Introduction to ML

What is Machine Learning?

- Using algorithms to build models from a large amount of examples (training) data.
- Area at the intersection of several scientific domains
  - statistics
  - optimization
  - computer science
  - image/signal processing

What is Machine Learning?

Some examples of ML tasks

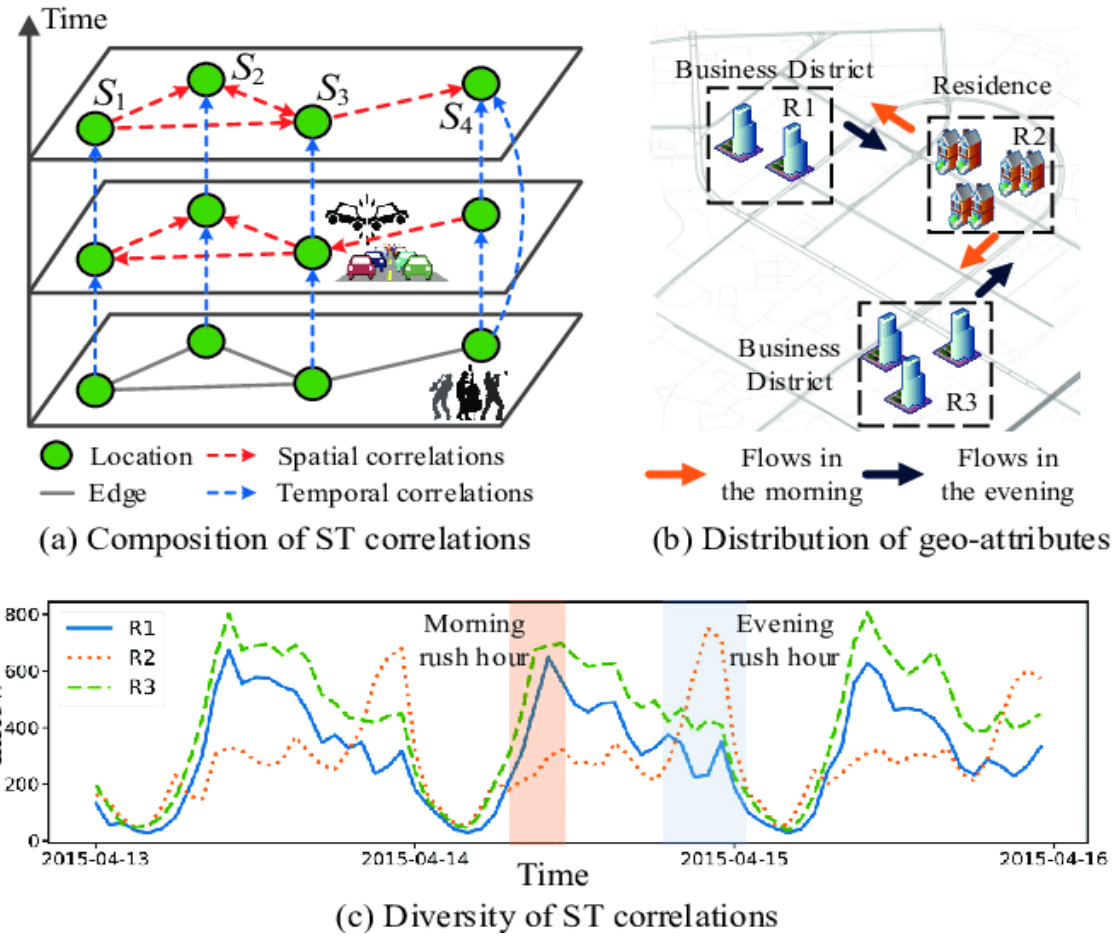
Sentiment analysis

Sentiment	Tweet mention
Positive	Maybe I'm mad but I'm now the proud owner of a potentially #bendy #iPhone6, it's so much bigger than the #4s
	Finally got to see an iPhone 6 today. Not revolutionary at all but it's absolutely gorgeous. (And I want one). #iPhone6
Negative	I'm not sure I want it. It's too big to fit in my back pocket! lol #iphone6
	I'm really disappointed with the #iPhone6. It took them 2 years to change the screen & size. Let down.

What is Machine Learning?

Some examples of ML tasks

Urban traffic prediction



(a) Composition of ST correlations

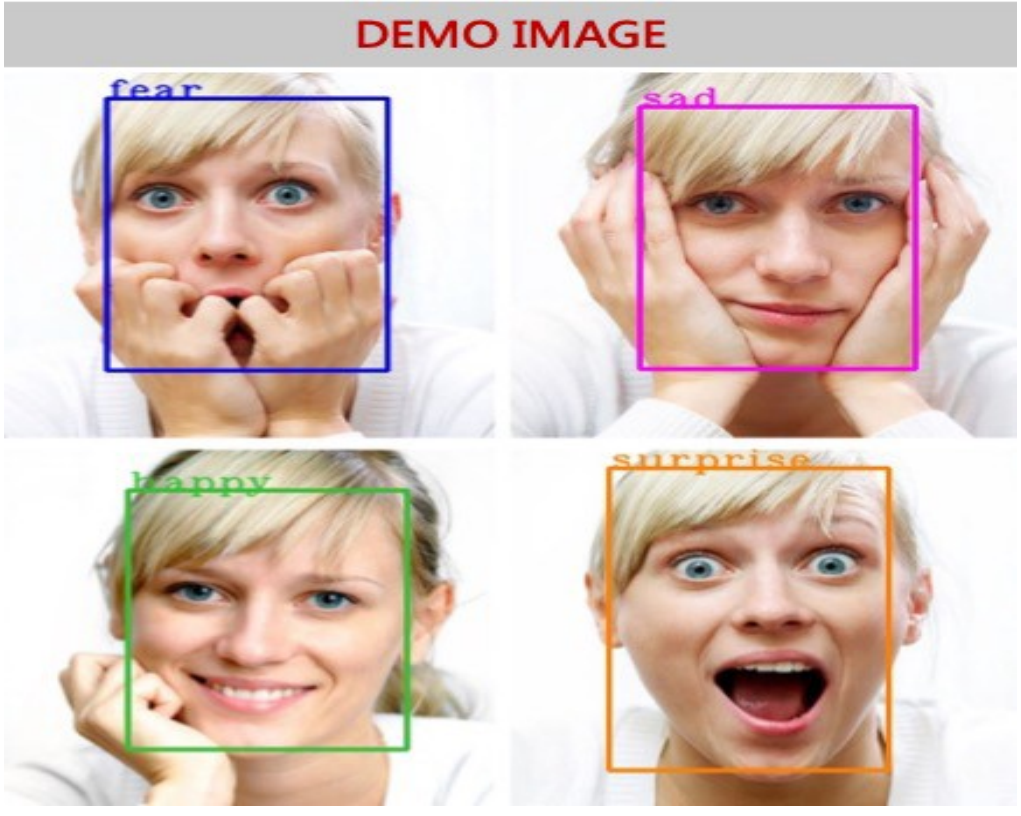
(b) Distribution of geo-attributes

(c) Diversity of ST correlations

What is Machine Learning?

Some examples of ML tasks

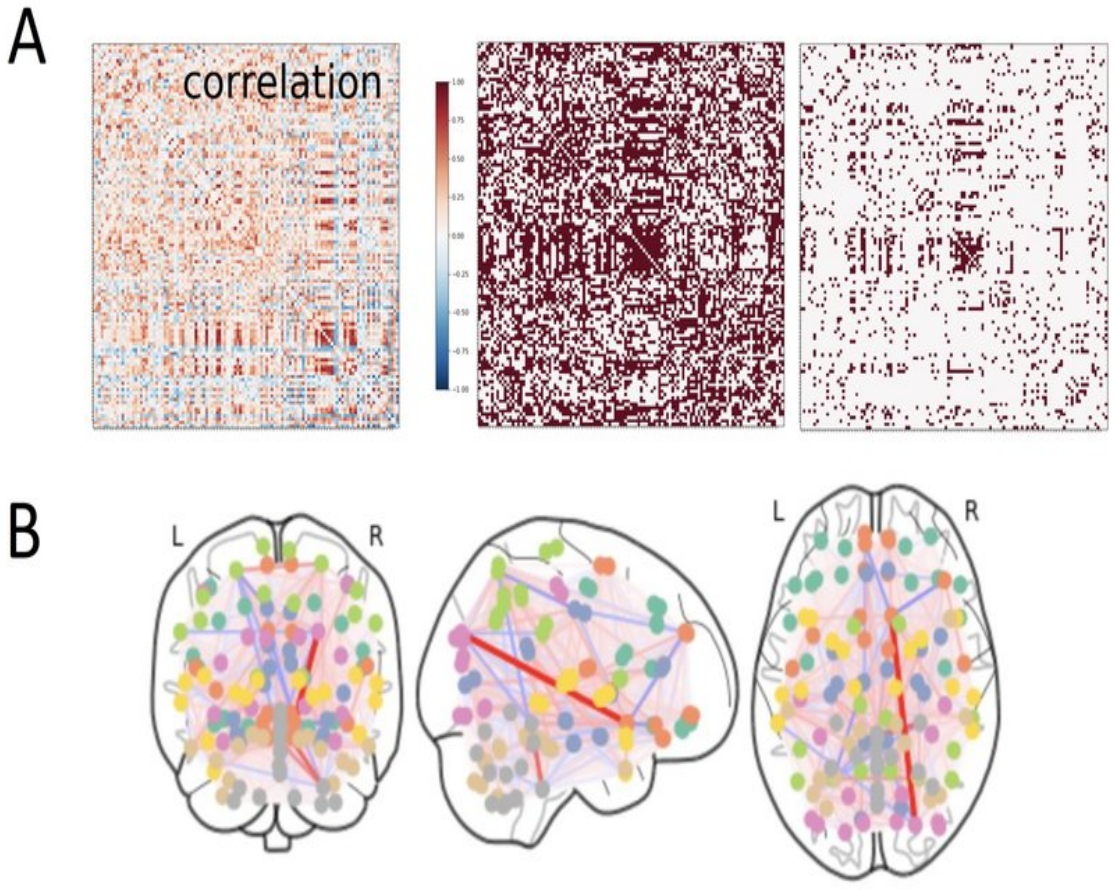
Face recognition and emotion detection



What is Machine Learning?

Some examples of ML tasks

Identification of fMRI patterns related to specific tasks



What is Machine Learning?

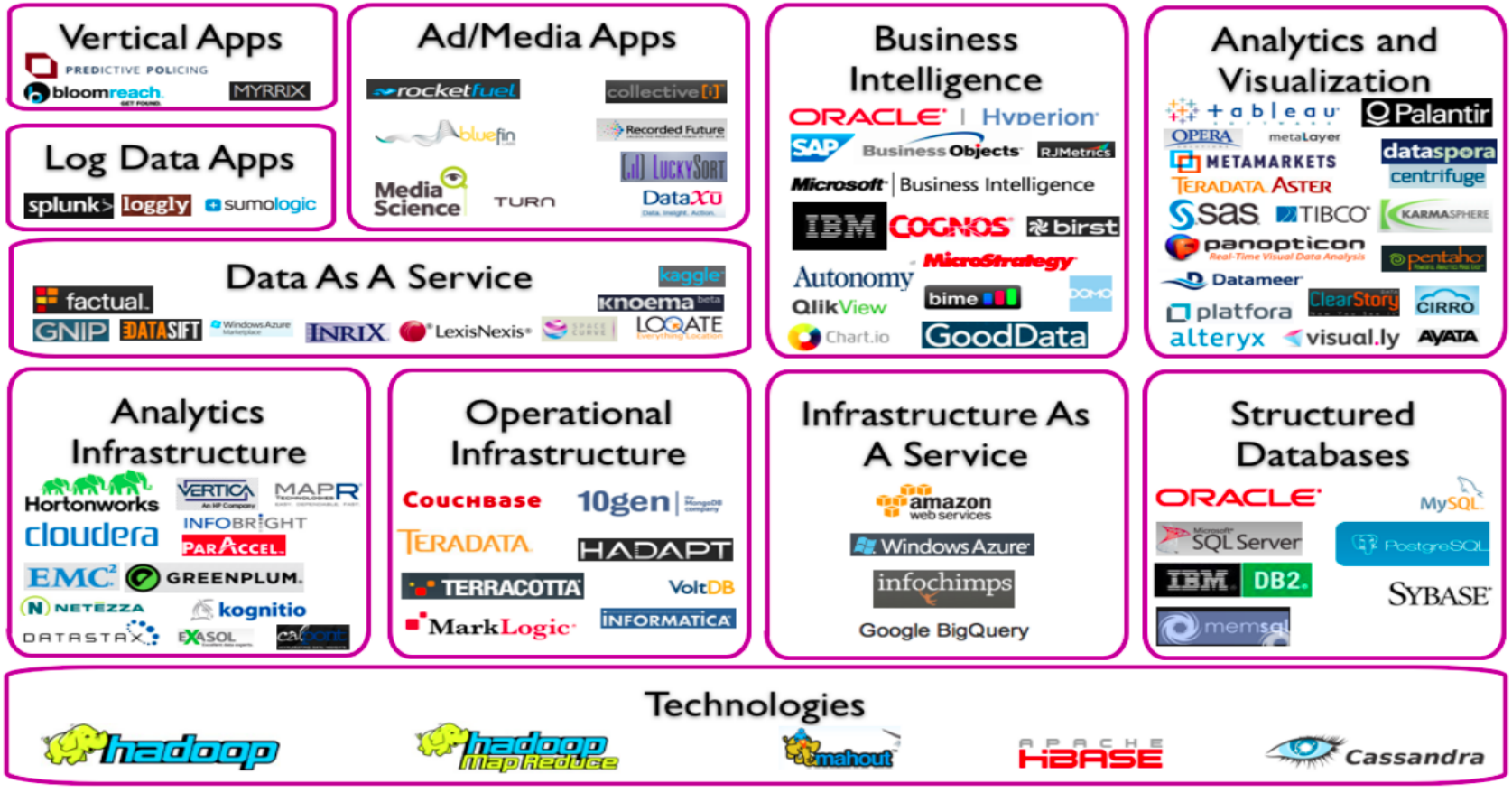
Challenges

- Deal with large scale data ?
  - Data that does not fit in RAM;
  - Data streams;
  - Algorithms do not run in a reasonable time on a single machine.
- Deal with complex data : unstructured textual data, time series, graphs

What is Machine Learning?

A large range of applications

A rich ecosystem

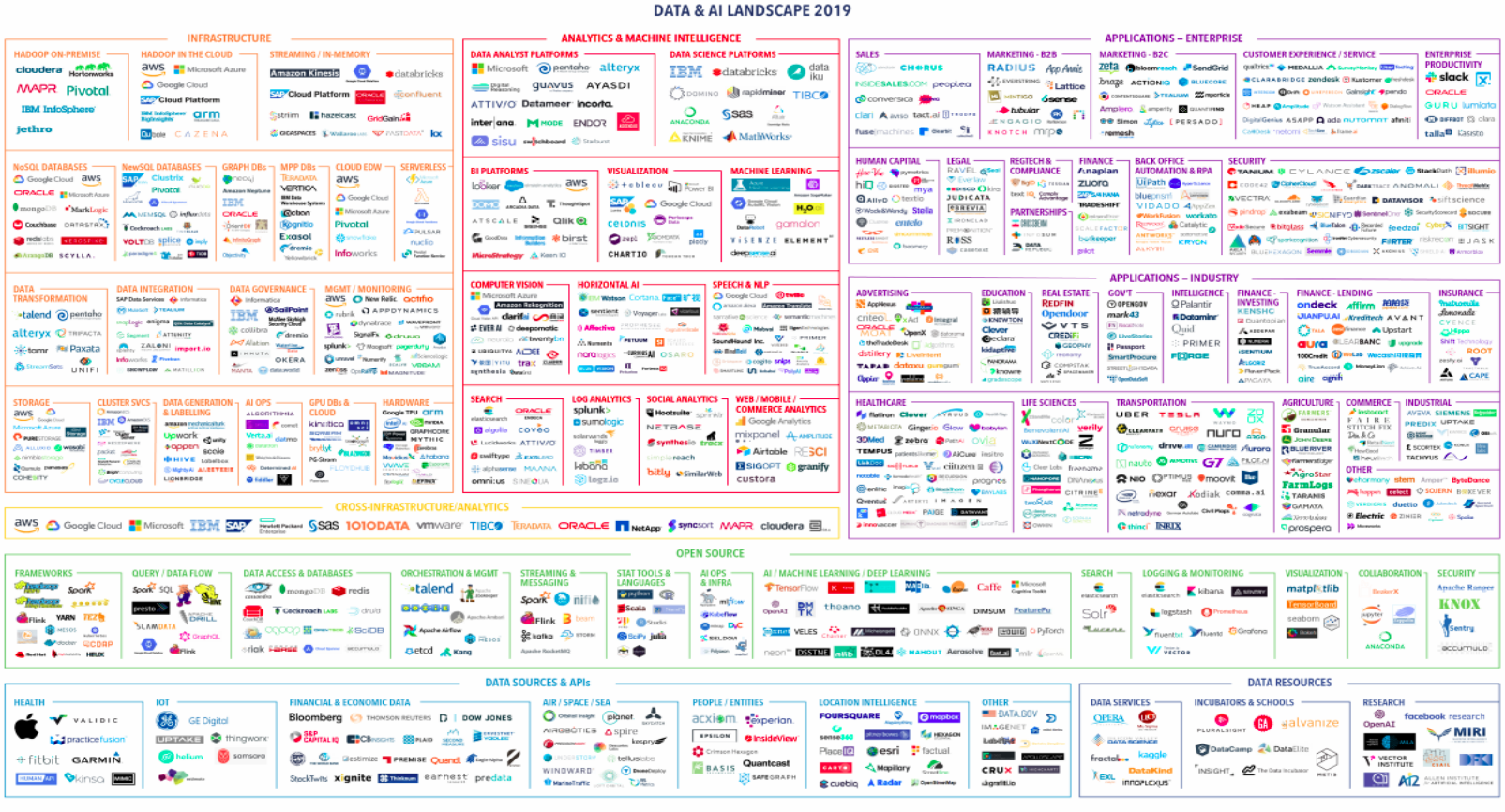


dave@vcdave.com
blogs.forbes.com/davefeinleib

What is Machine Learning?

A large range of applications

A rich ecosystem



A brief review of ML problems

Unsupervised learning

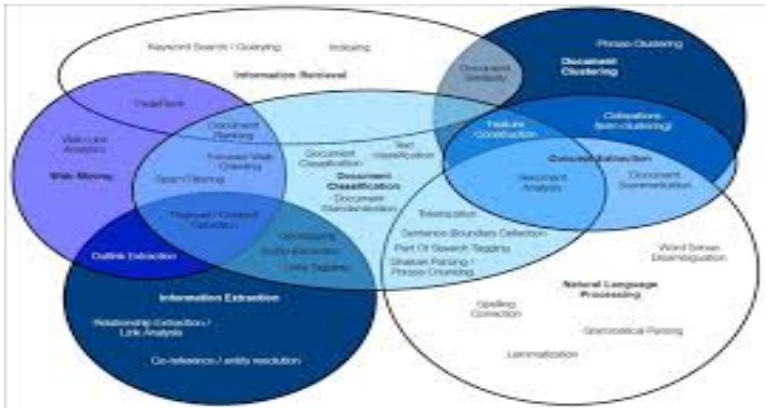
Aim : learn the structure of the data without no pre-existing labels to discover patterns

Some class of unsupervised problems

- Clustering : find groups in the data
- Dimension reduction : how can we represent the data in a lower dimension space, either in an automatic manner, either using a priori knowledge (for e.g. dictionaries ?

A brief review of ML problems

		1	2	3	4	5	6	7	.....	
1	Apple	0.9898	0.7865	0.5645	0.7509	0.4534	0.5467	0.6498	0.7613	0.8
2	Banana	0.4533	0.8644	0.1538	0.4313	0.3511	0.2422	0.2422	0.3553	0.2
3	Cat	0.8734	0.8363	0.4821	0.1378	0.2341	0.2122	0.6775	0.3432	0.
4	Dog	0.9873	0.4836	0.1342	0.19564	0.2131	0.3433	0.2244	0.7453	0.5
5	Eag	0.9473	0.4836	0.4343	0.9211	0.1221	0.4634	0.7464	0.2424	0.5
6	Google	0.7634	0.4836	0.1313	0.1344	0.1232	0.6222	0.6564	0.3522	0.3
7	Home	0.8463	0.9732	0.4411	0.1333	0.6453	0.3435	0.3535	0.2442	0.3
.....	.	0.8653	0.4835	0.1343	0.4421	0.7567	0.2424	0.5241	0.3221	0.3
00	Zoo	0.4736	0.9473	0.1453	0.1134	0.6564	0.1749	0.1892	0.1344	0.3



The example of words embeddings

A brief review of ML problems

Supervised learning

Aim : make a prediction.

We can predict

- either discrete predictions : classification
- either continuous ones : regression

A brief review of ML problems

Classification vs regression

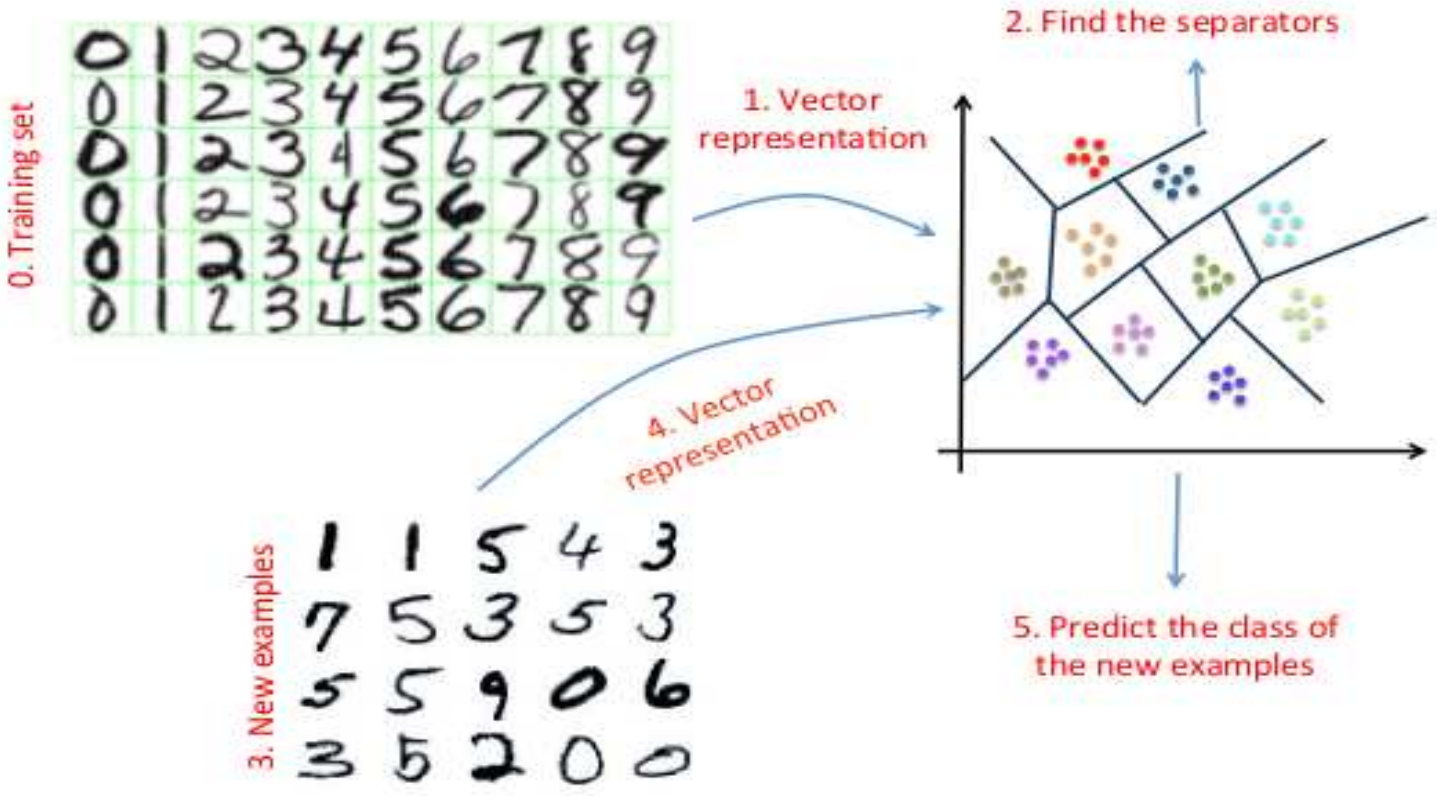
A brief review of ML problems

More ML paradigms

- Unsupervised learning:
  - Dimensionality reduction
  - Clustering;
- Supervised learning:
  - Regression
  - Classification
- Semi-supervised learning.
- Reinforcement learning.

A brief review of ML problems

A classical pipeline in ML



A brief review of ML problems

In what follows we shall focus on two main topics

- Dimensionality reduction
- Supervised learning