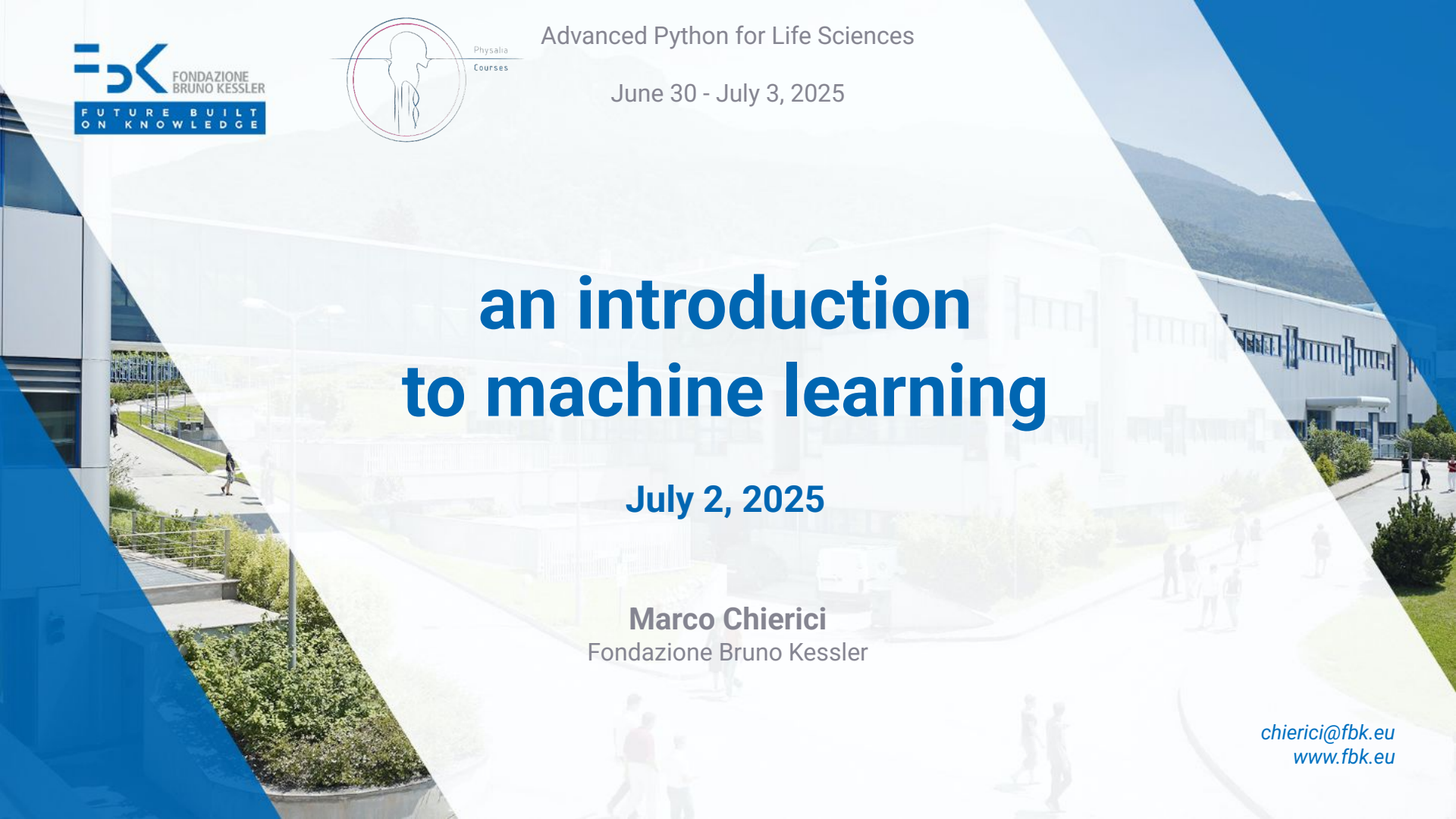


Advanced Python for Life Sciences

June 30 - July 3, 2025

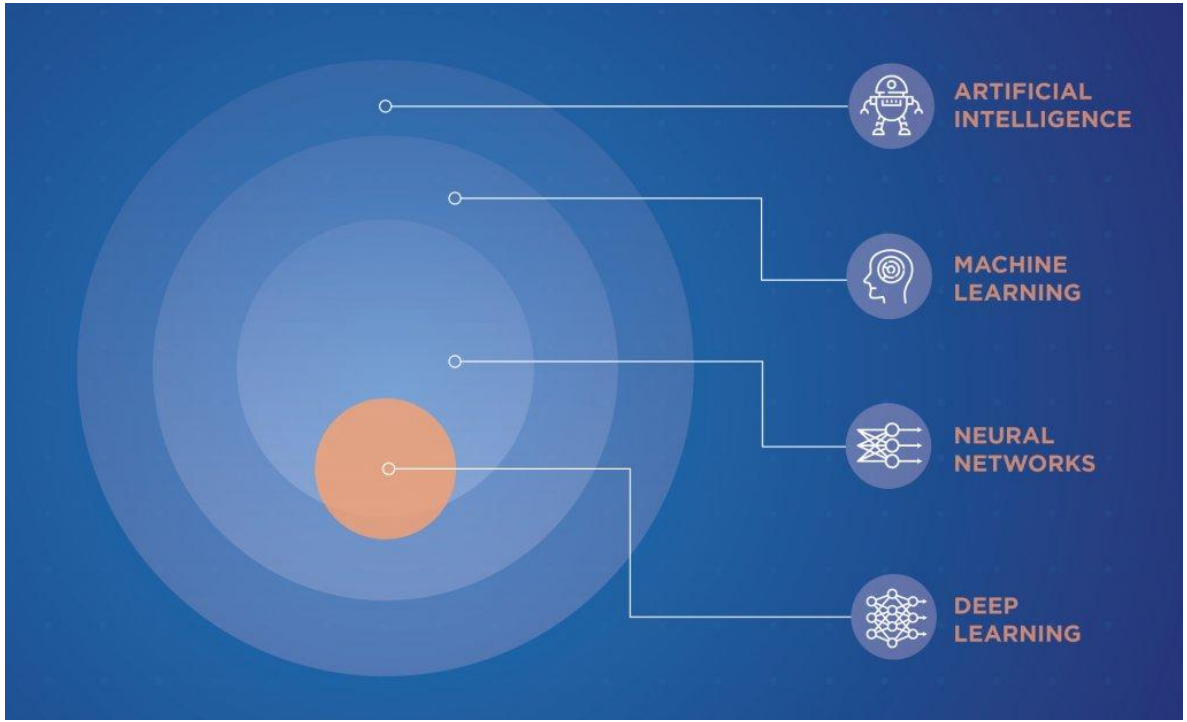
an introduction to machine learning

July 2, 2025

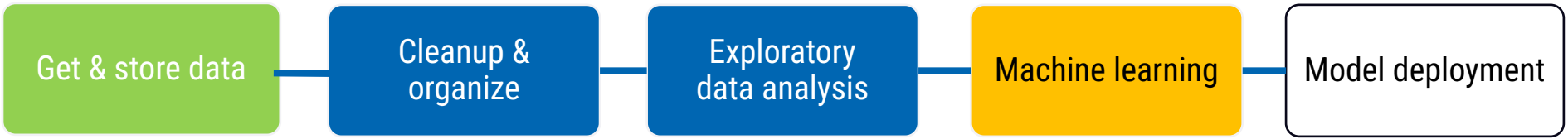
Marco Chierici
Fondazione Bruno Kessler

chierici@fbk.eu
www.fbk.eu

AI > machine learning > deep learning



machine learning
general workflow



Supervised
(predict an outcome)

Unsupervised
(discover patterns)

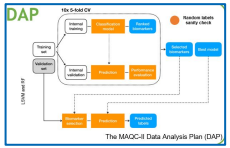
computational biology
a conceptual pipeline

UPSTREAM

DOWNSTREAM

SAMPLES

Bioinformatics



Knowledge Bases

STAT / AI
DEEP LEARNING

Predictive Medicine

Data preprocessing

Sense-making

how are things learned?

- Through memorization
  - **Memorize** as many facts as you can
  - **Declarative knowledge**, statements of truth
  - Limited by time necessary to observe facts and memory to store them
- Through inference
  - **Deduce** new information from old knowledge
  - **Imperative knowledge**
  - Limited by the accuracy of the deduction process: it's essentially a predictive activity and the underlying assumption is that the past can predict the future
- Machine Learning
  - We give data to a program
  - We want the program to **infer useful information from implicit patterns** in the data
  - And we want to use this to **make predictions** on data we haven't seen yet

learning

universal process

learning seems a characteristic aspect of human ingenuity, but it is a **general** process, based on the assimilation of examples and experiences and the synthesis of information

empiric process

in order to learn how to identify objects and acquire basic activities, it is essential to have **examples** and to be able to integrate skills

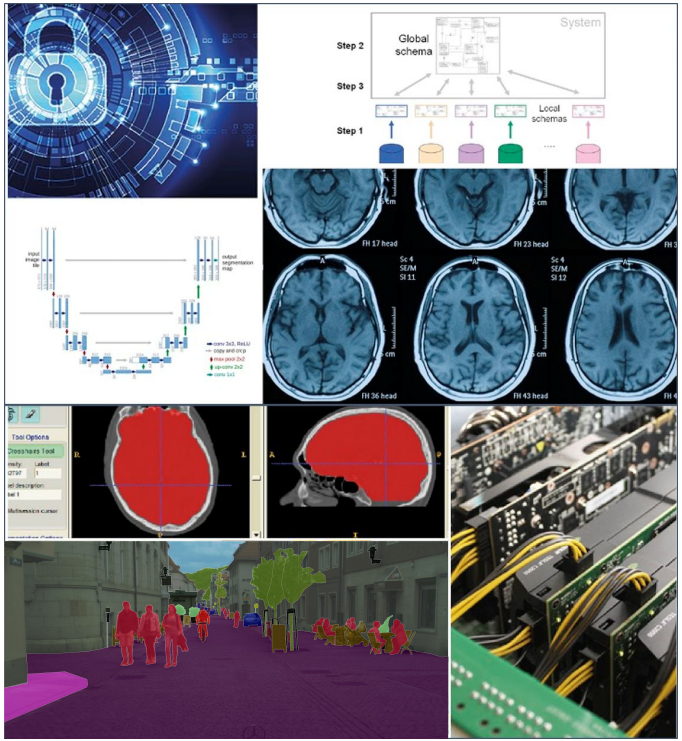
we learn from examples, often through a teacher, modifying actions and decision from our mistakes

*a **continuous** process, building capacity to respond to **change** and not just the mere execution of rules*

machine learning

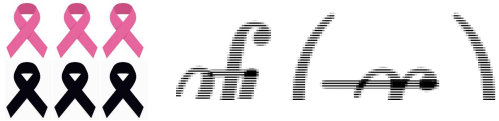
“the ability to learn making predictions and improve from experience without being explicitly programmed”

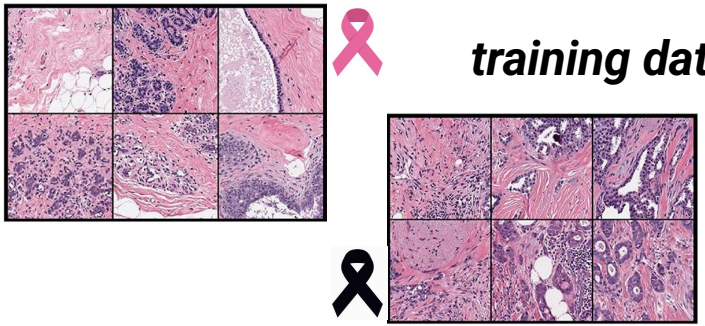
arthur samuel, 1959



AI can solve problems such as autonomous driving, facial recognition, and simultaneous translation. **Health improvement is now among its most ambitious goals.**

machine learning paradigm
"train to predict"



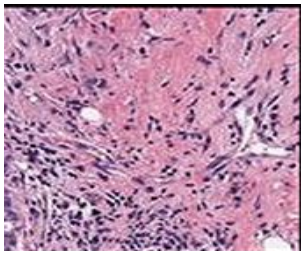


training data

model

trained model

training



new data

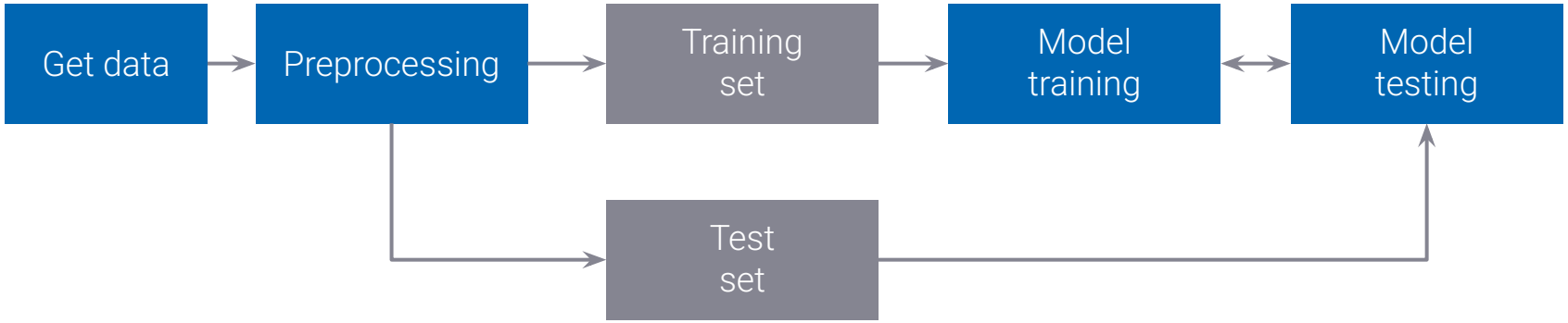
1%

99%

prediction

abridged from Giuseppe Jurman

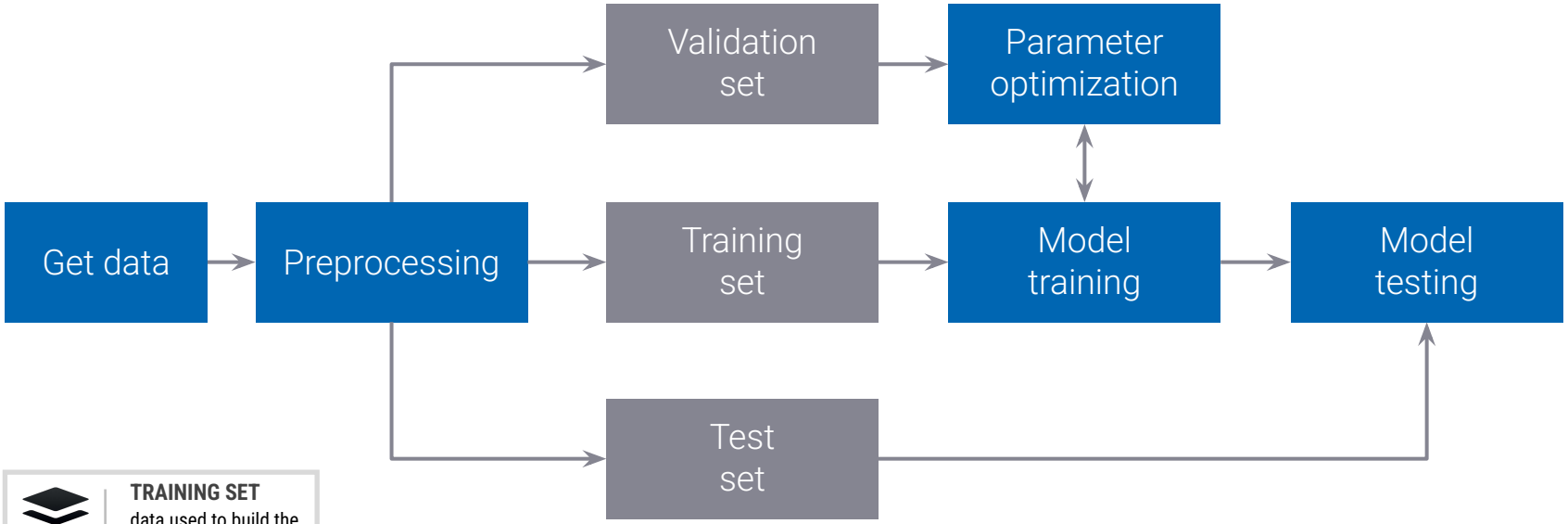
machine learning paradigm



TRAINING SET
data used to build the model

TEST SET
to test the model on unseen data

machine learning paradigm



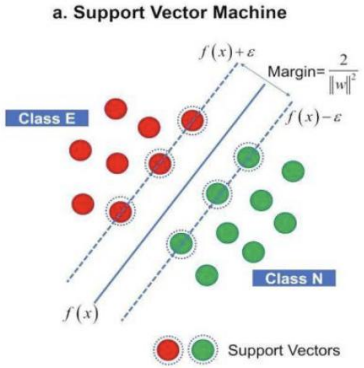
TRAINING SET
data used to build the model

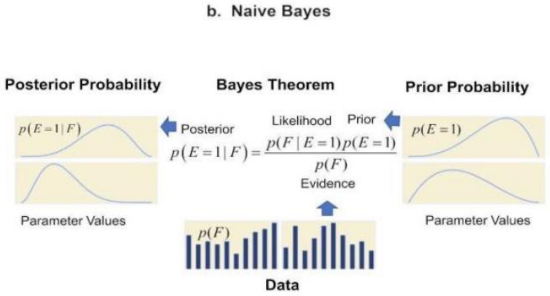
VALIDATION SET
to find model's optimal parameters

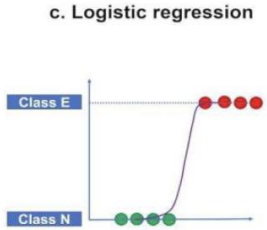
TEST SET
to test the model on **unseen** data

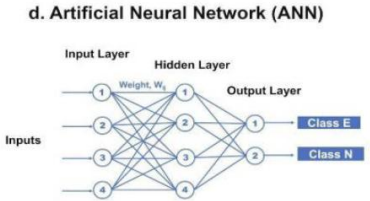
machine learning algorithms

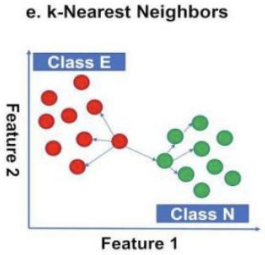
- Logistic Regression
- Tree-based methods
- K-Nearest Neighbors
- Support Vector Machines
- Naïve Bayes
- K-Means clustering
- Neural Networks

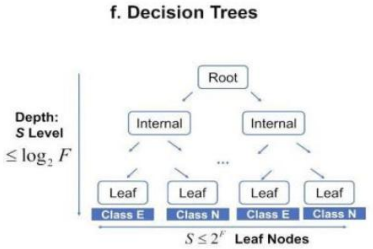


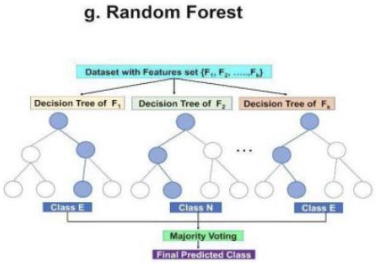


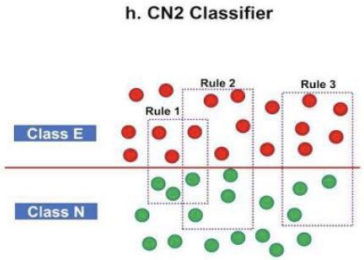


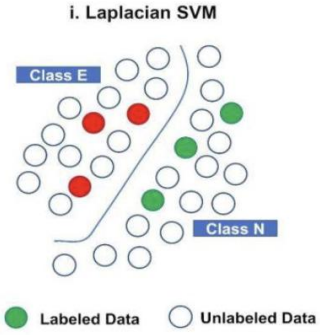












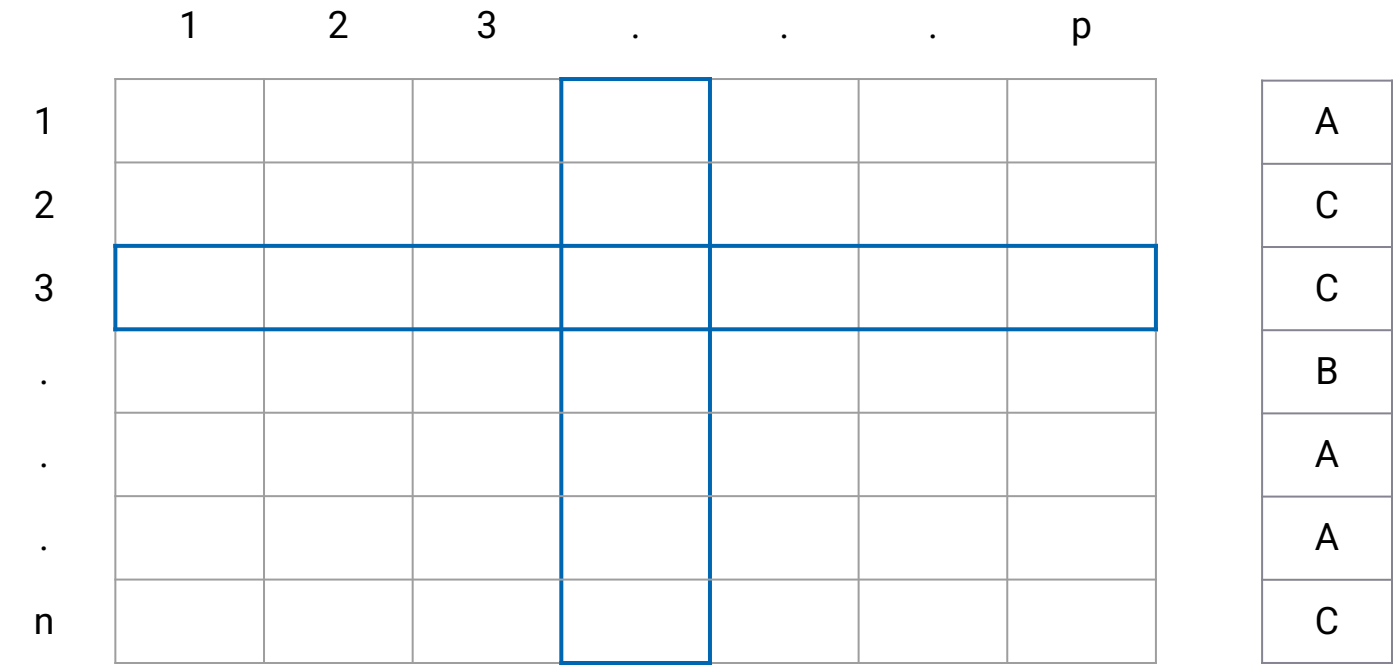
Gawade P et al, 2022

machine learning terms

columns
features, variables, predictors, components, dimensions

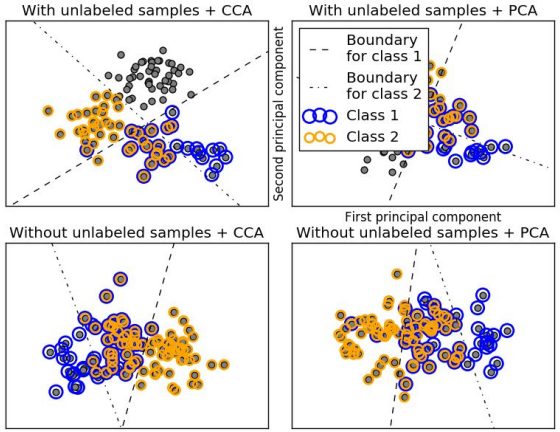
labels
groups, class, category

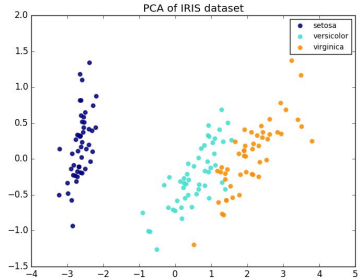
rows
samples, observations



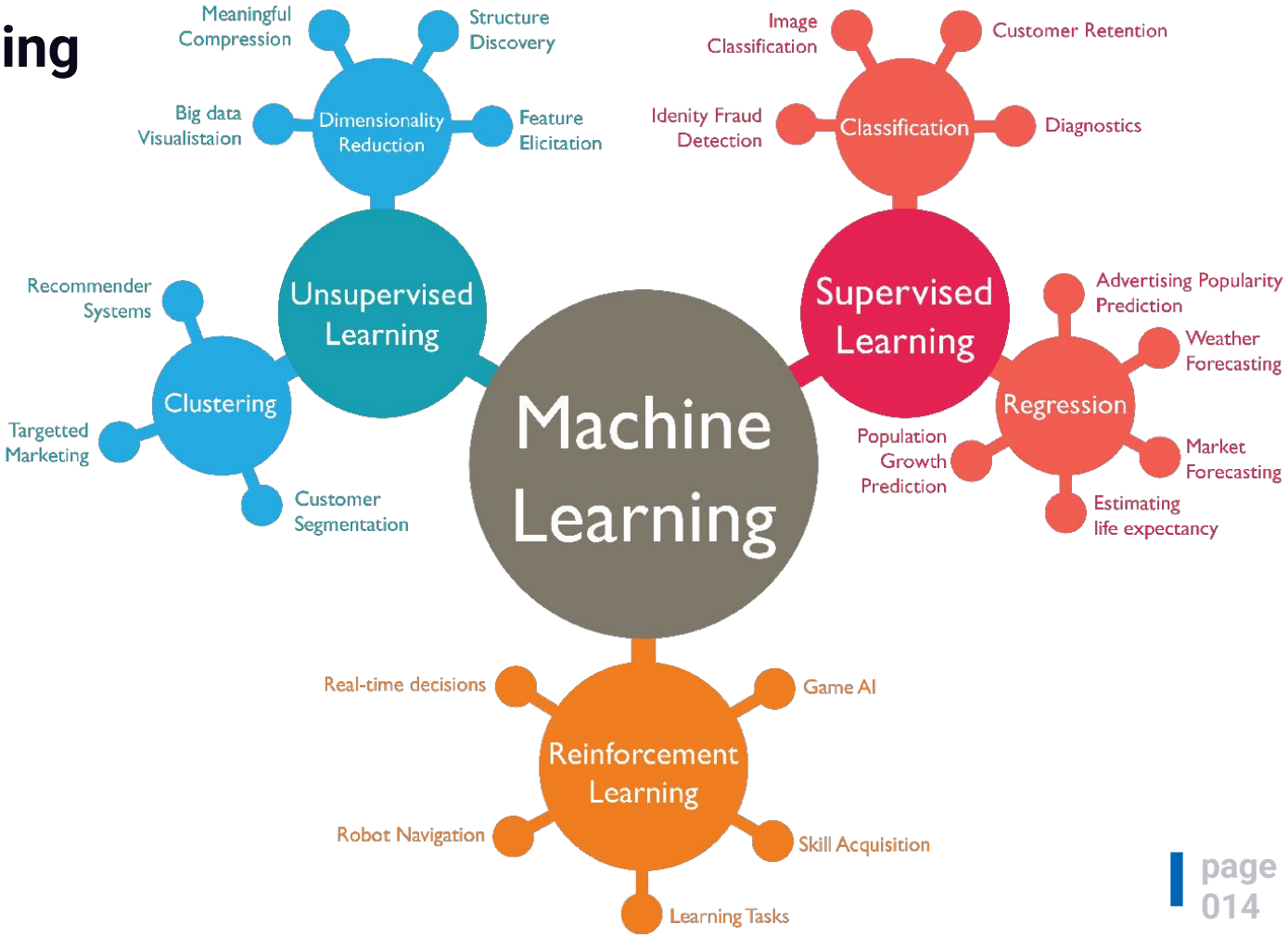
machine learning tools

- Free machine learning library for **Python**
- Open Source (BSD)
- Simple **fit / predict / transform** methods
- Python / NumPy
- Features various classification, regression and clustering **algorithms** including support vector machines, random forests, gradient boosting, k-means…





machine learning types



reinforcement learning

- Agents learn to **take actions in an environment** so as to **maximize** some notion of cumulative **reward**

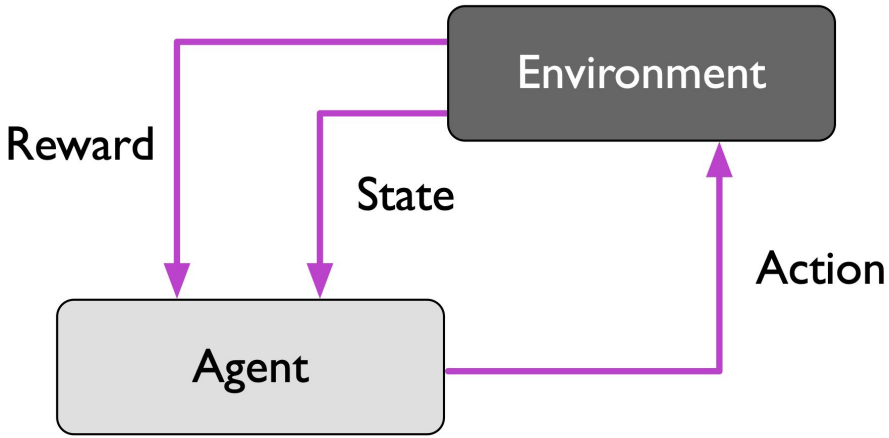
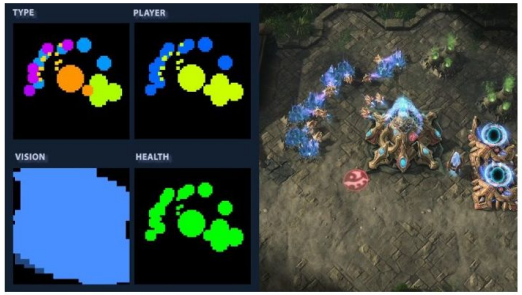
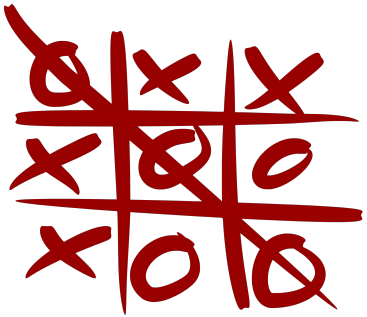


Figure credits: Raschka S, “Machine learning with pytorch and scikit-learn (Packt publishing, 2022)

reinforcement learning

AlphaGo





multi-armed bandits

unsupervised learning

Data has **no labels** 👉 no external teacher to supervise

A. **Group observations into clusters** (clustering)

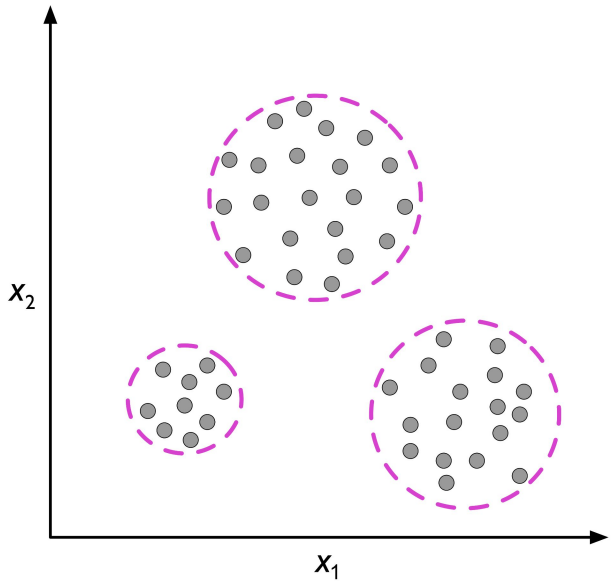


Figure credits: Raschka S, “Machine learning with pytorch and scikit-learn (Packt publishing, 2022)

unsupervised learning

Data has **no labels** 👉 no external teacher to supervise

B. **Reduce the number of features** (dimensionality reduction)

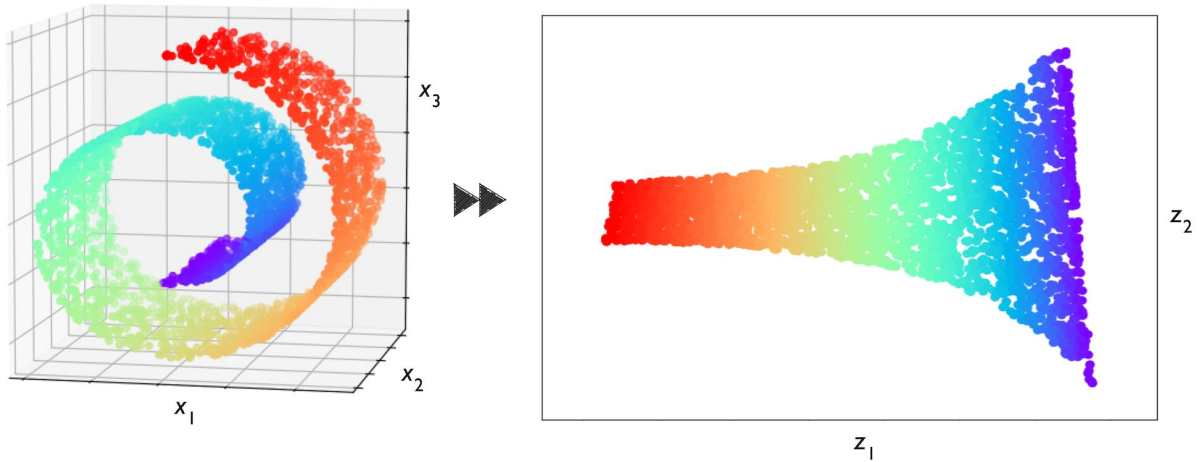


Figure credits: Raschka S, “Machine learning with pytorch and scikit-learn (Packt publishing, 2022)

supervised learning

Data has **labels** 👉 external teacher supervises the training

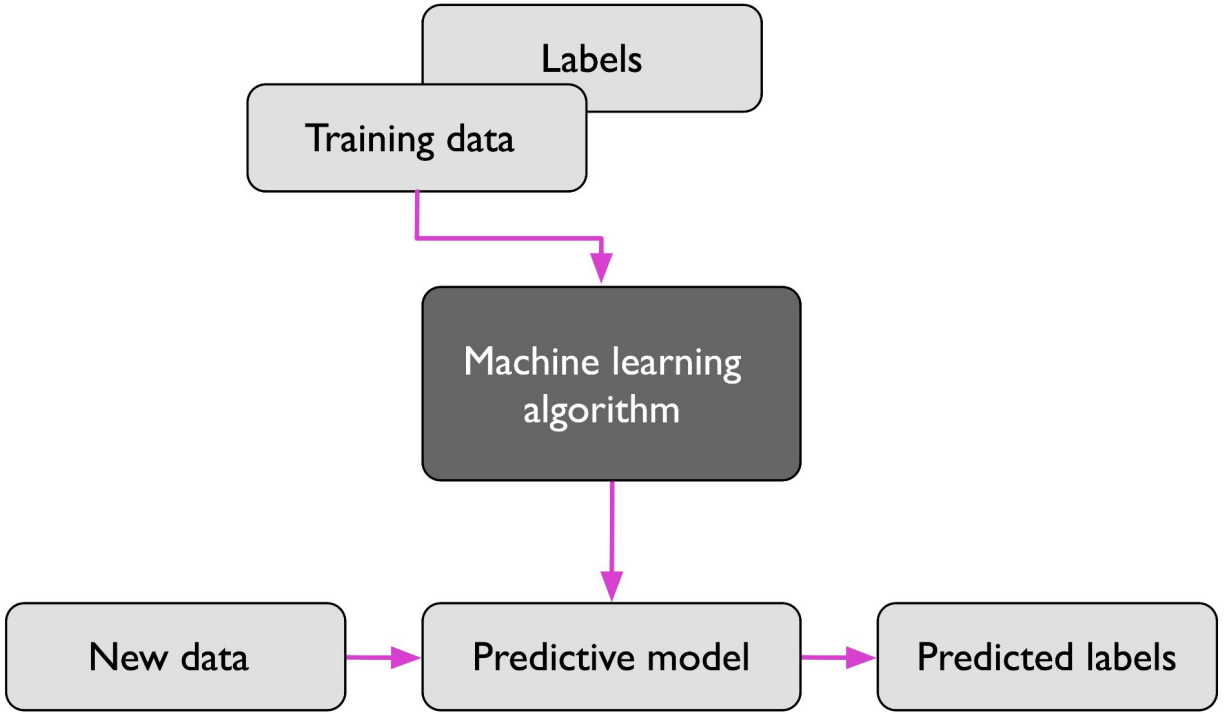


Figure credits: Raschka S, Packt publishing, 2022

supervised learning

Predict categorical class labels: **classification**

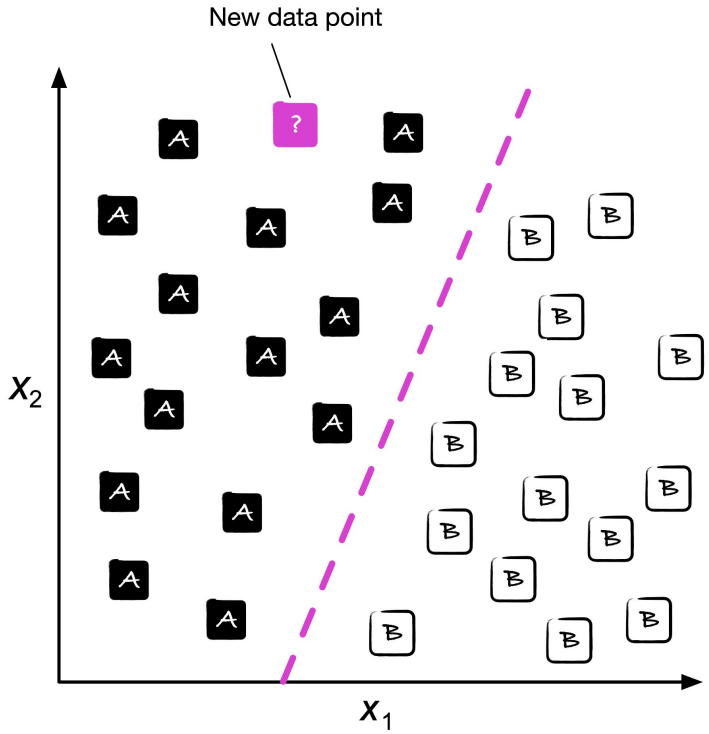


Figure credits: Raschka S, Packt publishing, 2022

supervised learning

Predict categorical class labels: **classification**

supervised learning

Predict continuous outcomes: **regression**

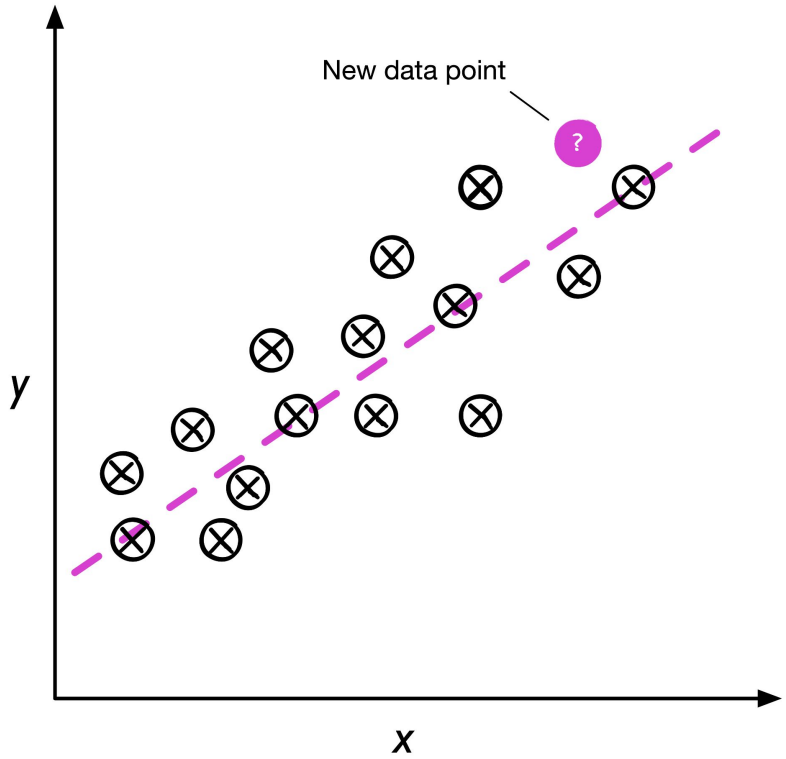
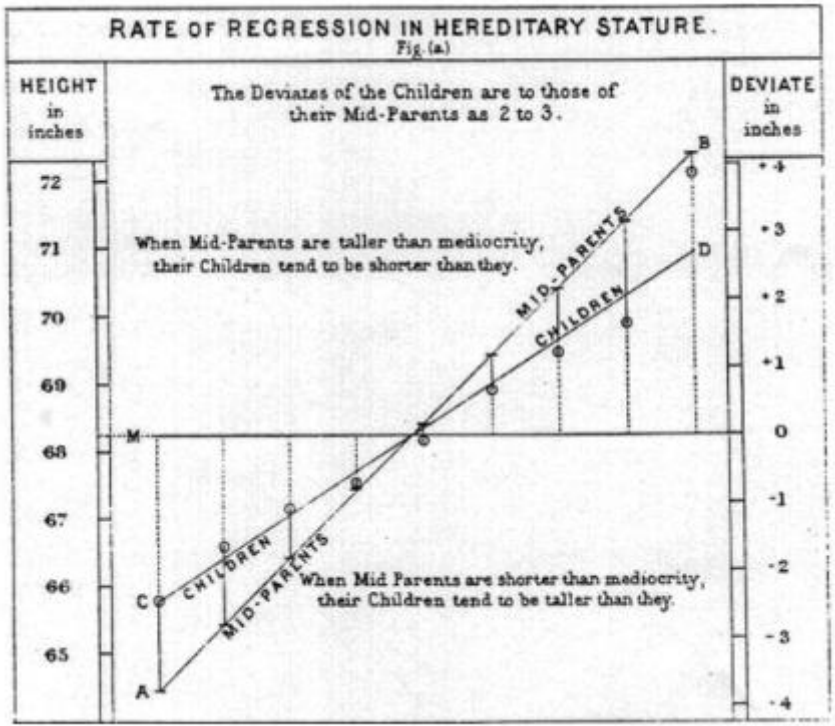


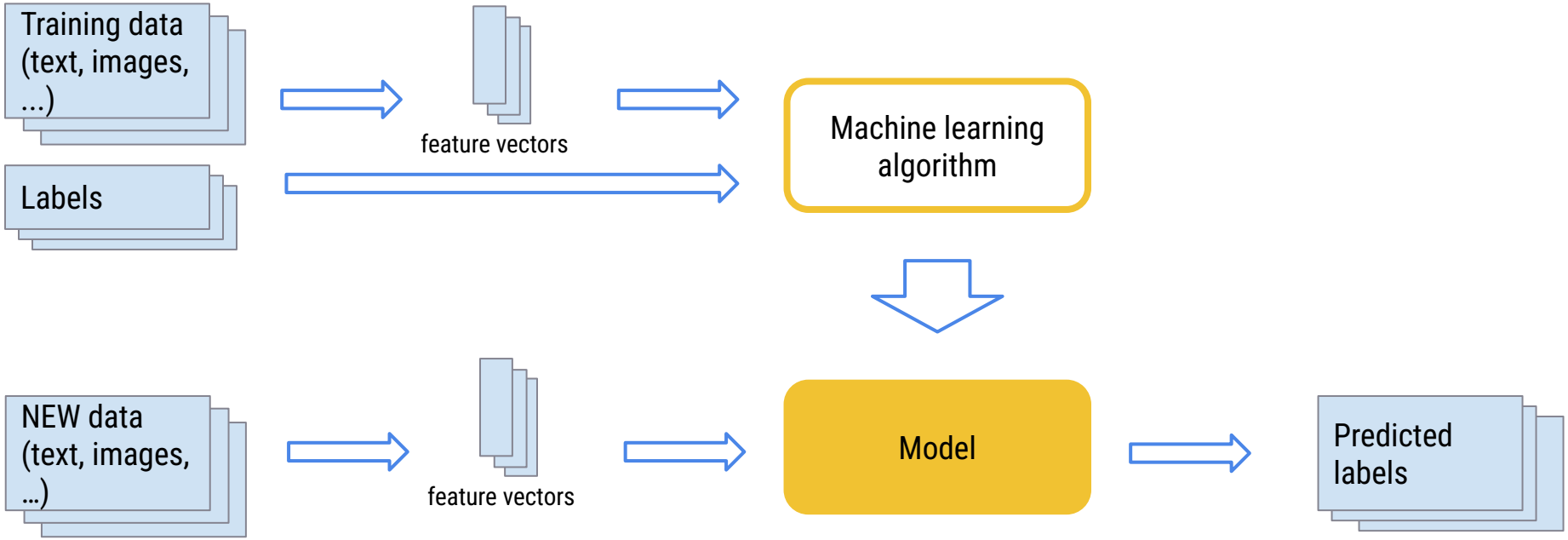
Figure credits: Raschka S, Packt publishing, 2022

Predict continuous outcomes: **regression**



Galton F. “Regression towards mediocrity in hereditary stature” (1886)

supervised learning
a conceptual workflow

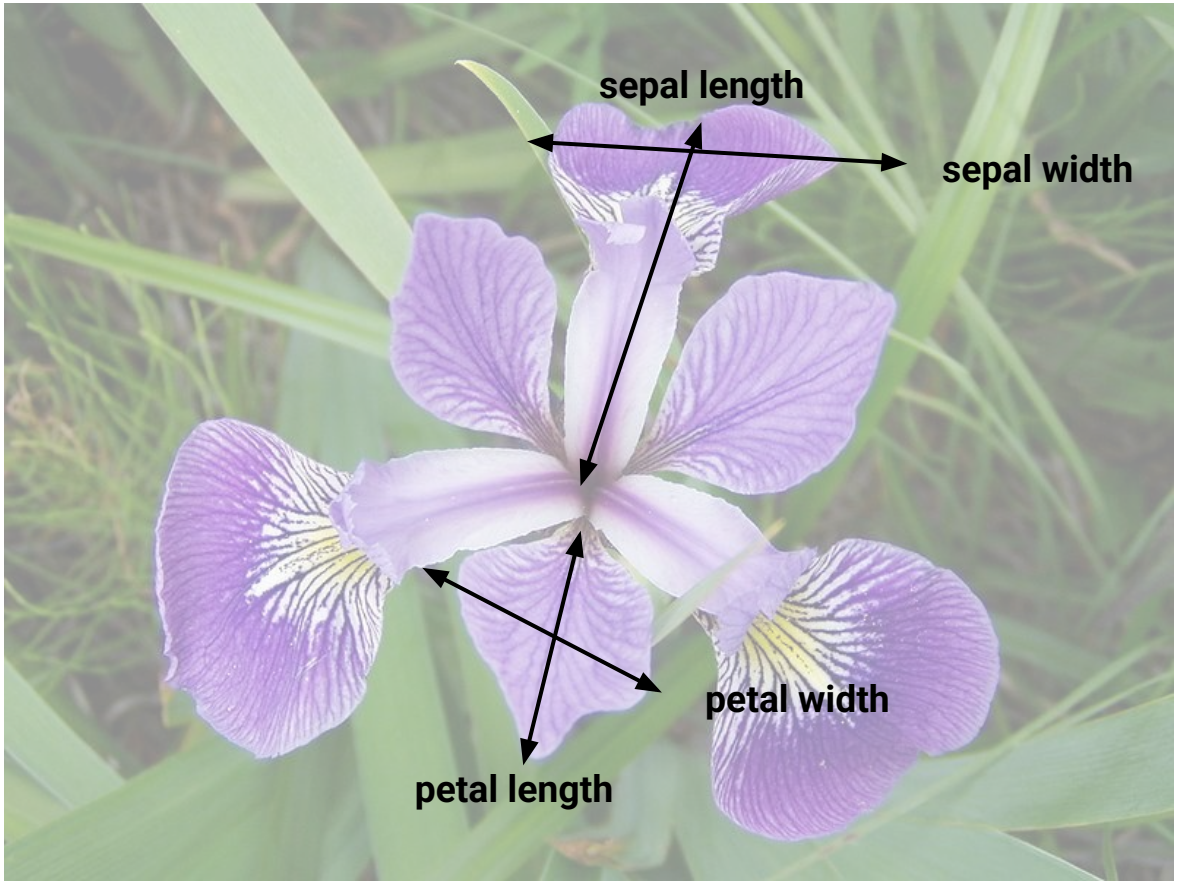


(adapted from Olivier Grisel, 2015)

working example
iris flower data set

- Or Fisher's *Iris* data set, or Anderson's *Iris* data set
- Describes a set of measurements collected for three distinct Iris species (*Iris setosa*, *Iris virginica* and *Iris versicolor*)
  - Sepal width and length
  - Petal width and length
- Introduced to describe Fisher's linear discriminant model, became a classical machine learning example

petal vs sepal



iris dataset in practice



features:
attributes of the data

target:
what has to be predicted

sepal length	sepal width	petal length	petal width	species
6.7	3.0	5.2	2.3	virginica
6.4	2.8	5.6	2.1	virginica
4.6	3.4	1.4	0.3	setosa
6.9	3.1	4.9	1.5	versicolor
4.4	2.9	1.4	0.2	setosa
4.8	3.0	1.4	0.1	setosa
5.9	3.0	5.1	1.8	virginica
5.4	3.9	1.3	0.4	setosa
4.9	3.0	1.4	0.2	setosa
5.4	3.4	1.7	0.2	setosa

to the notebook!

`day3/machine_learning_partial.ipynb`

thank you / grazie !

Acks

Giuseppe Jurman
Margherita Francescatto