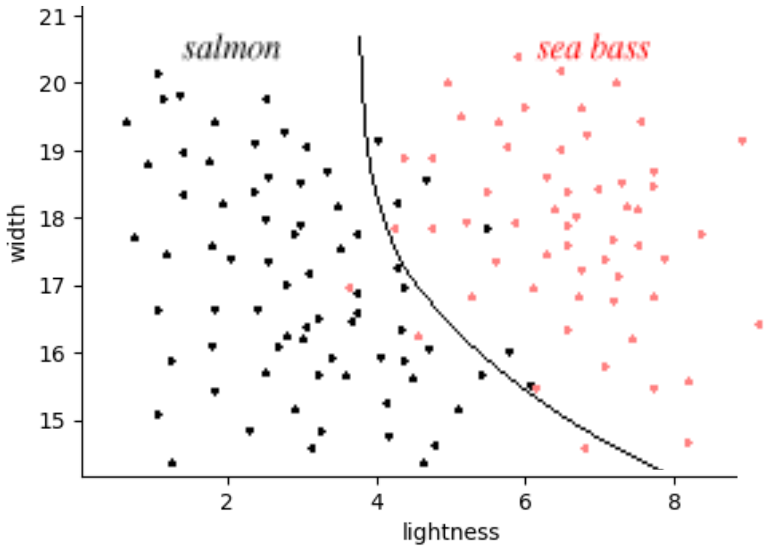
- However, our satisfaction is premature because the central aim of designing a classifier is to correctly classify novel inputs
  – issue of ***generalization***!

Classification: *salmon?*

- Generalization ability linked with:
  - underfitting risks
  - overfitting risks
- Aim: find a balanced model capacity



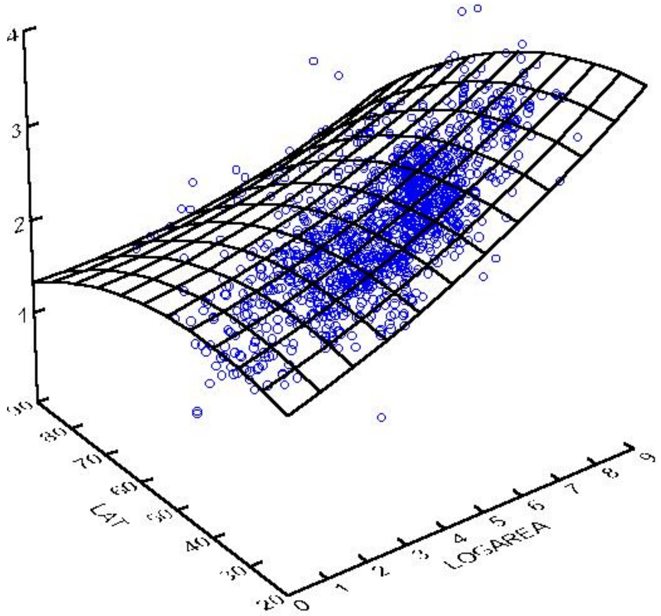




Regression

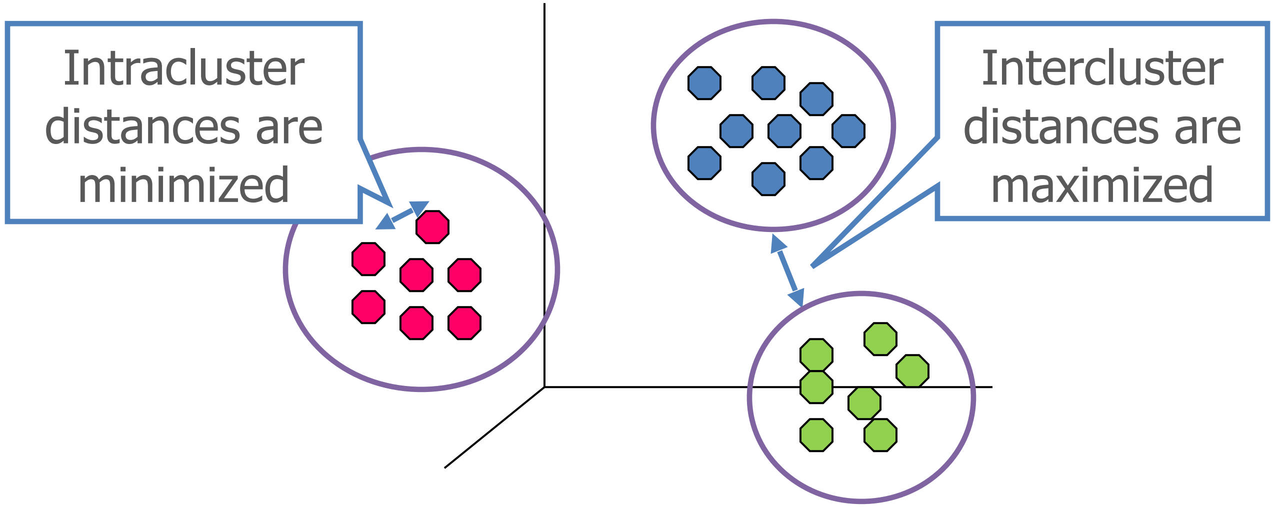
– *Descriptive setting*: given a set of observations, $\{(\mathbf{x}_1, z_1), \ldots, (\mathbf{x}_n, z_n)\}$ where $z_n \in \mathbb{R}$, describe a relation between a set of (explanatory) variables and a target real-valued variable

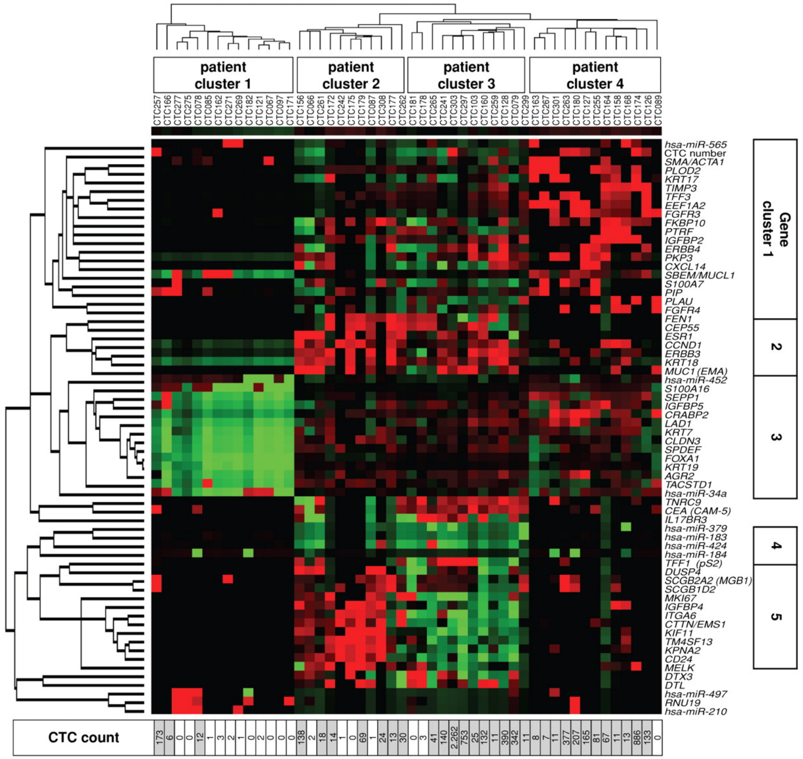
– *Predictive setting*: given a set of observations with a real-valued outcome, $\{(\mathbf{x}_1, z_1), \ldots, (\mathbf{x}_n, z_n)\}$ where $z_n \in \mathbb{R}$, learn a mapping, $M : X \rightarrow Z$, to estimate the outcome of a new observation



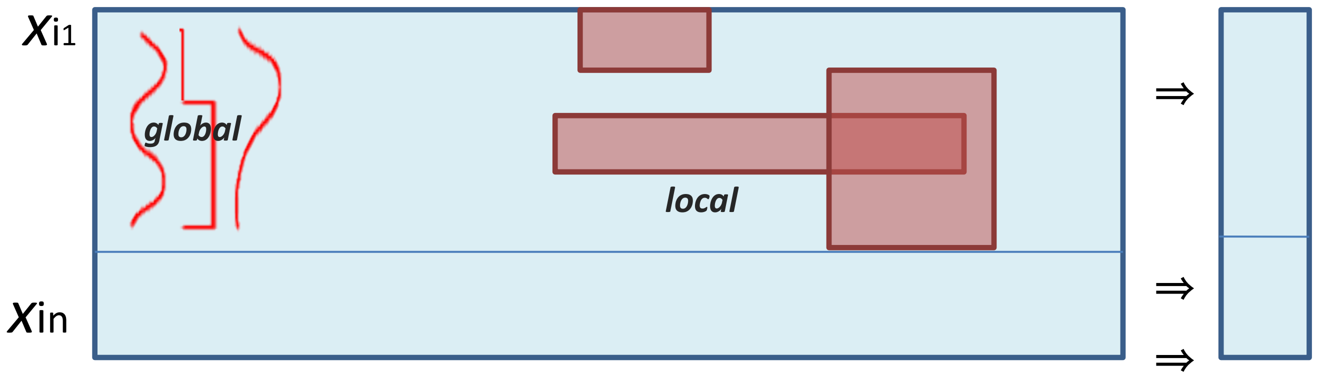
Clustering

Given a set of data observations, $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, 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, testBpositive} ⇒ condition1 [sup=10%, conf=80%, lift=1.4, sig=1E-4]

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

– statistical significance criteria (min number of observations to deviate from expectations)

– discriminative power (qualitative targets) or correlation (numeric targets) criteria

Example: learning from biomedical data

- ***Descriptive modeling***: models of disease/treatment (e.g. health progression)
- ***Clustering***: group individuals in accordance with health profile
- ***Pattern mining and subspace clustering***: discover meaningful patterns and associations with impact on disease/treatment study and discrimination
- ***Classification***: diagnostics/prognostics, treatment recommendation
- ***Regression***: estimate risk, drug dosage or efficacy, quantifiable phenotypes

Example: learning from biomedical data

- **observations** generally correspond to:
  - individuals
    - **input variables**: health-related features (multi-omics, clinical records, exposomics…)
    - **output variable**: outcome annotation
      - qualitative clinical condition (diagnostics, prognostics, therapies, traits)
      - quantifiable phenotypes (impairments, molecular levels, severity, survivability, drug dosage)
  - hospitals, undertaken procedures, care professionals, drugs…
- clinical trials (cohort studies) with enough, precise data observations, e.g. case-control populations
- ability to **generalize** from a population to new patients
  - prevent overfitting (including non-relevant relations in the learned models)
  - prevent underfitting (excluding relevant relations from learned models)

Outline

- Machine learning
  - intelligence and learning
  - data science and AI
  - symbolic learning
  - terminology
  - descriptive and predictive tasks
- **Univariate data analysis**
  - **numeric and categoric variables**
  - **empirical and theoretical distributions**
  - **summary statistics**
  - **outlier removal**
  - **discriminant analysis**
  - **correlation**

Univariate data analysis

- **Random**/aleatory **variable**
  - function $Y\colon \Omega \rightarrow \mathrm{E}$ from a **sample space** Ω to a **measurable space** E
  - e.g. height variable is a function which maps a person from a population Ω to her height in $\mathbb{R}^{+}$ ($\mathrm{E} = \mathbb{R}^{+}$)
    - the observed height is referred as a **measurement**
  - from now on, we will refer *random variable* simply as *variable*

- **Univariate data**
  - single input variable
  - comprises univariate data statistics and, in the presence of an output variable, **bivariate data statistics**
- **Multivariate data**
  - multiple (input) variables
  - **multivariate order** = number of (input) variables

Variables

- **Categorical** (or qualitative) variables
  - values are categories
  - can either be **nominal**/symbolic or **ordinal** (e.g. low, average, high)
  - **binary** variables are variables with two categories (whether nominal or ordinal)
  - variable **cardinality** = number of categories

- **Numeric** (or quantitative) variables
  - values are quantities
  - can be either be **discrete** (e.g. integers) or **continuous** (e.g. real values)

- Exercise: classify the following variables – gender, age, height

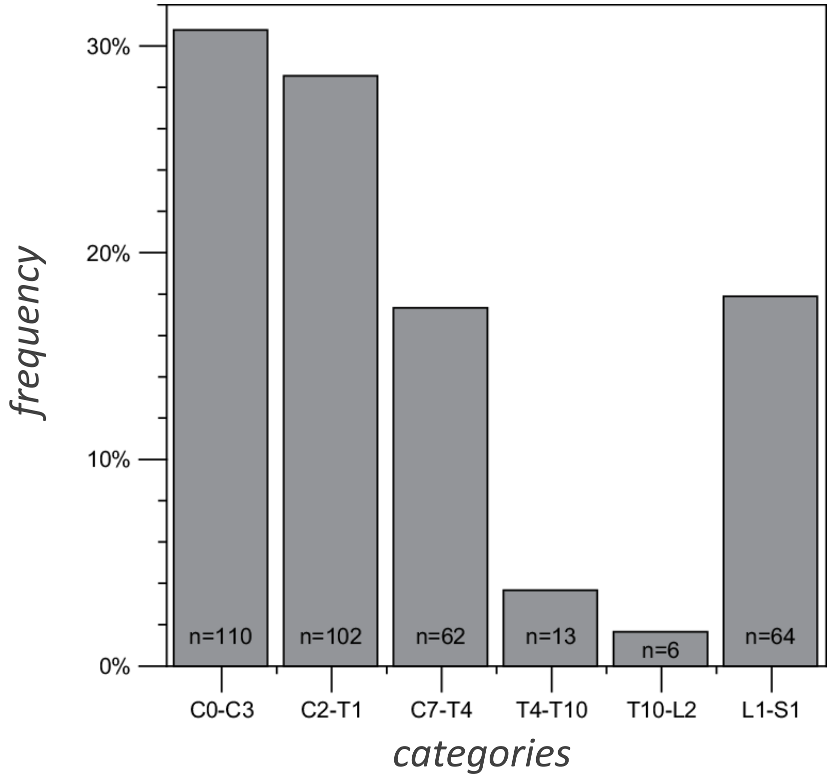
Variables

- [**discretization**] numeric variables can be discretized into ordinal variables
  - e.g. age categories of 0-10, 11-20, 21-30, 31-40…
  - trade-off: loss of information versus utility for subsequent data analysis
- [**normalization**] numeric variables can be normalized
  - comparability between variables with different domains E
- [**aggregation**] categorical variables with high cardinality can be aggregated
  - 100 colors can be aggregated into coarser categories in accordance with hue
- [**imputation**] missing values can occur
  - unobserved, error or noisy measurements
  - missing values can be imputed using variable expectations

Data profiling

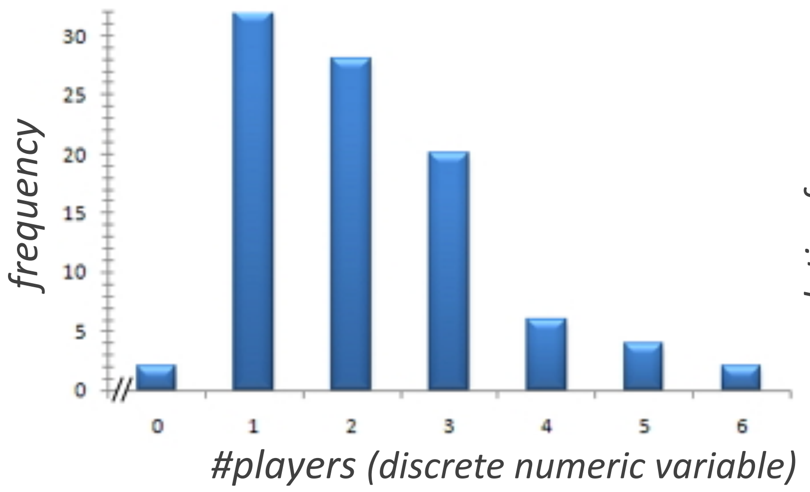
- **Data profiling ≡ data exploration**
  - essential step to know and learn from data

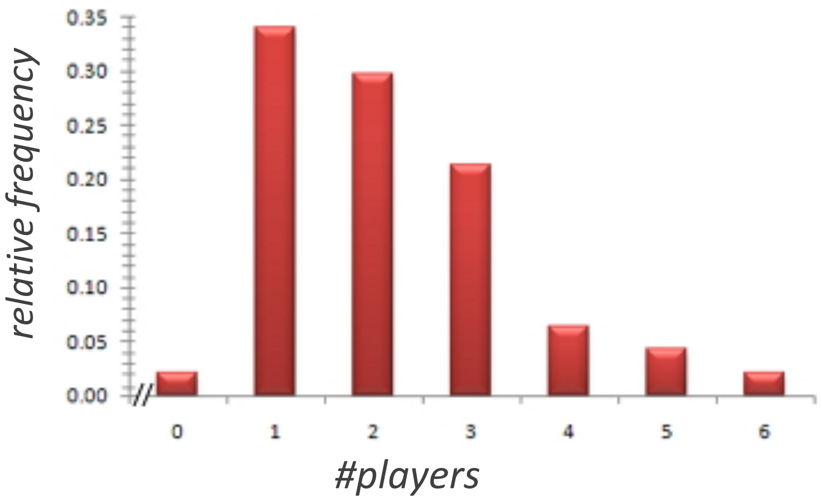
- **Frequentist statistics**
  - Categorical variables
    - summary statistics (e.g. mode)
    - category frequencies
    - category probabilities



Data profiling

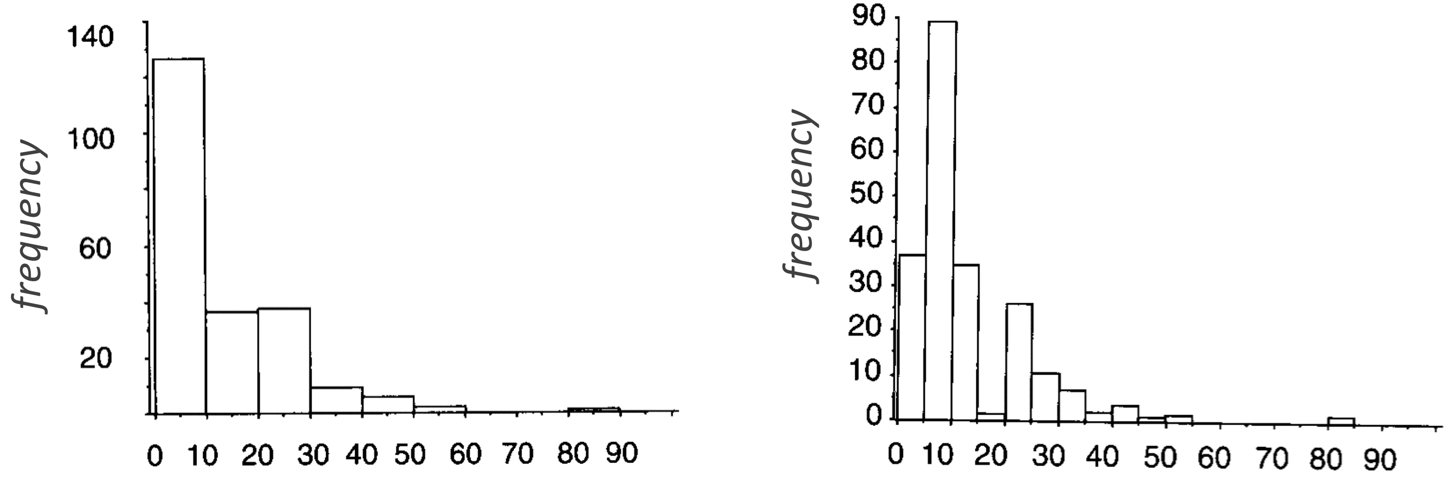
- **Frequentist statistics**
  - numeric variables
    - summary statistics (e.g., percentiles)
    - classic histograms (bin frequencies)
    - empirical probability distribution (bin probabilities)
      - density function for continuous variables
      - mass function for discrete variables, example:





Data profiling: histograms

- How?
  - divide the range of values in a distribution into several bins of equal size
  - toss each value in the appropriate bin
- The choice of bin size can strongly affect the frequency histogram
  - revealing details when we lower bin sizes, yet at times a result of overfitting
  - bin size also affect one‘s perception of the shape of distribution



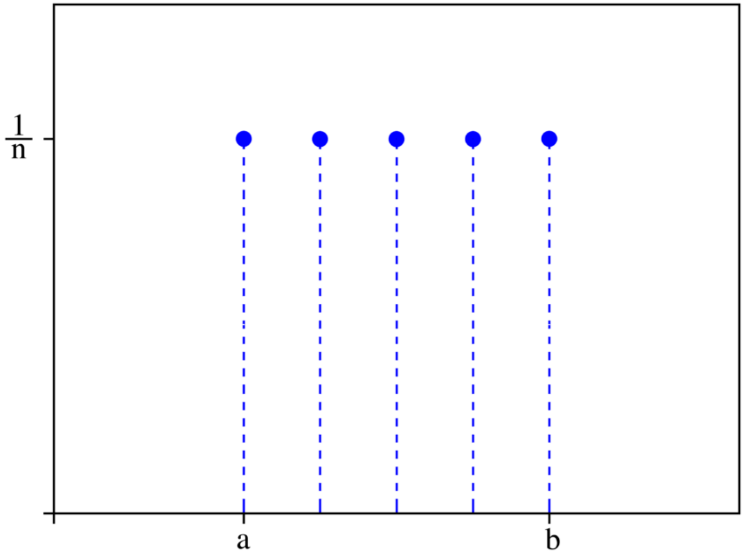
Data profiling

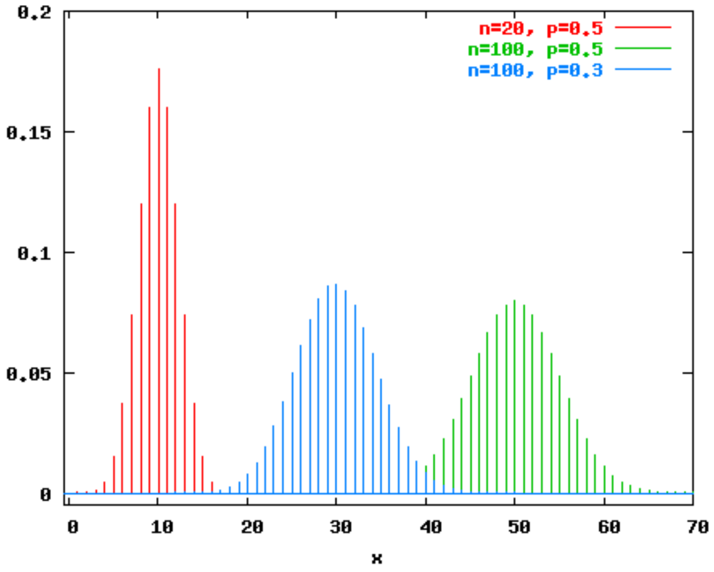
- **Theoretical statistics**
  - summary statistics
    - mean and deviation statistics (Gaussian assumption)
  - fitting theoretical distributions
    - discrete numeric variables: fitted probability mass function
    - continuous numeric variables fitted probability density function

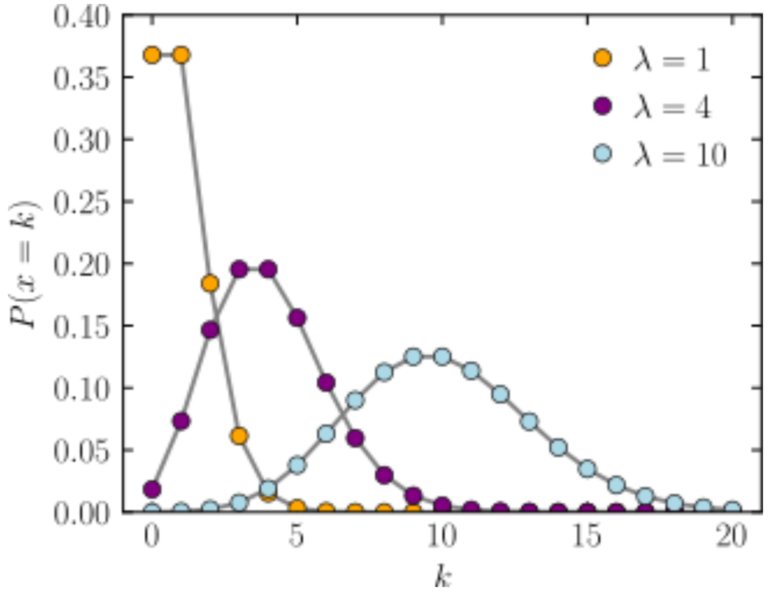
- Empirical *versus* theoretical distributions
  - empirical distribution are perfectly overfitted to observed data
    - this is problematic for low-to-moderate data sample size, otherwise preferable

Data profiling: theoretical distributions

- **Discrete distributions**
  - uniform
  - Binomial
  - Poisson

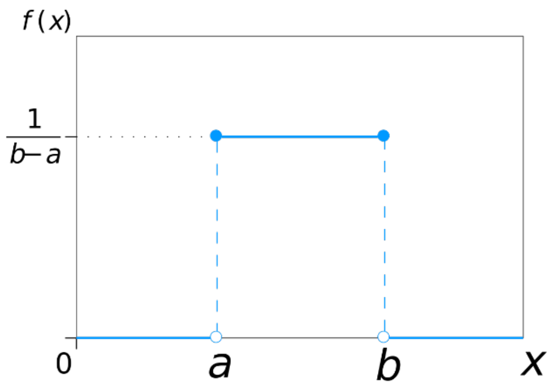


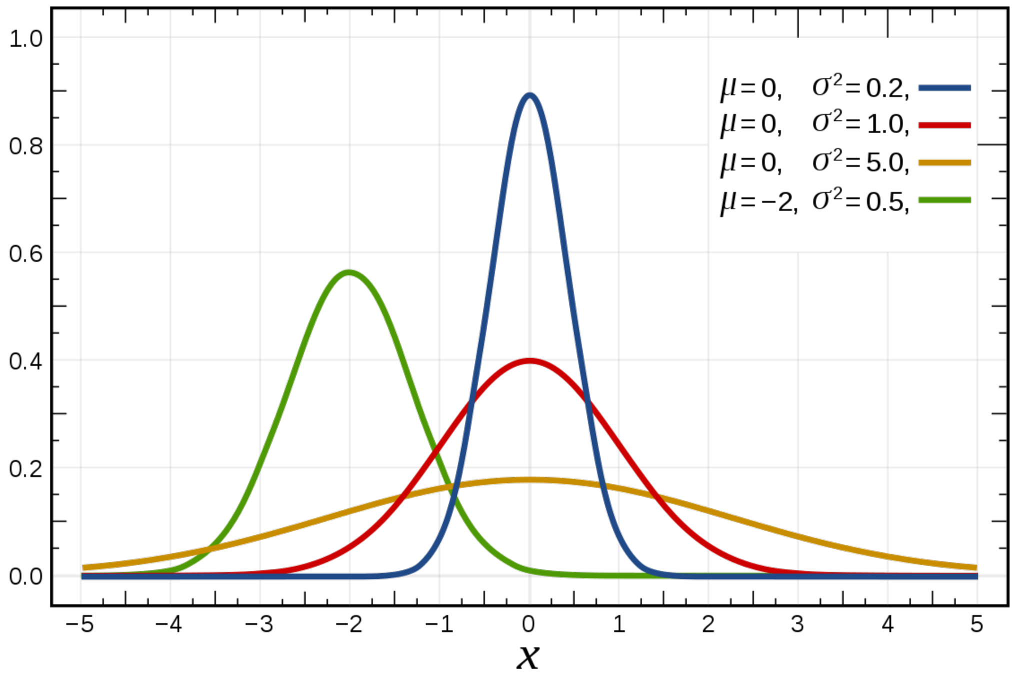


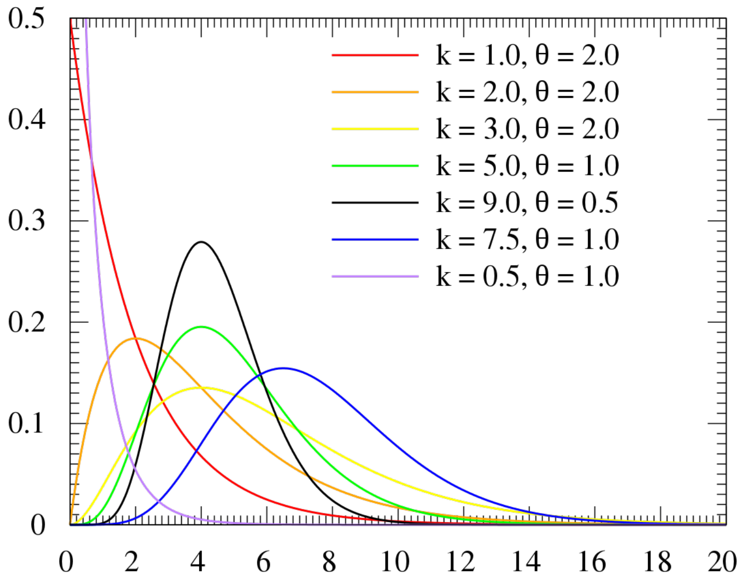


Data profiling: theoretical distributions

- **Continuous distributions**
  - uniform
  - Gaussian
  - Gamma

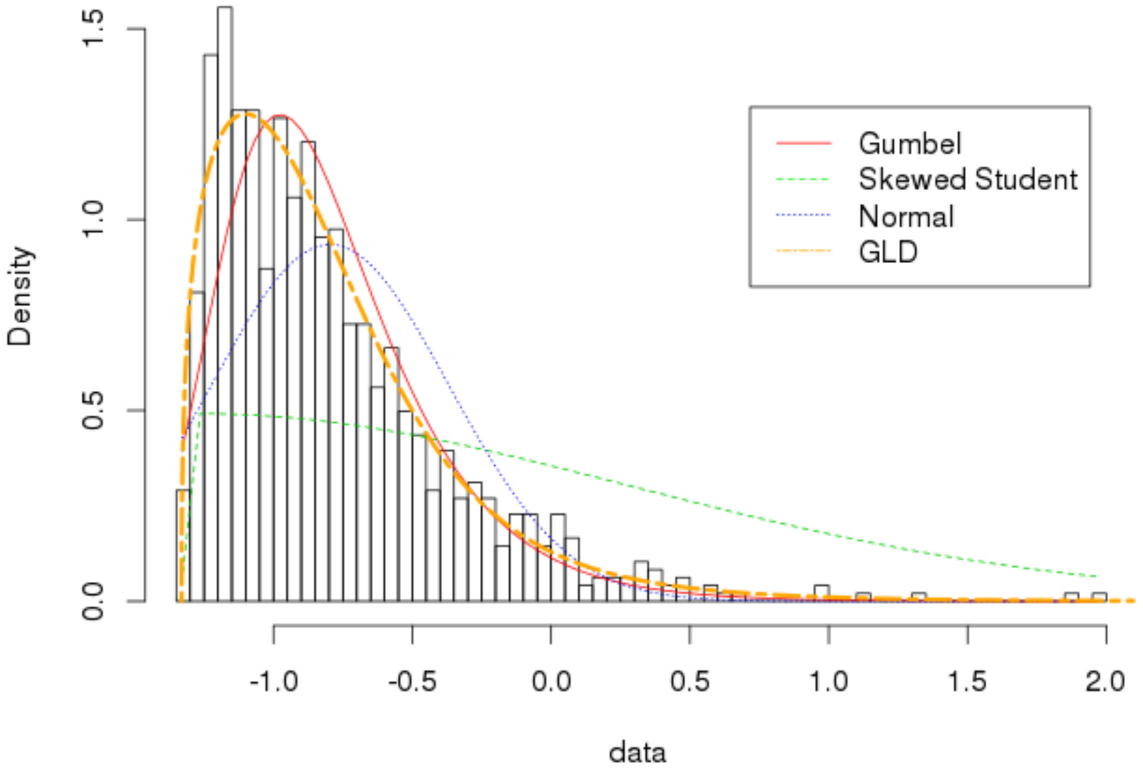






Data profiling: fitting

- Learn parameters from sample to describe the variable
- Kolmogorov-Sminorv statistical test to assess fitting between sample and theoretical distribution
  - https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html

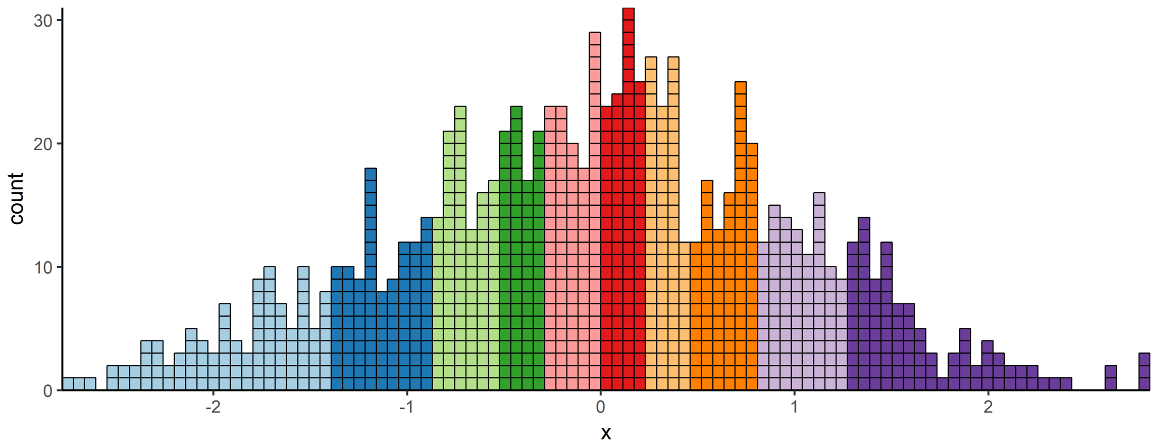


Univariate summary statistics

- *sample size*: number of data observations, n
- *mean*: arithmetic mean is the average value

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- *maximum, minimum, range* $(\max - \min)$
- *percentiles*
  - median, max and min correspond to the 50, 100 and 0 percentiles
  - 5, 10, 25 (first quantile), 75 (third quantile), 90, 95 are also informative



Univariate summary statistics

Center statistics are relevant to understand expectations

- *harmonic mean*
- *median*
  - sorted values, the median is the value that splits the distribution in half
  - $median(1,1,1,2,3,4,5) = 2$
  - If n is even, the median can be found by interpolating them
- *mode* for categorical and discrete numeric values
  - $mode(1,2,2,3,4,4,4) = 4$
  - application in continuous variables: after rounding, bin sorting, discretization
- *trimmed mean*
  - lop off a fraction of the upper and lower ends of the distribution, and take the mean of the rest
  - Example with lop off two: $0,0,1,2,5,8,12,17,18,18,19,19,20,26,86,116$
    - trimmed mean = 13.75
    - arithmetic mean = 22.75

Univariate summary statistics

Deviation statistics are important to assess the variability of variable measurements

- *Standard deviation*: square root of the variance

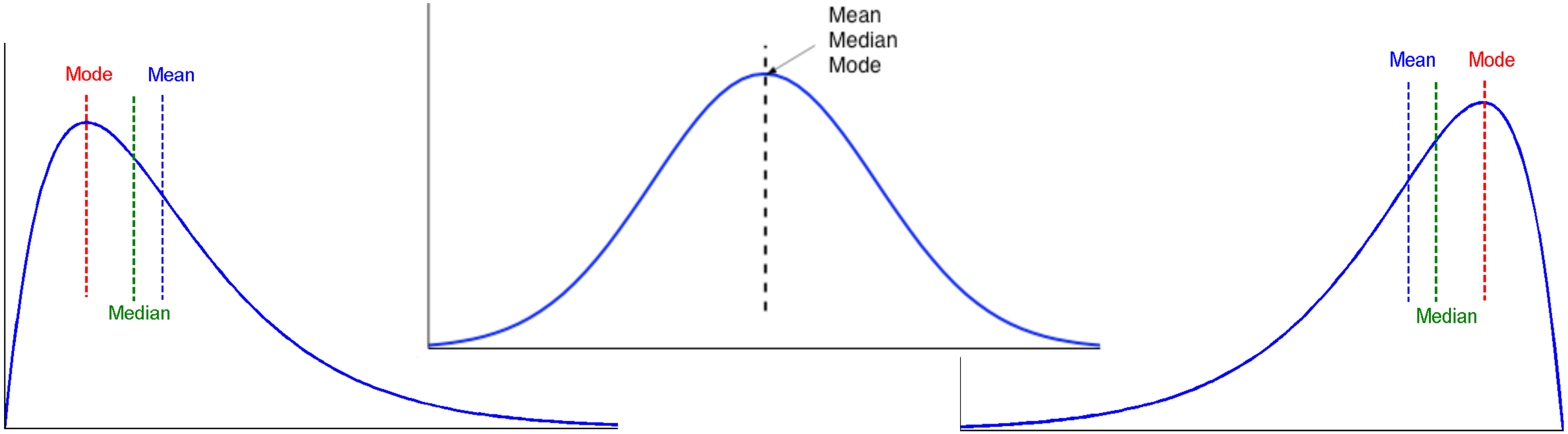
$$\sigma_{population} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{\mathbf{x}})^2}, \quad \sigma_{sample} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{\mathbf{x}})^2}$$

  – **population-based** deviation
    – divided by n
  – **sample** deviation
    – divided by $n - 1$
    – conservative estimate (higher variance) because as we are unable to observe the whole population

  – *example*: 1, 2, 15 measurements
    – $\mu = 6, \sigma_{population} = 6$ (approx.), $\sigma_{sample} = 7.3$ (approx.)

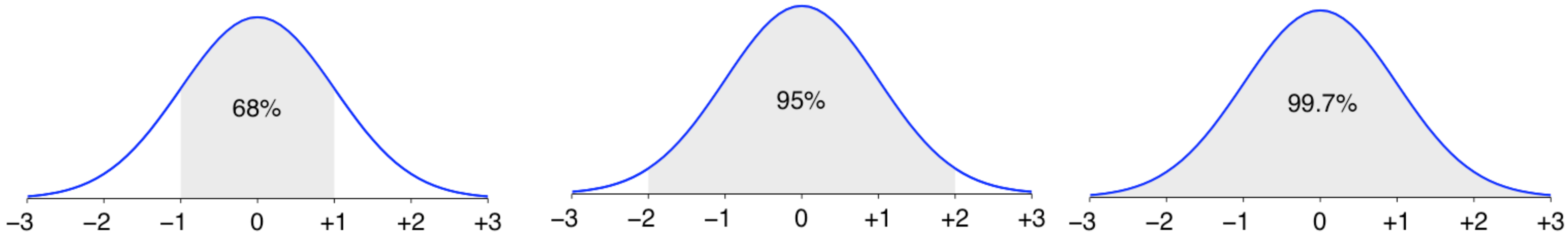
Univariate data statistics: skew

- In a skewed distribution the bulk of the data are at one end of the distribution
  - If the bulk of the distribution is on the right (tail is on the left): **left skewed** or negatively skewed distribution
  - If the bulk of the distribution is on the left (tail is on the right): **right skewed** or positively skewed distribution
- **Symmetric** distributions are not skewed
- Percentile statistics are not distorted by outliers



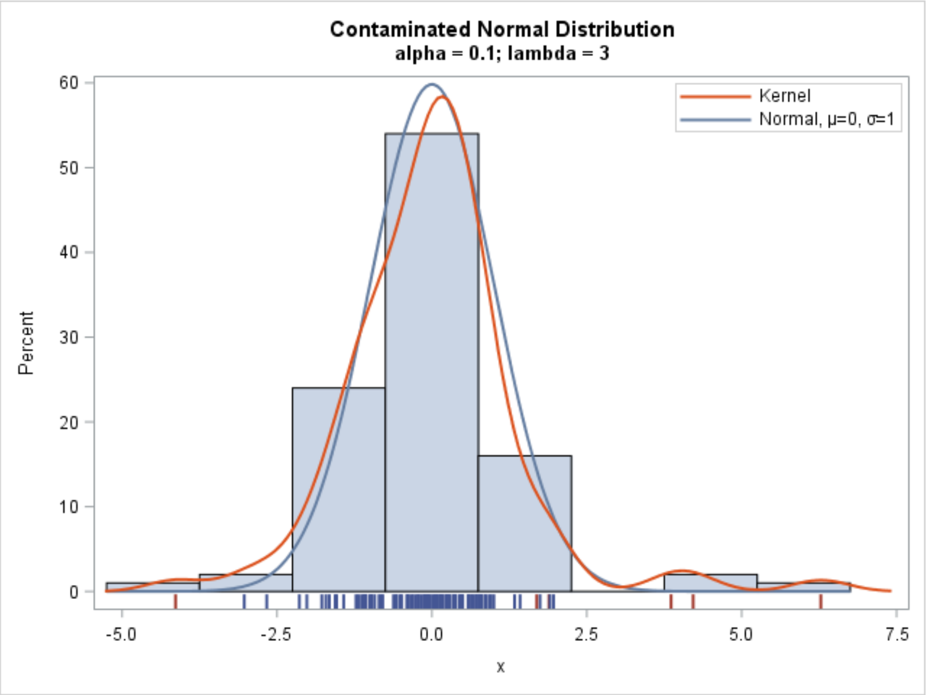
Properties of Normal distribution

- Many real-world variables are well-approximated to a Gaussian curve
- How to check if one variable satisfies the Gaussian assumption?
  - use the introduce Kolmogorov-Sminorv or, more suitably, Shapiro-Wilk test
  - remember the central limit theorem: 30 measurements are often necessary to check this assumption
- Interesting properties of the Normal curve:
  - from μ–σ to μ+σ: contains about 68% of the measurements (μ: mean, σ: standard deviation)
  - from μ–2σ to μ+2σ: contains about 95% of it
  - from μ–3σ to μ+3σ: contains about 99.7% of it



Outliers

- **Outlier** values = uncommon values
  - unexpected measurements against a variable distribution
- Mean and the variance are based on averages, hence sensitive to outliers
- Outliers can cause strong effects that can wreck our interpretation of data
  - for example, the presence of a single outlier can render some statistical comparisons insignificant
- Detecting and removing outlier values requires judgment and depend on one‘s purpose



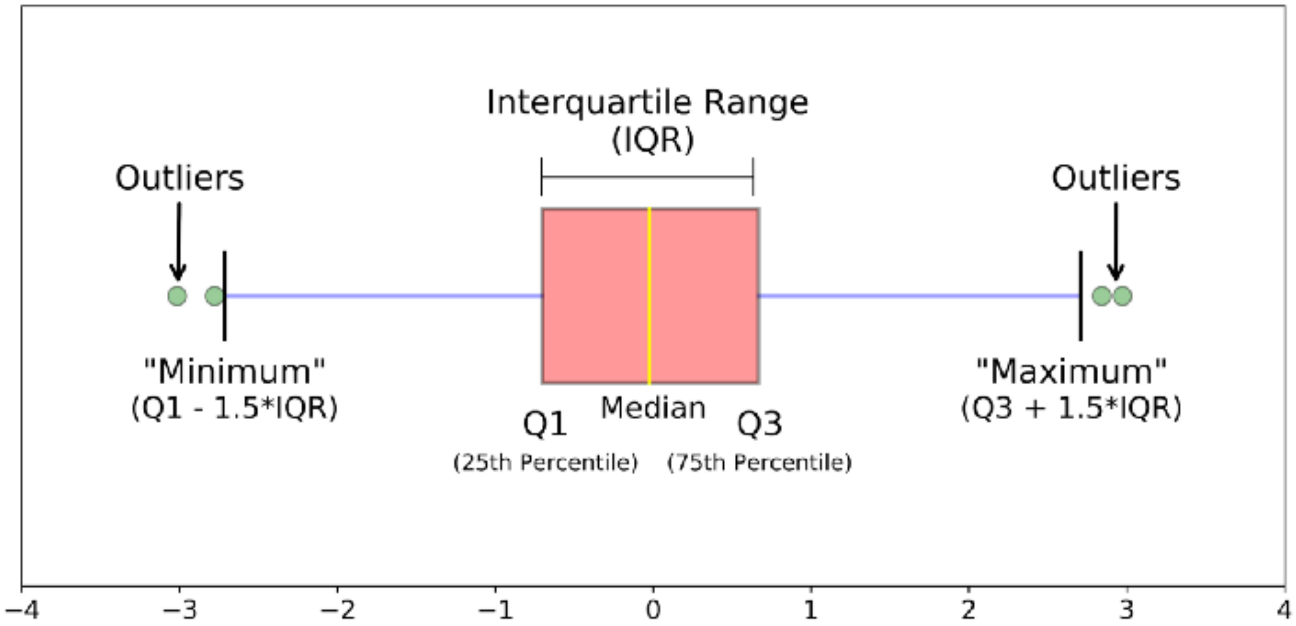
Interquartile range (IQR)

- **Interquartile range** is used to measure value expectations
  - distribution can be divided into **quartiles**, each containing the same number of observations
  - the difference between the highest value in the third quartile and the lowest in the second quartile is the **interquartile range**
  - example:
    - $quartiles(1,1,2,3,3,5,5,5,5,6,6,100) = \{(1,1,2),(3,3,5),(5,5,5),\ (6,6,100)\}$
    - interquartile range *5-3=2*

- IQR is empirically known to be robust against outliers
  - observations falling outside $[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR]$ are seen as outliers

Boxplot

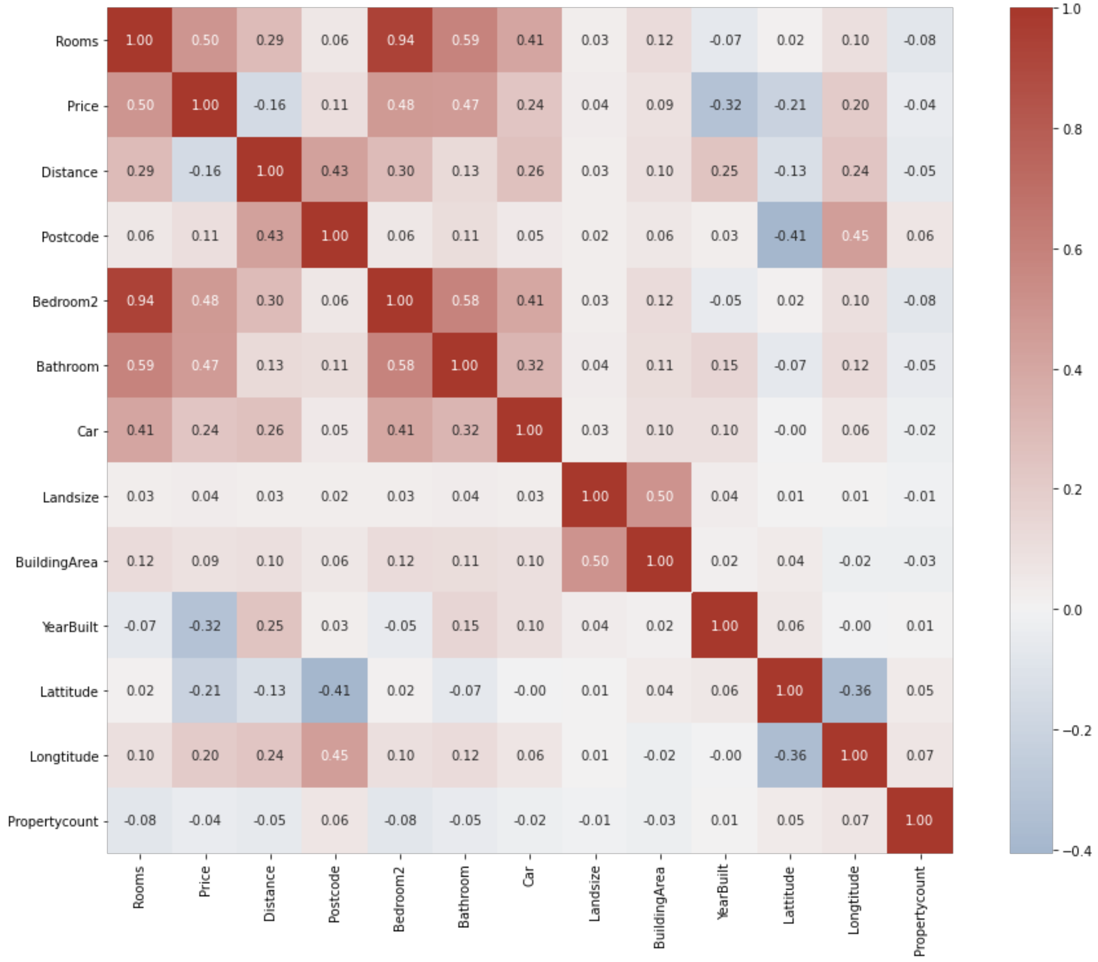
- The distribution of variable can be visualized:
  - bar charts (categorical variables) and histograms (numeric variables)
  - theoretical probability distributions
  - boxplots: particularly useful to visually detect outlier values



Bivariate data statistics

- Considering pairwise input variables:
  - check whether the two distributions are **correlated**
  - if highly correlated, variables may be **redundant**
    - select the one with the highest variability

- *Exercise*: select non-redundant variables on the provided right example



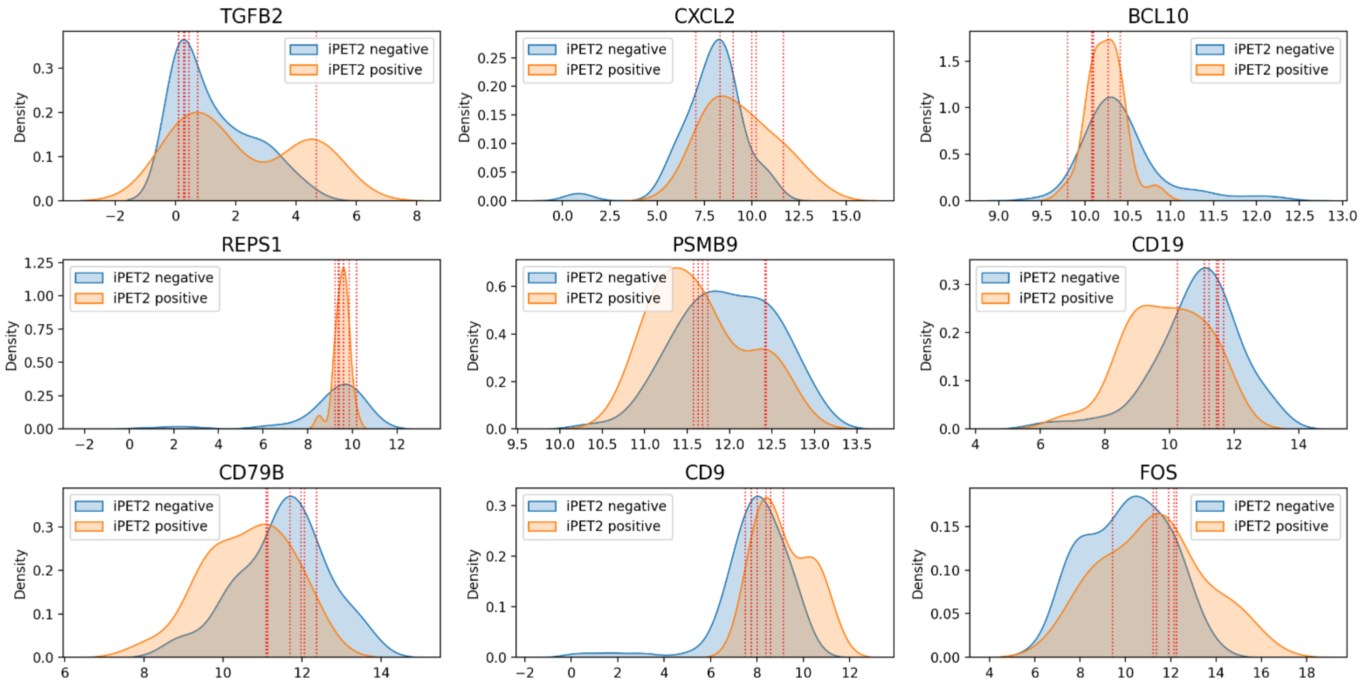
Bivariate data statistics

- Considering one input and one output variable
  - when referring to class variables: we want to assess the ***discriminative power*** of the input variable
  - when referring to numeric output variables: we want to assess the ***correlation*** with the input variable
    - the higher the correlation, the higher the relevance of the input variable to characterize the targets

  - if both input-output variables are numeric
    - linear correlation given by **Pearson** correlation coefficient (PCC)
    - rank-based correlation given by **Spearman** tau prioritizes ranks instead of magnitude
  - if variables are either ordinal or numeric: **Spearman** tau is suggested
  - if one variable is nominal and other numeric: **analysis of variance** (ANOVA)
  - if both variables are nominal: χ^2

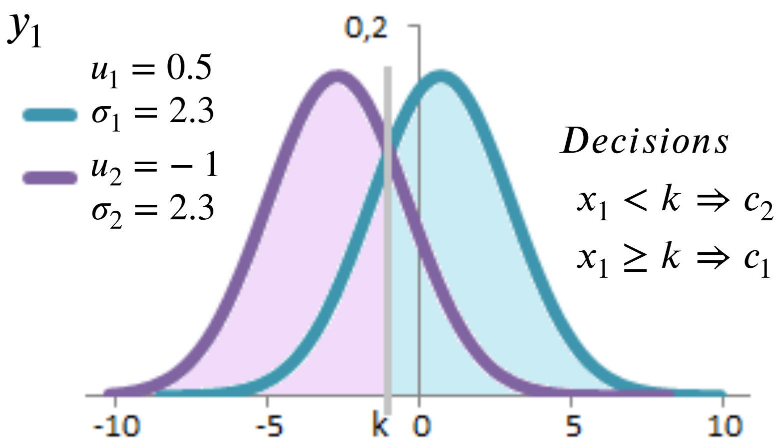
Discriminative power

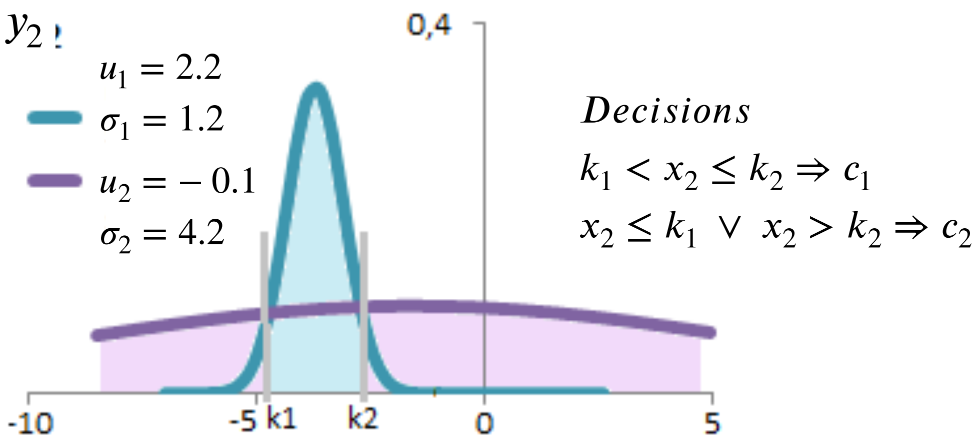
- **Class-conditional distributions**
  - the higher the dissimilarity between class-condition distributions: the higher the discriminative power
  - *exercise*: consider a dataset composed by the following 9 numeric input variables and binary class
    - are we in the presence of a simple or difficult classification task?



Discriminative power

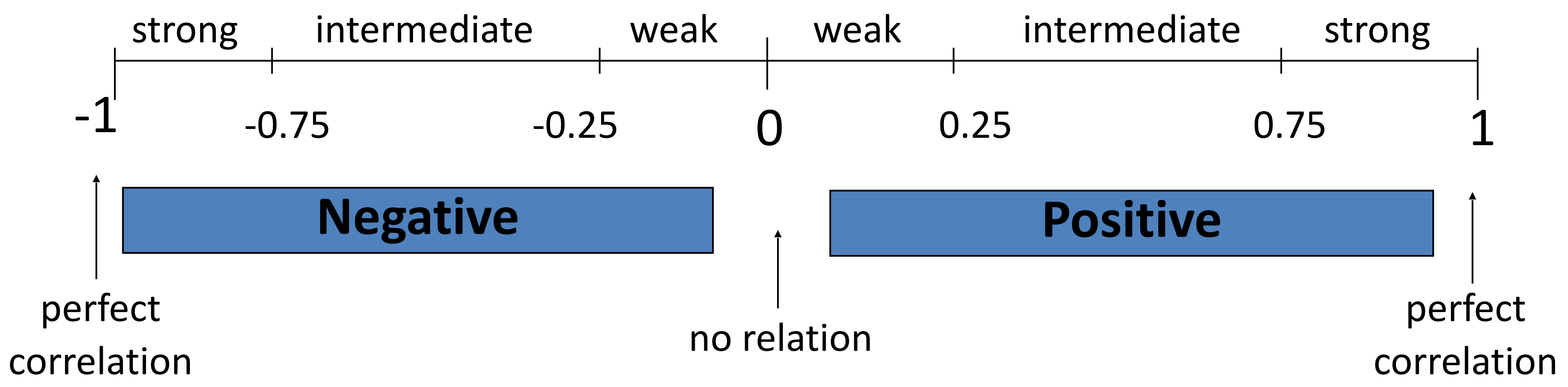
- Using class-conditional distributions:
  - **discriminative rules** can be inferred by identifying the class of higher probability along the input values
    - this classifier is termed **univariate discriminant**





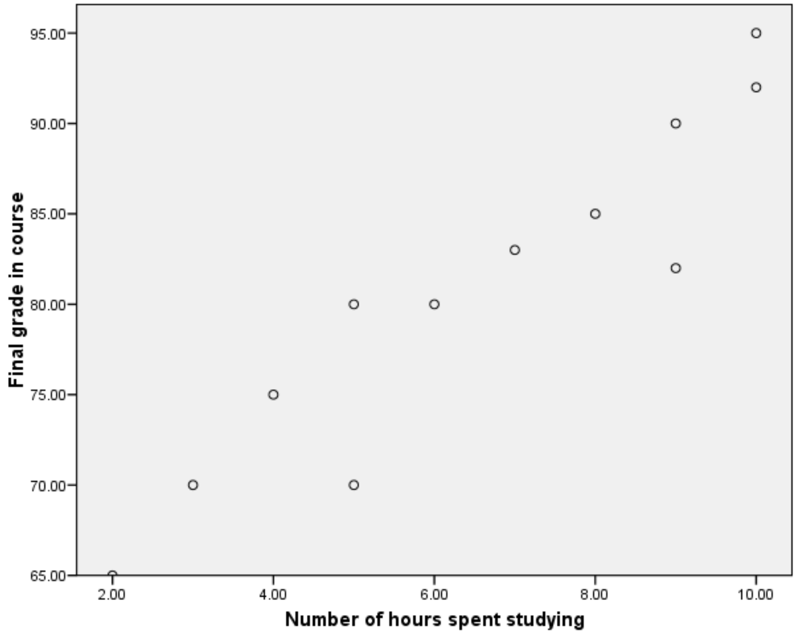
Correlation...

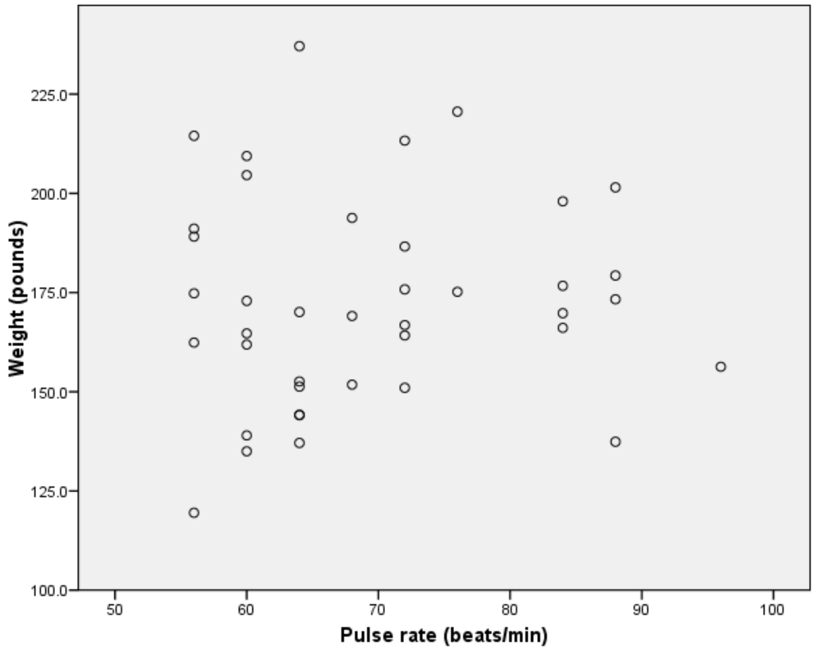
- Relationship between two quantitative attributes
  - correlation: degree to which two attributes are related (in [-1,1])
    - the *sign*: nature of association (> 0 direct; < 0 inverse)
    - the absolute *value* of r: strength of association
    - unable to infer causal relationships



Correlation...

Scatter diagrams can be used to visually assess correlation



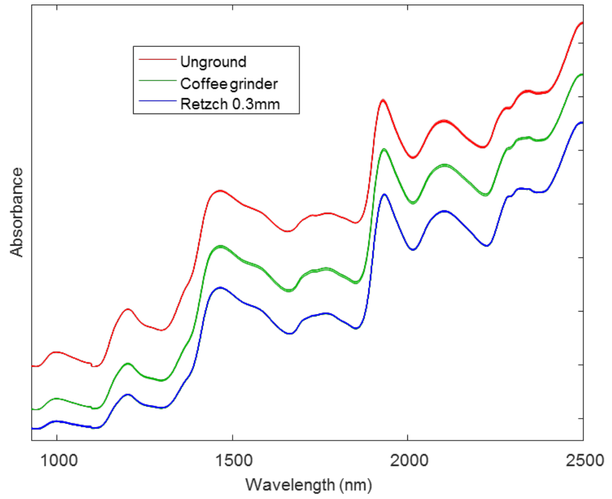


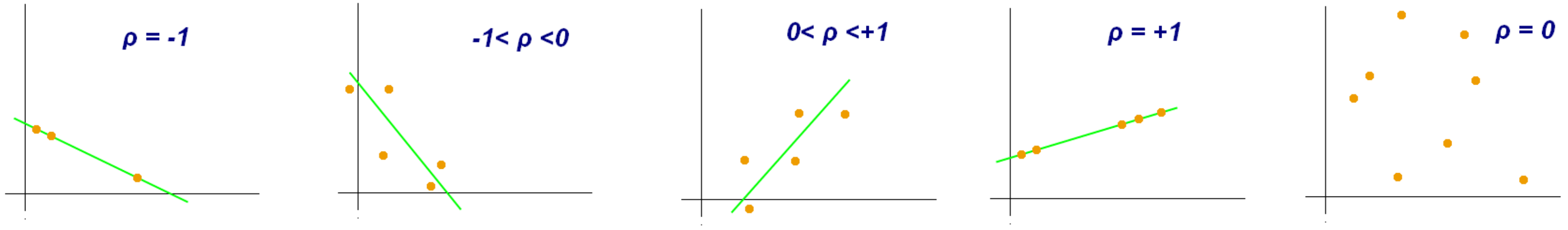
- each pair of values is treated as a pair of coordinates and plotted as points in the plane
- provides a first look at bivariate data to see clusters, outliers, etc.

Pearson correlation

- Pearson correlation (or product moment correlation) coefficient
  - only suitable for numeric attributes
  - able to handle scales and shifts

$$\mathbf{r} = \frac{cov(y_1, y_2)}{\sqrt{var(y_1)}\sqrt{var(y_2)}} = \frac{\Sigma y_1 y_2 - \frac{\Sigma y_1 \Sigma y_2}{n}}{\sqrt{\left(\Sigma y_1^2 - \frac{(\Sigma y_1)^2}{n}\right)\left(\Sigma y_2^2 - \frac{(\Sigma y_2)^2}{n}\right)}}$$





Pearson correlation

Anxiety ()	**Test score ()**			
10	**2**	100	4	20
8	**3**	64	9	24
2	**9**	4	81	18
1	**7**	1	49	7
5	**6**	25	36	30
6	**5**	36	25	30
$\sum y_1 = 32$	$\sum y_2 = 32$	$\sum y_1^2 = 230$	$\sum y_2^2 = 204$	$\sum y_1 y_2 = 129$

$$r = \frac{(6)(129) - (32)(32)}{\sqrt{\left(6(230) - 32^2\right)\left(6(204) - 32^2\right)}} = -.94$$

indirect strong correlation

Spearman rank

- Non-parametric coefficient
  - works with rankings instead of absolute values
- **How?**
  1. Rank the values of y_1 and y_2
  2. Apply the Pearson correlation

In the given example:

$$r_s = PCC((5\ 6\ 1.5\ 3.5\ 3.5\ 7\ 1.5), (3\ 5.5\ 7\ 5.5\ 4\ 2\ 1))$$

$$r_s = -0.17$$

r_s denotes the magnitude of association

education level (y_1)	**income** (y_2)	rank y_1	rank y_2
Preparatory	25	5	3
Primary	10	6	5.5
University	8	1.5	7
Secondary	10	3.5	5.5
Secondary	15	3.5	4
Illiterate	50	7	2
University	60	1.5	1

Outline

- **Machine learning**
  - intelligence and learning
  - data science and AI
  - symbolic learning
  - terminology
  - descriptive and predictive tasks
- **Univariate data analysis**
  - numeric and categoric variables
  - empirical and theoretical distributions
  - summary statistics
  - outlier removal
  - discriminant analysis
  - correlation

Thank You

miguel.j.couceiro@tecnico.ulisboa.pt

andreas.wichert@tecnico.ulisboa.pt