Hes·so
Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences
Western Switzerland

MSE | MASTER OF SCIENCE IN ENGINEERING

Machine Learning

T-MachLe

1. Fundamentals on Machine Learning

Jean Hennebert
Andres Perez Uribe

Plan

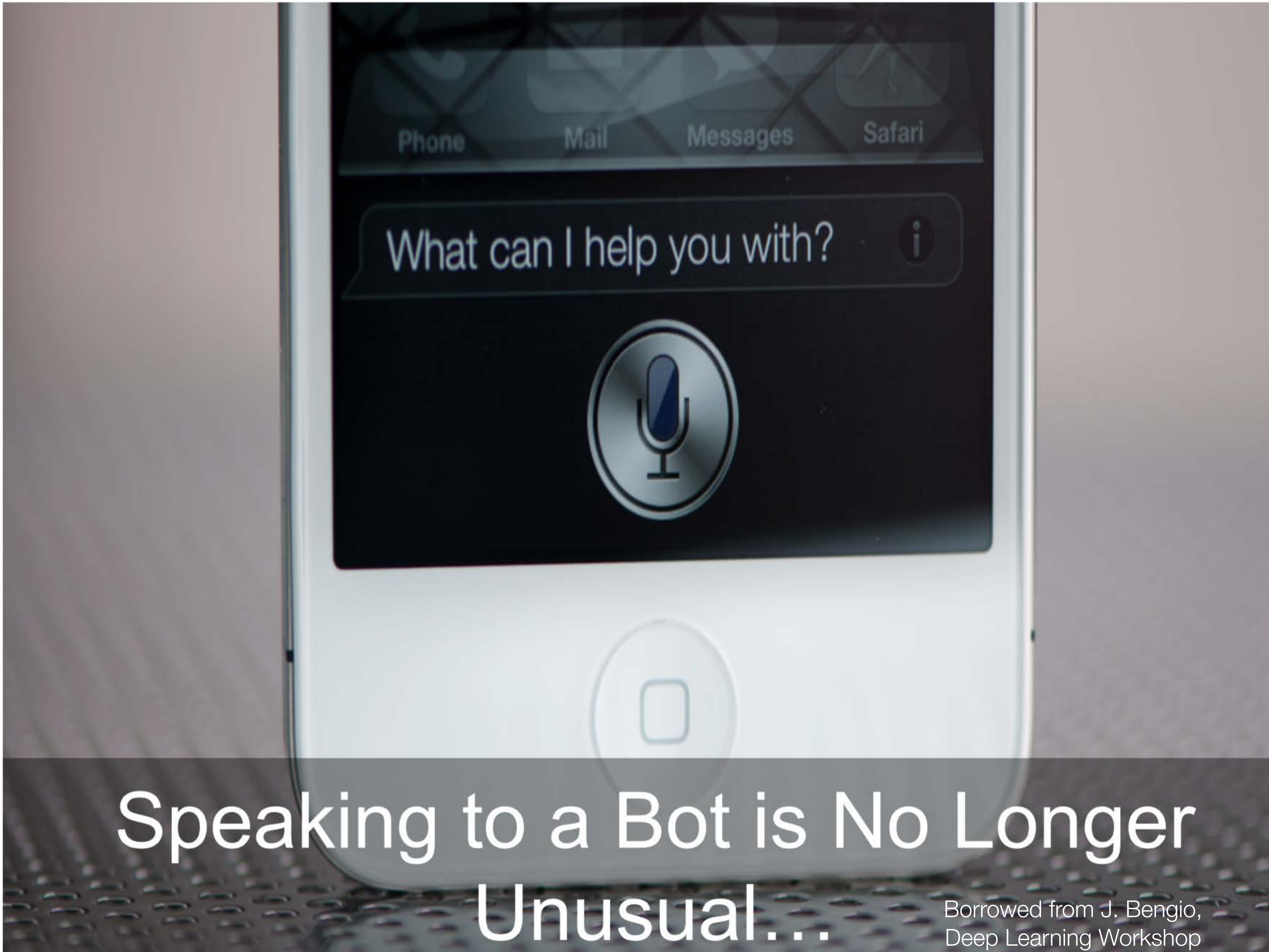
Why machine learning?

Activity

- Address the following points to produce keywords summary.
  - How to define simply machine learning? Imagine you want to explain it to your grandmother!
  - Why do we need such systems?
  - Did you use some machine learning applications today, this week?
  - What are the dangers of machine learning?

1.1
Some initial examples

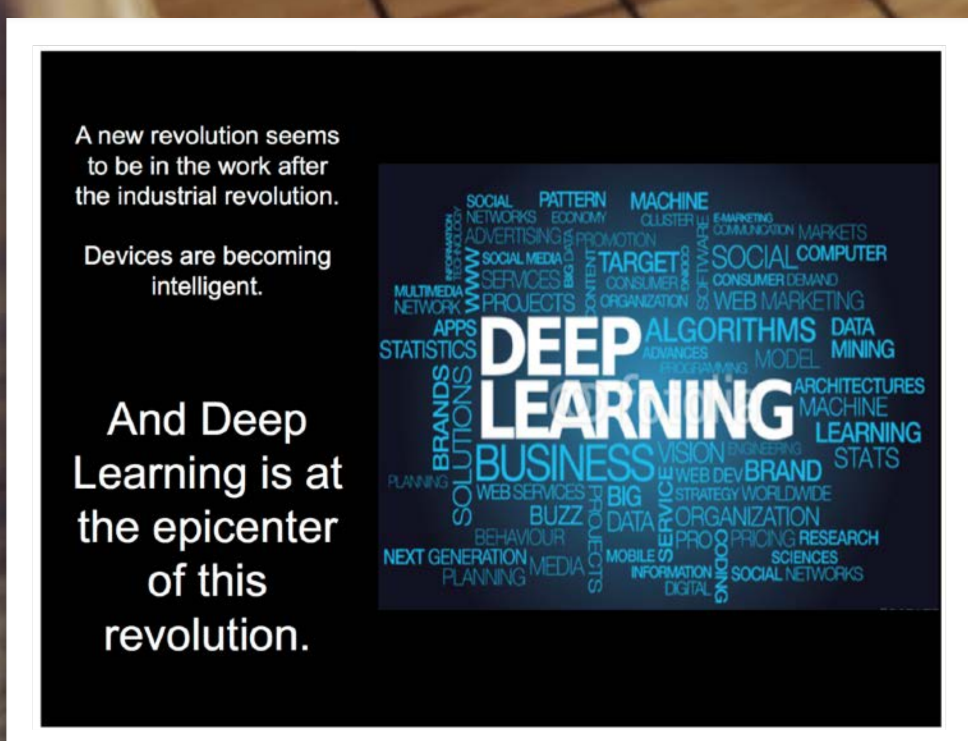

Cars are now driving themselves…
(far from perfectly, though)
NVIDIA
Borrowed from J. Bengio, Deep Learning Workshop



Speaking to a Bot is No Longer Unusual…

Borrowed from J. Bengio, Deep Learning Workshop

March 2016: World Go Champion Beaten by Machine



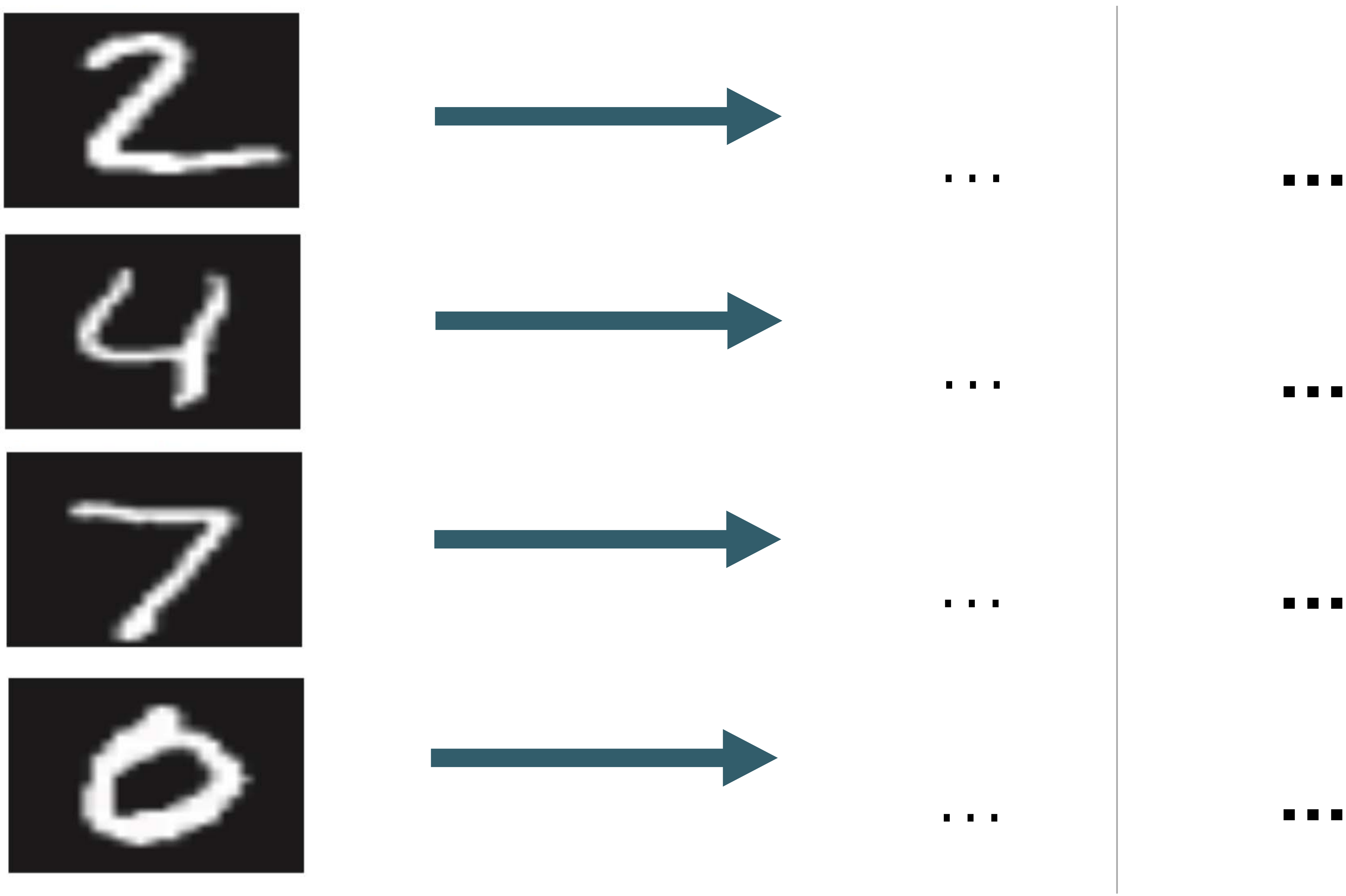
Borrowed from J. Bengio, Deep Learning Workshop

Let's assume we want to build a handwriting recogniser only digits to simplify the discussion

Activity

- How would you proceed?
  - Formalise the system in terms of inputs and outputs
  - What are the difficulties of the system?
  - How can we evaluate such a system?

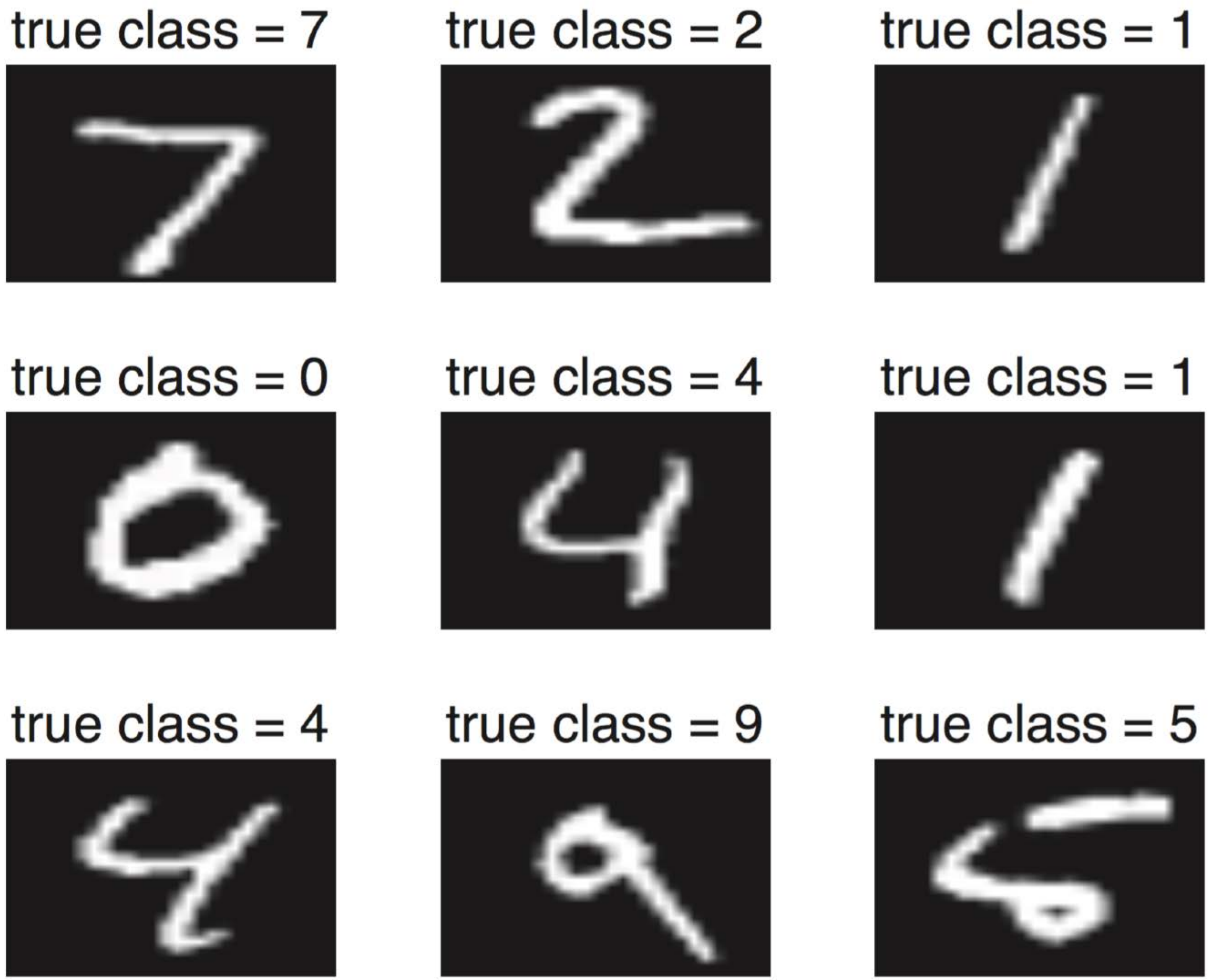
Let's take the place of the machine for one moment…



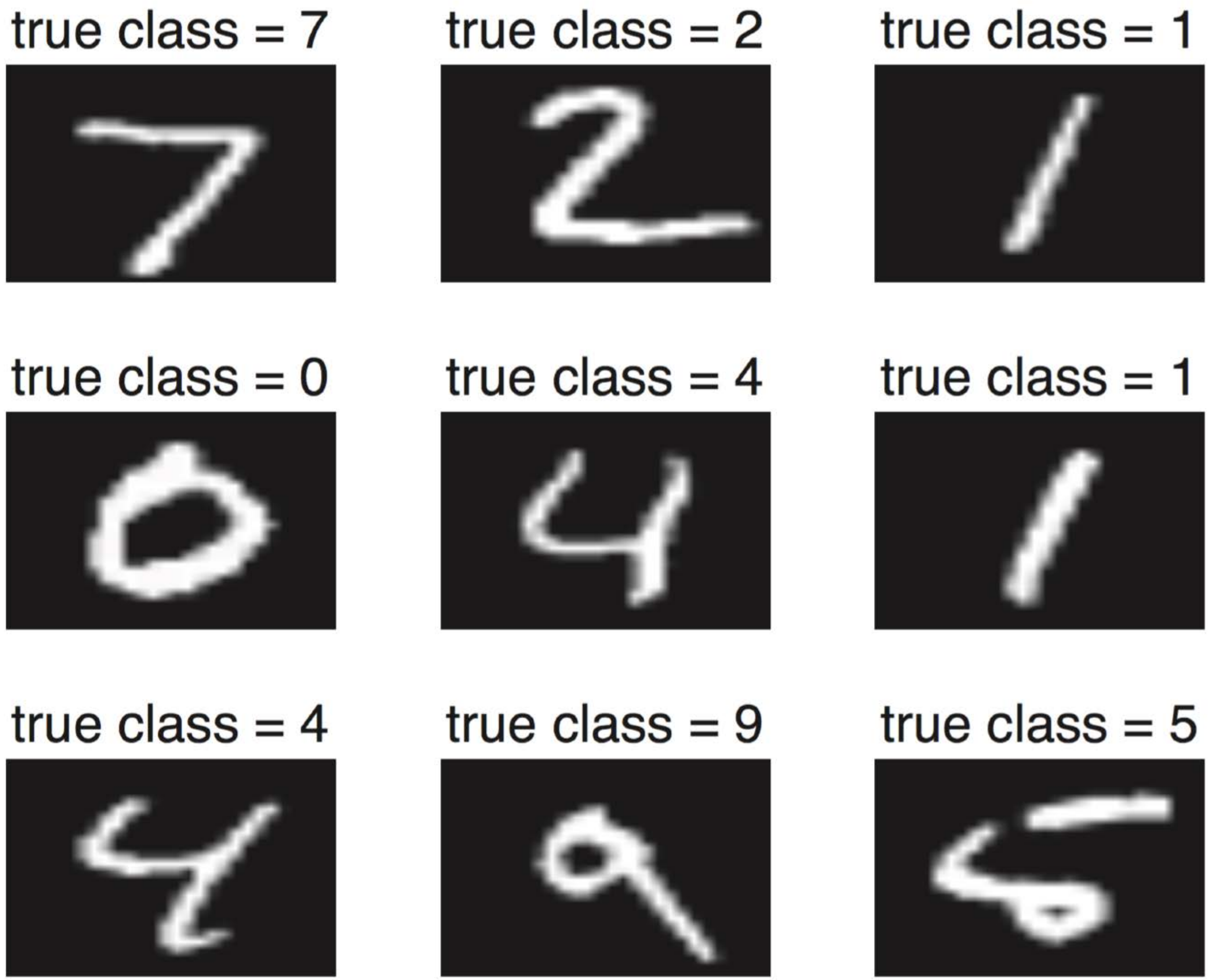
Let's take the place of the machine for one moment…

x input variable	***h(x)*** mapping function	$\hat{y}$ output variable "gotten output"	***y*** output variable "desired output"
		2	**2**
		4	**4**
		7	**7**
		6	**0**

The ground truth !


true class = 7
true class = 2
true class = 1
true class = 0
true class = 4
true class = 1
true class = 4
true class = 9
true class = 5

The output in this case is a class label from 0 to 9

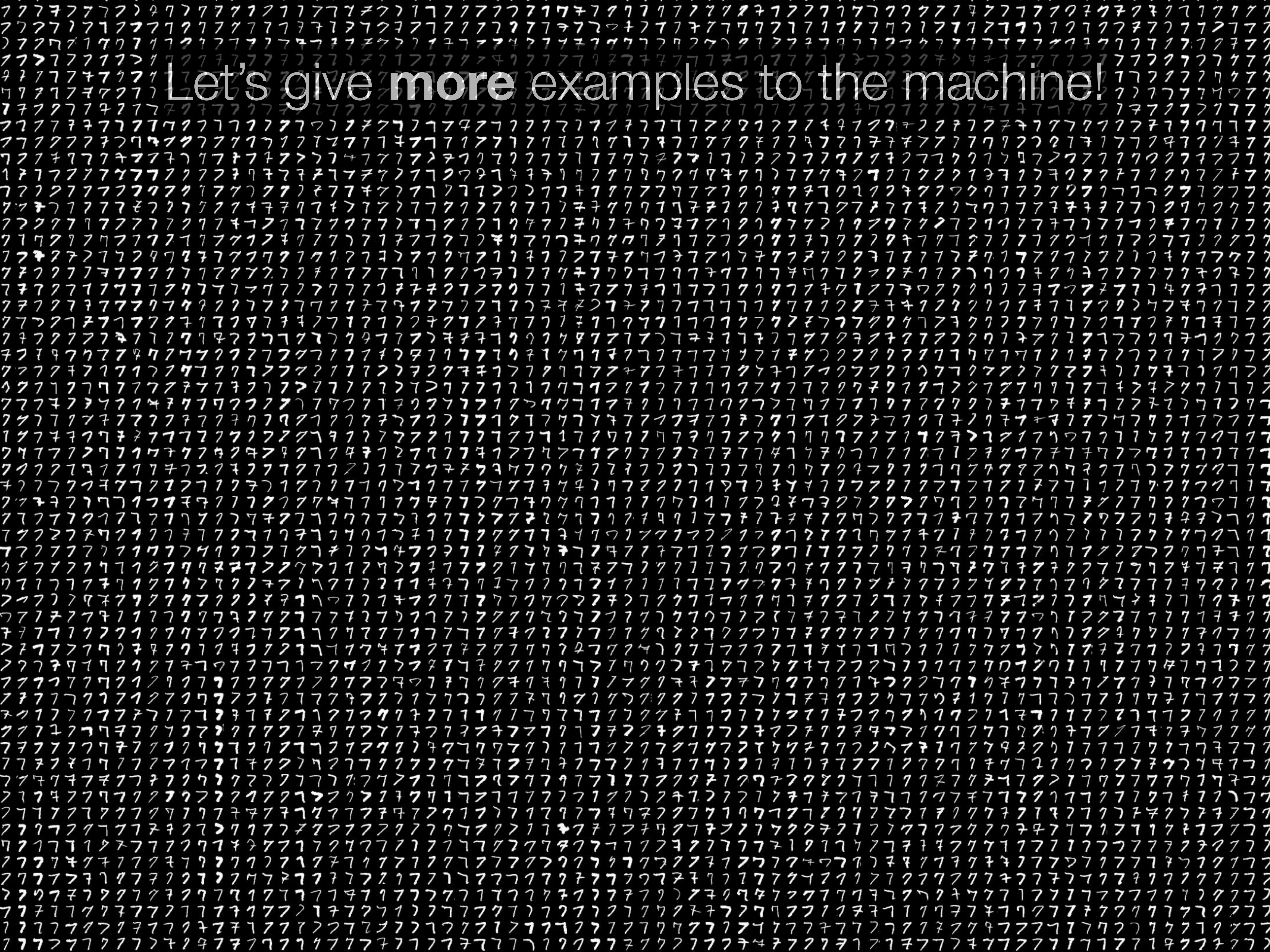


This is a **classification** problem

Let’s give examples to the machine!

We observe **variabilities**
that we want to capture

Let’s give **more** examples to the machine!



What's happening in the learning process?

- The machine will try to find in a huge set of potential mapping functions h(x), one that is (hopefully) the best at reducing the *distortion* (kind of) between

$\hat{y}$ and y

$\hat{y}$: output variable "**gotten** output"

y: output variable "**desired** output" "**target** output"

- Open points that we will see later in this class:
  - What kind of *h(x) family* can we use?
  - How to explore these families?
  - What do we mean by *distortion*?

Let’s assume we want to build a face recognition system

Activity

- How would you proceed?
  - Formalise the system in terms of inputs and outputs
  - What are the difficulties of the system?
  - How can we evaluate such a system?

Who is in this picture?

https://en.wikipedia.org/wiki/The_Big_Bang_Theory

Who is in this picture?

Sheldon

Penny

Raj

Leonard

Amy

Howard

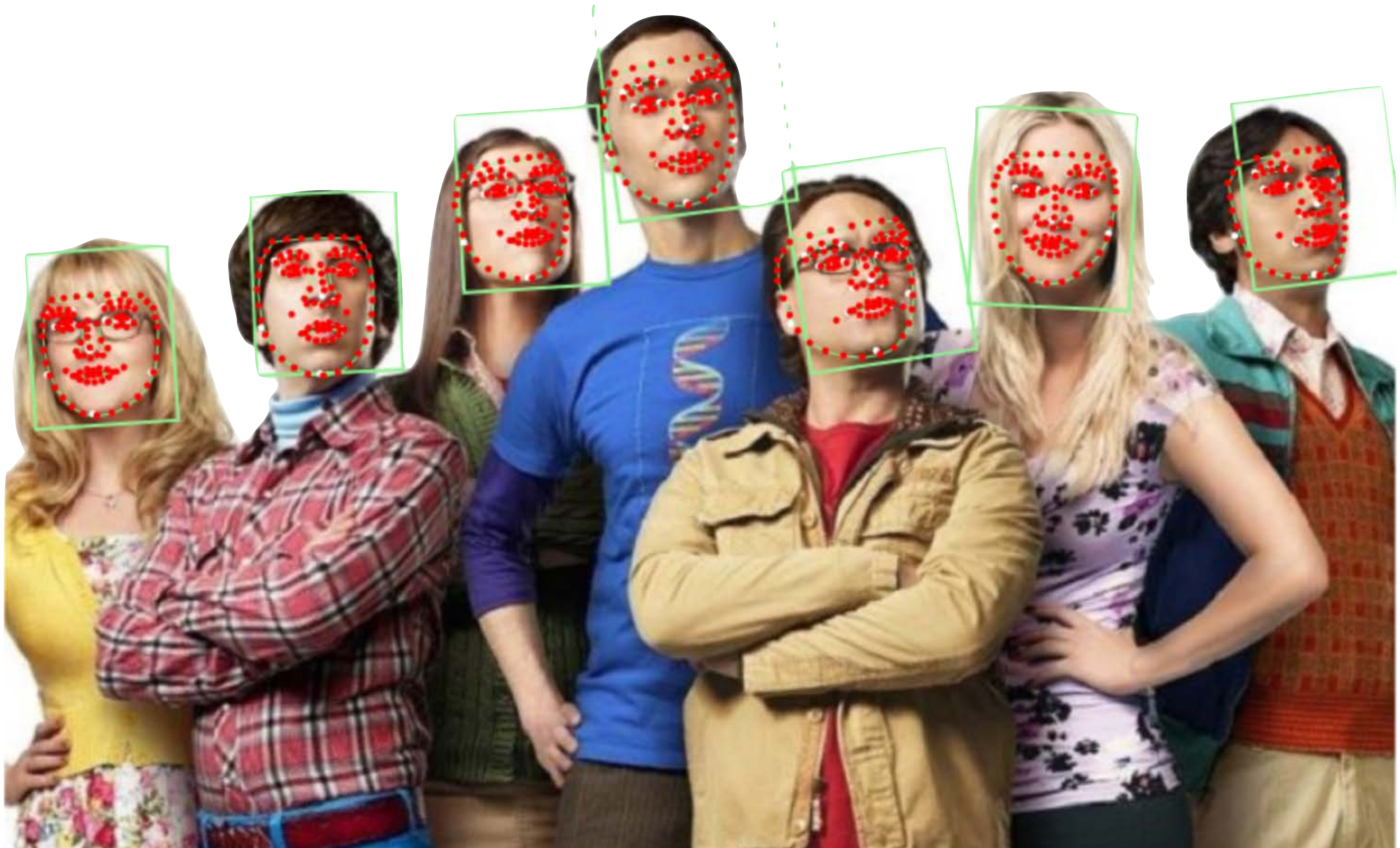
Bernadette

The problem has actually two parts

- Face detection — Where are the faces?
- Face identification —To which id corresponds each detected face?

1st part: face detection

- This can be viewed as a "**feature extraction**" it could be debated if it is really a feature extraction — we'll introduce a more formal definition of feature extraction later on, at least let's agree it is a kind of needed **pre-processing**
- Objective = detect the bounding box (in green)
- The used API also provides the face landmarks (in red) https://www.betafaceapi.com/demo.html



2nd part: face identification

- Assuming a detected face in its bounding box: who is it?

  …

- We have again here a **classification** problem as we need to give a **label** to the face
- We can train it as for the digits, giving as many example as we may find of pictures of Sheldon, Leonard, Howard, Penny, …

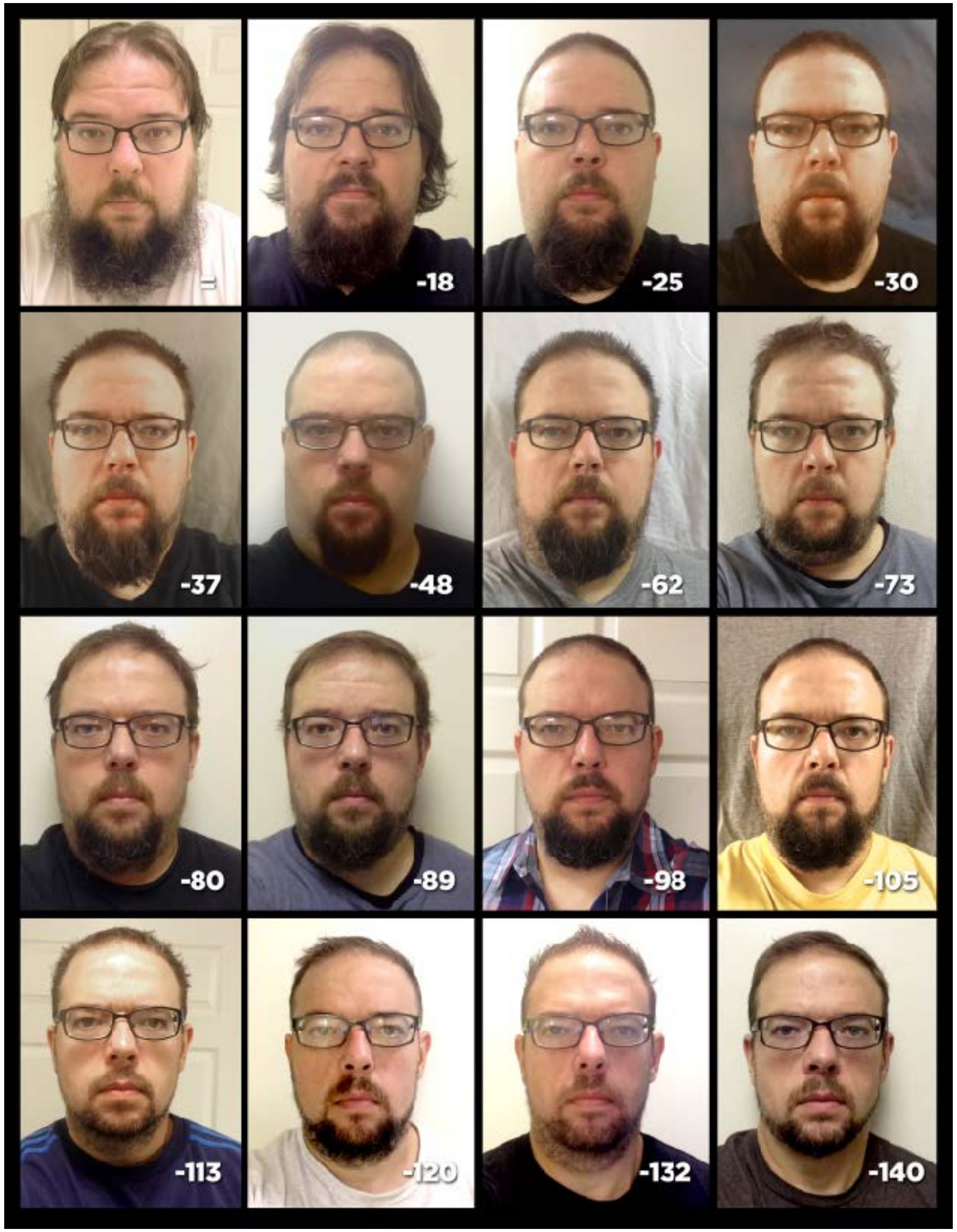
2nd part: face identification

- Who is it?

…

…

- Are we able to **generalise**?
- We have here more complex **variabilities** such as growing a barb, changing hair style, …
- This means we will have to look for *invariant features …* or to increase dramatically the training set.



Source: https://www.quora.com/How-drastically-can-your-face-change-through-exercise-and-by-getting-lean

2nd part: face identification - reasoning on class cardinality

- **How many classes** do we have?

…

2nd part: face identification - reasoning on class cardinality

- **How many classes** do we have?

...

- How many classes do we have?
  - 7 - considering the picture
  - 20 - considering S05E04 of BBT
  - 200 - considering S05 of BBT
  - 1000 - considering BBT with all seasons
  - 7.6 billion - considering planet earth

Remark: the problem gets more difficult when the number of classes increase

Remark: a simpler problem is face verification — "*Is this the face of Sheldon?*", actually a 2-class problem.

Other examples from iCoSys

- Institute of Complex Systems
  - Created end 2012
  - For now: clear focus on computer science arm of Complex Systems
  - 5 profs - 11 scientific collaborators - 4 PhD students

> *All things being in a chain of influence and in a chain of causes, i deem it impossible to know the whole without knowing the parts or to know the parts without knowing the whole.*
>
> *Blaise Pascal, Thinkings.*

Distributed Computing

> Large-scale parallel and distributed architectures
> Middelware for parallel and distributing programming
> Parallel and distributed high performance applications
> Large mobile and sensor networks

Intelligent Data Analysis

> Machine learning
> Big data analysis
> Signal processing
> Algorithms

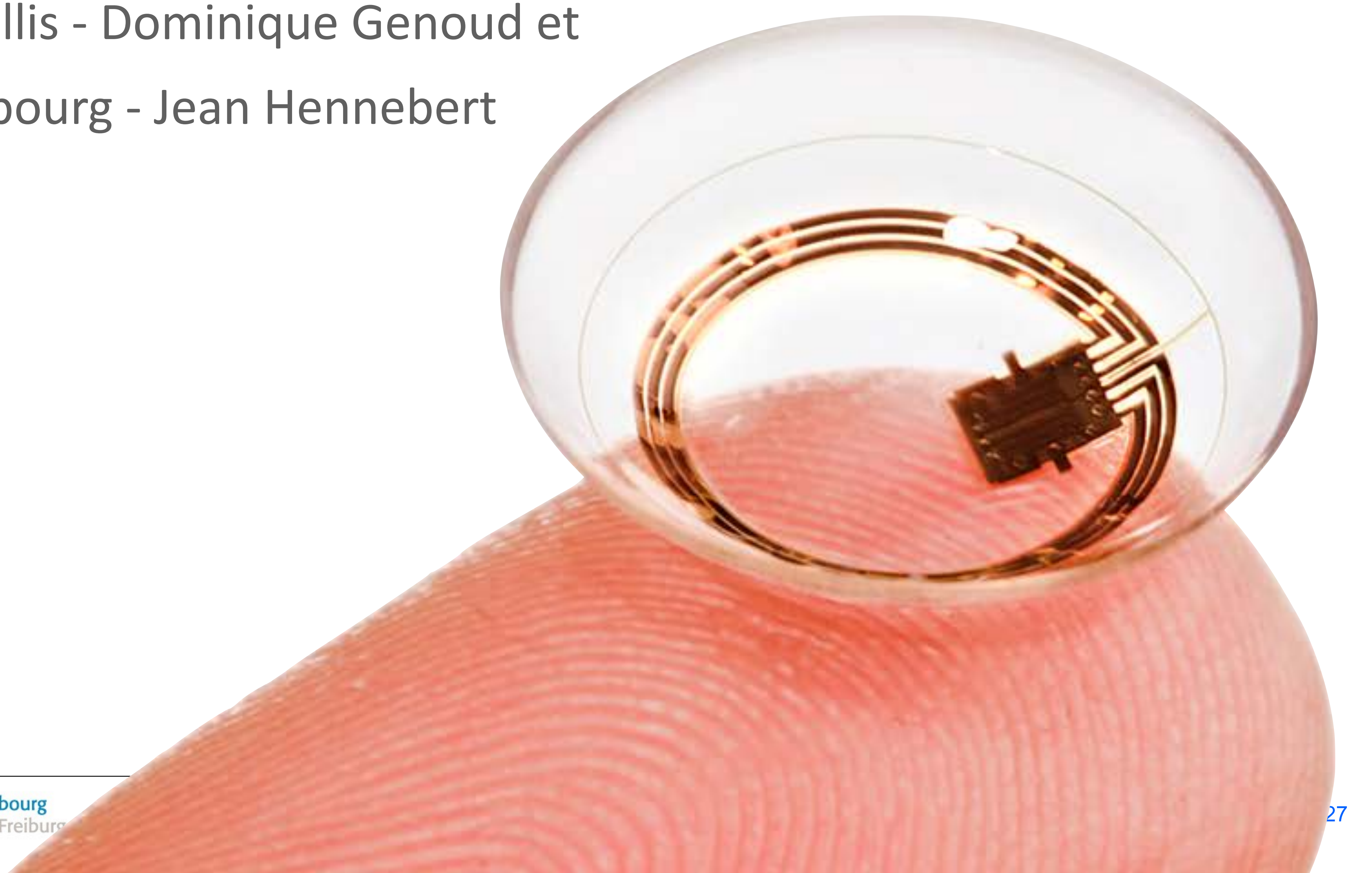
Sustainable ICT for Smart Living

> Data management and processing for sensor networks
> Web of Things
> Energy efficient IT
> IT for energy efficiency

Ecole d'ingénieurs et d'architectes de Fribourg
Hochschule für Technik und Architektur Freiburg

Projet CTI SENSIMED PLUS

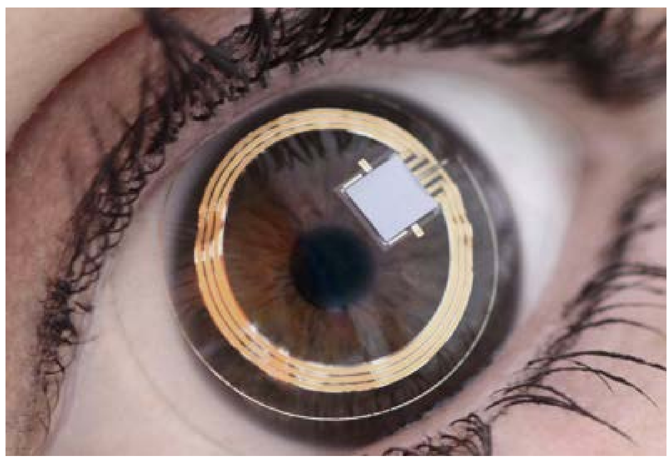
- "Early detection of glaucoma"
  - Entreprise Sensimed,
  - HES-SO//Wallis - Dominique Genoud et
  - HES-SO//Fribourg - Jean Hennebert




TEDx
San Diego
x = independently organized TED event
Shared with you by
OnStar®
Institut des Systèmes Complexes - Projet Sensimed Plus
Projet CTI Franc Fort 2012-2013 - Jean Hennebert, Pierre Kuonen

Sensimed Plus

From sensors to cloud to intelligence

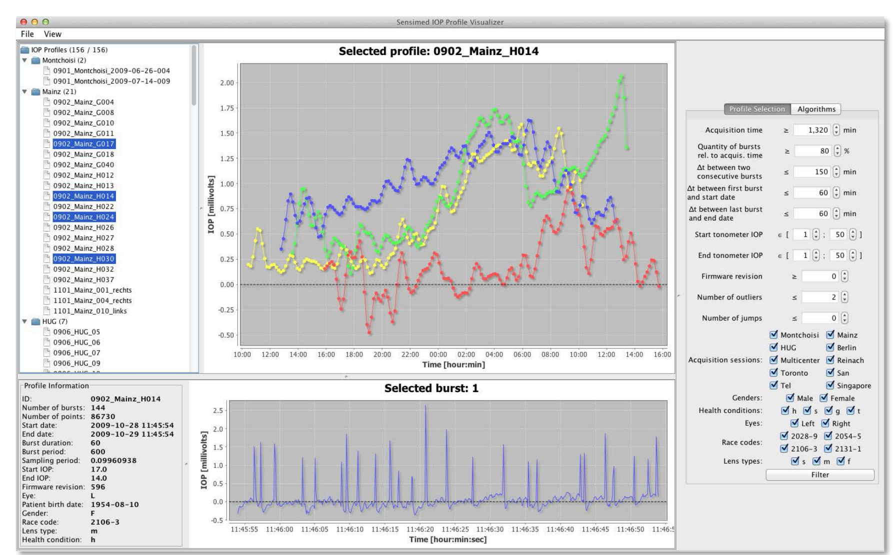


Patient

Embedded system

Visualisation and basic analysis
(PC of the doctor)

Scalable server architecture
(private at Sensimed)



Machine Learning to detect evolution patterns
on healthy and glaucomatous patients

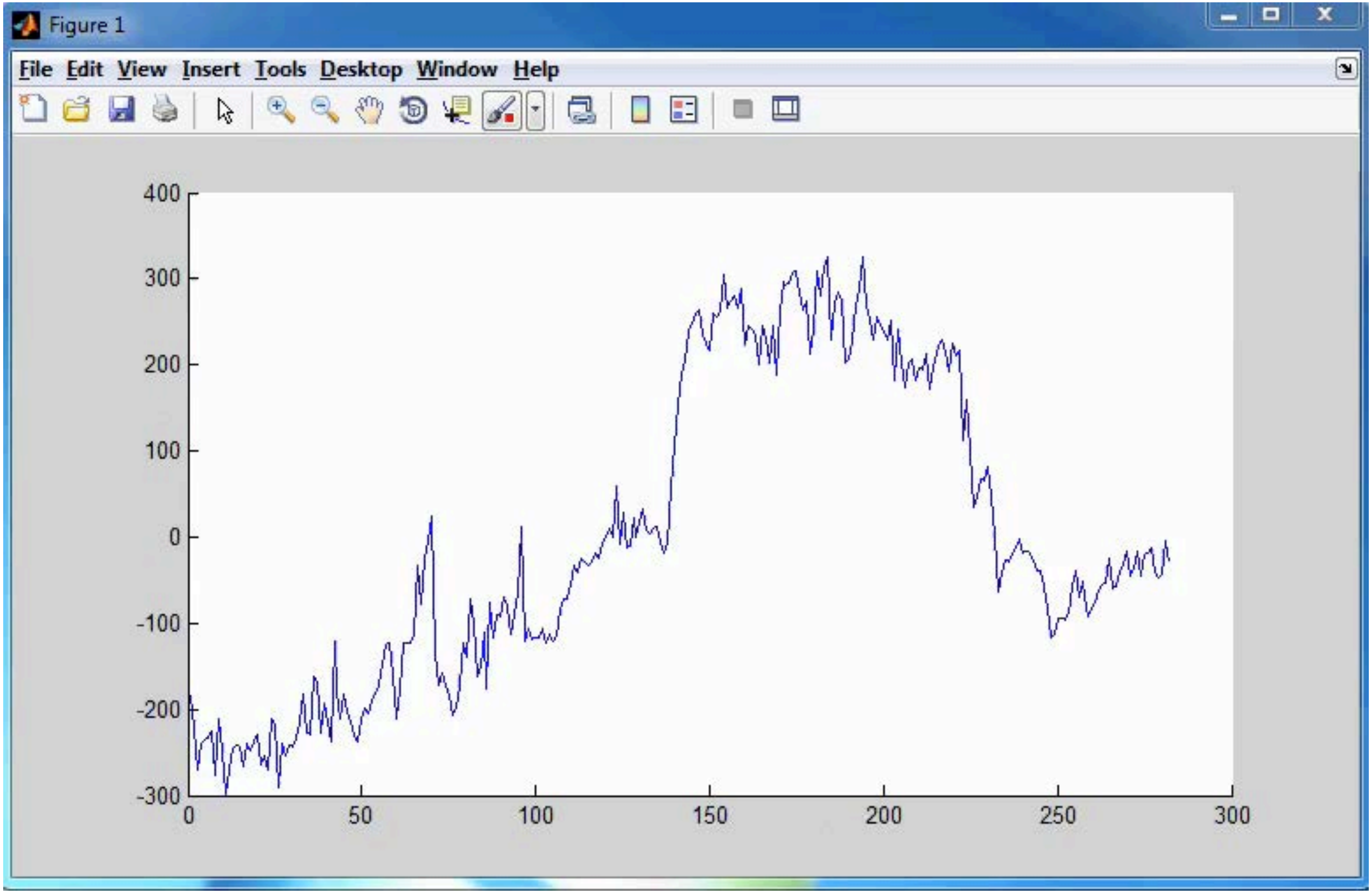
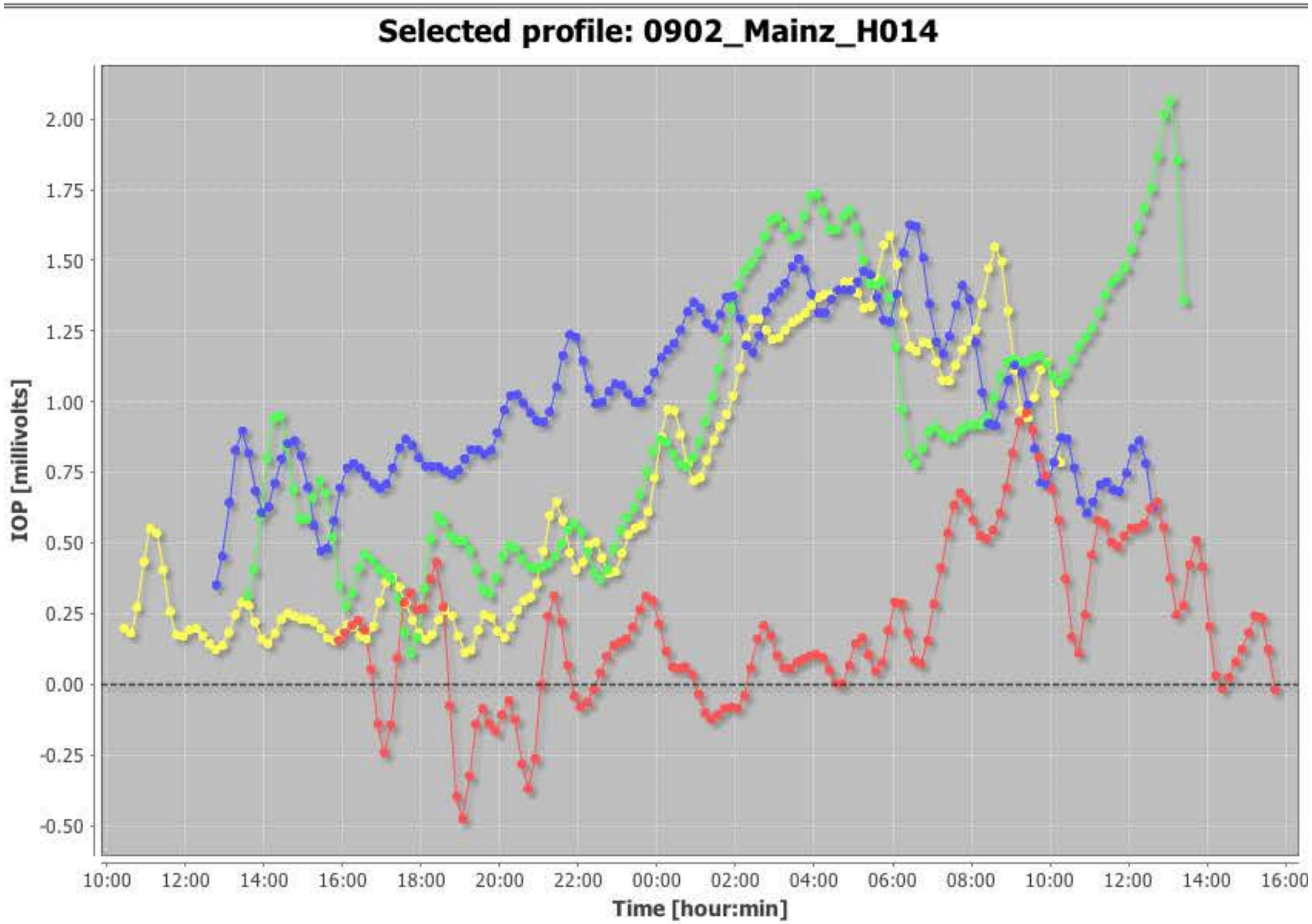
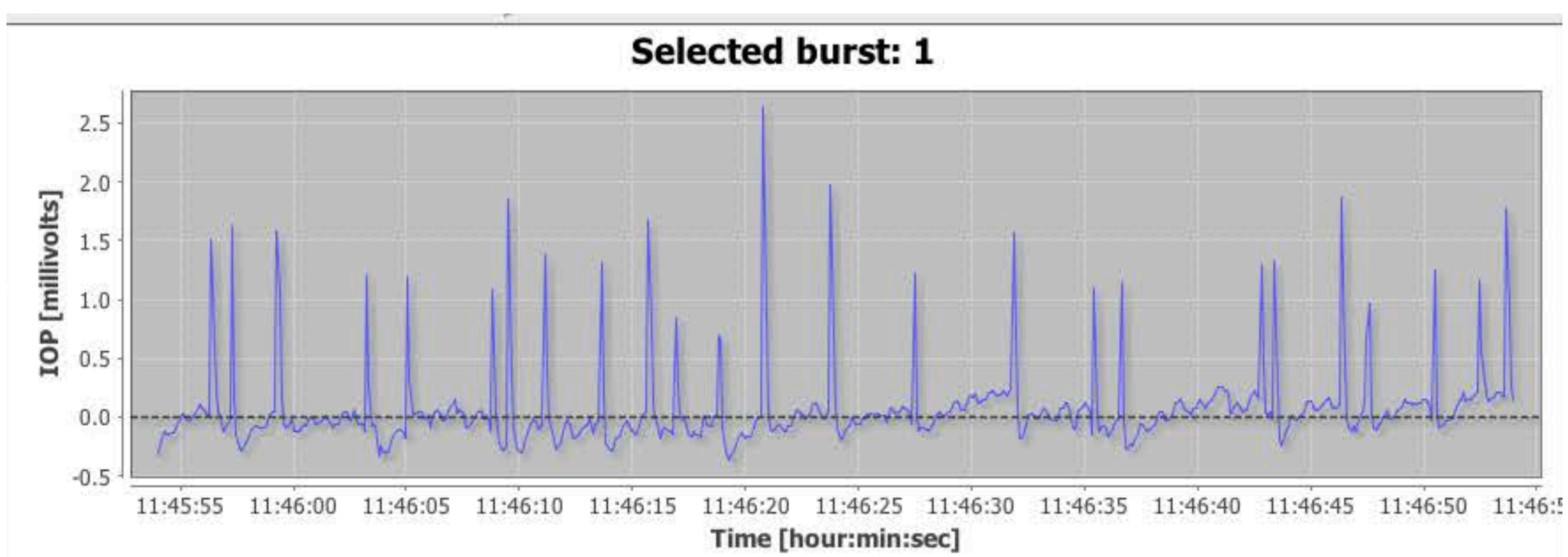


Figure 1
File Edit View Insert Tools Desktop Window Help
400
300
200
100
0
-100
-200
-300
0
50
100
150
200
250
300

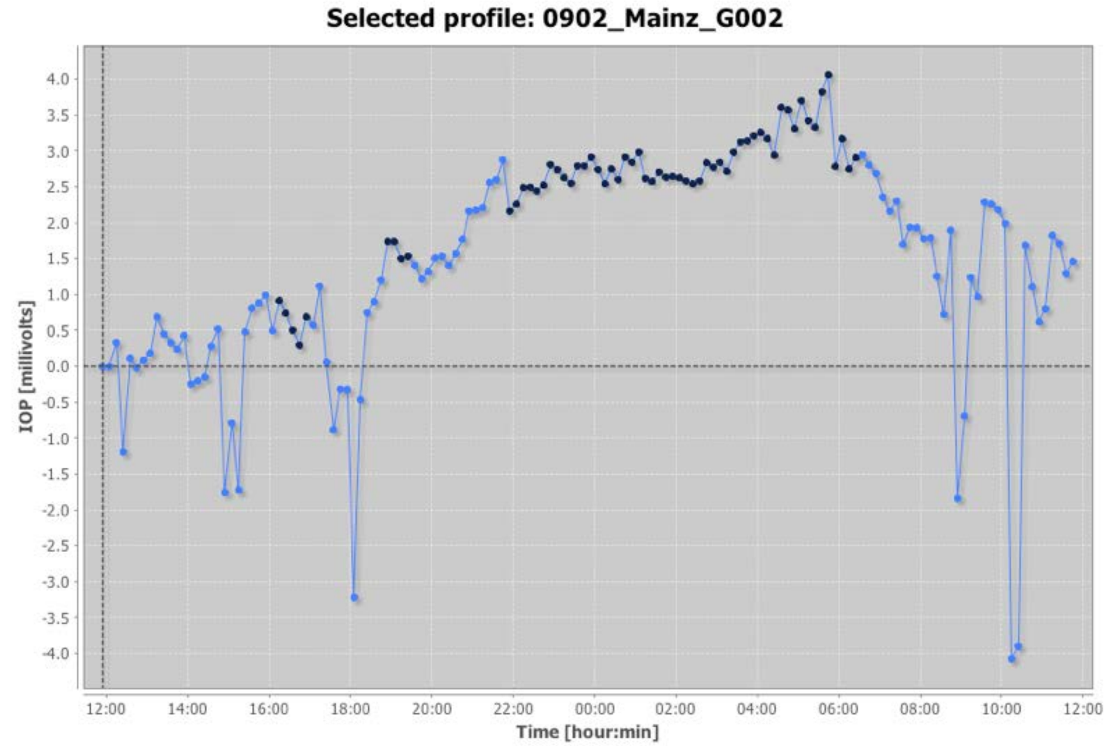


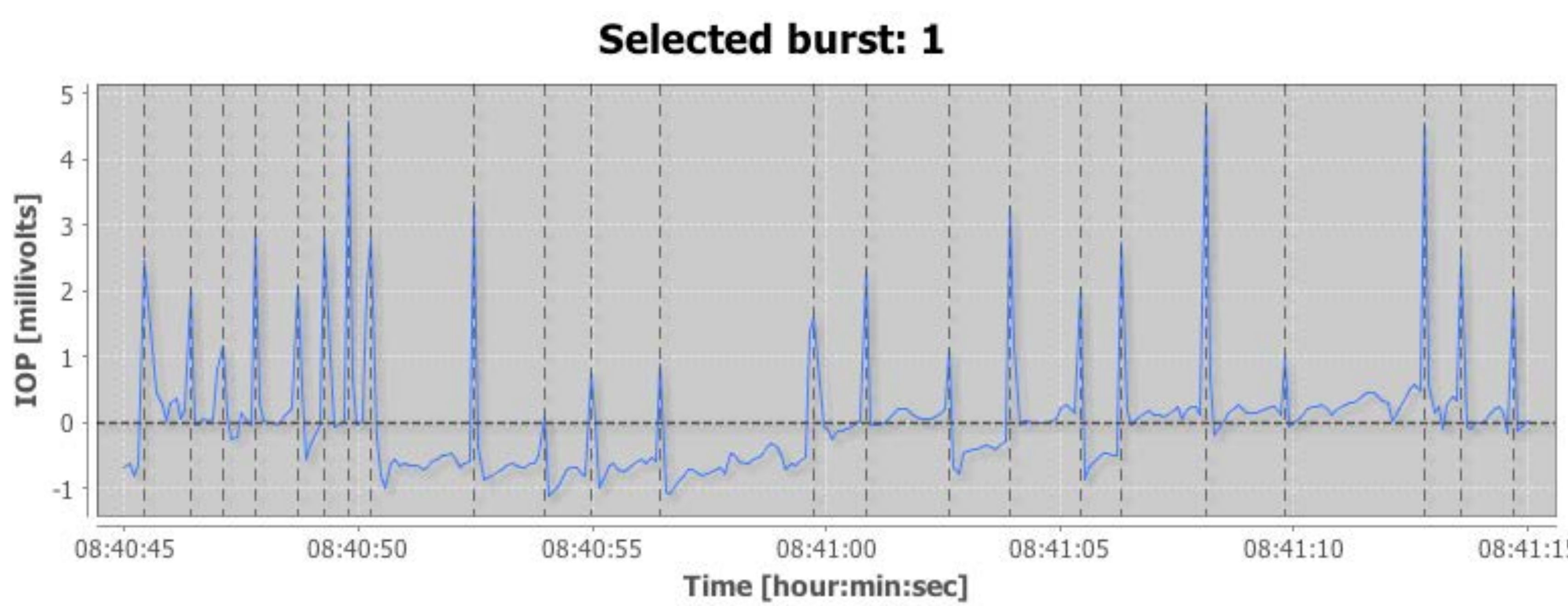
2 levels of knowledge



Sensimed Plus

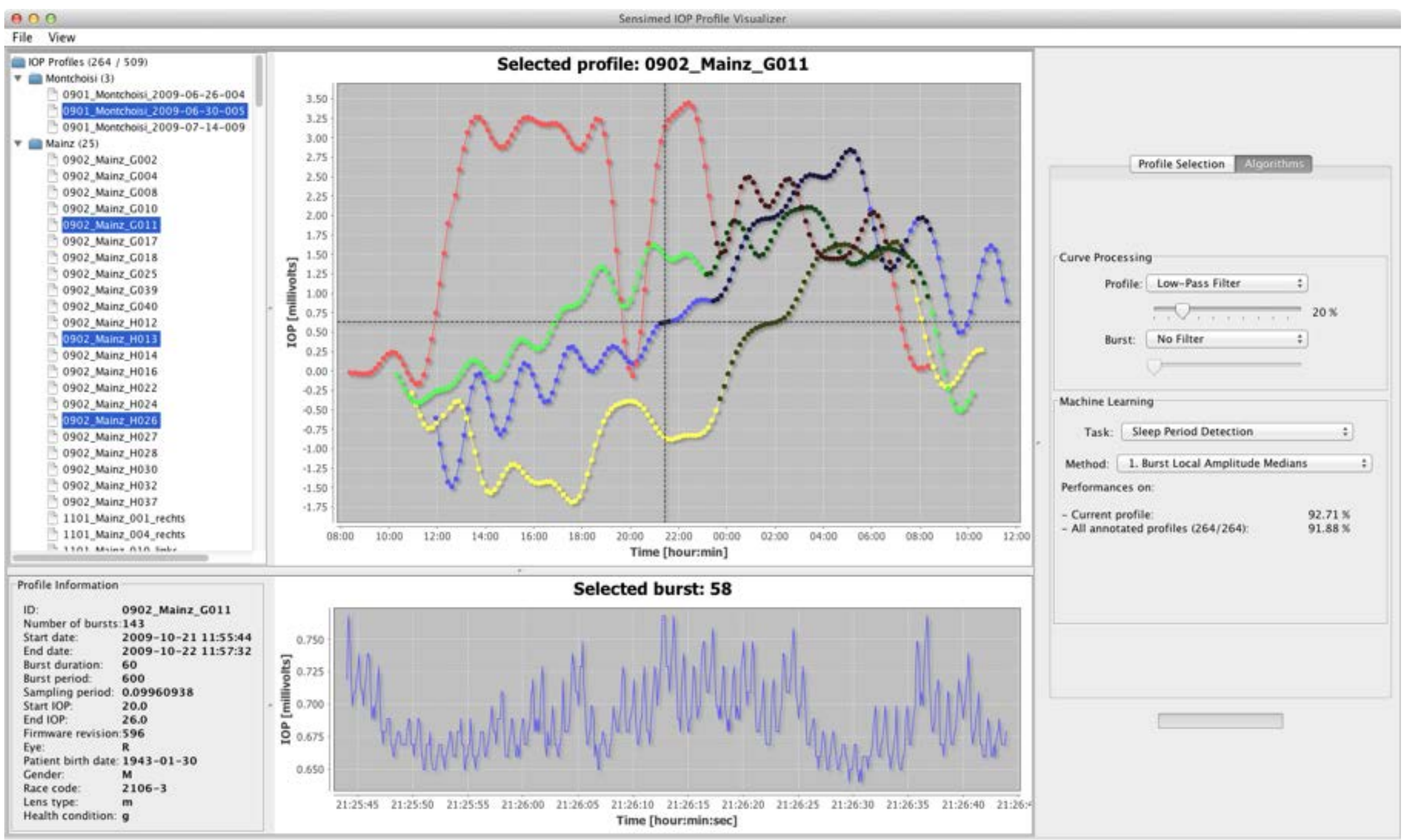
- **First level of knowledge = “sensor based knowledge”**
  - Blink detection
  - Awake - sleeping region detection
  - Occular pulse analysis





Sensimed Plus

- **Second level of knowledge = statistical analysis across populations and machine learning**
  - Glaucomatous vs healthy **classification**
  - **Clustering** of glaucoma
  - **Similarity search** across patients

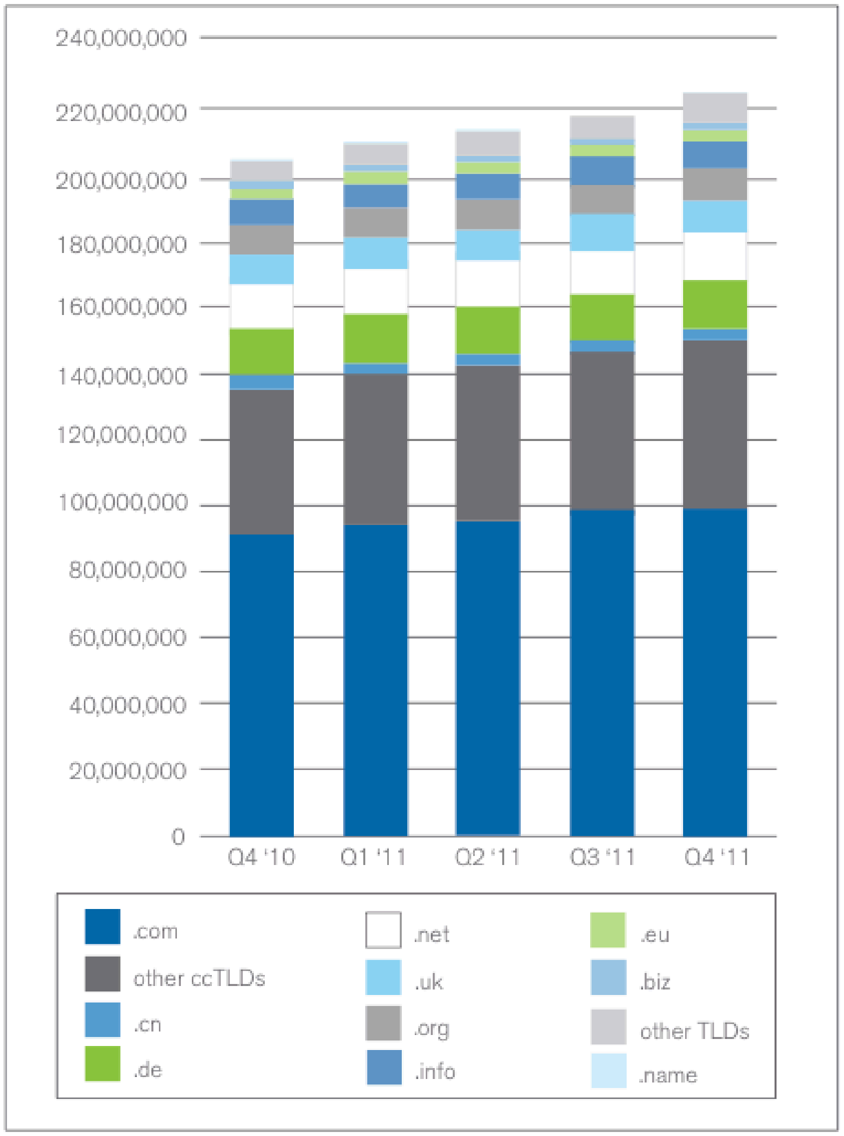


VeriName:
Advanced Domain Name Suggestion Services

- .com and .net
  - 113.8 million domains
  - 4% growth

- Objective: use machine learning to elaborate new name suggestion algorithms

Source: Zooknic, January 2012; Verisign, January 2012



Propose names that sound similar = new registrations

keywords: air mix

or domain name: airmix.com

suggestions	score
www.rmix.com	100%
www.airmeex.com	50%
www.airnix.com	45%
...	...

Catch probable phonetic spelling = brand protection

keywords: der Spiegel

or domain name: spiegel.de

spoken by	suggestions	score
DE	www.spiekel.com	85%
EN	www.speegel.com	75%
EN	www.speekl.com	20%
FR	www.chpigel.com	60%

Latinization of names = international markets

keywords: Лужниках

or domain name: -

spoken by	suggestions	score
RU	www.luznikah.com	95%
EN	www.luznika.com	85%
EN	www.loosnika.com	75%
FR	www.lusnica.com	80%

Apps › Plex It!

DomainDialect Service Testing Page

Phonetizer

Language: Choose a language...

Filter: 0.0

Acronym (g2p)

Acronym (g2g)

Fix missing phonemes

Grapheme

Graphetizer

Language: Choose a language...

Phonetize Graphetize

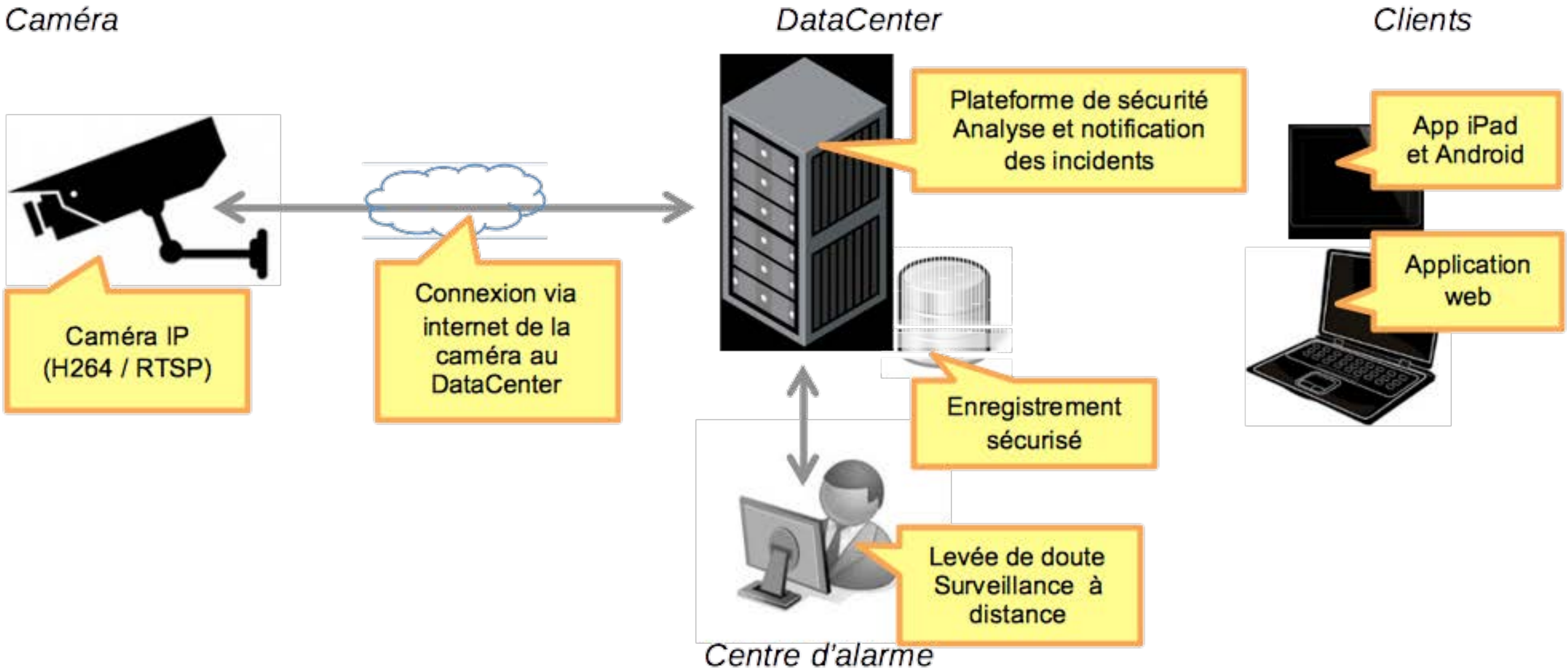
Phonemes

Phoneme	Score	Language

Graphemes

Grapheme	Phoneme	Score	Language

VideoProtector project



- ***From video surveillance to video protection***
- Using machine learning technologies to detect abnormal events in video streams.

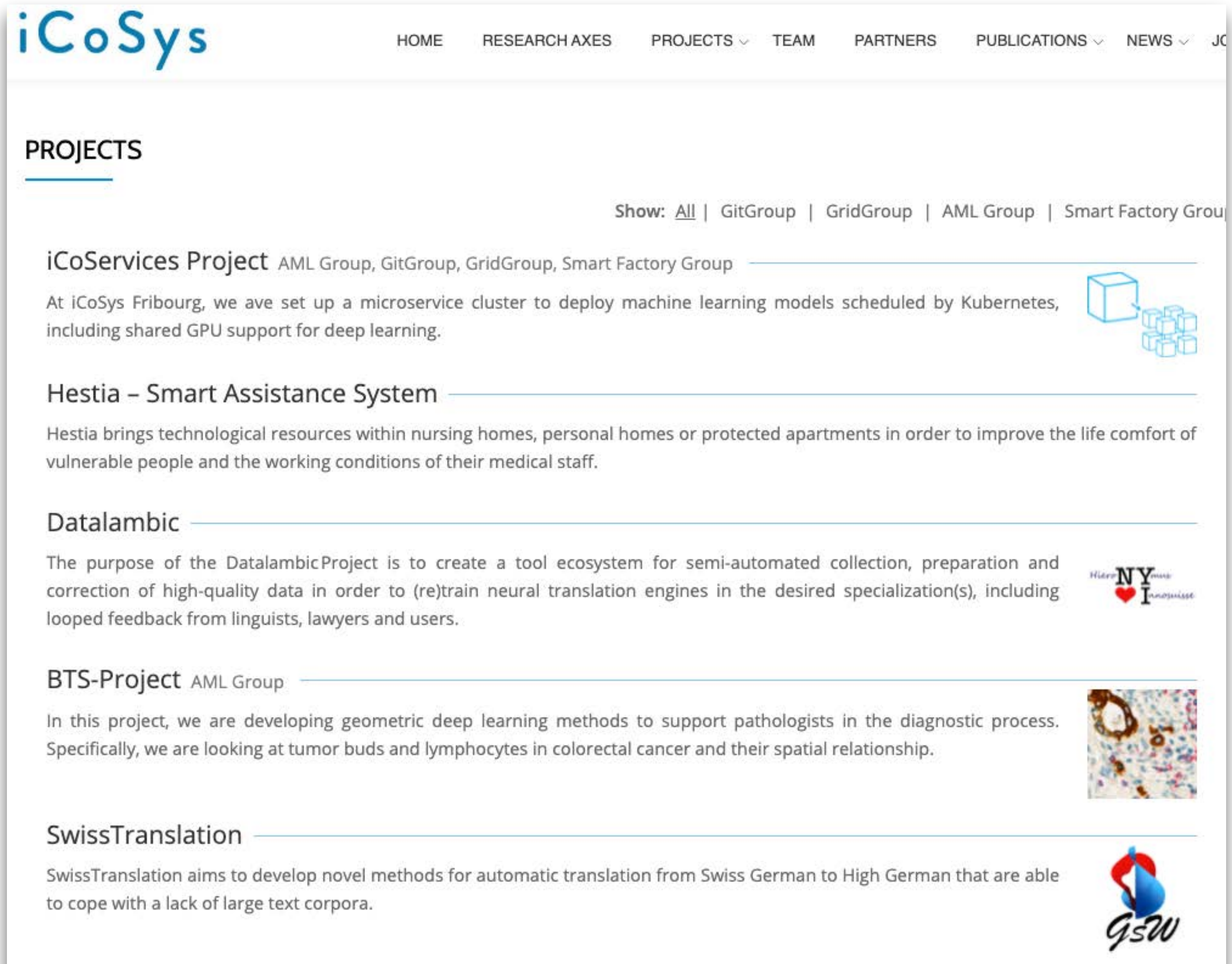
Fall Detection

Vehicule Classification

Ecole d'ingénieurs et d'architectes de Fribourg
Hochschule für Technik und Architektur Freiburg

Check out other cool projects on

- https://icosys.ch/projects



1.2 Definitions

- Learning
- Machine learning
    - 1st formal definition
    - 2nd formal definition
    - Our *ultimate* definition

Machine learning is at the convergence of

Mathematics
probability theory, statistics, regression, ...

Signal processing
filtering, feature extraction, time series, fft, ...

Algorithmics

Machine Learning

Software Engineering
very large DB, intensive cpu, distributed programming, ...

Application domain
finance, medicine, energy, biometrics, ...

Preliminary question: what is learning?

> **learning** | ˈlərniNG |
> noun
>
> the acquisition of knowledge or skills through experience, study, or by being taught: *these children experienced difficulties in learning* | *[as modifier]* : *an important learning process.*

- A definition closer to machines could be:
  - “Using past **experiences** to **improve** future **performance**”
  - An *experience* will be in the form of numeric data
  - The **performance** will always be associated to a **quantified** objective such as a loss to be minimised *or a gain to be maximised*

First formal definition

Machine learning - Field of study that gives computers the ability to learn without being explicitly programmed. Arthur Samuel (circa 1959)

- The Samuel Checkers-playing Program appears to be one of the world's first self-learning program, and as such a very early demonstration of the fundamental concept of artificial intelligence (AI).

Second formal definition

Machine learning - A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
Tom Mitchell (1998) — http://www.cs.cmu.edu/~tom/

- As we will see later on, the performance measure is a key concept to:
  - derive the mathematical equations used for learning
  - tune the system and find the best configurations for a given problem

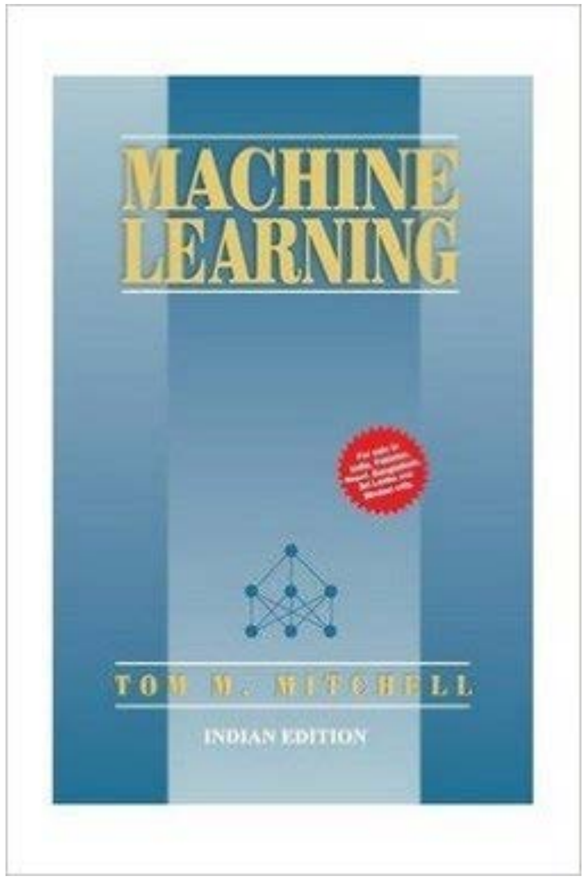


Tableau 1-1. Quelques exemples de machine learning proposés par Mitchell

Cas d'application	Checkers learning	Handwriting recognition	Robot driving learning
Tasks T	Playing checkers	Recognizing and classifying handwritten words within images	Driving on public four-lane highways using vision sensors
Performance measure P	Percent of games won against opponents	Percent of words correctly classified	Average distance traveled before an error (as judged by human overseer)
Training experience E	Playing practice games againts itself	A database of handwritten words with given classifications	A sequence of images and steering commands recorded while observing a human driver

Source: Data Science - fondamentaux et études de cas, Michel Lutz et Eric Bernât, Eyrolles

Our definition

Machine learning - set of computer methods that analyse observation data to automatically detect patterns, and then use the uncovered patterns to perform functions on new un-observed data. Examples of functions include : prediction, classification, clustering and more generally decision making

- Machine learning is usually divided into two main types:
  - **Supervised learning**
  - **Unsupervised learning**

- Other types:
  - Reinforcement learning
  - Semi-supervised learning
  - Recommender systems

1.3 Supervised Learning

Examples

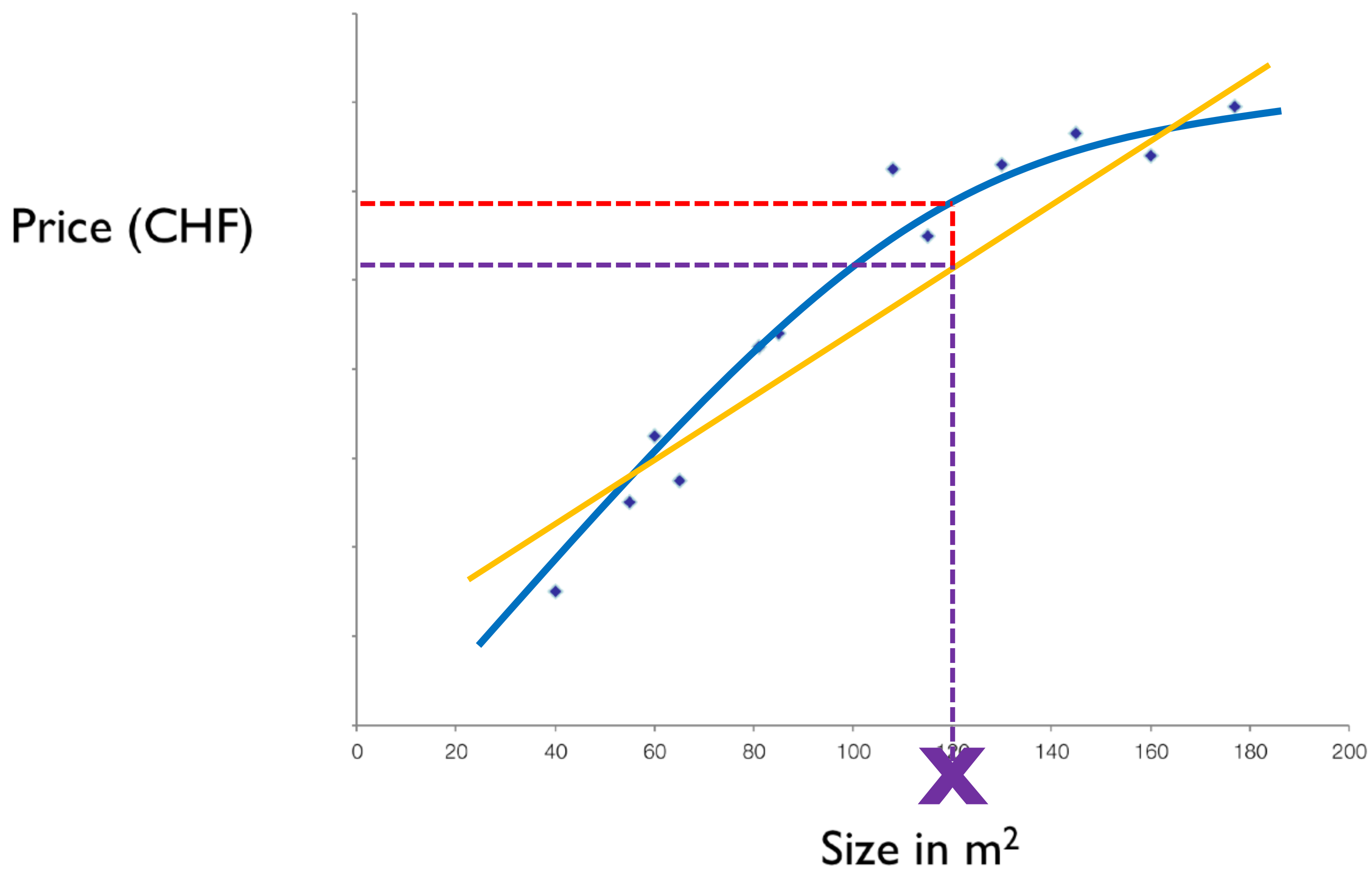
Definitions

Notations

Classification tasks

Regression tasks

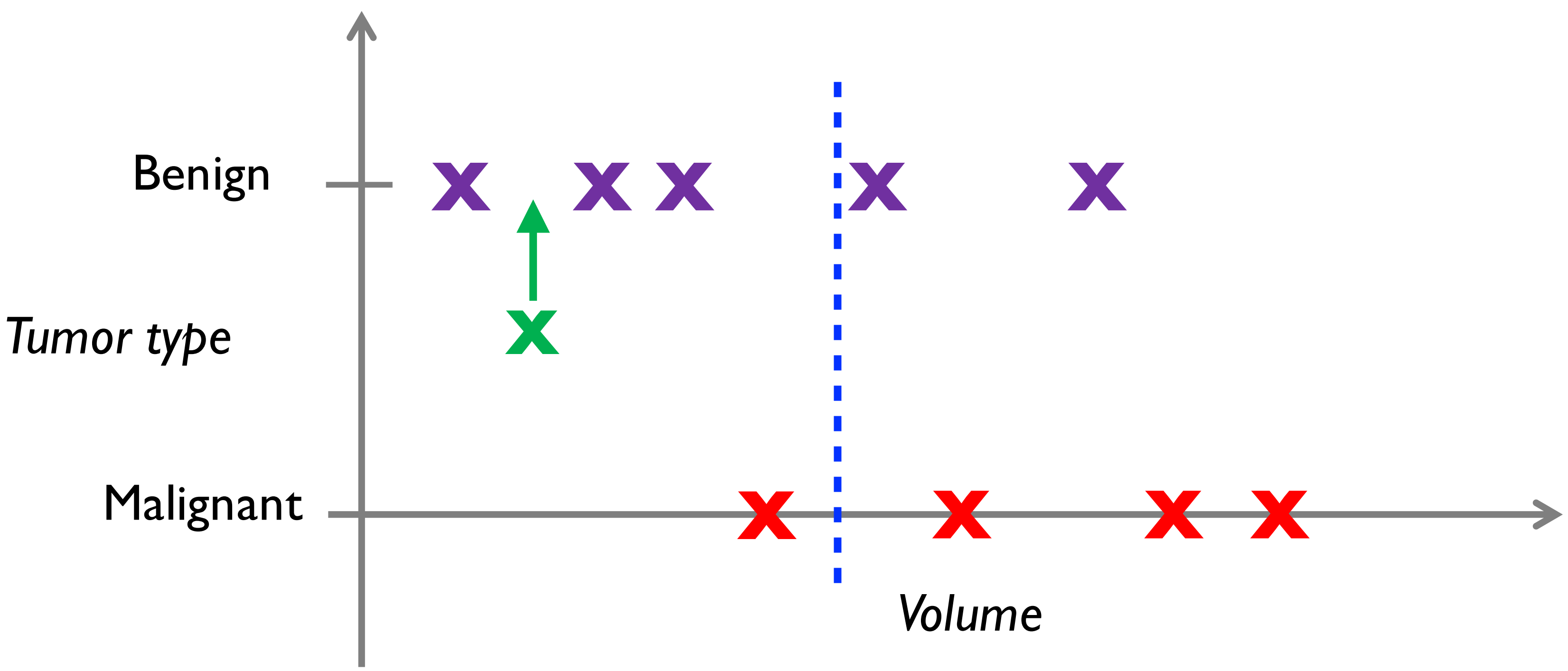
Supervised learning example 1 - house price prediction



The output of the system is a continuous value: this is called a **regression**

- We want to **predict** the price from the size.
- We have a set of price values with the *right answers*
- We build a mathematical model explaining the mapping: *price = h(size)*
  - yellow: linear model
  - blue: non linear model
- We can use the model to predict new house prices

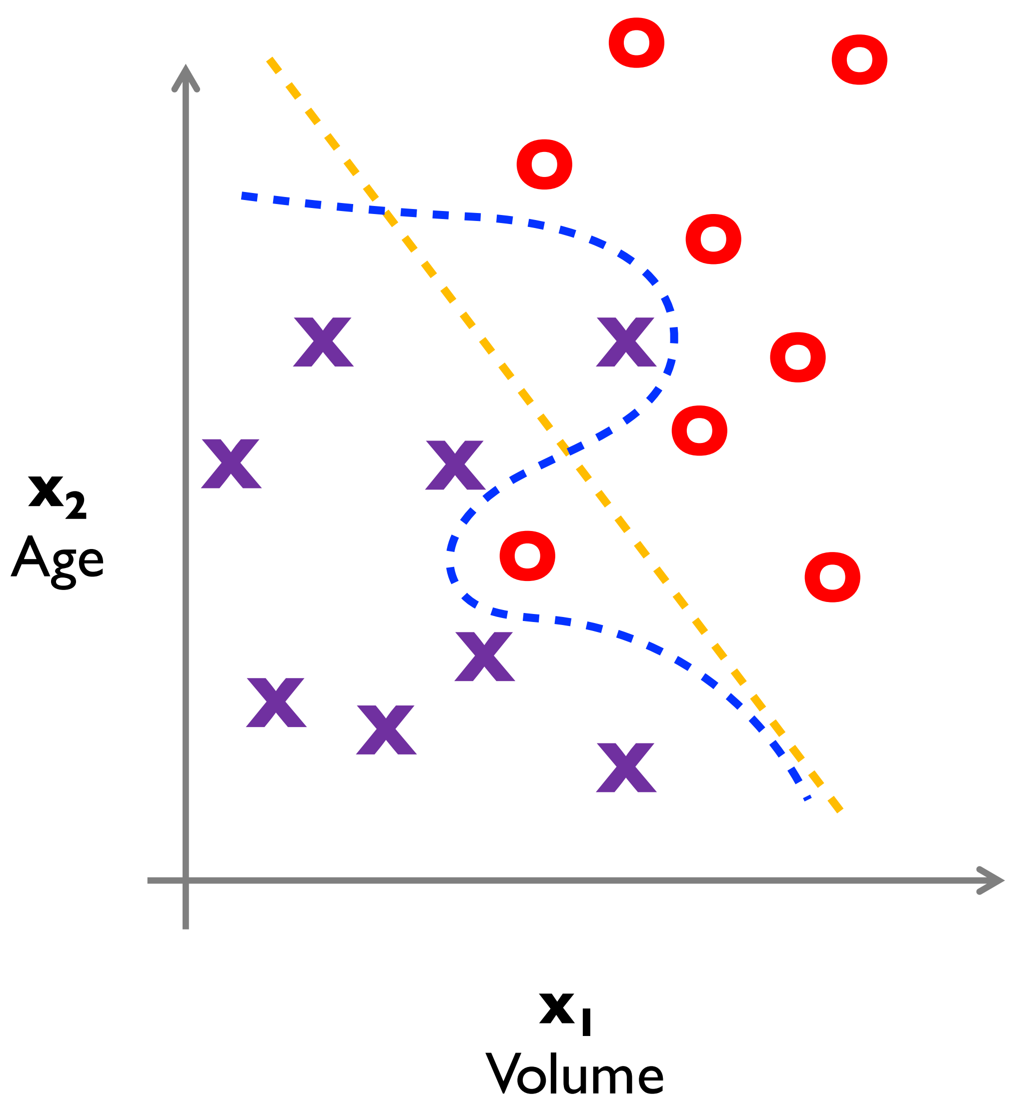
Supervised learning example 2 - tumour classification



- We want to **categorise** a tumour into two classes: B or M
- We have a set of examples with the *right answers*
- We build a mathematical model explaining the mapping: *class = h(volume)*
  - In this case a simple threshold
- We can use the model to **classify** a new tumour

The output of the system is discrete: this is called a **classification**

Supervised learning example 2 - tumour classification



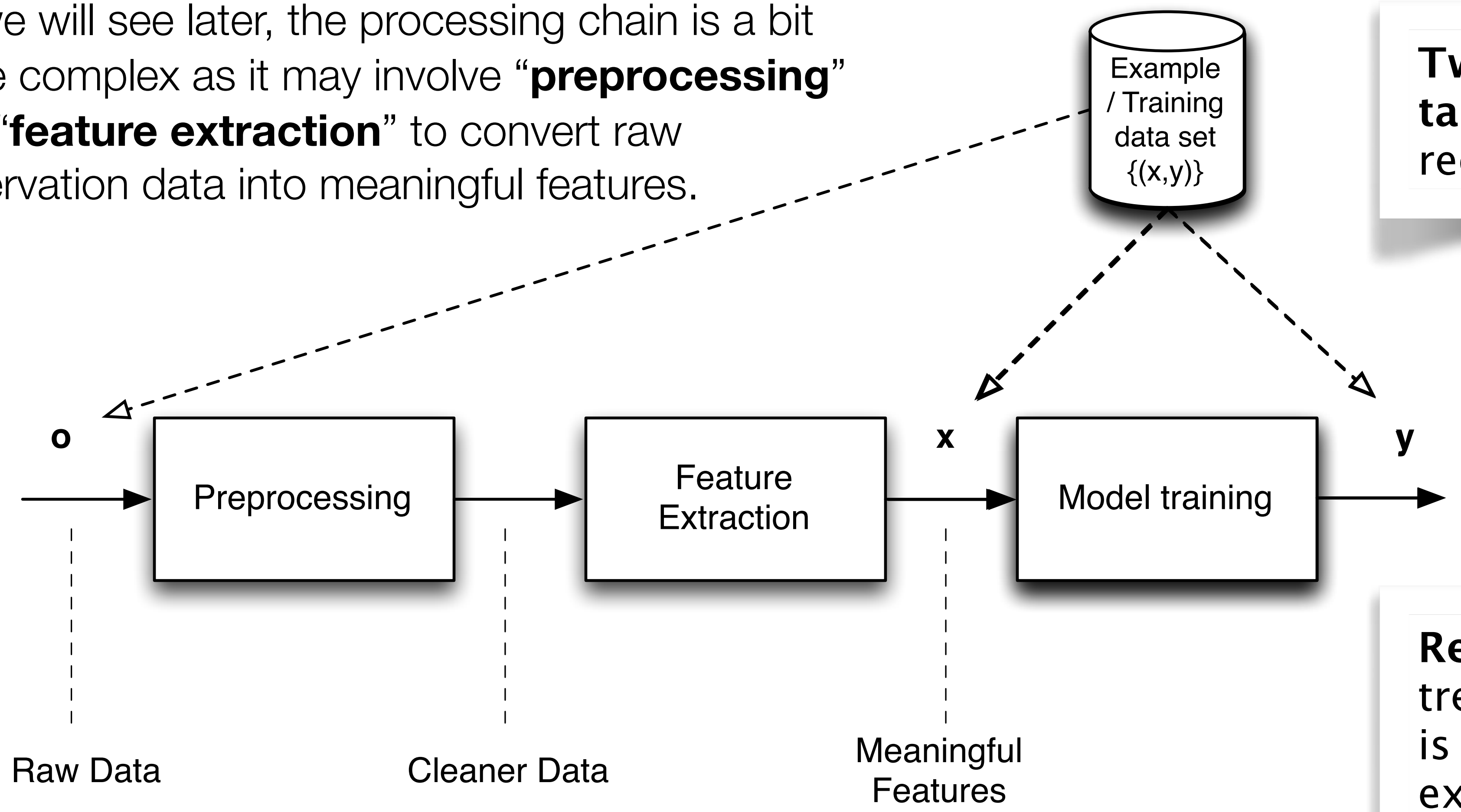
- A typical way to increase the performance is to take into account more **features**
  - the input is becoming a vector with 2 components in this example
- The mapping function is *class = h(x1,x2)*
  - in yellow a linear model
  - in blue a non-linear model
- What other features could we bring in?
  - smoking quantity, family history, …

Tip: adding relevant features may increase performance. However this usually requires to have more data. See curse of dimensionality https://en.wikipedia.org/wiki/Curse_of_dimensionality

Supervised learning

With **supervised learning**, the goal is to learn a mapping from inputs **x** to outputs **y** given a set of example data called the training set.

As we will see later, the processing chain is a bit more complex as it may involve "**preprocessing**" and "**feature extraction**" to convert raw observation data into meaningful features.



Two main types of tasks: classification or regression

Remark: the recent trend of deep learning is to train the feature extraction step

Training - test set

The **training set** is the set of data used for the learning process, i.e. to *train* the system.

- Usually big — *more data better data*
- Should be collected in the same conditions as when really using the system
- A part of the training set, called the cross-validation set is usually put apart to *tune* the system. We will come back to this in few weeks

The **test set** is an *independent* set of data used to evaluate the performance of the system

Some notations

$\mathbf{x}$	input variable of the system, also called **input feature**
y	output variable of the system, also called **target output**
$\widehat{y}$	**gotten output**, as computed by the mapping function
$(\mathbf{x}, y)$	a **training sample** (pair of feature and corresponding target)
$(\mathbf{x}_n, y_n)$	the training sample at index n in the training set ("in row" n)
N	the total number of training samples in the training set

The **training set** can then be noted

$$\{(\mathbf{x}_n, y_n); n = 1, \ldots, N\}$$

See https://moodle.msengineering.ch/pluginfile.php/147625/mod_resource/content/6/ML-notations.pdf

Mapping function

- We make the assumption that there is an unknown function *f(**x**)* that we want to "learn" from the set of training examples given in the training set.
- Our goal is to approximate this real function *f* by a function *h* that we learn on the data set.

$$\widehat{y} = h(\mathbf{x})$$

- The "hat" on the y value is to indicate that the system will be able to compute approximations of the target variable from new unseen ***x*** values.
- The objective of the learning is actually to make the $\widehat{y}$ values "tend" or "converge" in average to the real y values of the training set.

Remark: "*h*" actually means hypothesis — this is to emphasise the main principle of machine learning (against statistics) which is to let the machine explore a very large set of *hypothesis* functions and determine itself the one that best fits to the problem.

Supervised learning - classification tasks

A **classification task** maps inputs **x** to a finite set of discrete outputs **y**. The outputs are the class labels corresponding to the different categories we want to predict.

- Usually classes are **mutually exclusive**, i.e. only one label is output of the system.
  - However, some systems are said **multi-label** when a given input x belongs to more than one class.
- 2-class systems are sometimes called **detection** or **verification** systems, where the objective is to answer yes/no questions
  - Biometric example: *Is this the face of Sheldon? Is the identity verified?*
  - Warship example: *Is a torpedo going to hit us? Is a torpedo detected?*

Supervised learning - regression tasks

A **regression task** maps inputs **x** to an infinite set of continuous outputs **y**. The outputs are numeric values corresponding to the variable we want to predict.

- x are sometimes said to be "**explicative variables**"
- y are sometimes said to be the "**response variables**" or "dependent variables"
- in $\widehat{y} = h(\mathbf{x})$
  - if **x** is a scalar (vector of dimension 1), it is a **monovariate regression**, i.e. the response is influenced only by one other factor
  - if **x** is a vector (of dimension D), it is a **multivariate regression** i.e. the response is influenced only by multiple factors

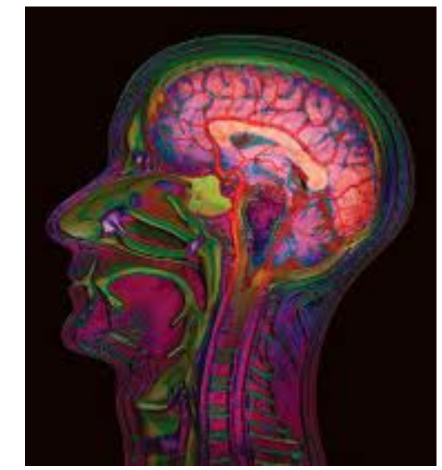
Transforming regression into classification

- It is possible to transform a regression into classification by associating ranges of prediction into classes
- Examples
  - 2 classes: is the stock value going **up** or **down**
  - 5 classes: is the age of the web shop visitor below 20, between 20 and 30, between 30 and 40, between 40 and 50, above 50
- Ultimately, by considering an infinite number of classes, we fall back into a regression
  - Generally, transforming a regression into classification makes the system simpler to tune.

Supervised learning - application domains

Recognition

Planning



Diagnosis

Robot Control

Prediction

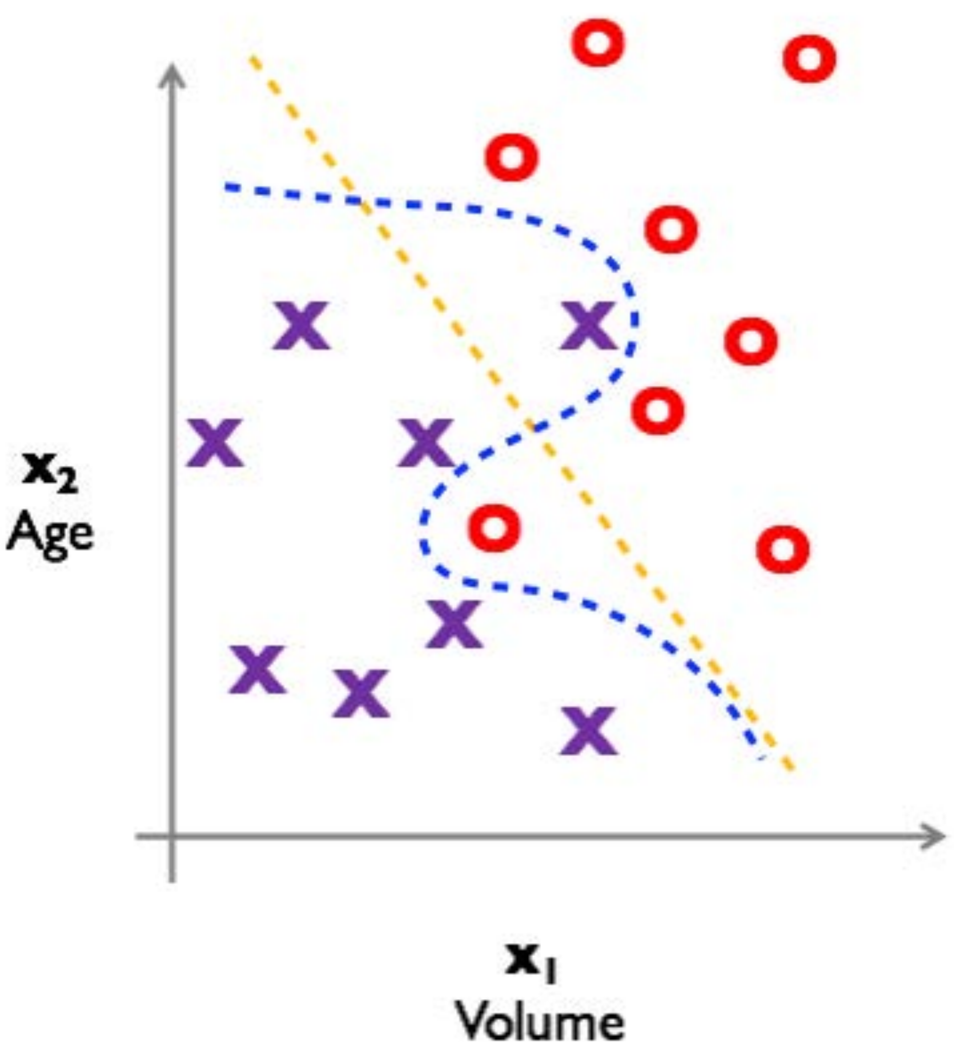
1.4 Unsupervised learning

Examples

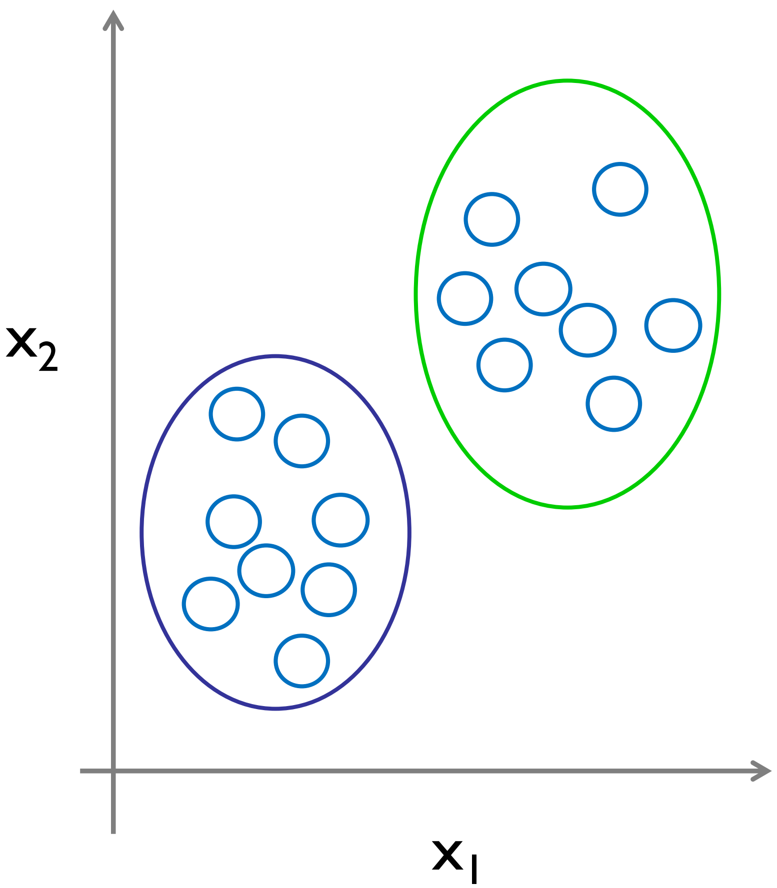
Definition

Application domains

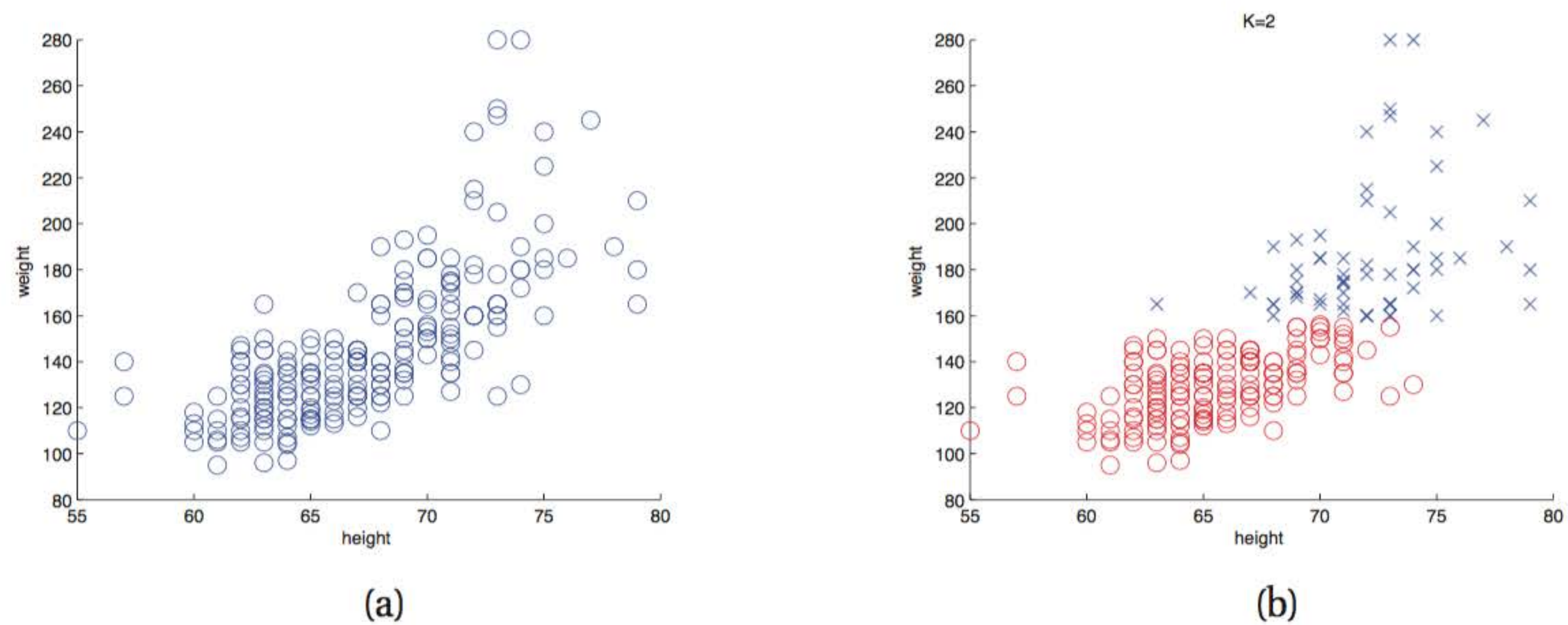
Supervised learning



Unsupervised learning



Example of clustering



Source: K. Murphy, "Machine Learning – A Probabilistic Perspective", 2012

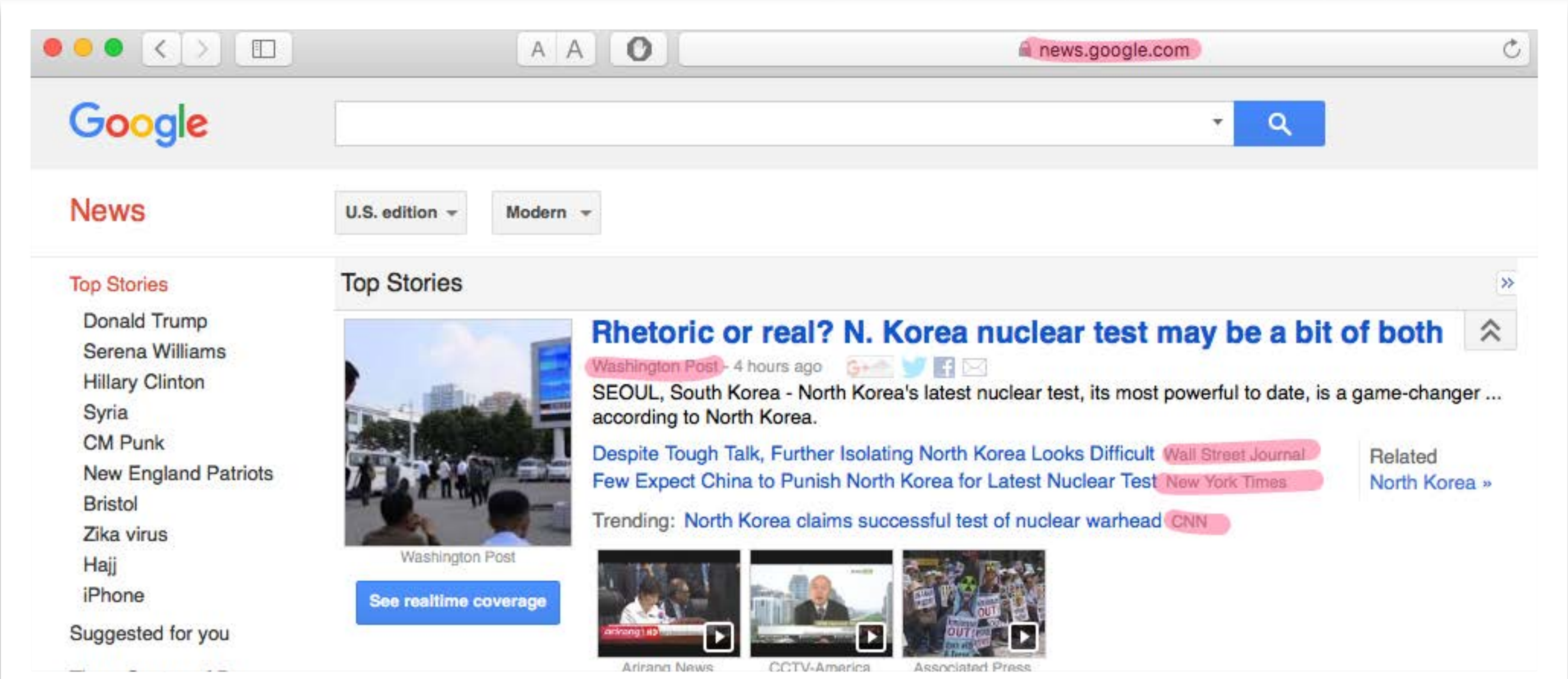
Figure 1.8 (a) The height and weight of some people. (b) A possible clustering using $K = 2$ clusters.

- A typical problem with unsupervised learning is to discover structure in the data, such as the presence of clusters. Once the clusters discovered, we may attempt to given them a signification.

Example of text clustering

Activity

- Analyse google news. How can they “group” similar news articles?
- Is it unsupervised learning? Under which conditions?
- How would you build such a system?
- How can we evaluate such a system?



Unsupervised learning

With **unsupervised learning**, the goal is to discover interesting structures from inputs **x** given a set of data called the training set.

- Typical tasks include:
  - Discovering **clusters**
  - Performing **dimensionality reduction**
  - Discovering **graph structures**
  - Performing **matrix completion** — inferring plausible values for missing entries
    - Collaborative filtering
    - Image inpainting
    - Market basket analysis

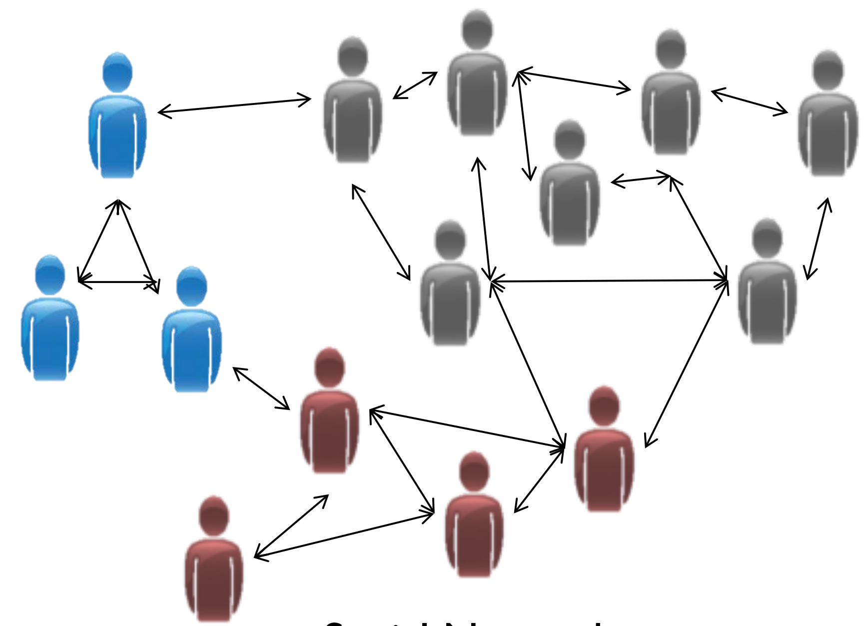
users

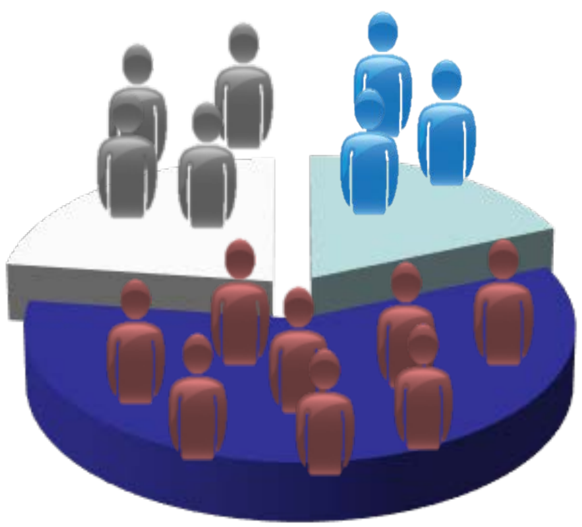
movies						
	1		?	3	5	?
	?	1				2
		4		4	5	?

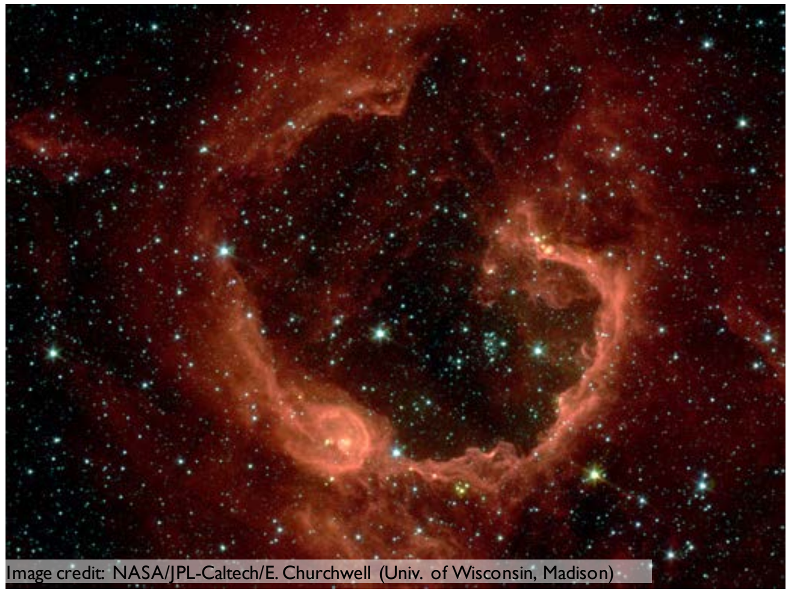
Remark: A family of recommender systems can be trained with matrix completion algorithms which can be qualified as *unsupervised*. However recommender systems can also be trained in supervised ways. See for example http://netflixprize.com

Unsupervised learning - application domains


Organize computing clusters


Social Network


Market segmentation



Astronomical data analysis

Conclusions

- Machine learning regroups a set of computer methods that **analyse automatically data to detect patterns**, in order to **perform functions on new un-observed data**.
- Supervised learning = learn to produce an output when given an input vector
  - Two flavors:
    - Regression: the target output is a real number or a whole vector of real numbers.
    - Classification: the target output is a class label
- Unsupervised learning = discover a good internal representation of the input from which we can make some sense