Statistical Machine Learning

Luca Bortolussi

Department of Mathematics and Geosciences
University of Trieste

328
~~Office 303, third floor, E1~~ 3
luca@dmi.units.it

Trieste, Summer Semester 2018

Some facts worth considering

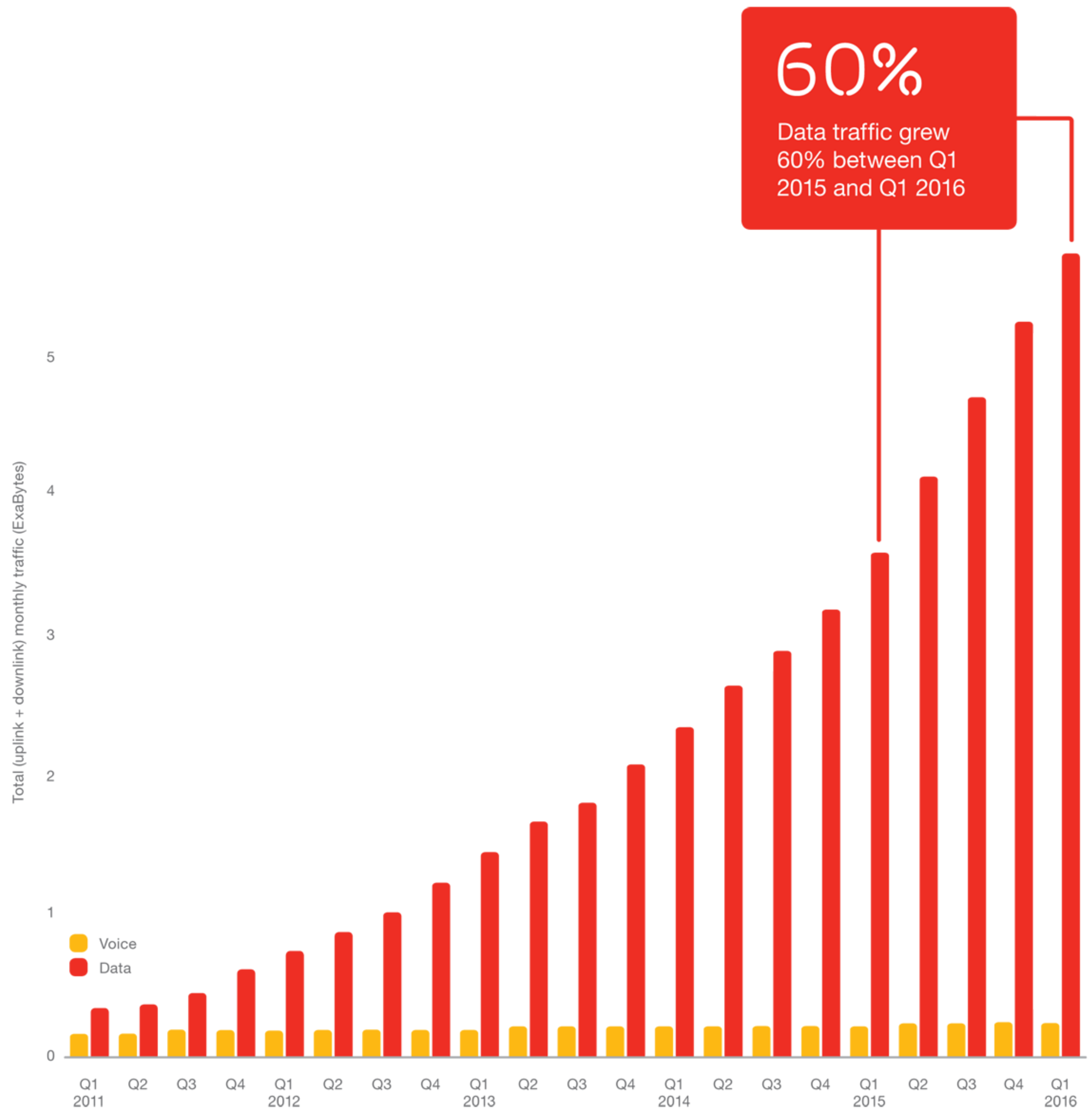
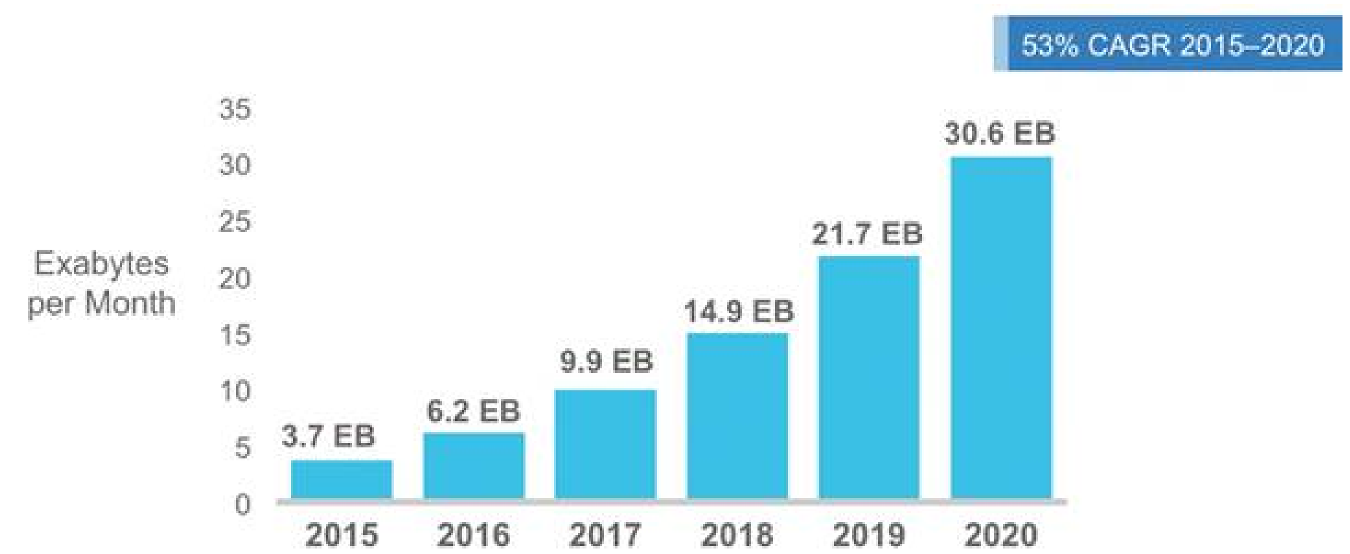


Table 1. The Cisco VNI Forecast—Historical Internet Context

Year	Global Internet Traffic
1992	100 GB per day
1997	100 GB per hour
2002	100 GBps
2007	2000 GBps
2014	16,144 GBps
2019	51,794GBps

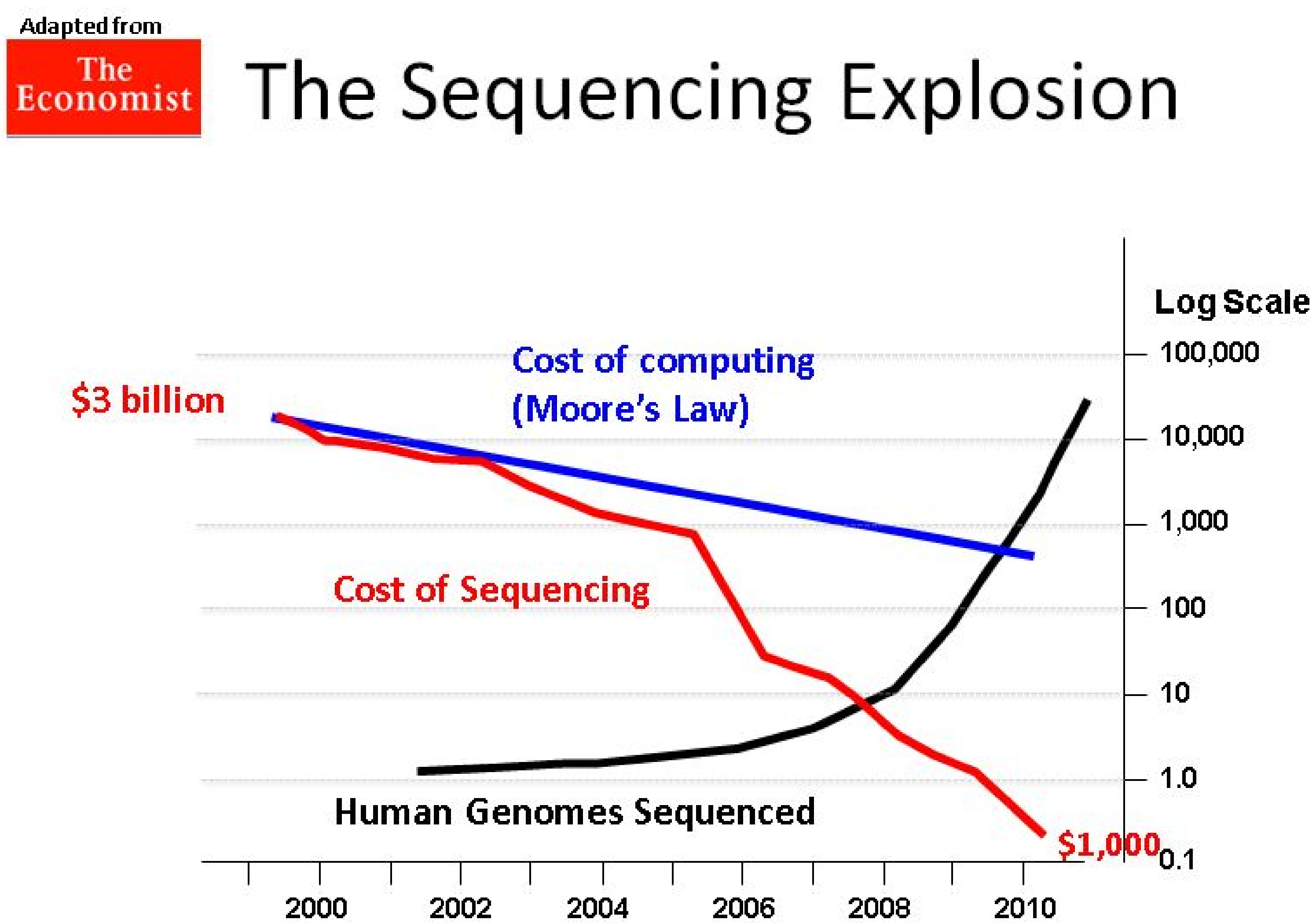
Source: Cisco VNI, 2015

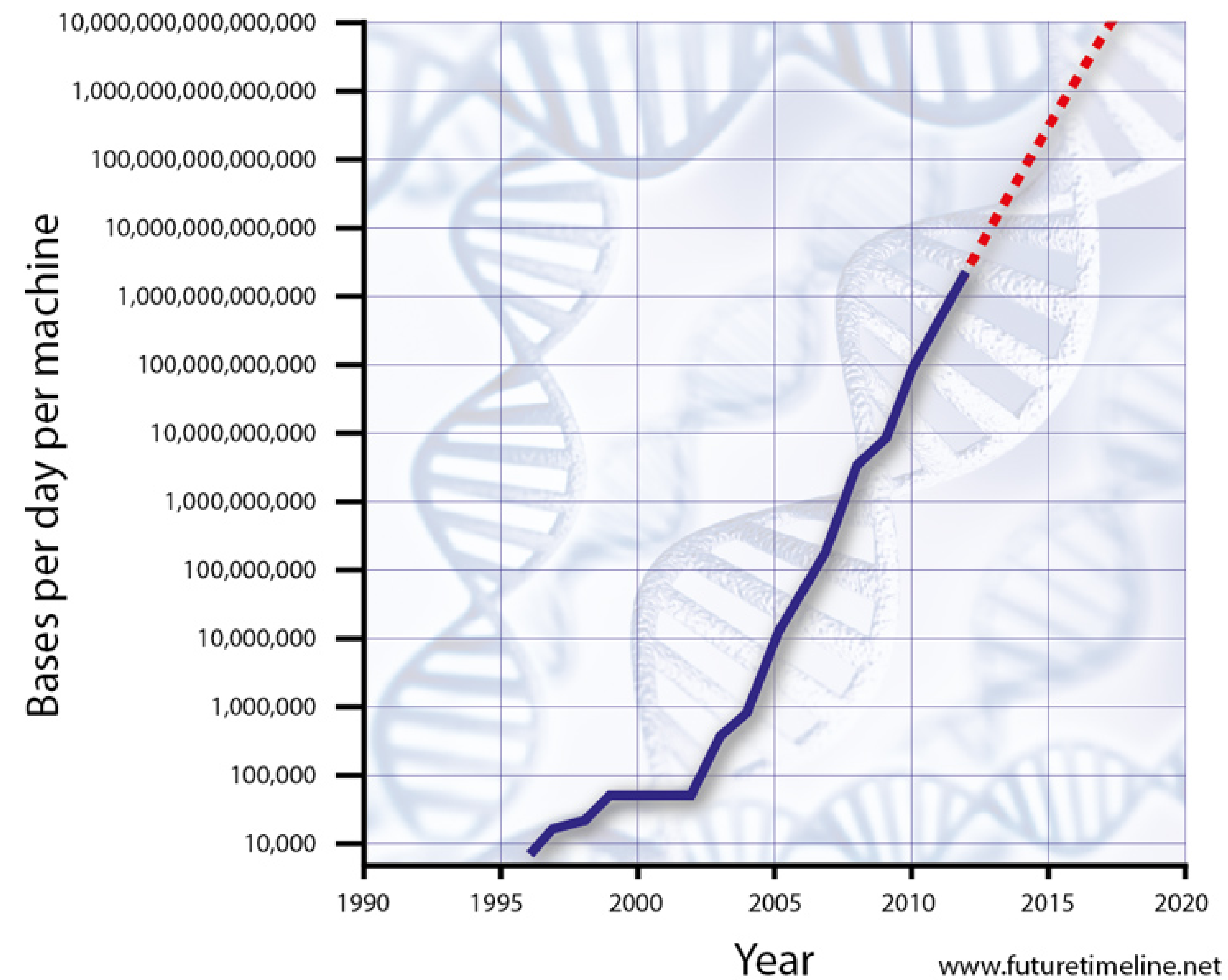


Mobile traffic in 2013 = 18 × total internet traffic in 2000
Mobile traffic in 2021 = 12 × traffic in 2015.

We are living in a world pervaded by data (information?)

Some facts worth considering





UK National Health Service plans to sequence genome of 750.000 cancer patients in the next ten years

How to make sense of all this data?
How to extract knowledge from it?

Some facts worth considering

ARTICLE

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. Here we introduce a new approach to computer Go that uses 'value networks' to evaluate board positions and 'policy networks' to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Without any lookahead search, the neural networks play Go at the level of state-of-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play. We also introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks. Using this search algorithm, our program AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0. This is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away.

Google purchased DeepMind (after 1 year of operation) for 450M GBP

And DeepMind is now one of the most important AI research centres int he world... cf. AlphaZero Go and AlphaZero Chess.

Some facts worth considering

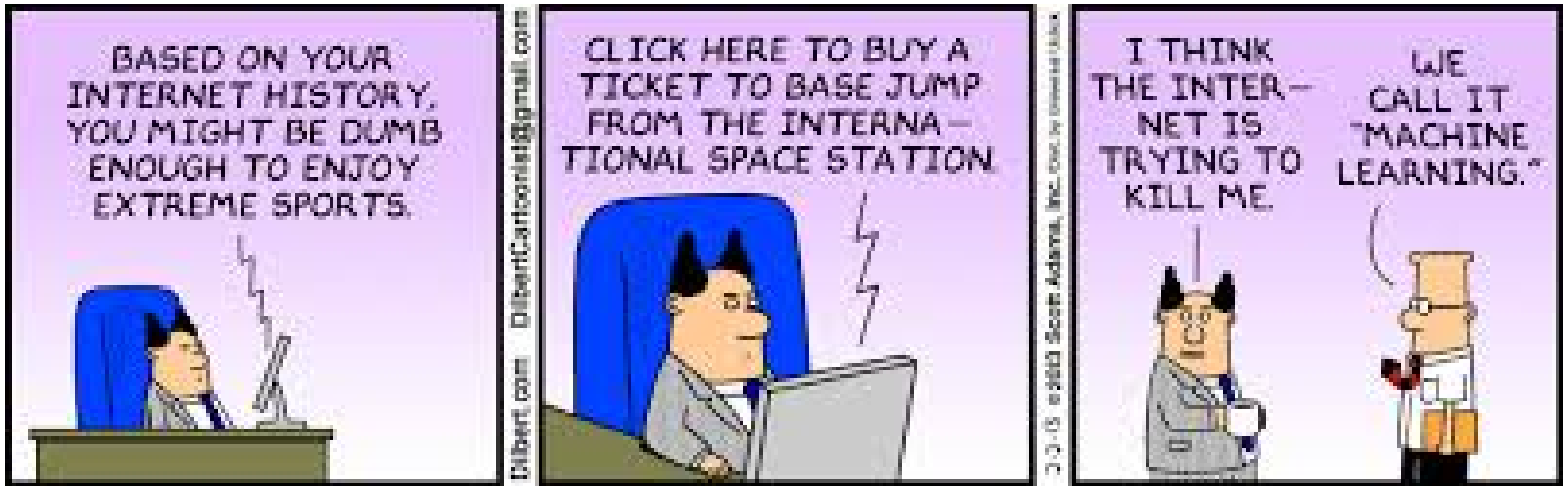
Data Science, as a term, "was first coined in 2001. Its popularity has exploded since 2010, pushed by the need for teams of people to analyze the big data that corporations and governments are collecting." (Wikibook on data science)

Number of job postings for data scientists increased globally by 20.000% between 2009 and 2015...
73% growth of job offers in data science in Italy, from jan-mar 2015 to jan-mar 2016.

Machine Learning

If you google it...

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. [source: wikipedia]



A ROUGH CLASSIFICATION

Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. [source: wikipedia]

- Supervised learning: learn a model from input-output data. The goal is to predict a the (most-likely) output value for a new, unobserved, input. We distinguish
  - Regression (continuous output)
  - Classification (binary/ discrete output)
- Unsupervised learning: extract information/ learn a model from input-only data
- Reinforcement Learning: find suitable actions to take in a given situation in order to maximize a reward.

IT'S ALL ABOUT THE MODELS

- Machine Learning is all about learning models...
- But, what is a model? Discuss for 5 minutes and provide 3 examples

My own answer

- A model is a hypothesis that certain features of a system of interest are well replicated in another, simpler system.

My own answer

- A model is a hypothesis that certain features of a system of interest are well replicated in another, simpler system.
- A *mathematical model* is a model where the simpler system consists of a set of mathematical relations between objects (equations, inequalities, etc).

My own answer

- A model is a hypothesis that certain features of a system of interest are well replicated in another, simpler system.
- A *mathematical model* is a model where the simpler system consists of a set of mathematical relations between objects (equations, inequalities, etc).
- A *stochastic model* is a mathematical model where the objects are probability distributions.

My own answer

- A model is a hypothesis that certain features of a system of interest are well replicated in another, simpler system.
- A *mathematical model* is a model where the simpler system consists of a set of mathematical relations between objects (equations, inequalities, etc).
- A *stochastic model* is a mathematical model where the objects are probability distributions.
- All modelling usually starts by defining a *family* of models indexed by some parameters, which are tweaked to reflect how well the feature of interest is captured.

My own answer

- A model is a hypothesis that certain features of a system of interest are well replicated in another, simpler system.
- A *mathematical model* is a model where the simpler system consists of a set of mathematical relations between objects (equations, inequalities, etc).
- A *stochastic model* is a mathematical model where the objects are probability distributions.
- All modelling usually starts by defining a *family* of models indexed by some parameters, which are tweaked to reflect how well the feature of interest is captured.
- Machine learning deals with algorithms for automatic selection of a model from observations of the system.

Generative and Discriminative models

$\boxed{p(x, y)}$ $x \in \mathcal{X}$ $y \in \mathcal{Y}$ $D = \{(x_i, y_i)\}_{i=1,\dots,N}$

- Supervised learning can have two flavours
- Two different types of question can be asked:
  - what is the joint probability of input/ output pairs?
  - given a new input, what will be the output?

$\{P_\theta(x,y)\}$ THE "BEST" P_θ GIVEN D

- The first question requires a model of the population structure of the inputs, and of the conditional probability of the output given the input → **generative modelling**
- The second question is more parsimonious but less explanatory → **discriminative learning** $\boxed{P(y|x)}$ $\{P_\theta(y|x)\}$
- Notice that the difference between generative supervised learning and unsupervised learning is moot

$\boxed{p(x)}$

Inference and Estimation

WE HAVE $p(x,y)$ $\vec{x} = (x_1 \ldots x_m)$ $\vec{y} = (y_1, \ldots y_d)$ $y_j \in \{0,1\}$

$$p(x) = \int p(x,y)\,dy \qquad p(x) = \sum_{y \in \mathcal{Y}} p(x,y) \qquad \mathcal{Y} = \{0,1\}^d$$

$$p(y|x) = \frac{p(x,y)}{p(x)} \qquad |\mathcal{Y}| = 2^d$$

INFERENCE

$d = 1000$

$2^{1000} \approx 10^{300}$

atoms in universe 10^{90}

ESTIMATION

$D = \{(x_i, y_i)\}_{i=1,\ldots N}$

$p(y|x,\theta)$ FIND 'THE BEST' θ that explains your data

OR $p(\theta|D)$ (GIVEN $p(\theta)$)

BAYESIAN : ESTIMATION $\simeq$ INFERENCE

Course plan

- Primer on Bayesian statistics, plus a review of some probability distributions and inference.
- Then we will touch the following topics:
  1. (Bayesian) Linear Regression and Classification, Laplace approximation, Model Selection;
  2. Clustering, Mixtures of Gaussians and Expectation Maximisation
  3. Probabilistic graphical models: definitions and inference
  4. Hidden Markov Models for sequential data
  5. Kernel Methods: Gaussian Processes for Regression and Classification
  6. Deep Learning

  ? If we have time: Variational inference (main ideas, revisiting Bayesian Linear Regression and classification, and Belief Propagation); Active Learning and Bayesian optimisation, ...
-

LAB+EXAM

LABORATORY

The Lab will account for roughly 50% of the course. In the Lab, we will experiment with Machine Learning in Python, playing with datasets and libraries like Pandas, Scikit-learn, keras (tensorflow), PyTorch, ...

Bring your own laptop...

Lab will be learn by doing, with a lot of self learning. Working in groups is welcome. Propose your own data and problems (from Kaggle, from your past courses).

EXAM

- Final team project, with presentation - possibly on datasets coming from companies.

Coordinates

Moodle

There is a moodle page of the course. Register, it is where you will get all the material.

Coordinates

Moodle

There is a moodle page of the course. Register, it is where you will get all the material.

Where can you find me?

Coordinates

Moodle

There is a moodle page of the course. Register, it is where you will get all the material.

Where can you find me?

- Around the World.

COORDINATES

MOODLE

There is a moodle page of the course. Register, it is where you will get all the material.

WHERE CAN YOU FIND ME?

- Around the World.
- Room 328, 3rd floor - email me first at `lbortolussi@units.it`.

OTHER STUFF

- question time at the end of each lecture
- Requests?