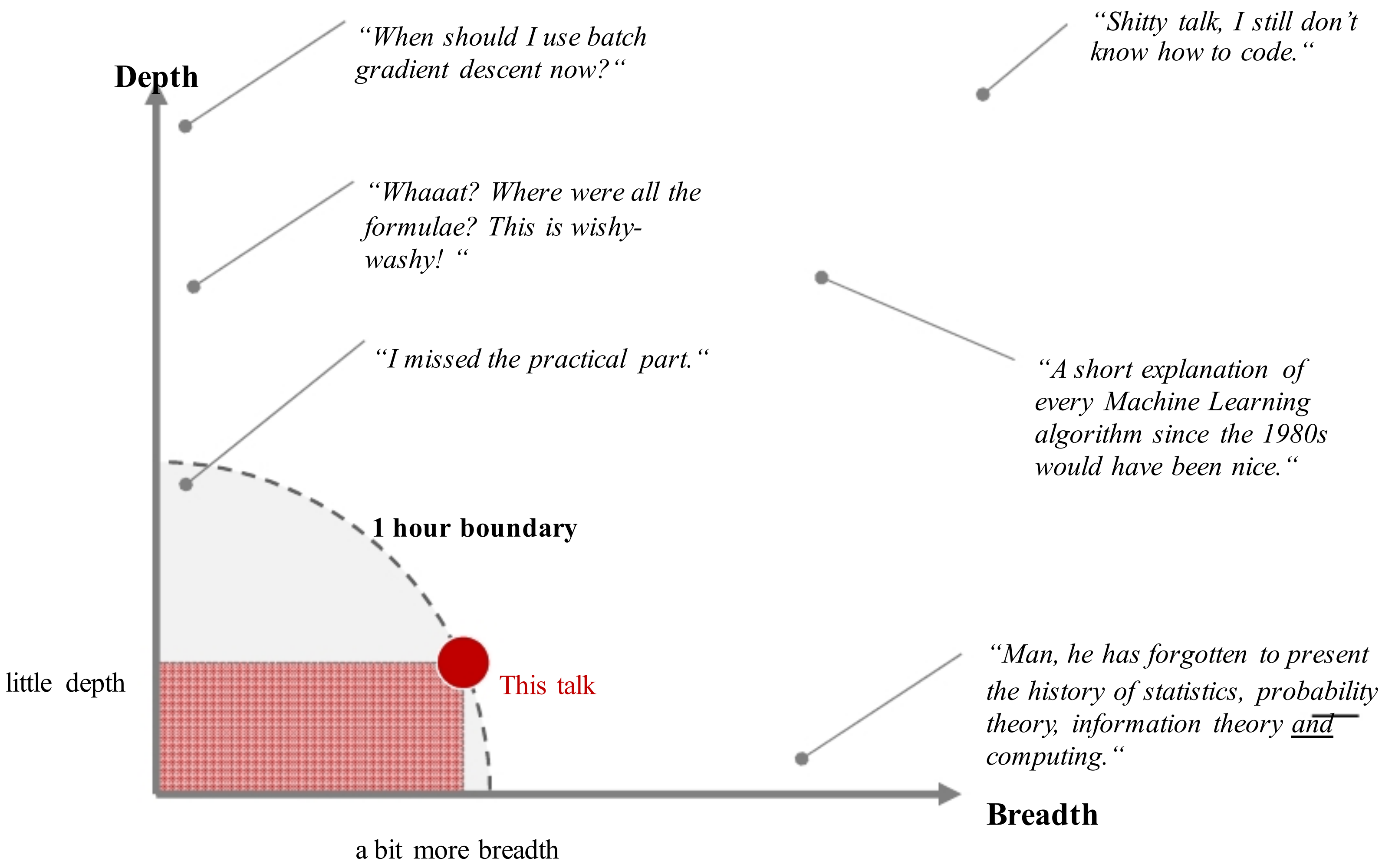
Introduction to Machine Learning

These slides are adapted from

Pascal Wichmann,
Machine Learning – a gentle and structured introduction, 2016/01/27.

Contents

A short talk on Machine Learning can only fall short of some people‘s expectations – this will be a gentle introduction only



Contents

Recent examples of Machine Learning

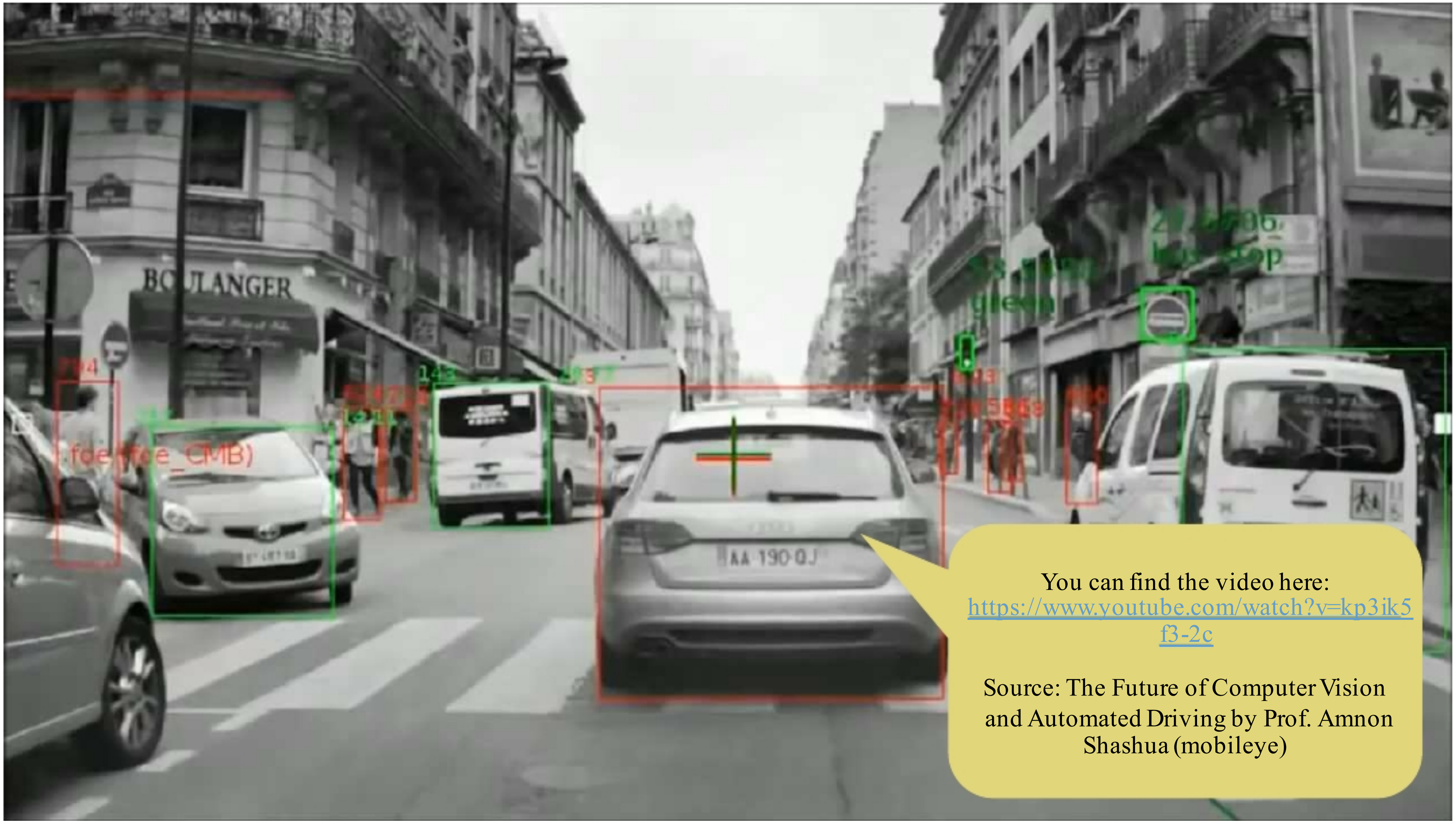
An algorithm that has learnt to play arcade games – better than any human…

You can find the video here: https://www.youtube.com/watch?v=08C17ii6viY

Please also note that DeepMind recently also published a video on the game "Go"; you can find it here: https://www.youtube.com/watch?v=g-dKXOlsf98

Recent examples of Machine Learning

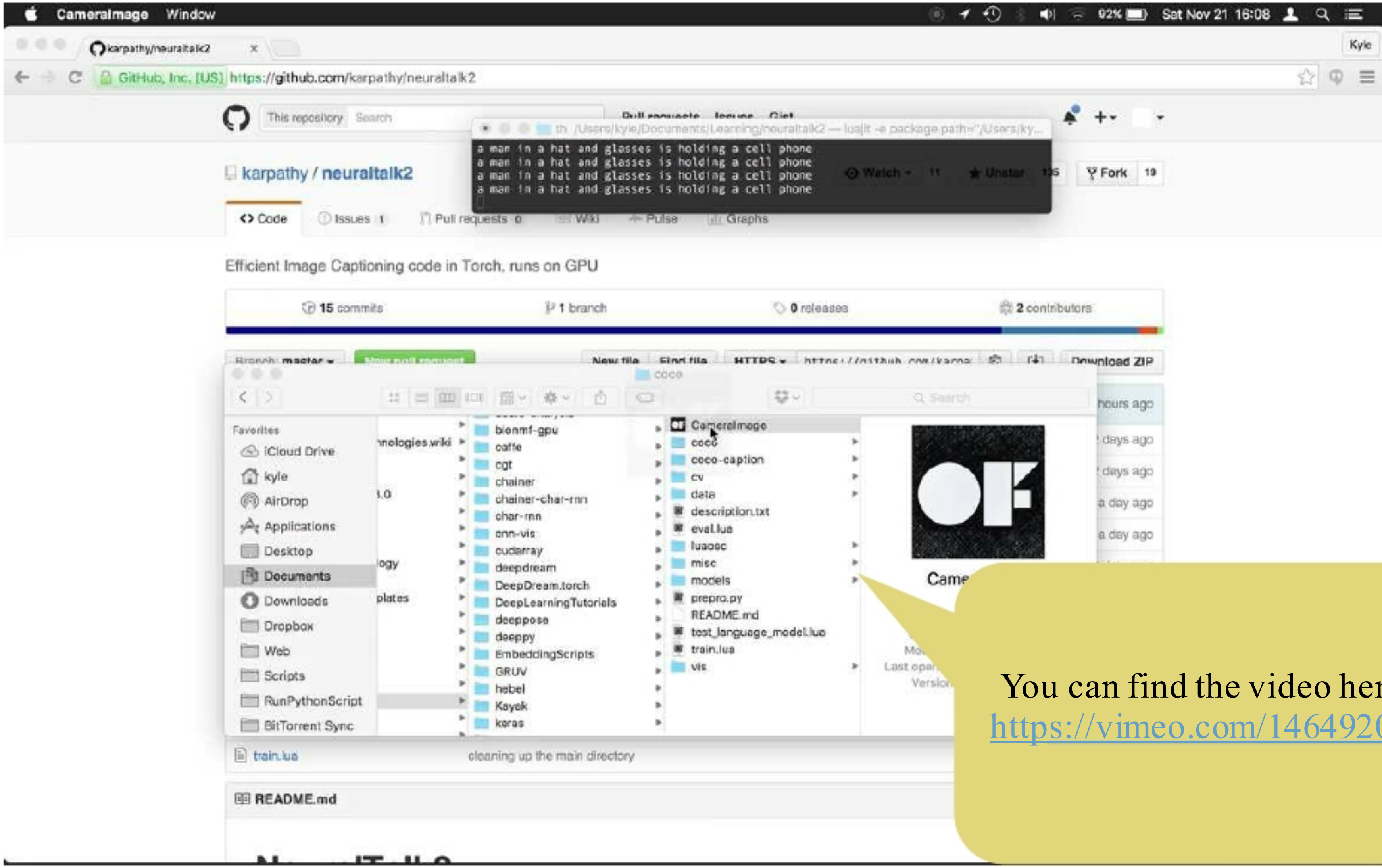
What an autonomous car sees…



You can find the video here:
https://www.youtube.com/watch?v=kp3ik5f3-2c

Source: The Future of Computer Vision and Automated Driving by Prof. Amnon Shashua (mobileye)

Descriptions generated in realtime by a neural network during a brief walk around Amsterdam…



You can find the video here:
https://vimeo.com/146492001

Contents

ML allows computer programs to improve their performance with experience – without being explicitly programmed

A useful definition of Machine Learning

"Learning is any process by which a system improves performance from experience."
Herbert A. Simon (Nobel laureate and computer scientist)

"[Machine Learning is the] Field of study that gives computers the ability to learn without being explicitly programmed."
Arthur Samuel (computer gaming and A.I. pioneer), 1959

"A computer program is said to learn to perform a task T from experience E, if its performance at task T, as measured by a performance metric P, improves with experience E over time."
"Machine Learning", Tom Mitchell, 1997

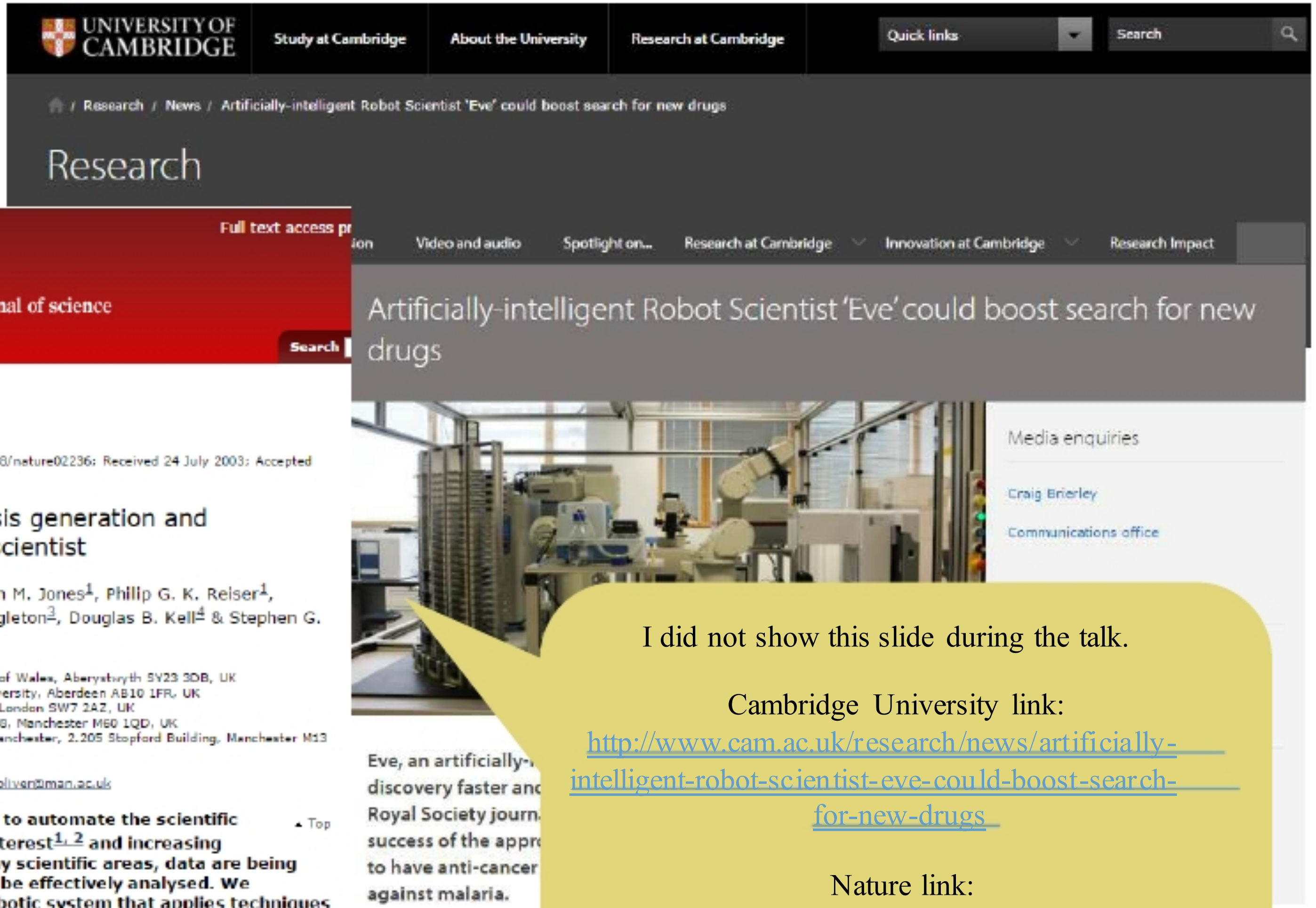
The promises of Machine Learning range from "automating discovery" in science…

*"Machine learning is the **scientific method on steroids**. It follows the same process of generating, testing, and discarding or refining hypotheses.*

*But while a scientist may spend his or her whole life coming up with and testing a few hundred hypotheses, a machine-learning system can do the same in a fraction of a second. **Machine learning automates discovery.***
*It's no surprise, then, that **it's revolutionizing science as much as it's revolutionizing business**."*

"The Master Algorithm", Pedro Domingos (University of Washington)

Researchers use “Robot scientists” in an attempt to automate the scientific process



Letters to Nature

Nature 427, 247-252 (15 January 2004) | doi:10.1038/nature02236; Received 24 July 2003; Accepted 14 November 2003

Functional genomic hypothesis generation and experimentation by a robot scientist

Ross D. King[1], Kenneth E. Whelan[1], Ffion M. Jones[1], Philip G. K. Reiser[1], Christopher H. Bryant[2], Stephen H. Muggleton[3], Douglas B. Kell[4] & Stephen G. Oliver[5]

1. Department of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK
2. School of Computing, The Robert Gordon University, Aberdeen AB10 1FR, UK
3. Department of Computing, Imperial College, London SW7 2AZ, UK
4. Department of Chemistry, UMIST, P.O. Box 88, Manchester M60 1QD, UK
5. School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Manchester M13 9PT, UK

Correspondence to: Stephen G. Oliver[5] Email: steve.oliver@man.ac.uk

Top

The question of whether it is possible to automate the scientific process is of both great theoretical interest[1, 2] and increasing practical importance because, in many scientific areas, data are being generated much faster than they can be effectively analysed. We describe a physically implemented robotic system that applies techniques from artificial intelligence[3, 4, 5, 6, 7, 8] to carry out cycles of scientific experimentation. The system automatically originates hypotheses to explain observations, devises experiments to test these hypotheses, physically runs the experiments using a laboratory robot, interprets the results to falsify hypotheses inconsistent with the data, and then repeats the cycle. Here we apply the system to the determination of gene

I did not show this slide during the talk.

Cambridge University link:
http://www.cam.ac.uk/research/news/artificially-intelligent-robot-scientist-eve-could-boost-search-for-new-drugs

Nature link:
http://www.nature.com/nature/journal/v427/n6971/full/nature02236.html

… to “solving intelligence” itself

“Our mission at DeepMind is very easy to articulate – but obviously quite hard to do. And we usually describe it as a two-step process:
Step 1: Solve intelligence; ...and then…
Step 2: Use it to solve everything else.”

Demis Hassabis (Google DeepMind), 2015

Bonus quiz: What are reasons why Machine Learning has gained popularity in recent years?

Reasons I mentioned during the talk (probably not exhaustive):

Data availability

availability of massive amounts of data (unlabelled & labelled via MTurk)

cost-effective storage of huge amounts of data

Realisation that the stored data is actually valuable → so massive amounts of data are actually stored

Computing power

Faster processors

Parallel processing

use of GPUs instead of CPUs

computing clusters / cloud computing (e.g. EC2) – computing as a service

With the rise of efficient GPU computing, it has become possible to train neural networks with many layers (deep learning).

Advances in algorithms / toolkits

e.g. the old idea of neural networks that has been revived multiple times & is now probably one of the most impressive methods

fully-fledged and highly optimised libraries that can be used (Python, R, …)

New libraries published every few days: TensorFlow by Google; etc

Contents

To prevent from getting lost, we use this framework to structure the landscape of Machine Learning

Problem side

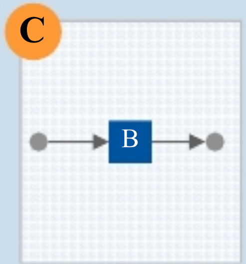
A Types of (Machine) learning

B Types of (Machine Learning) problems

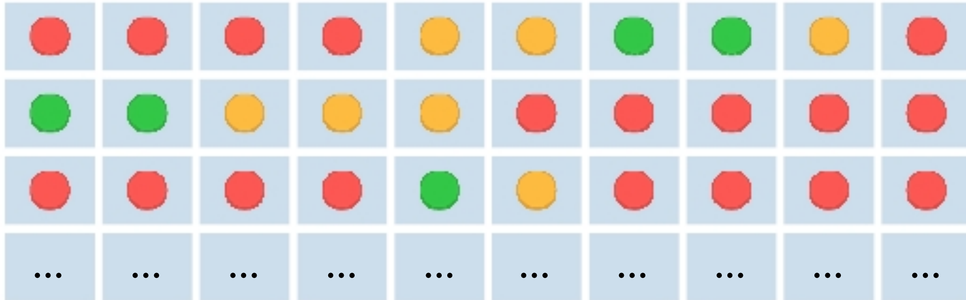
Solution side

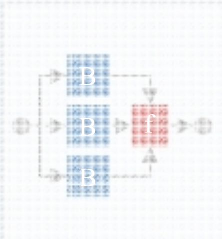
Types of (Machine) learning methods

C



"Base learners"
(= individual learning methods); families of algorithms are inspired by different schools of thought





"Ensemble methods"
(= use multiple different models of the same or diifferent base learner and combine their outputs)

Contents

Question to you

How does one learn?

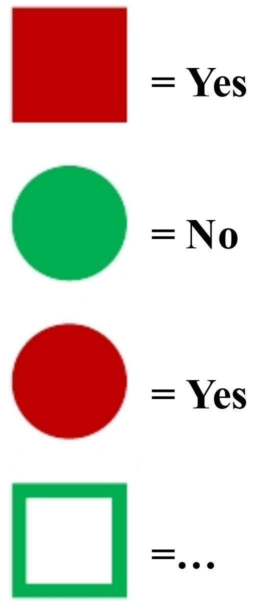
How did you learn as a child?

How do you teach an animal what to do?

How does one learn

1 The first type of learning

3319 = 1
8831 = 3
8810 = 5
2467 = 2
9821 = 4

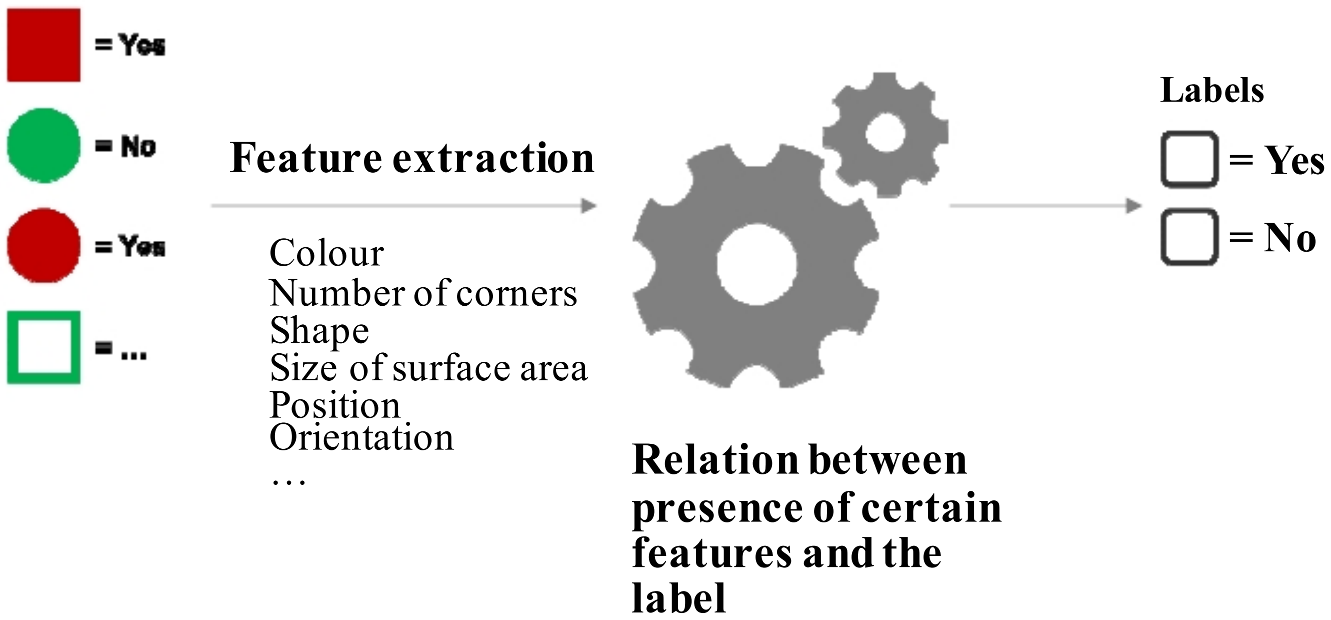


What could be a possible answer for the new example?

What could be a possible answer for the new example?

How does one learn?

1 The first type of learning



1 By generalising from examples + correct answers

You are given examples with the correct answer

From this you infer some form of rules (you generalise)

– Actually, before you infer the rules, you extract *features* (like colour, shape, number of edges, etc.)

Then you apply these rules to a new example in order to predict the answer

If you get a new correct answer, you can correct your rules and get even better

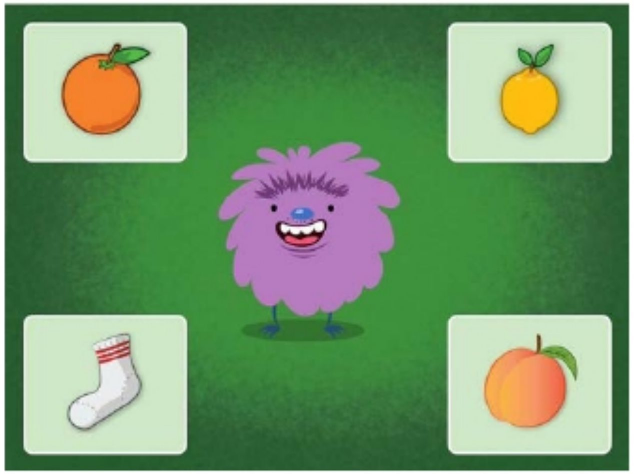
Remarks:

– Do I have enough examples to derive the relation?

– Have I considered the right 'features' to derive the relation?

How does one learn?

2 The second type of learning



Source: education.com

Please create groups of similar keys

One of the things is not like the others
Find the thing that doesn’t belong.

2 By comparing

You look at the things around you, compare them, arrange them according to similarity and then gain some insights (groups of similar items, odd ones, somehow important ones …)

For this, you do not need the „right“ answer; it might even be difficult to define the “right” answer

Remarks:

- Do I have enough examples to understand what similar means?
- Do I consider the right things (the right *features*) when I say two things are similar?
- How do I know how many groups you want?

How does one learn

3 The third type of learning

You can find the video here:
https://www.youtube.com/watch?v=TtfQlkGwE2U

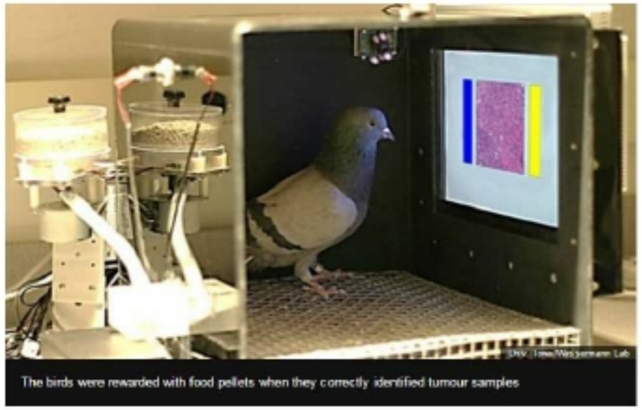
The paper 'Superstition' in the pigeon got published by Skinner in 1948:

Skinner B.F. (1948). 'Superstition' in the pigeon., Journal of Experimental Psychology, 38 (2) 168-172. DOI: 10.1037/h0055873

How does one learn?

3 The third type of learning

You can find the video about rats sniffing out land mines here: https://www.youtube.com/watch?v=nEm5zR1IND0



Pigeons identify breast cancer 'as well as humans'

By Andrea Szöllössi
Science writer

20 November 2015 | Science & Environment

The birds were rewarded with food pellets when they correctly identified tumour samples

You can find the video about pigeons identifying cancer here: https://www.youtube.com/watch?v=flzGjnJLyS0

3 By feedback (reward signal)

I don't tell you what or how to do it.

I don't give you any examples at the beginning.

But I will tell you after you have done something good (delayed feedback)

I might also tell you *how* good you have been (smaller or bigger award)

So I use some reward to reinforce behaviour that should be maintained or increased

Other examples: Learning how to walk, riding a bike, …

Remarks:

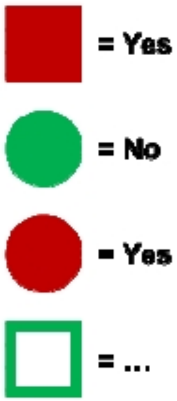
– Ok, this takes ages.

– How complex can the behaviour get if you just get a reward signal?
  Very complex.

The different types of learning are supervised and unsupervised learning – often reinforcement learning is treated as a separate type

Simplified

1 By generalising from correct answers

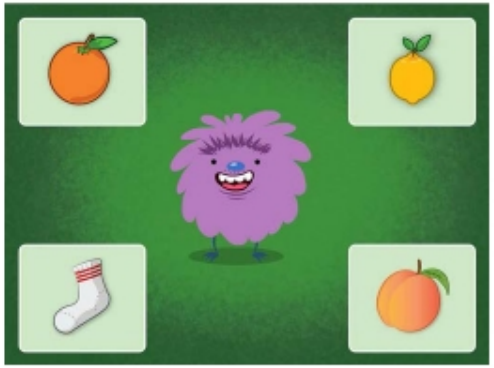


I will give you examples with the correct answer

From this you infer rules and then you apply these rules to something new

Supervised Learning

2 By comparing



I will provide you with examples

But I do not give you the correct answers

You use some metrics of similarity and compare the examples

Unsupervised Learning

3 By feedback (reward signal)

I don't tell you how to do it

I don't give you any examples or correct answers at the beginning.

But I will tell you when you have done something good (you maximise reward)

Reinforcement Learning

Quiz:
Which is the Machine Learning type for the following problems ?

Face recognition
("Who is the person on this photo?")

Customer segmentation
("What types of customers do we have?")

House price estimation
("How much is this house worth?")

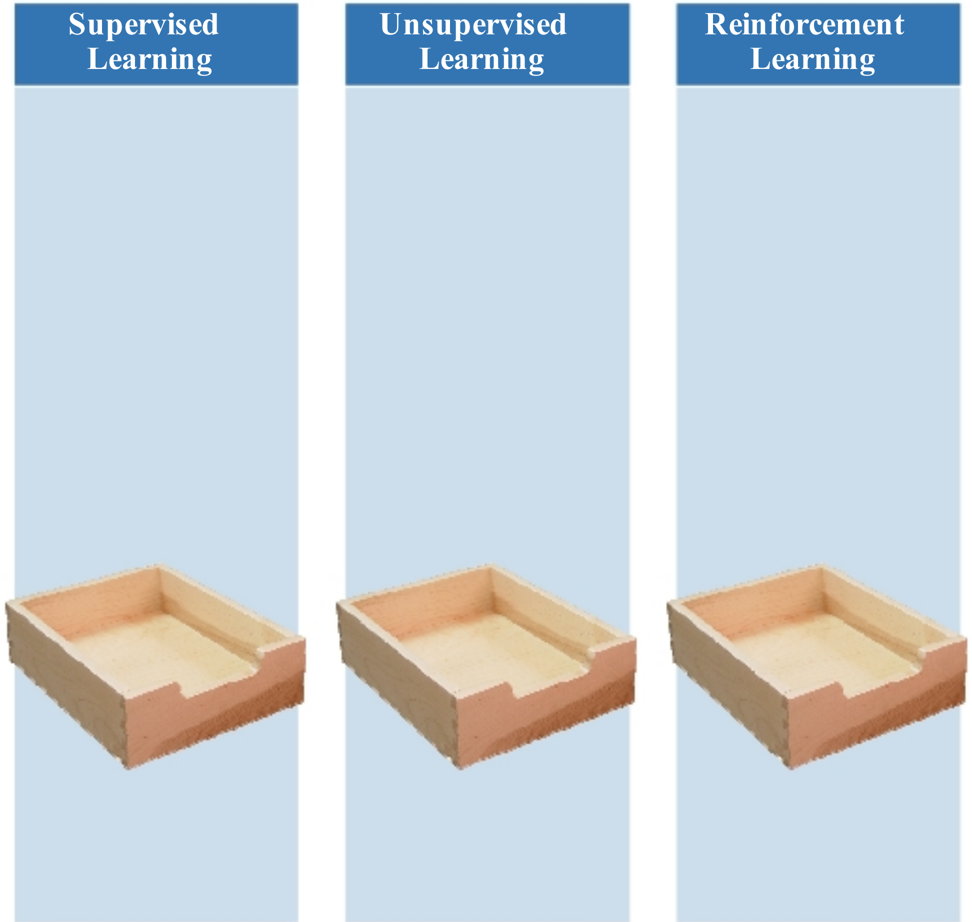
Fraud detection
("Is there anything fishy going on with this client's credit card transactions?")

Spam filter
("Is this email spam?")

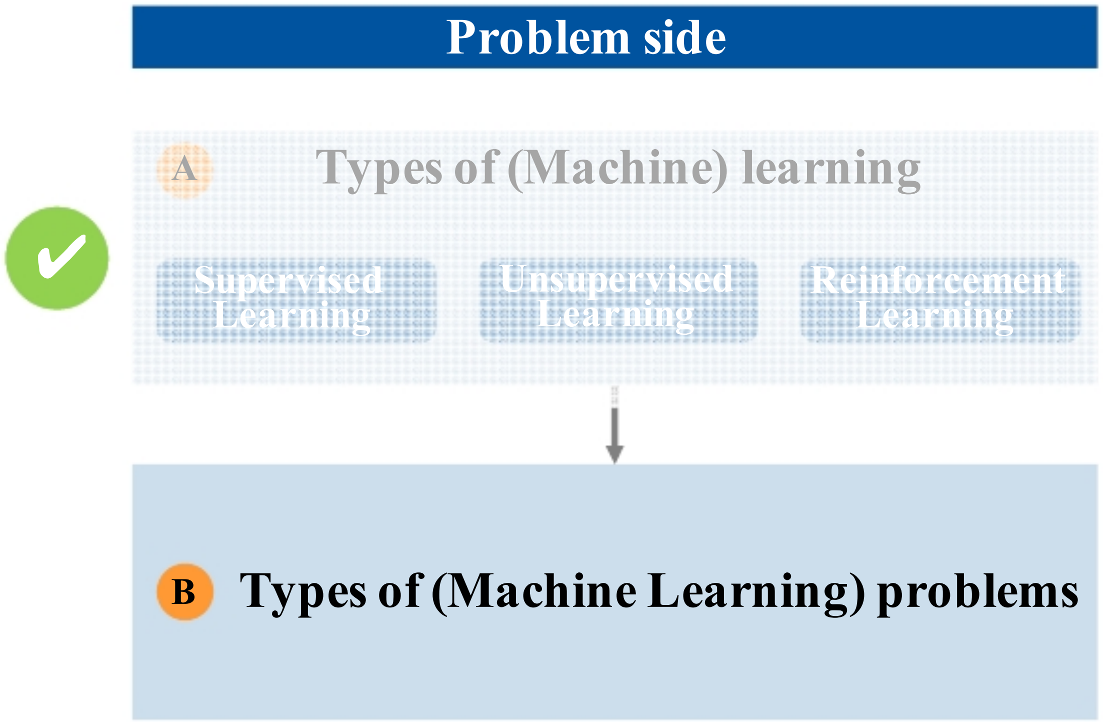
Recommendation system
("(How) will a customer like this movie?")

Identifying handwritten characters
("Which character is that supposed to be?")

Training a robot how to juggle or fly stunt manoeuvres in a helicopter



In our framework, we have now covered the 3 general types of (Machine) learning and can now move on to the most common types of problems



Contents

We go one level deeper and distinguish different Machine Learning problems

Overview of ML problems

Not exhaustive

Machine Learning

General types of learning

- **Supervised learning**
  “Here is the data set where the right answers (labels) are given for each example. Please produce more right answers.”
- **Unsupervised learning**
  “Here is the unlabelled data. Please find peculiarities, similarities or structures (e.g. clusters) in the data yourself.”
- **Reinforcement learning**
  “Learn to do something yourself purely by maximising your expected reward.”

General types of problems

- Supervised learning:
  - **Regression**
    “Predict a <u>continuous</u> valued output.”
  - **Classification**
    “Predict a <u>discrete</u> valued output (e.g. a label or a class)”
- Unsupervised learning:
  - **Clustering**
    „Group similar examples into subsets – called clusters“
  - **Dimensionality reduction**
    „Maybe you don‘t need all the data. What is the essence of the data?“
  - ...
- Reinforcement learning:
  - ...

This can be a goal on its own but is often used as a pre-processing step for other ML tasks

Quiz: What is the Machine Learning problem for the following problems?

Face recognition
("Who is the person on this photo?")

Image segmentation based on colour
("Tell me which areas have a similar colour")

Prediction of future stock prices
("What is this stock worth in the future?")

Image compression of medical images
("Please reduce image size without losing important information")

ICD-10 coding
("Given this this medical diagnosis, which are the right ICD-10 codes?")

Doughnut demand prediction
("How many donuts will I sell on a Monday, 02. January?")

Regression problem?
Classification problem?

Clustering problem?
Dimensionality reduction problem?

Let us have a closer look a regression…

Overview of ML problems

Not exhaustive

Machine Learning

General types of learning

Supervised learning
"Here is the data set where the right answers (labels) are given for each example. Please produce more right answers."

Unsupervised learning
"Here is the unlabelled data. Please find peculiarities, similarities or structures (e.g. clusters) in the data yourself."

Reinforcement learning
"Learn to do something yourself purely by maximising your expected reward."

General types of problems

Regression
"Predict a continuous valued output."

Classification
"Predict a discrete valued output (e.g. a label or a class)"

Clustering
„Group similar examples into subsets – called clusters"

Dimensionality reduction
„Maybe you don't need all the data. What is the essence of the data?"

This can be a goal on its own but is often used as a pre-processing step for other ML tasks

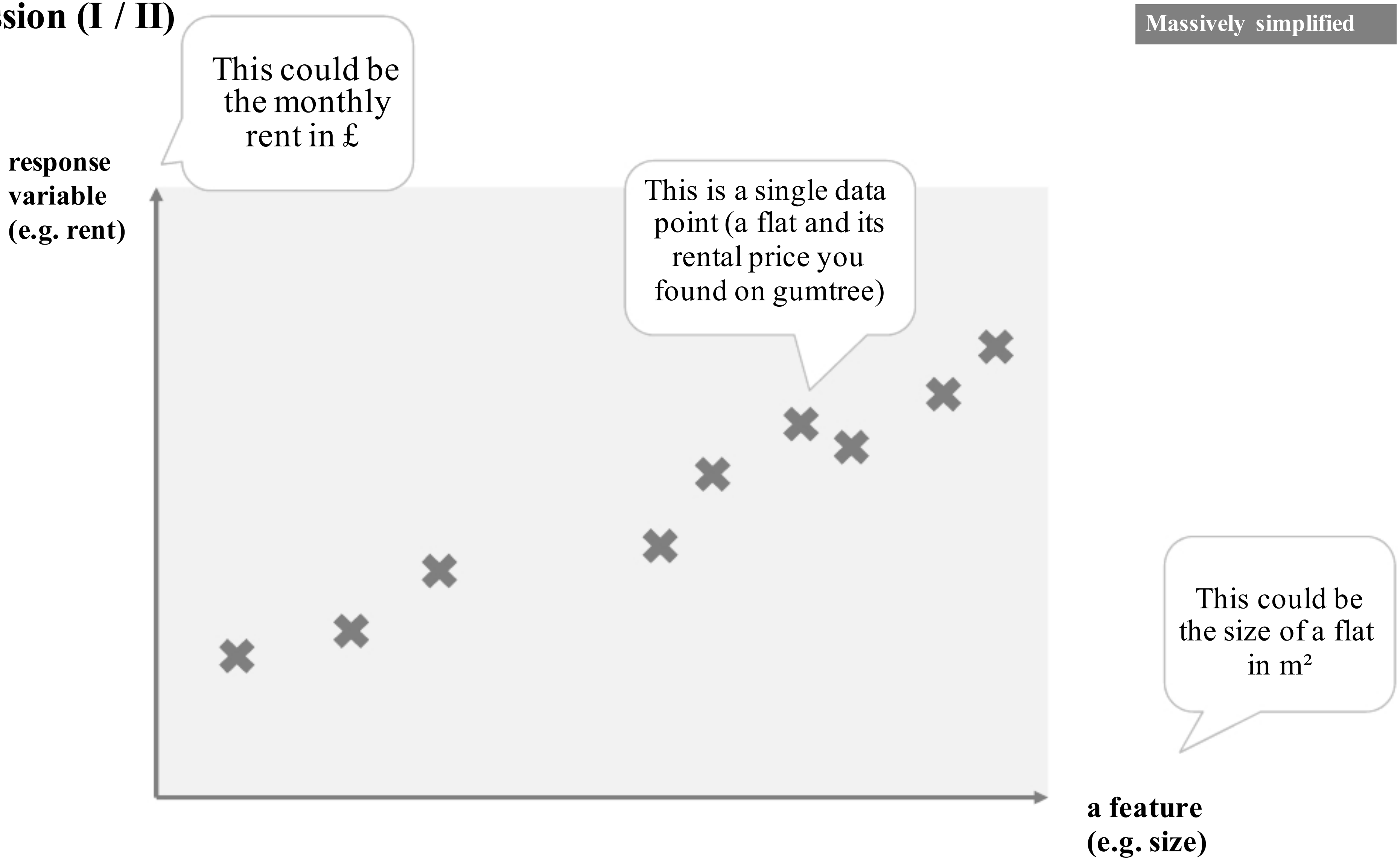
…

…

We illustrate regression problems by plotting the response variable as a function of some feature(s)

Regression (I / II)

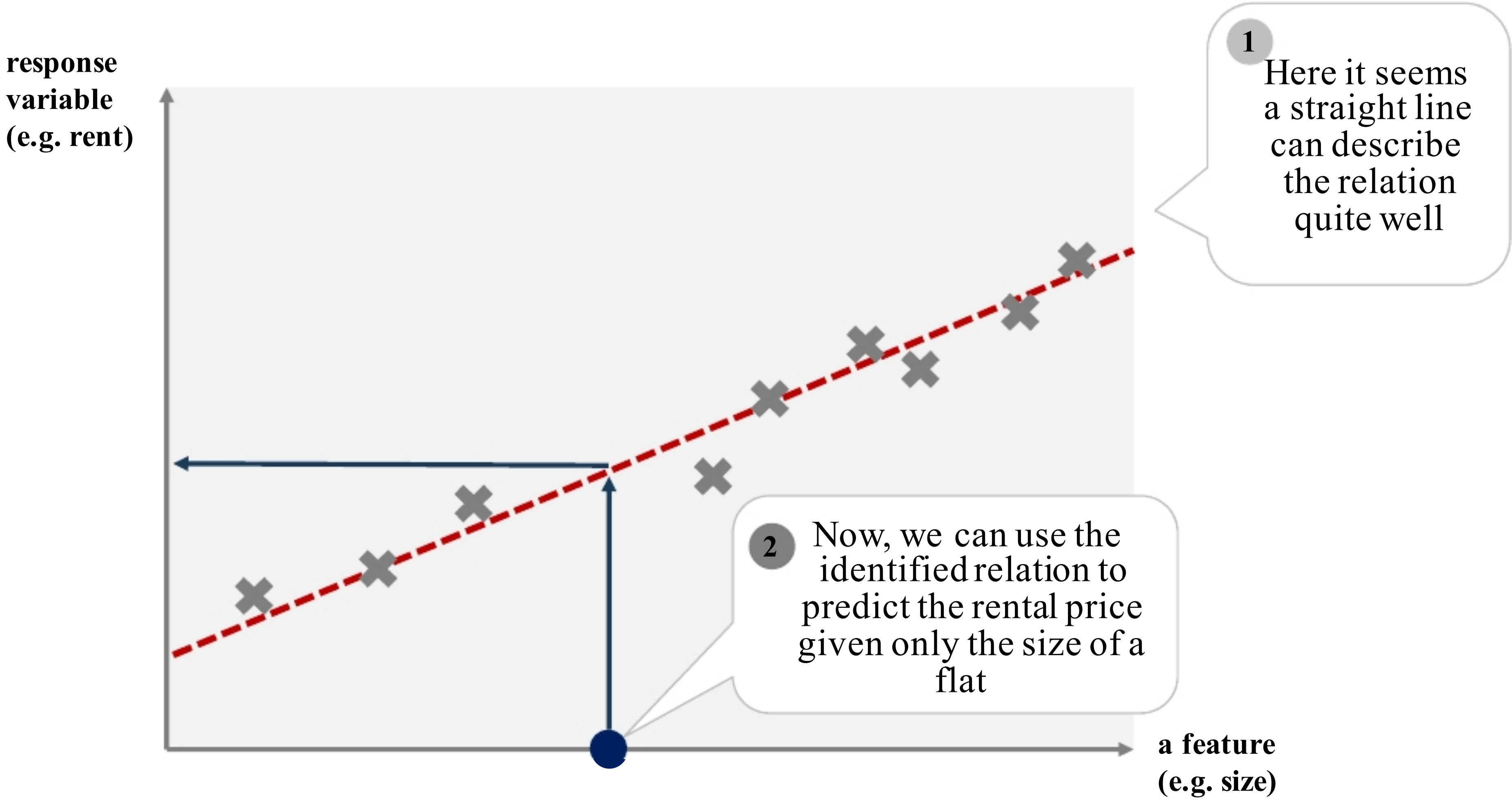
Massively simplified



We “fit the data by a function”
So we can use the function to predict (unseen) values

Regression (II / II)

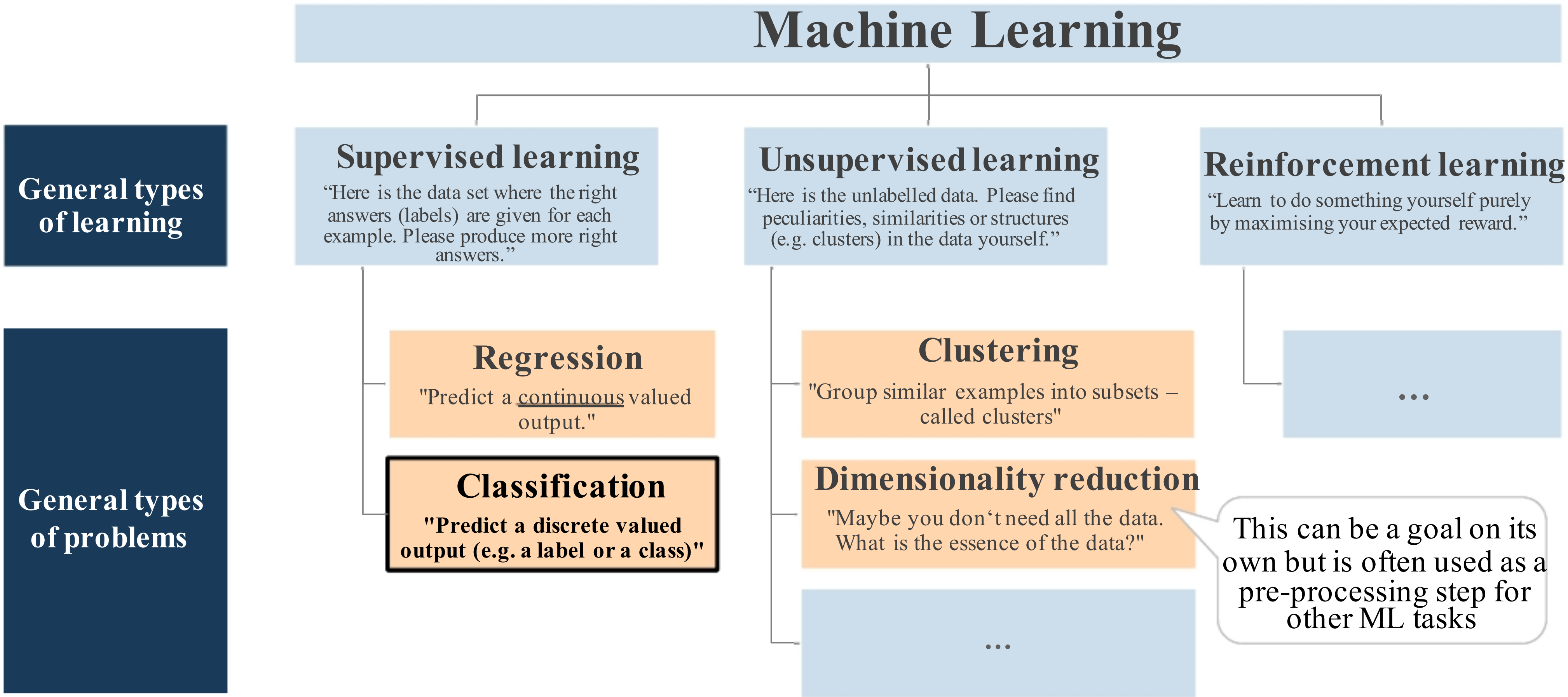
Massively simplified



Let us do the same with classification…

Overview of ML problems

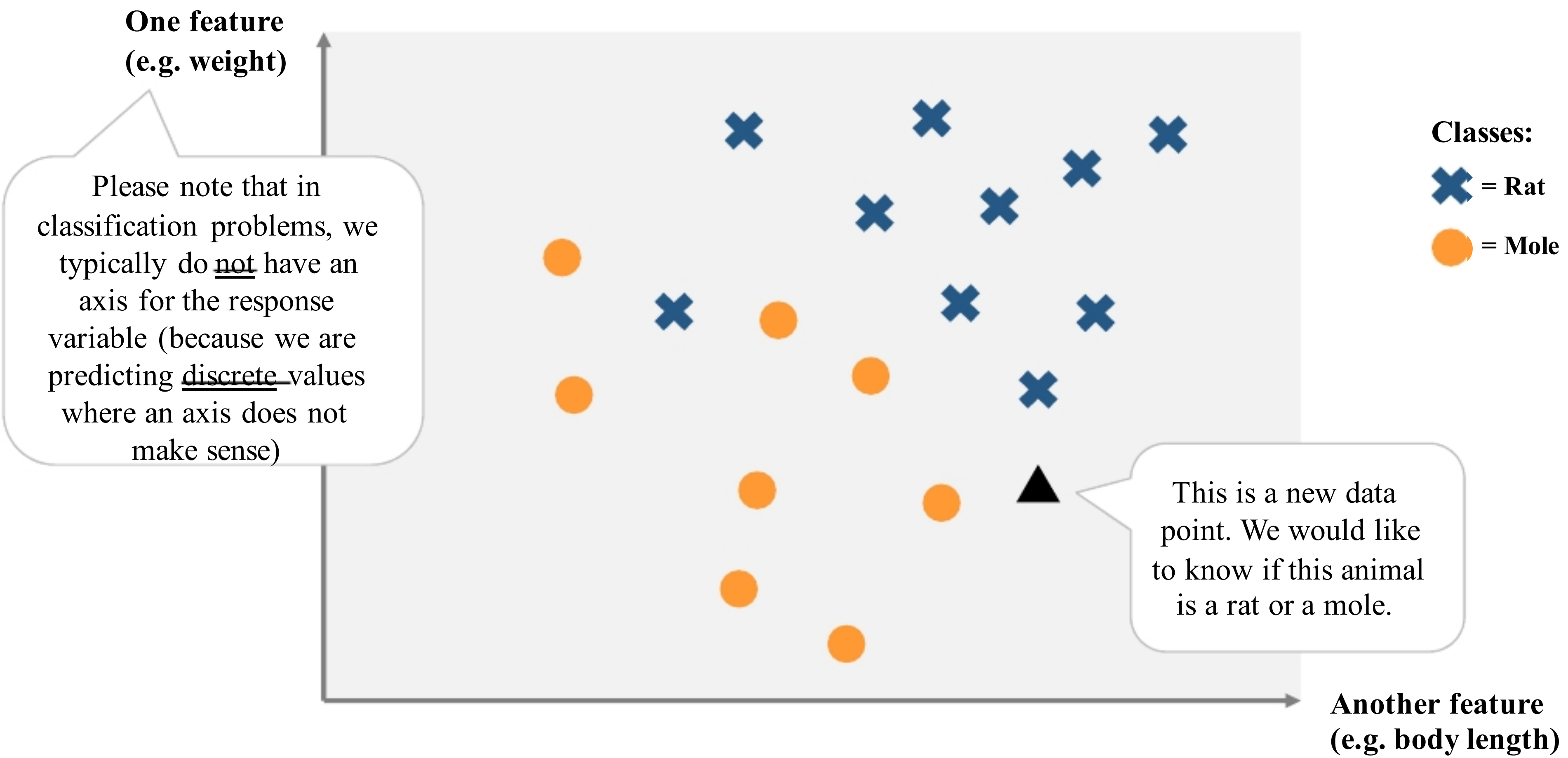
Not exhaustive



In a classification problem, you are given labelled data and need to predict the correct class for a new (unlabelled) example

Classification

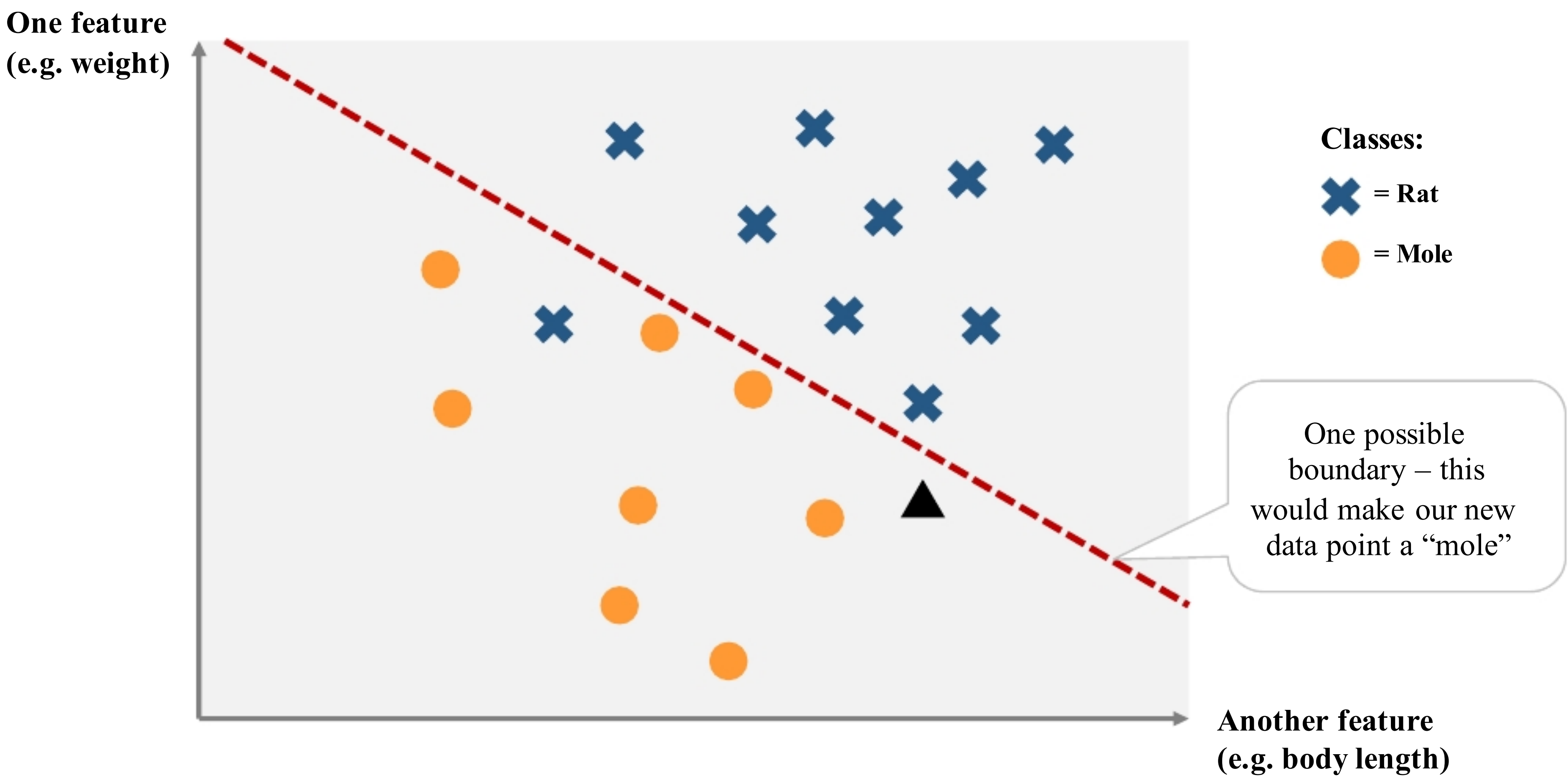
Massively simplified



Again, we need to "fit some function to the data" – but this time the function shall represent the boundary between the classes

Classification

Massively simplified



Finally, let us also look as clustering…

Overview of ML problems

Not exhaustive

Machine Learning

General types of learning

- **Supervised learning**
  “Here is the data set where the right answers (labels) are given for each example. Please produce more right answers.”
- **Unsupervised learning**
  “Here is the unlabelled data. Please find peculiarities, similarities or structures (e.g. clusters) in the data yourself.”
- **Reinforcement learning**
  “Learn to do something yourself purely by maximising your expected reward.”

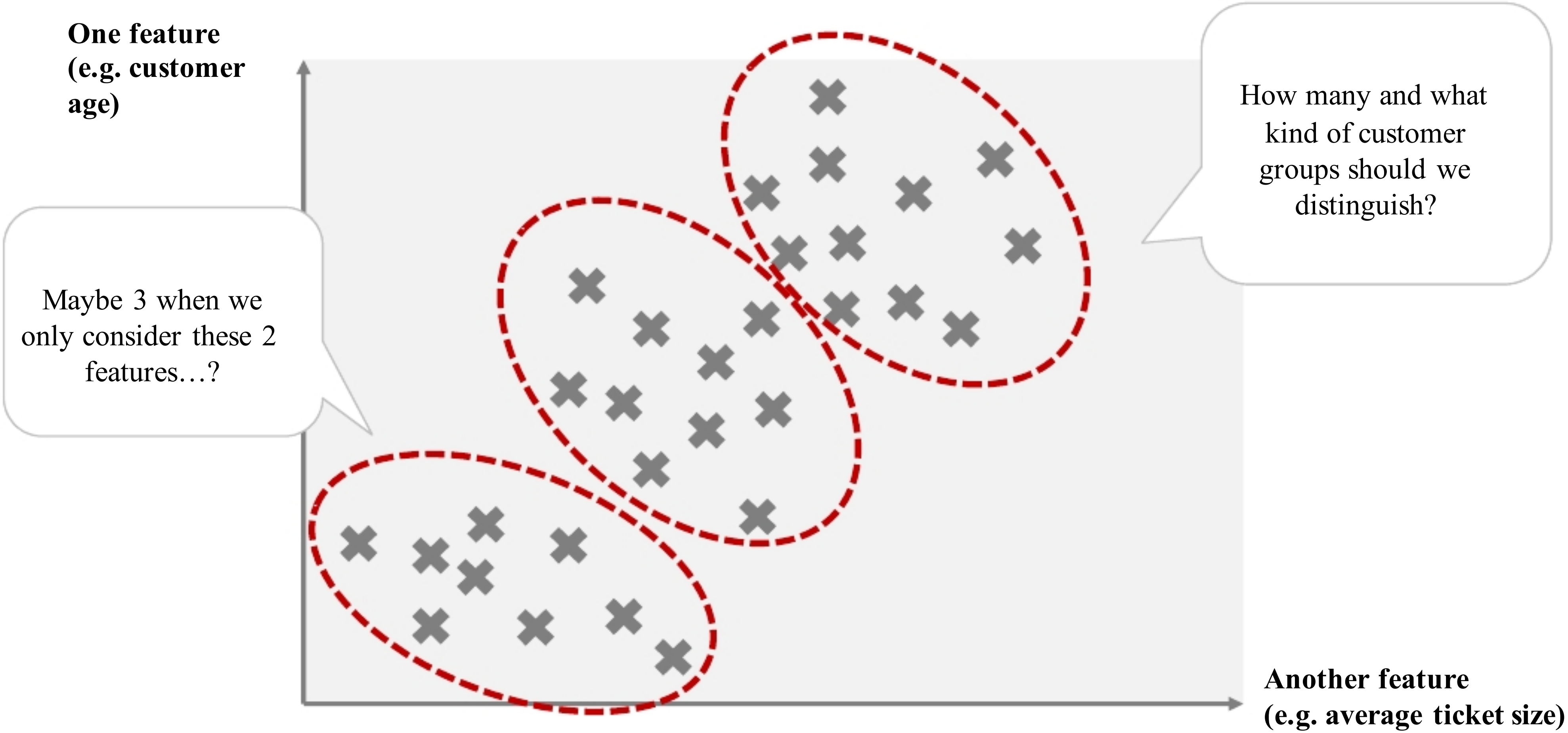
General types of problems

- Supervised learning:
  - **Regression**
    “Predict a continuous valued output.”
  - **Classification**
    “Predict a discrete valued output (e.g. a label or a class)”
- Unsupervised learning:
  - **Clustering**
    “Group similar examples into subsets – called clusters”
  - **Dimensionality reduction**
    „Maybe you don‘t need all the data. What is the essence of the data?“
    This can be a goal on its own but is often used as a pre-processing step for other ML tasks
  - …
- Reinforcement learning:
  - …

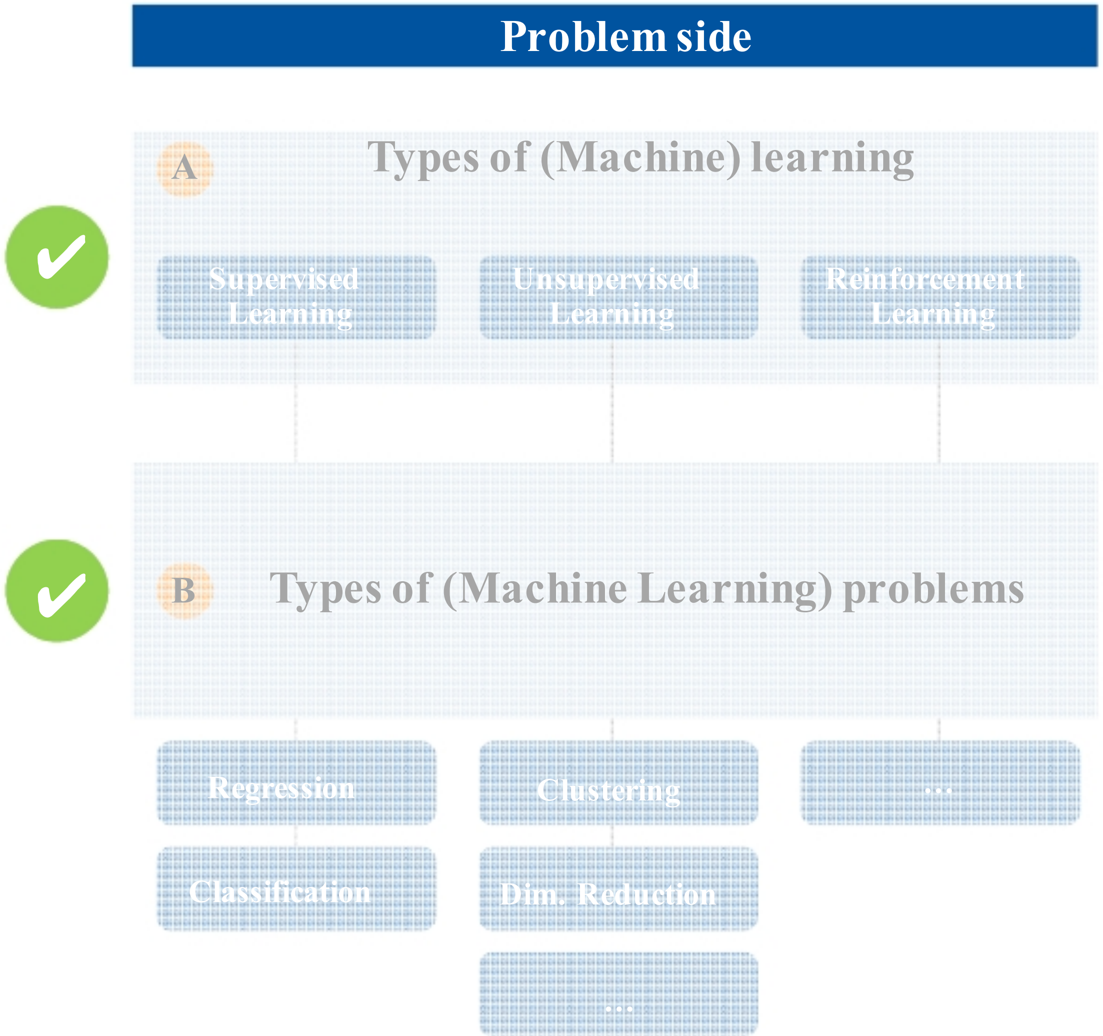
In a clustering problem, you do not have any labelled data – all you have is unlabelled data points

Clustering

Massively simplified



In our framework, we have now covered the problem side



Contents

There are thousands of Machine Learning algorithms – it is impossible to know and understand them all

Selection of Machine Learning algorithms and families

Far from being 'MECE'

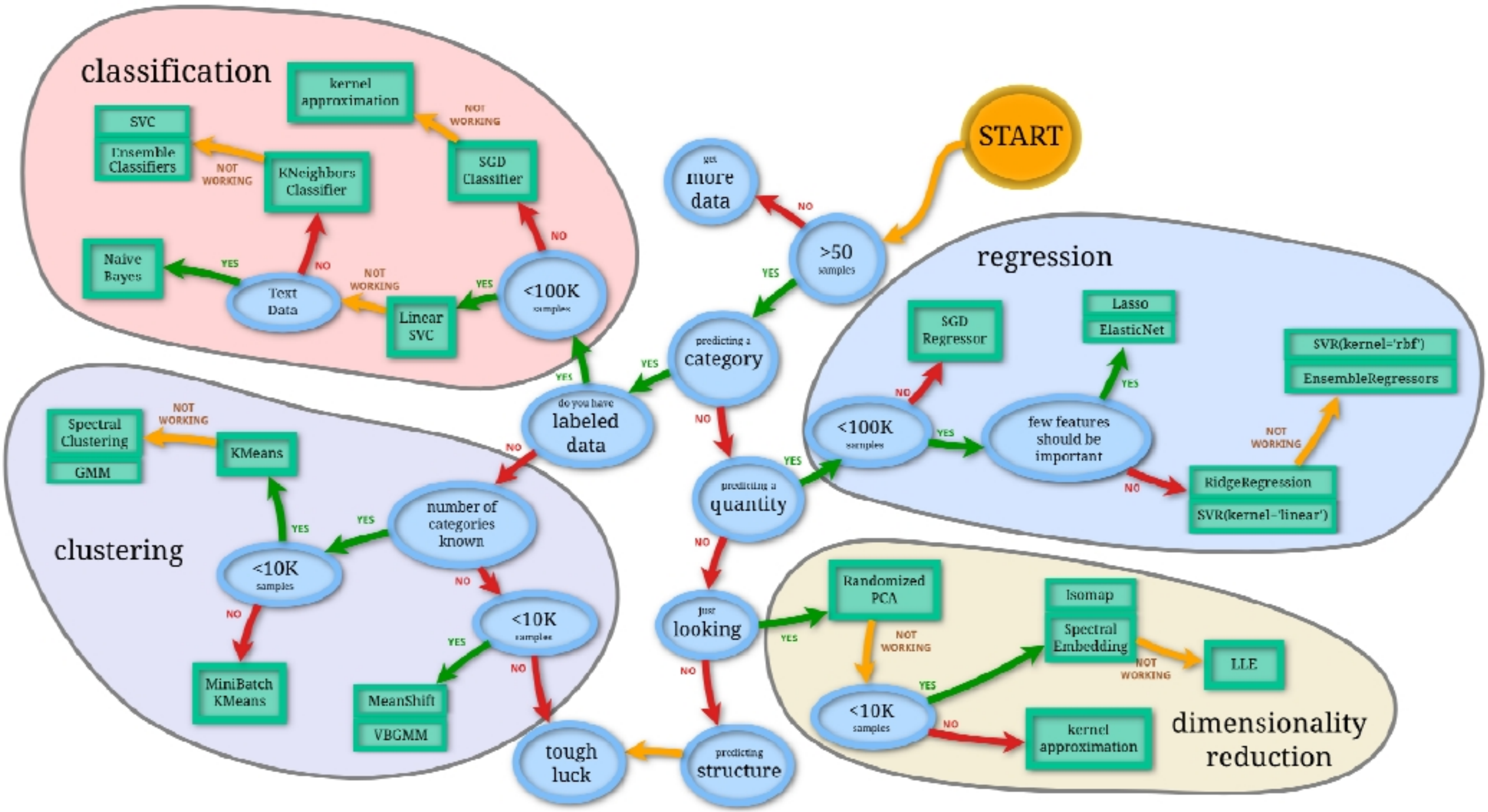
Decision trees
K-nearest neighbour (KNN)
Perceptron
Artificial Neural Networks (ANN)
Unsupervised Neural network models ("Restricted Boltzmann machines")
Deep belief networks
Random Forests
Linear Regression
Ordinary least squares (OLS)
Penalised regression
Principal Component Analysis (PCA)
Randomised PCA
Logistic Regression
(Linear / Quadratic) Discriminant Analysis
Support Vector Machines (SVM)
(Linear) Support Vector Classifier (SVC)
Support Vector Regression
Naive Bayes
K-means
Independent Component Analysis (ICA)
Non-negative matrix factorisation (NMF)

IsoMap
Association analysis
Hidden Markov Model
Kernel Approximation
MeanShift
Recurrent neural networks
Novelty and Outlier Detection
Density Estimation
Gaussian mixture models (GMM)
Manifold learning
Spectral Embedding ("Laplacian Eigenmaps")
Deep Learning
Locally linear embedding (LLE)
Hessian-based LLE ("Hessian Eigenmapping")
Multi-dimensional Scaling (MDS)
Bayes nets
Latent linear models
Sparse Bayesian Learning
Gaussian processes
CART

AdaBoost
LogitBoost
Polynomial Regression
State space models
Markov random fields
Convolutional neural networks
Conditional random fields (CRF)
Monte Carlo inference
Markov Chain Monte Carlo (MCMC) inference
Latent variable models
Latent Dirichlet allocation (LDA)
(Linear) Stochastic gradient descent (SGD) classifier
Gaussian Naive Bayes Classifier

…and thousands more…

'SciKit learn' (a Machine Learning library in Python) provides a useful cheatsheet for main algorithm families



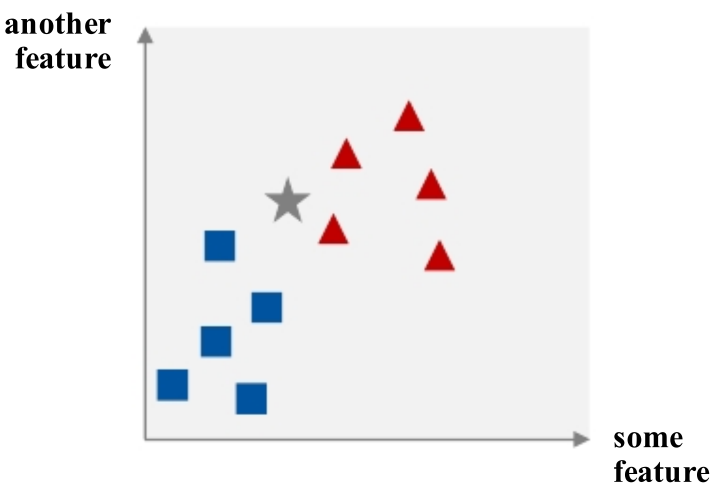
Source: *http://scikit-learn.org/stable/tutorial/machine_learning_map/*

Contents

Examples of Machine Learning algorithms: kNN

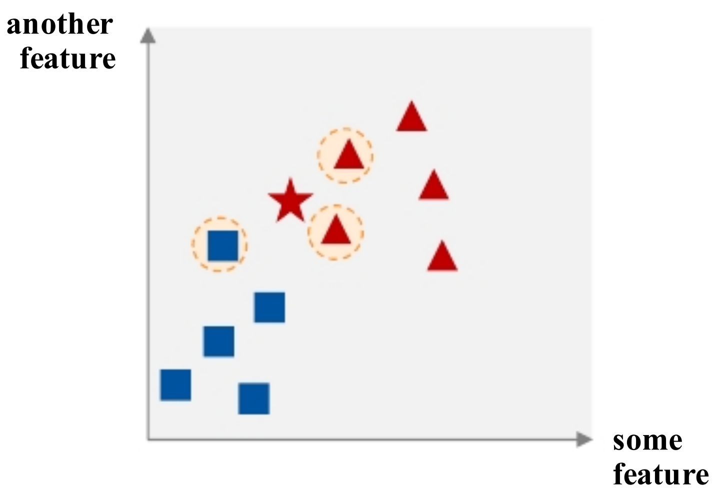
Massively simplified

Problem



I want to **classify** my data
I already have some correct classifications
(*what type of learning is this?)*
Now I got this new example that I need to classify
Which class should it be?

Basic idea to solve it



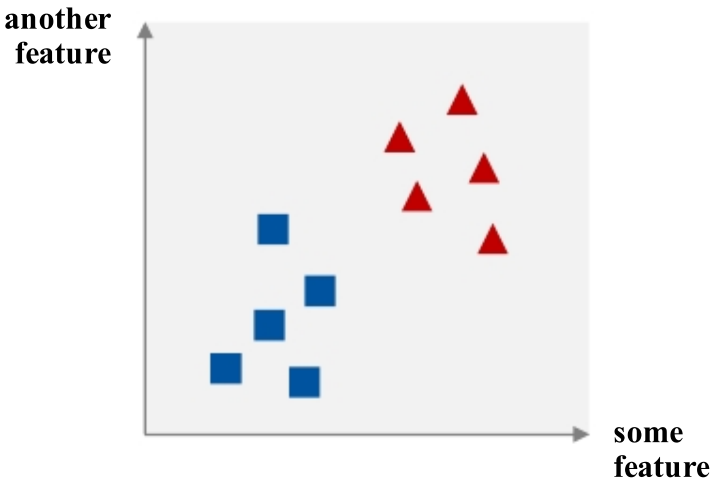
Why don‘t you use a majority vote of the nearest, let‘s say 3, labelled examples?

k nearest neighbours (kNN)

Examples of Machine Learning algorithms: SVM (Support Vector Machine)

Massively simplified

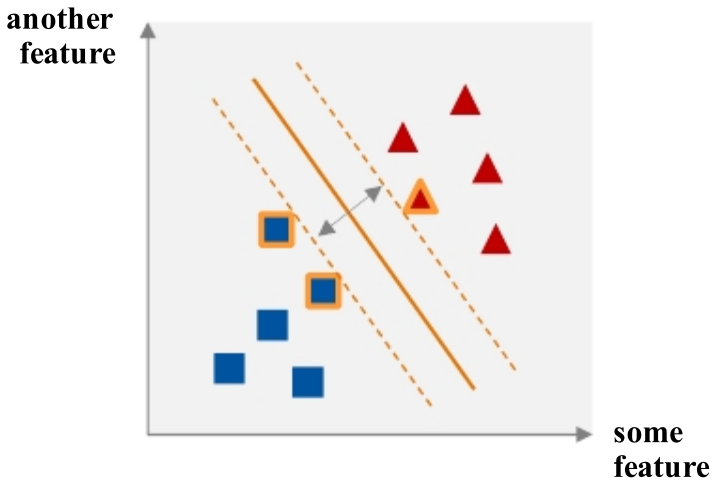
Problem



Listen, I have got these data points that have already been assigned to **classes** (*what type of learning is this?)*

Now I want to put a line between them so that I can classify new examples

Basic idea to solve it



Ok, why don‘t you put the line in there such that the margin between the closest points and the line is maximal?

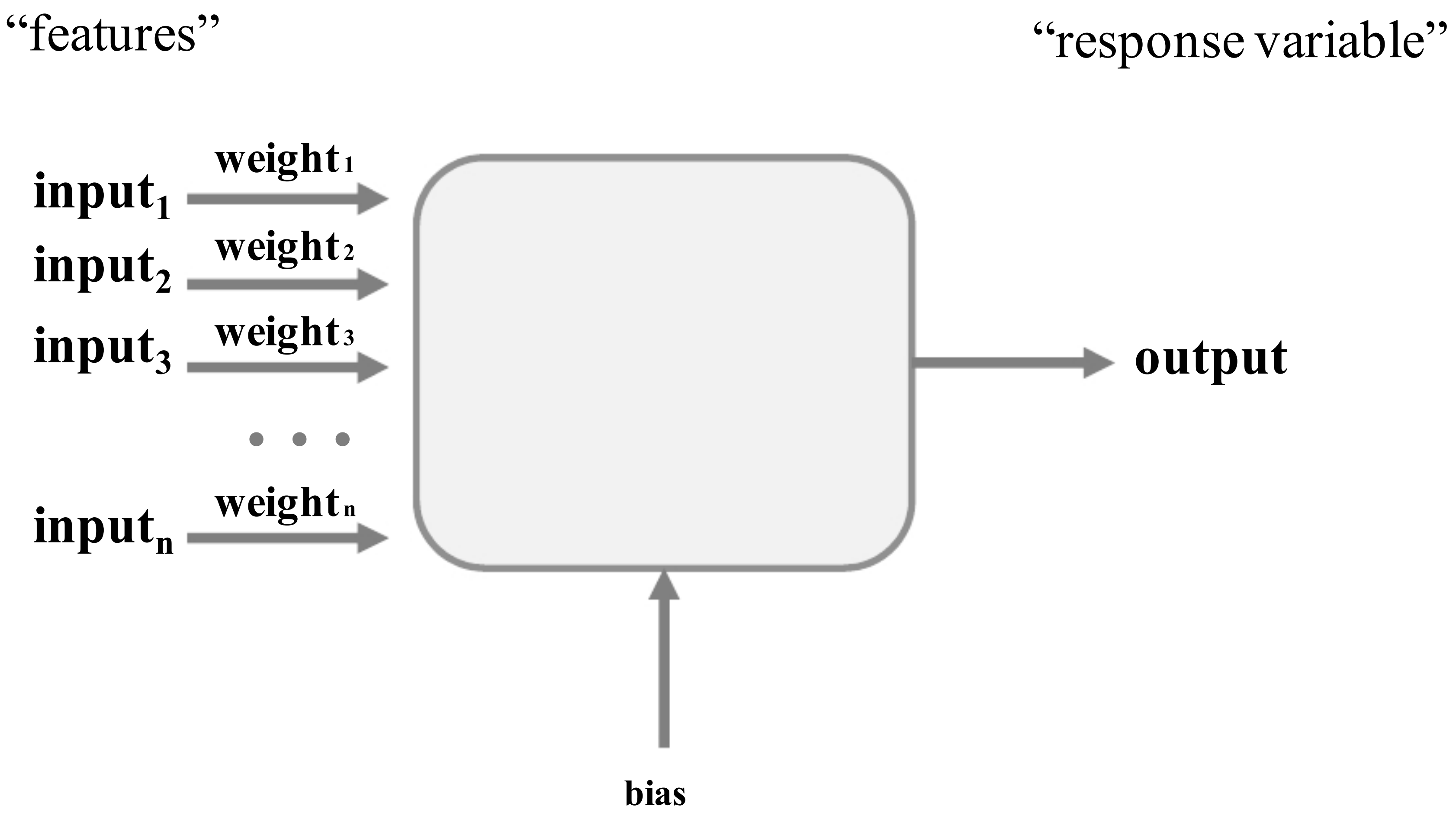
Linear Support Vector Classifier

1. This is for the linearly separable case only; please note that this is simplified: in reality we fit a hyperplane to the data

Neural Networks / Deep Learning An artificial neuron is the building block

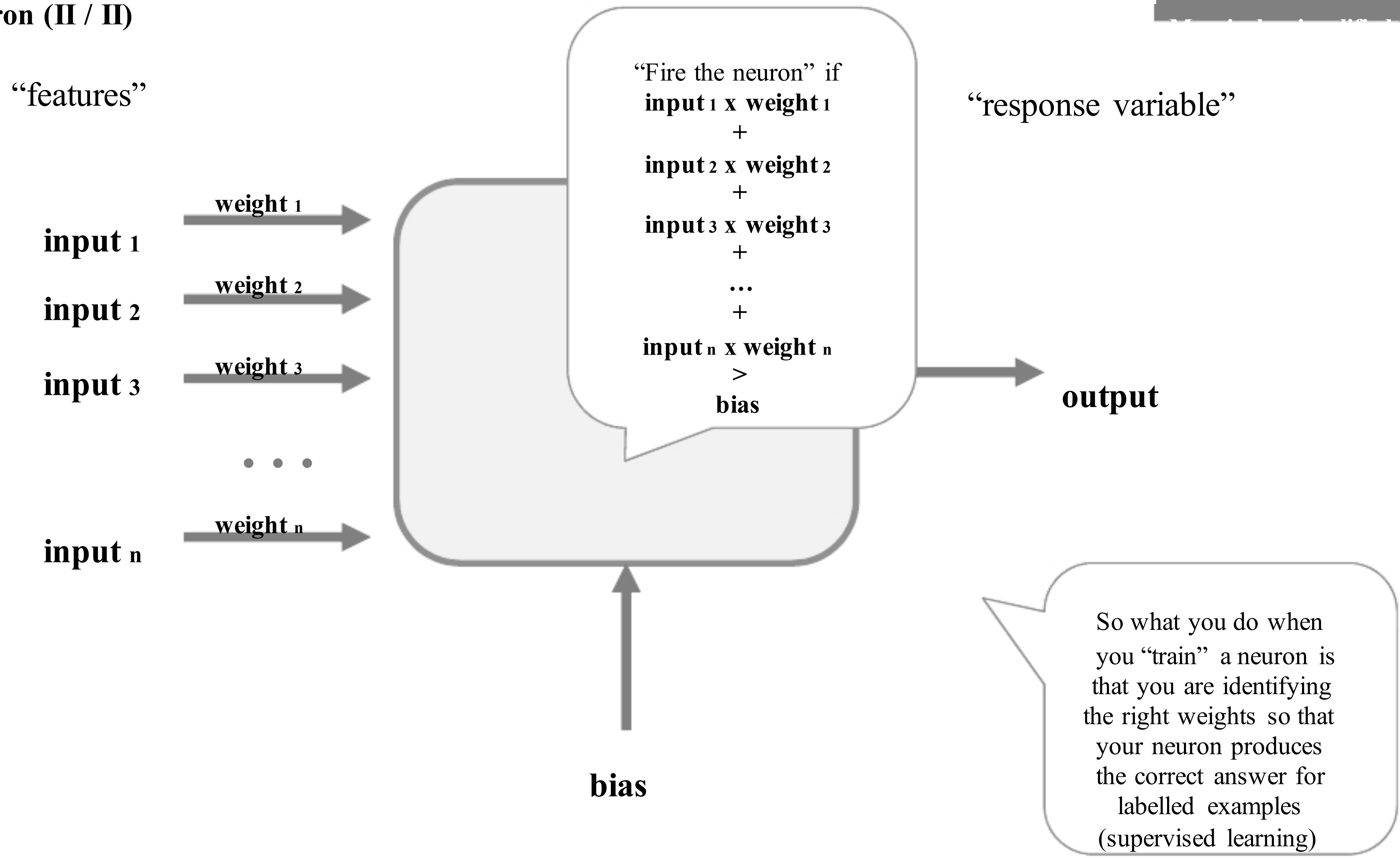
Artificial neuron (I / II)

Massively simplified

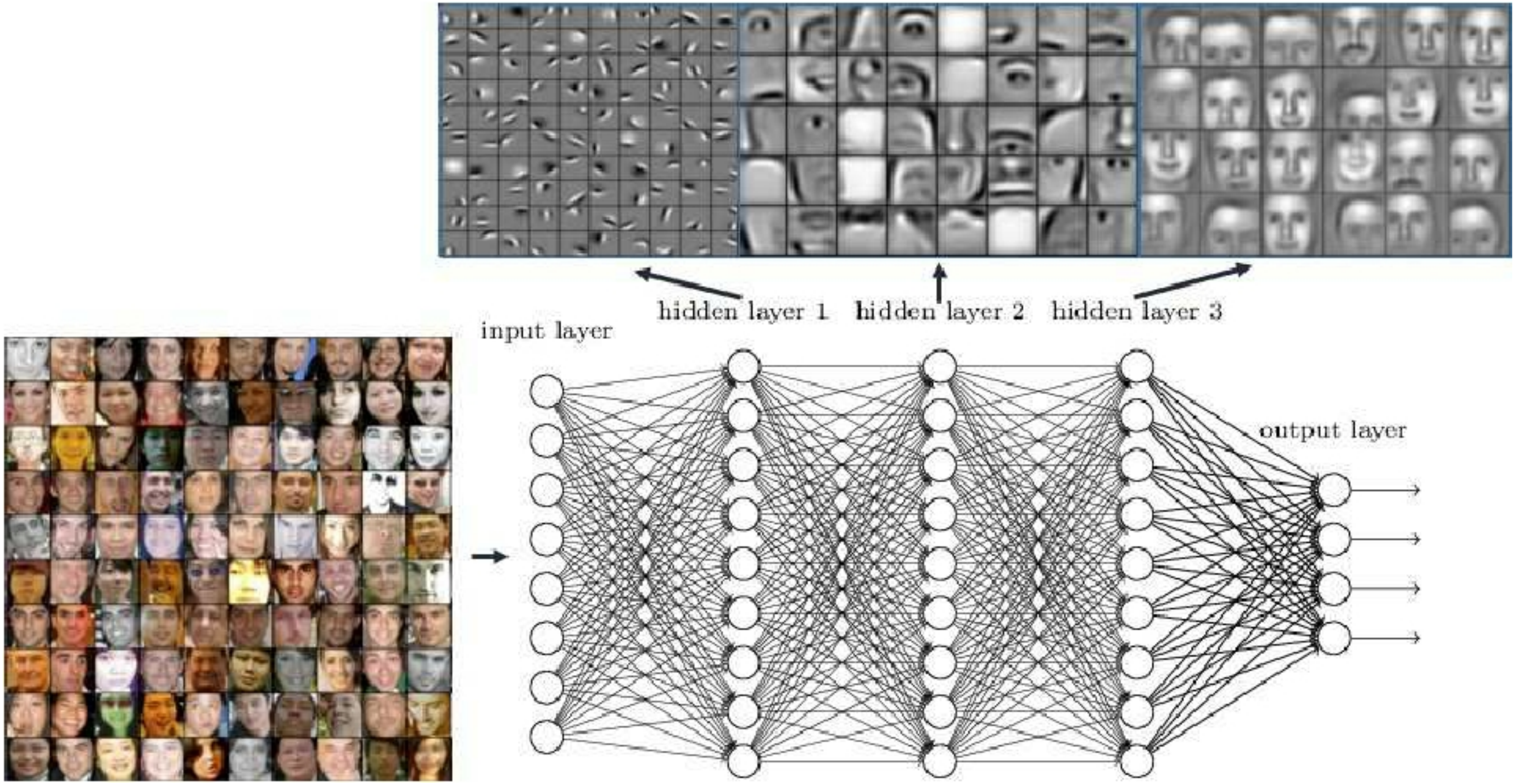


You can "train" a neuron by identifying the weights for each input in such a way that the neuron produces the correct answer given a set of inputs

Artificial neuron (II / II)



A single neuron itself is not very exciting – the magic happens when you use multiple neurons in parallel and then use multiple layers of these



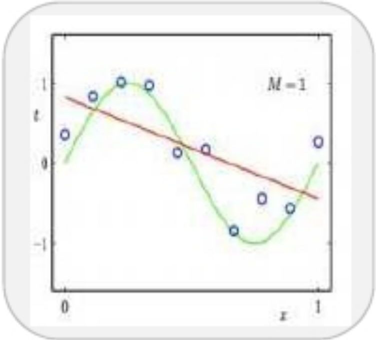
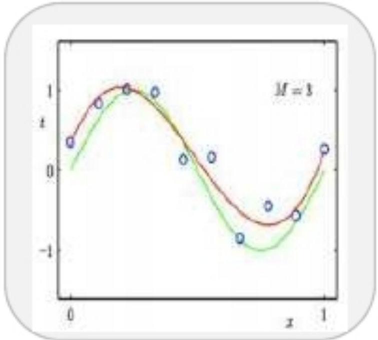
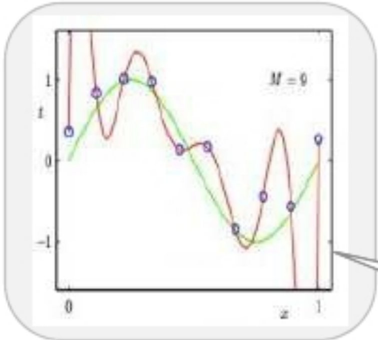
Source: Taken from http://www.rsipvision.com/

Contents

When fitting a regression model to the data, we can make the model infinitely complex simply by increasing the degree of the polynomial

Underfitting vs. overfitting (regression)

Supervised learning only

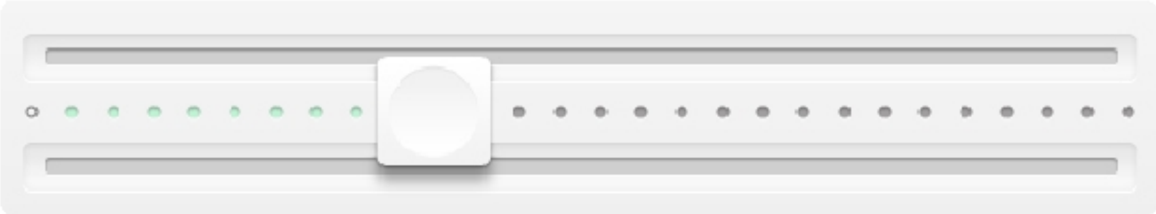
Underfitting	Appropriate	Overfitting
		
1 st degree (linear)	3 rd degree (cubic)	9 th degree

= actual data

= our model

Every single data point is on the line.

But does this reflect the underlying truth?



Model complexity

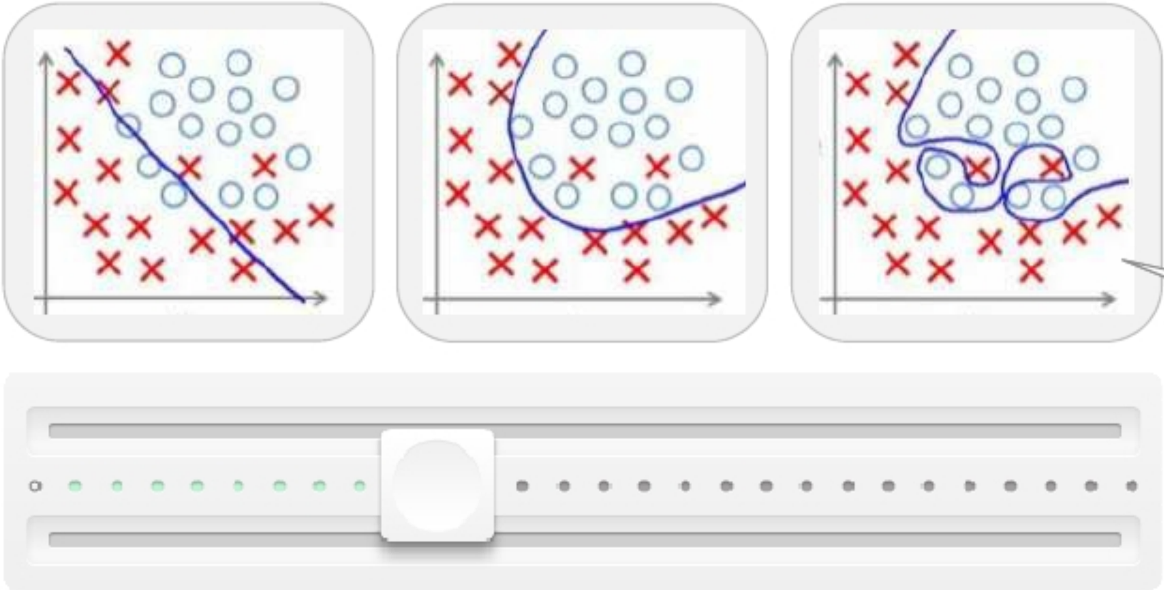
Source: Over- / underfitting drawings taken from http://www.inf.ed.ac.uk/teaching/courses/iaml/slides/eval-2x2.pdf

The same is true for classification – we can make our decision boundary infinitively complex

Underfitting vs. overfitting (classification)

Supervised learning only

Underfitting | Appropriate | Overfitting



= Example from class 0

= Example from class 1

Nice fit with the training data. We have not made a single mistake!

But what if we get new data?

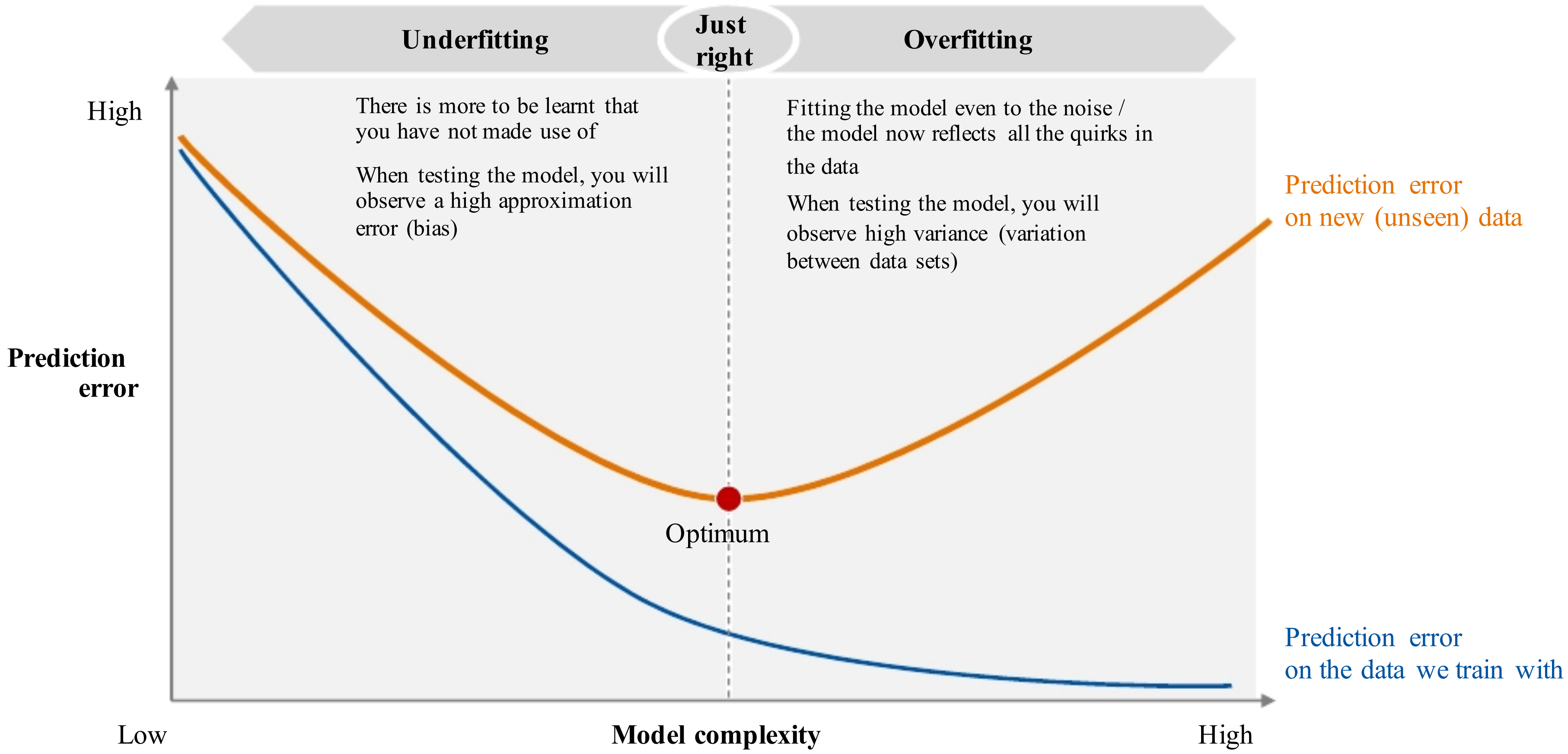
Model complexity

Source: *Over- / underfitting drawings taken from http://www.inf.ed.ac.uk/teaching/courses/iaml/slides/eval-2x2.pdf*

There is a (bias-variance) trade-off when fitting a model to the data – we can under- or overfit our learner to the data

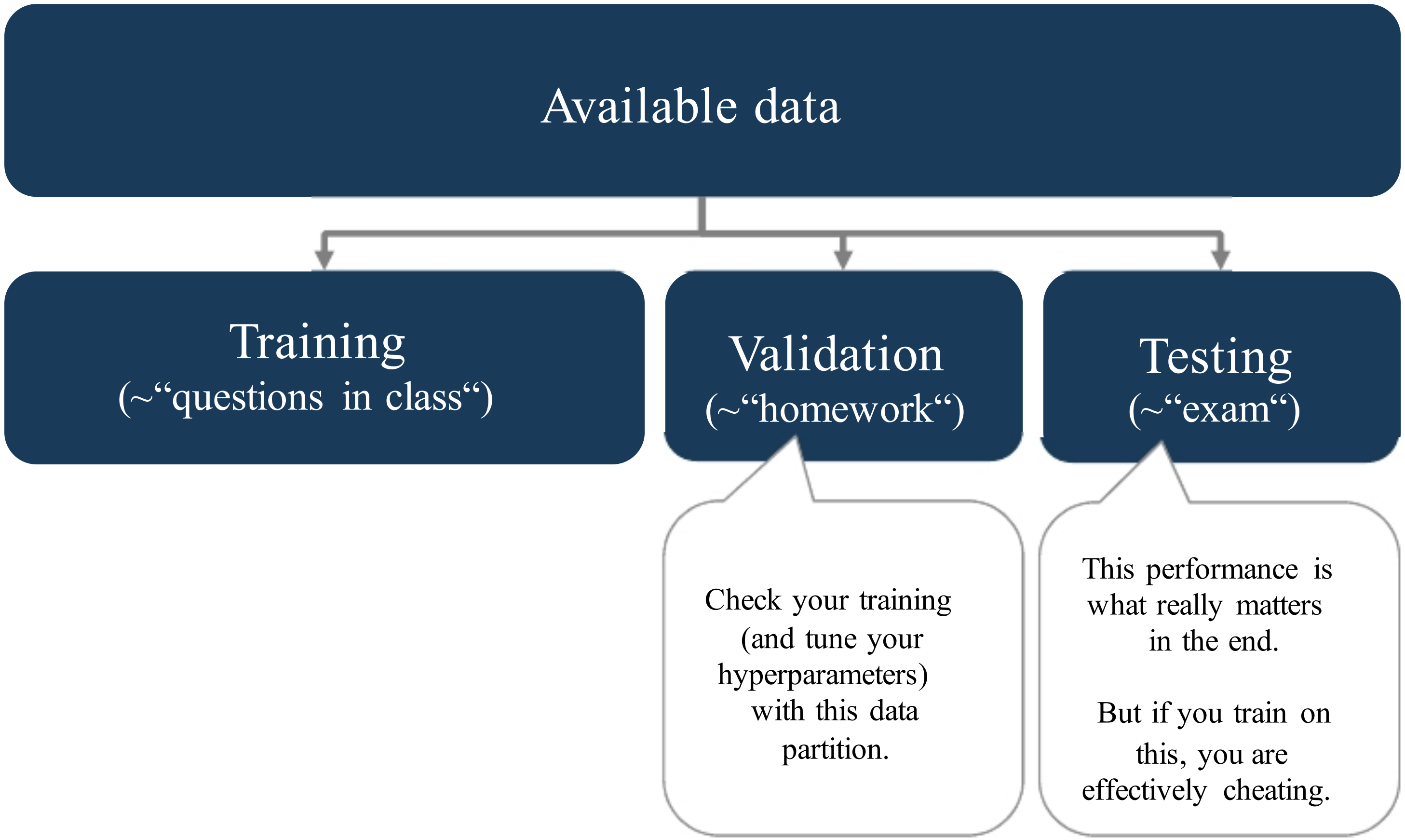
Overfitting vs. underfitting

Supervised learning only



In order to prevent overfitting, the available data is typically split into three partitions: for training, validation and for testing

Supervised learning only



Conclusion

Scope of this lecture
Recent examples of Machine Learning
Definition and promises of Machine Learning
The framework for this lecture
“The problem side”
“The solution side”
Training (“fitting”), validating and testing