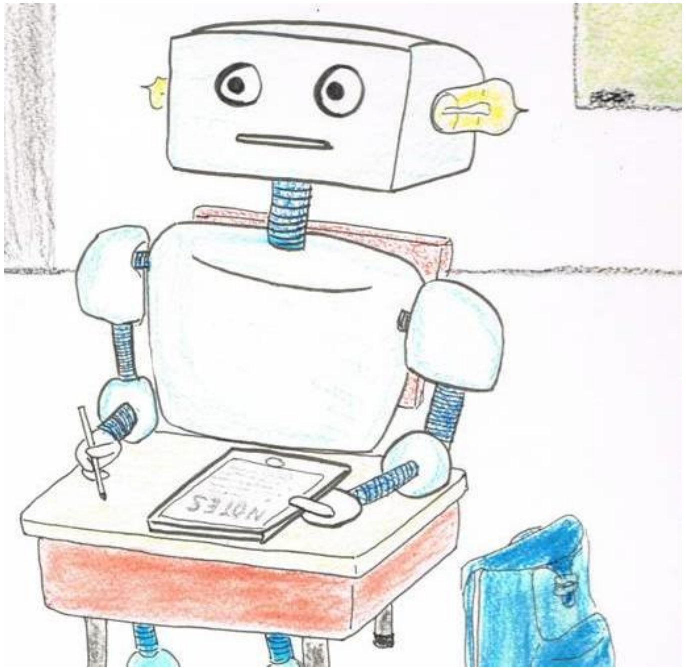
Machine Learning
An Overview

Konstantin Todorov

Machine Learning Seminar
February 2019

Machine Learning?

Could computers be made to *learn* and to *improve* automatically with experience?



Can we develop algorithms that can *learn from* and *make predictions* on data?

(Almost) like we humans do...

Defining the Machine Learning Problem

The Defining Question of ML

How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?

Computer Science

How can we build machines that solve problems, and which problems are inherently tractable?

Statistics

What can be inferred from data and a set of modelling assumptions, with what reliability?

A Multidisciplinary Field

Probabilities

Artificial Intelligence

Statistics

Machine Learning

Philosophy

Information Theory

Psychology & Neuroscience

Control Theory

Optimization & Computational Complexity

A Definition of Machine Learning

"A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in T, as measured by P, improves with experience E."

– *Tom M. Mitchell*

- An operational, not a cognitive or an etymological definition
- A. Turing: *Can machines **think**?* →
  *Can machines **do** what thinking beings can do?*

Depending on how we define T, P, and E, the learning task might also be called by names such as *data mining*, *classification*, *clustering*, *reinforcement learning*, *etc*...

A Definition of Machine Learning

An example: filtering spam from emails

T task: decide whether an email is spam or not

P performance measure: the percent of correctly filtered emails

E training experience: a dataset of emails with associated classes (spam / email)

A long list of applications...

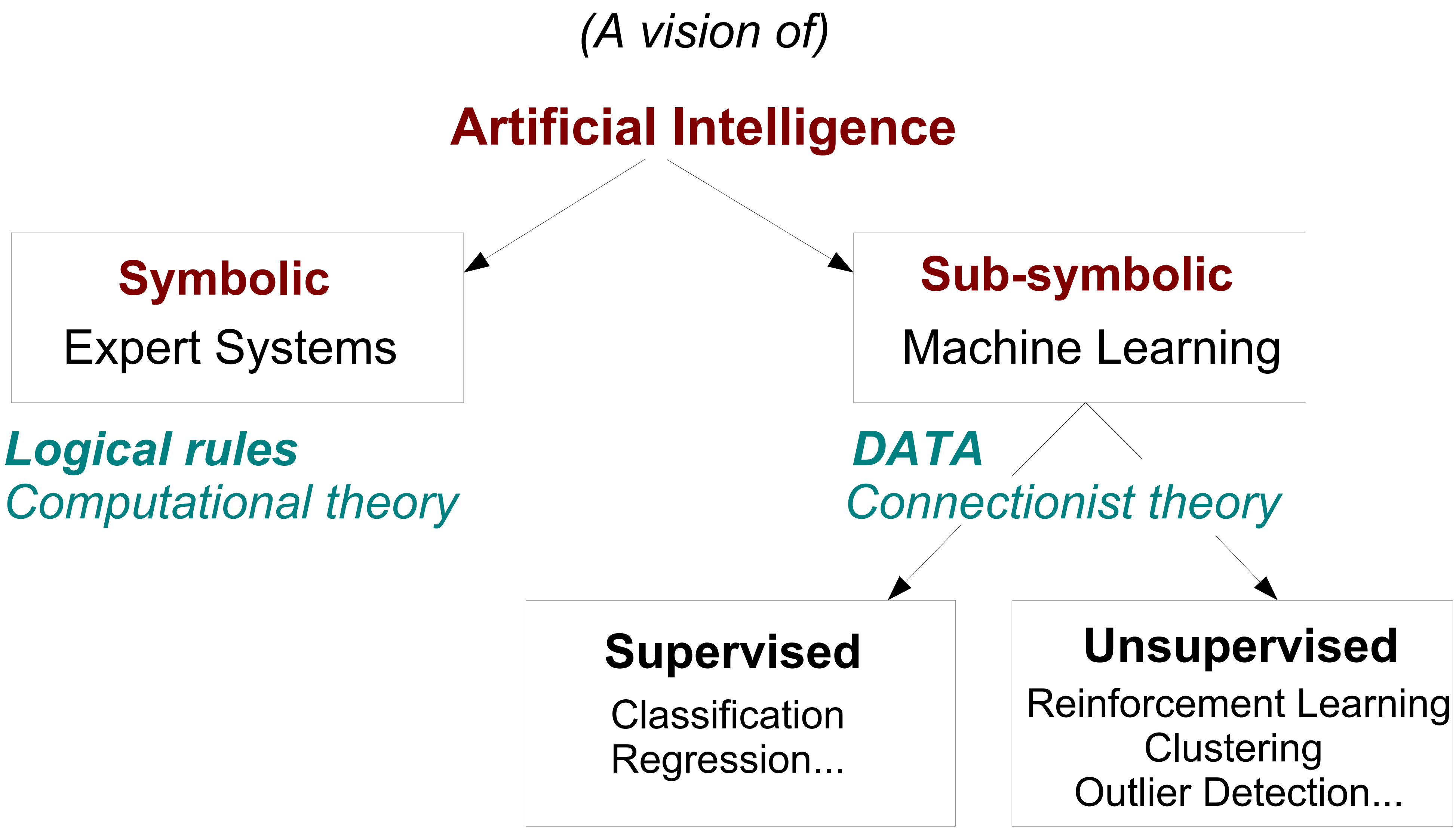
web page ranking, recommendation, automatic translation, autonomous cars, diagnostics, face recognition...

A Brief History Line

18^{th}– 20^{th}	Advances in probability theory (Bayes, Markov Chains,...)
1950	Turing: a learning machine that can become artificially intelligent
1951	Minsky: first neural network (NN) machine
1957	Rosenblatt: the perceptron
1967	Pattern recognition with nearest neighbours
1960s	The rise and fall of the perceptron
1970s	AI winter, due to unrealised promises of AI research
1982	Recurrent neural networks
1986	LeCun: back-propagation reinvented
1989	Reinforcement learning
1990s	Vapnik and Cortes: Support Vector Machines shadow NN
2000s	NN regaining popularity due to advanced computational powers
2010s	Rapid acceleration of Deep Learning research

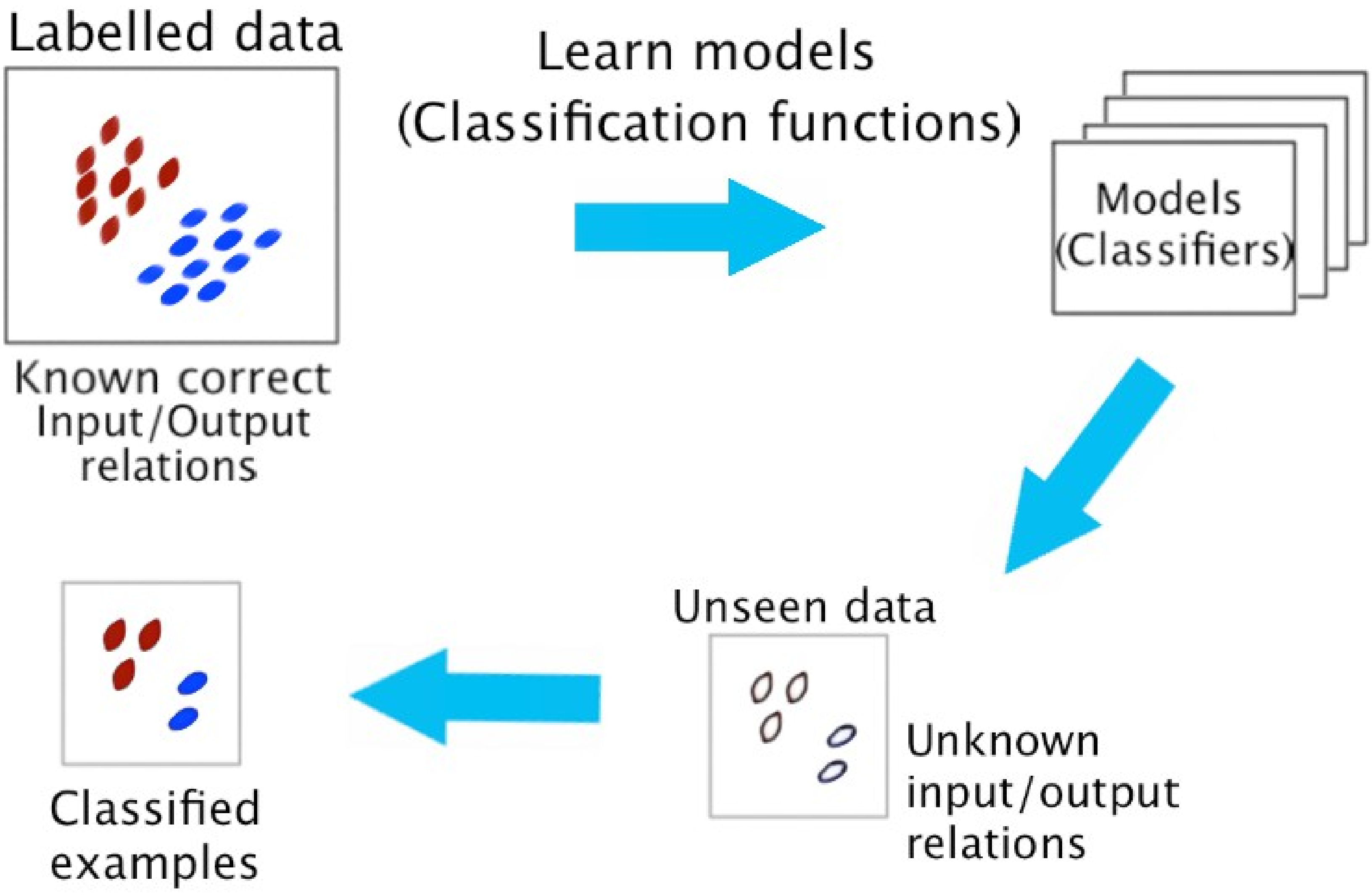
Kinds of Machine Learning

The Machine Learning Tasks



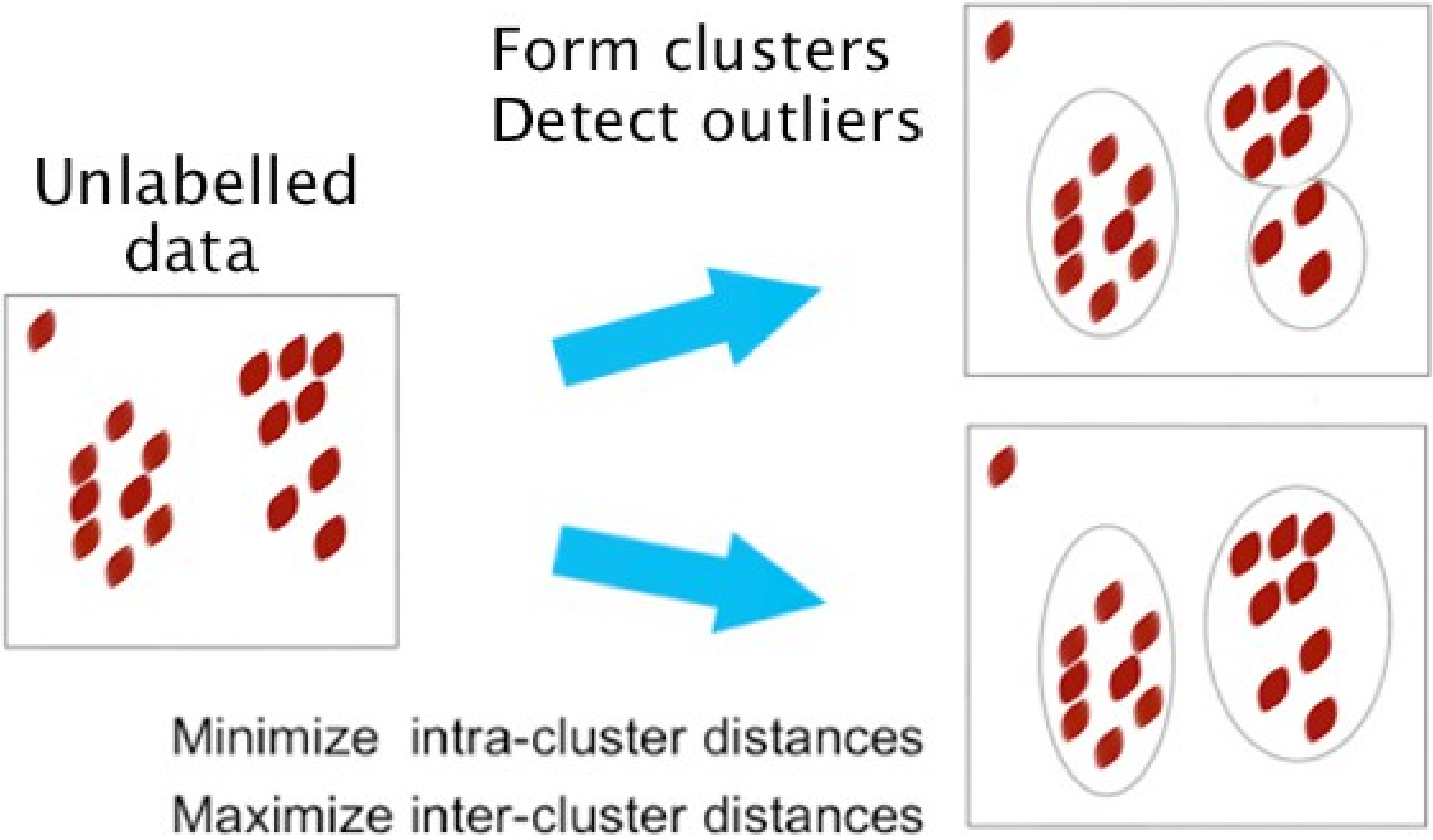
Supervised Machine Learning

Infer input/output functions from ***labelled*** data.



Unsupervised Machine Learning

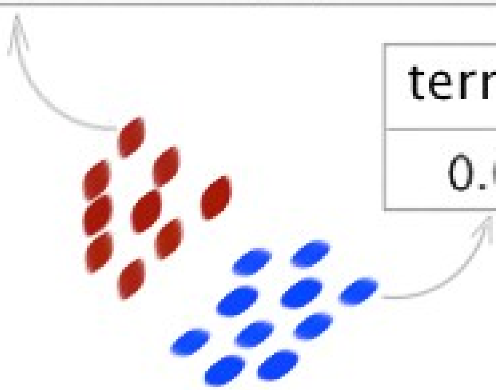
Infer a latent structure from ***unlabelled*** data.



Data Representation: Features

Remember our spam filtering example: data-points are emails.

term1	term2	...	termN	class
0.3	0.89	...	0.0	SPAM



term1	term2	...	termN	class
0.0	0.2	...	0.96	MAIL

Model instances as ***vectors*** described by a number of ***features*** **(*variables, attributes*).**

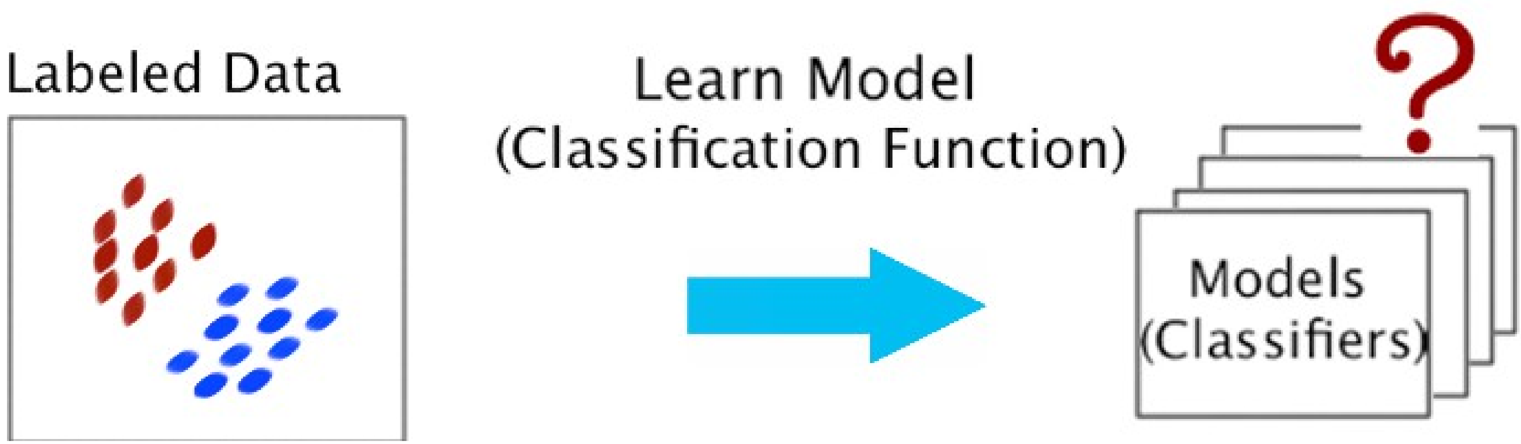
What features best describe instances and allow to separate classes or form clusters? → ***Feature (variable) selection***

- remove noisy features
- analyse their explanatory strength
- reduce dimensionality

Model Selection and Assessment

Overfitting *vs.* **Generalisation**

- how well the learned function performs on *unseen* data?
- select a model (a set of parameters) that *generalises* well
- *evaluate* and avoid overfitting

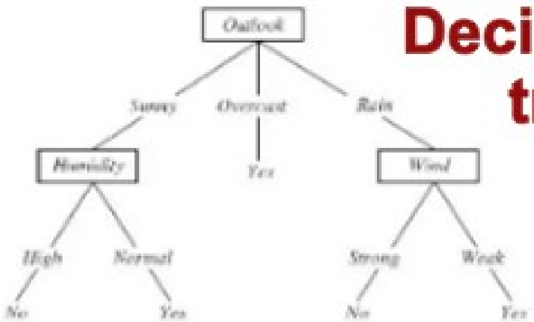


Model Selection, Model Validation

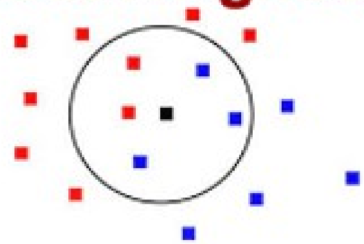
Methods, Tools and Applications (Examples)

Methods and Applications

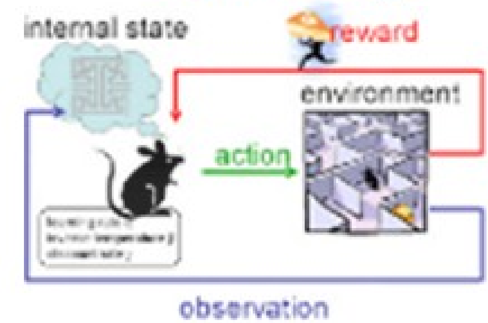
Decision trees



K Nearest Neighbours



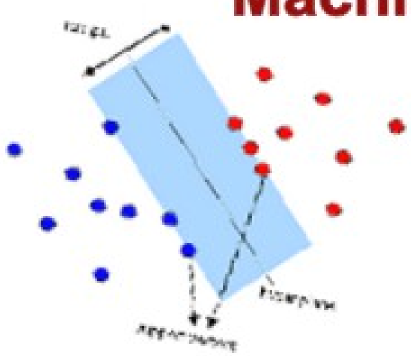
Reinforcement Learning



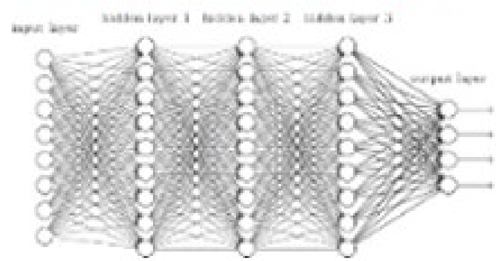
Bayesian Classifiers

$$P(C \mid A) = \frac{P(A \mid C)P(C)}{P(A)}$$

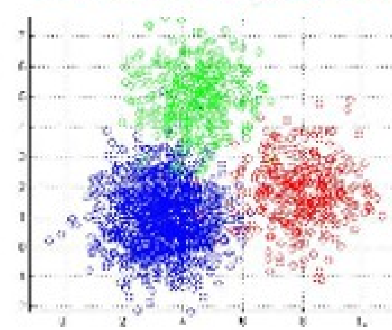
Support Vector Machines



Neural Networks / Deep Learning

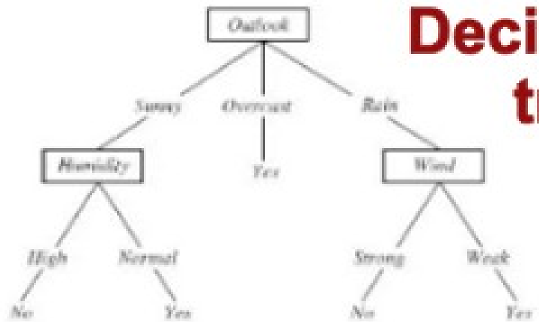


Cluster Analysis

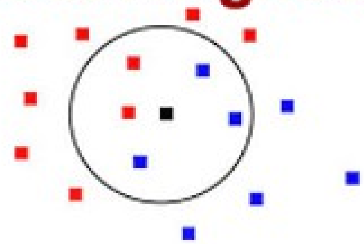


Methods and Applications

Decision trees



K Nearest Neighbours



Financial distress prediction

Process control

Fraud detection

Medical diagnosis

Bayesian Classifiers

$$P(C \mid A) = \frac{P(A \mid C)P(C)}{P(A)}$$

Text Categorization

Automatic translation

Web search

Computer vision

Image retrieval

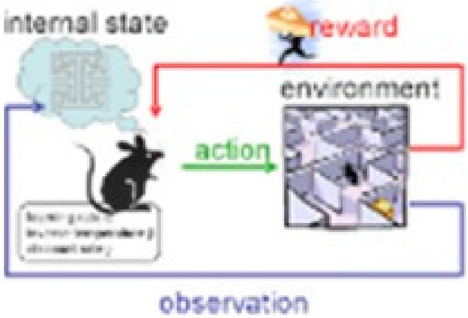
Robotics

Driving autonomous cars

Playing games

Trading stradegies

Reinforcement Learning

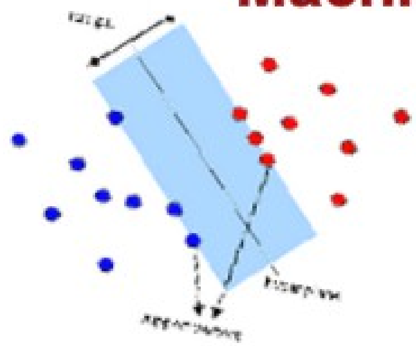


Gene clustering

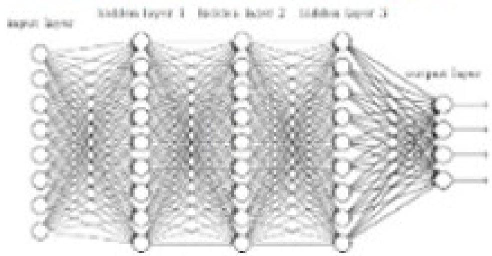
Topic discovery

Market segmentation

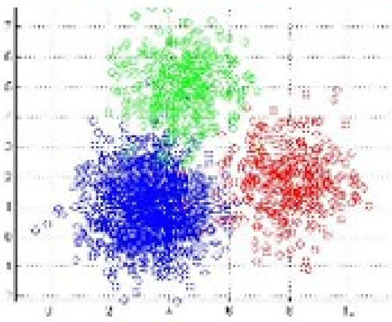
Support Vector Machines



Neural Networks / Deep Learning



Cluster Analysis



Tools and Their Usefulness

A long list of open source tools...

Weka, ELKI, R, Mahout, RapidMiner*,...

Often balckboxes for users.

→ How to implement a given ML solution (which API)?
Algorithms don't change from one API to another...

→ What and how much data is needed? How to select a model? Which method for what problem?
An empirical science... with some heuristics.

→ How deep an understanding of the algorithms is required?
Investing in statistical inference:
– hiring a statistician / data scientist, training engineers

Sources & Reading

Mitchell, T. M. (1997). Machine learning. WCB.

Mitchell, T. M. (2006). *The discipline of machine learning* (Vol. 9). Carnegie Mellon University, School of Computer Science, Machine Learning Department.

Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

Langley, P. (2011). The changing science of machine learning. *Machine Learning*, *82*(3), 275-279.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer Berlin Heidelberg.

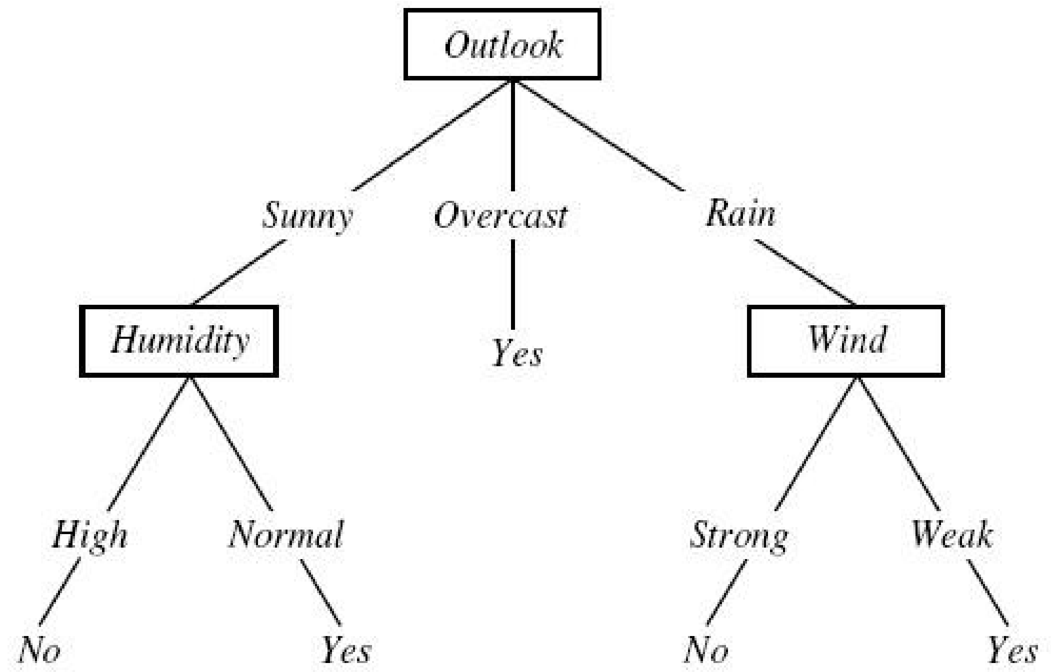
Thank you for listening.

ML Methods and Applications

Decision trees

Supervised / Classification

Fits data into a tree
Attributes → nodes
Values → branches
Easy to interpret
Overfitting occurs often



From T. Mitchell's "Machine Learning"

Applications

Biomedical engineering: selecting features for implantable devices
Manufacturing, production: process control
Molecular biology: analyzing amino acid sequences
Fraud detection

ML Methods and Applications

Bayesian Classifiers

Supervised / Classification

$$P(C \mid A) = \frac{P(A \mid C)P(C)}{P(A)}$$

Creates a model per class, using probability theory.
Attributes are assumed independent.
Probabilities are estimated from data.

Applications

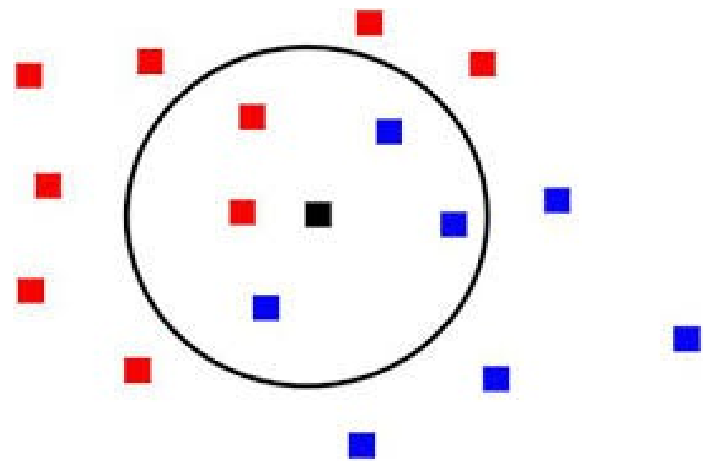
Text categorisation
Speech recognition
Automatic medical diagnosis

ML Methods and Applications

K Nearest Neighbours

Supervised / Classification

Lazy instance-based learners.
Uses distance calculation over all instance pairs.



Applications

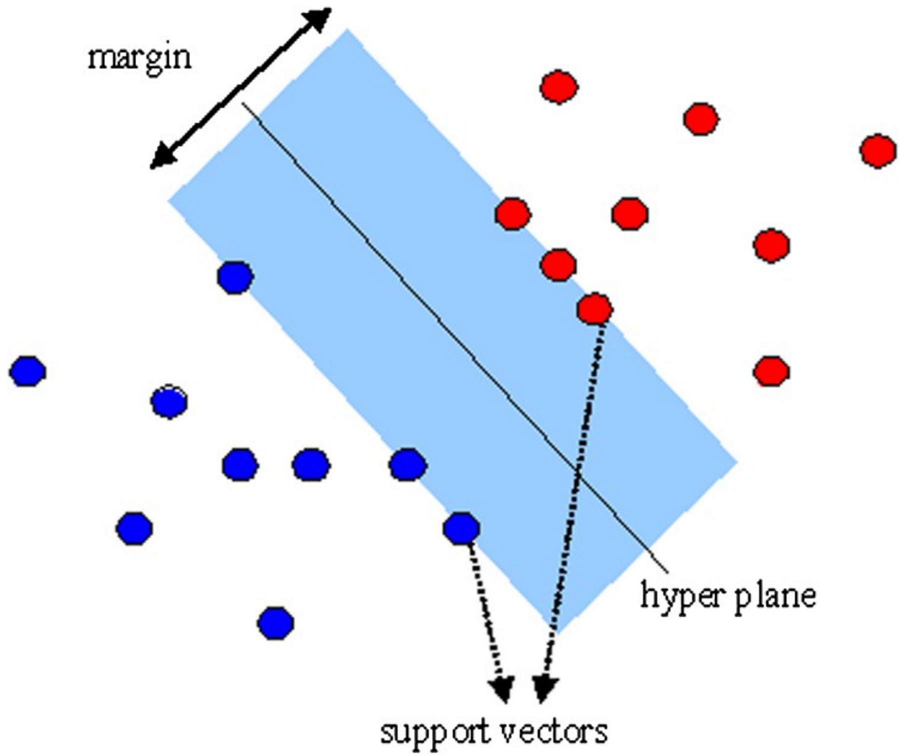
Cancer diagnosis
Financial distress prediction
Computer vision

ML Methods and Applications

Support Vector Machines

Supervised / Classification

Learns a maximum margin separation hyperplan.
Deals with non-lineary separable data
Uses kernels



Applications

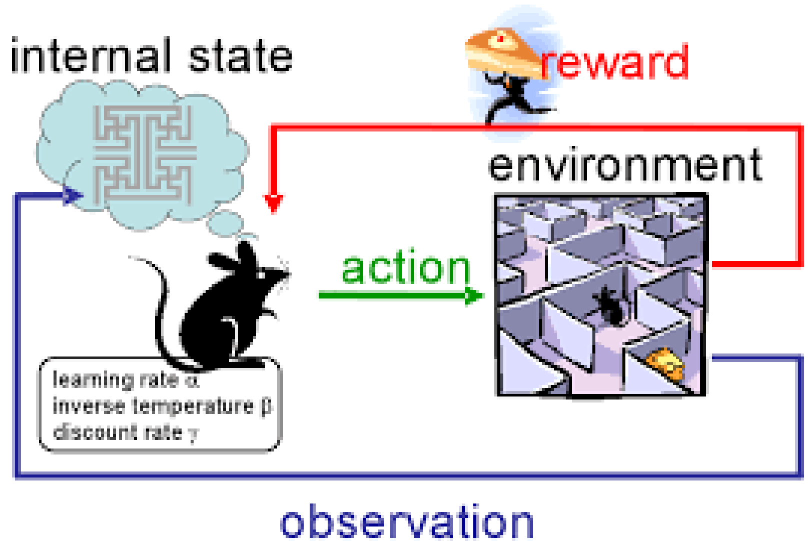
Text categorisation
Automatic translation
Computer vision
Handwriting / face / facial expression recognition
Content-based image retrieval

ML Methods and Applications

Reinforcement Learning

Unsupervised or Semi-supervised

Take actions according to rewards.
Behaviour optimisation with respect to the environment.



Applications

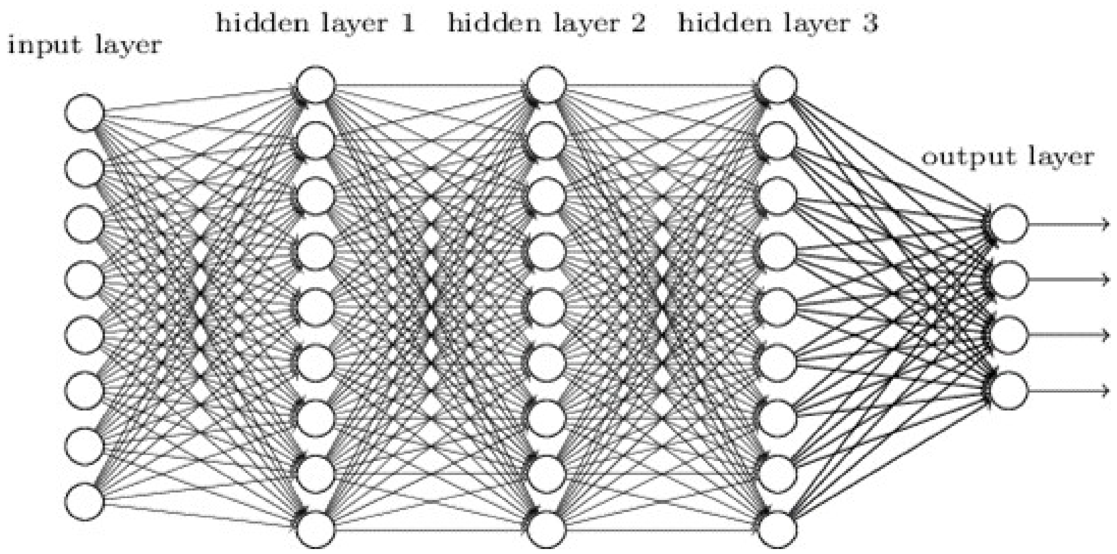
Driving autonomous vehicles
Robot vision
Playing games

ML Methods and Applications

Neural Networks / Deep Learning

Supervised and Unsupervised

Bio-inspired: a complex net of interconnected neurones



Applications

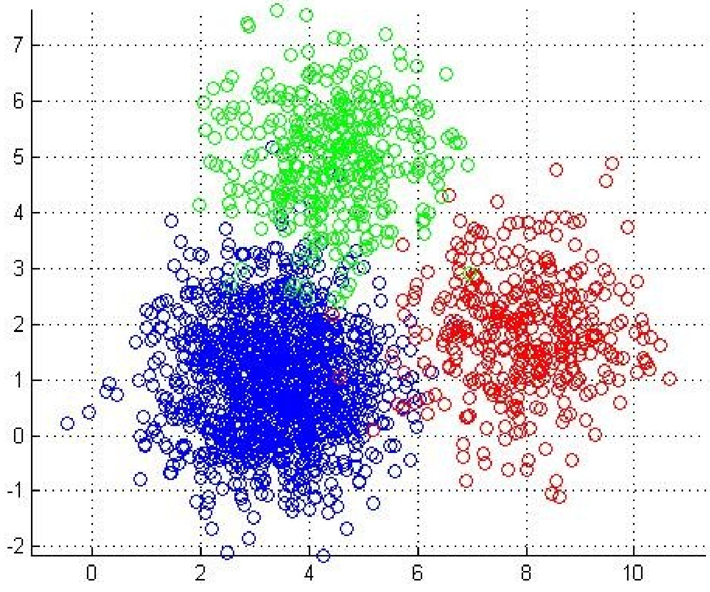
Driving autonomous vehicles
Computer vision
Speech / face / handwriting recognition
Sensor data interpretation
Image retrieval

ML Methods and Applications

Cluster Analysis

Unsupervised / Clustering

Group together instances into subsets
Maximise intra-cluster instance
similarities and inter-cluster distances.
K-means, DBSCAN,
Descriptive Statistics, ...



Applications

Market segmentation
Gene clustering
News summarisation
Topic discovery