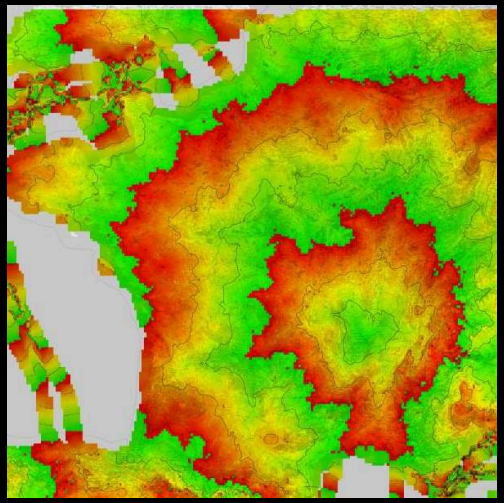
Machine Learning
What, how, why?

Rémi Emonet (@remiemonet)

2015-09-30

Web En Vert

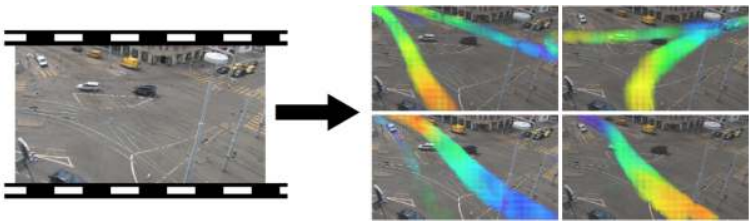


software carpentry

? ! :)

learn teach help enjoy

FOSS



$ whoami

$ whoami

- Software Engineer
- Researcher: machine learning, computer vision
- Teacher: web technologies, computing literacy
- Geek: deck.js slides, isochrones, ...

You are shrewd, skeptical and restrained.

You are independent: you have a strong desire to have time to yourself. You are calm-seeking: you prefer activities that are quiet, calm, and safe. And you are philosophical: you are open to and intrigued by new ideas and love to explore them.

Experiences that give a sense of prestige hold some appeal to you.

You are relatively unconcerned with both tradition and taking pleasure in life. You care more about making your own path than following what others have done. And you prefer activities with a purpose greater than just personal enjoyment.

IBM BlueMix (Watson)

What is Machine Learning?

Machine Learning Basic Principle

- Given a dataset $\{y_i, x_{i1}, \ldots, x_{ip}\}_{i=1}^{n}$
- Optimize the likelihood function
$$\mathrm{L} = \sum_{d} \sum_{w} \sum_{t_a} n(w, t_a, d) log \sum_{z} \sum_{t_s} p(w, t_r|z) p(z, t_s|d)$$
- Or using a sparse regularization
$$\mathrm{L} - \lambda_{sparse} \sum_{d} \sum_{z} KL(U||p(ts|z, d))$$
- By using a Gibbs sampler
$$p(W_{ji}, at_{ji}|o_{ji} = o, O^{-ji}) = \frac{N_{obs}^{-ji}(W_{ji}, rt_{ji}, z_{ji}) + \eta(W_{ji}, rt_{ji})}{\sum_{w', rt'} \left(N_{obs}^{-ji}(w', rt', z_{ji}) + \eta(w', rt') \right)}$$

Machine Learning in the Wild

Which One of These Services Uses Machine Learning?

Machine Learning in Future Tech?

What is Machine Learning?
an example motivation

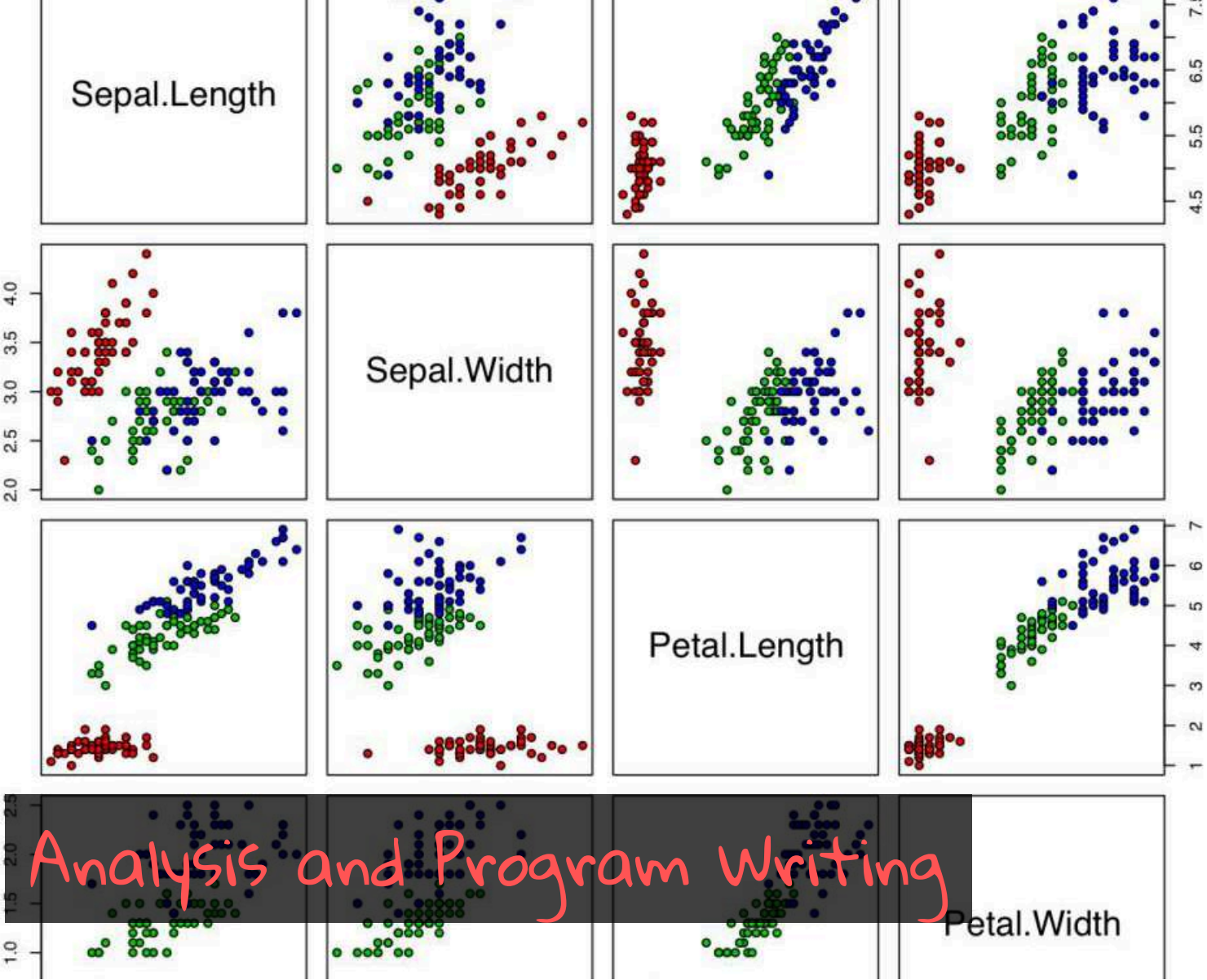
CC by Asa-maya (flickr)

Feature Extraction

⇒

```
Sepal length: 5.1
Sepal width: 2.5
Petal length: 4.2
Petal width: 1.0
```

Expected Label: "Iris Setosa"



Analysis and Program Writing

IFTTT...
CC by -Parsyph- (Flickr)

Predictive Machine Learning

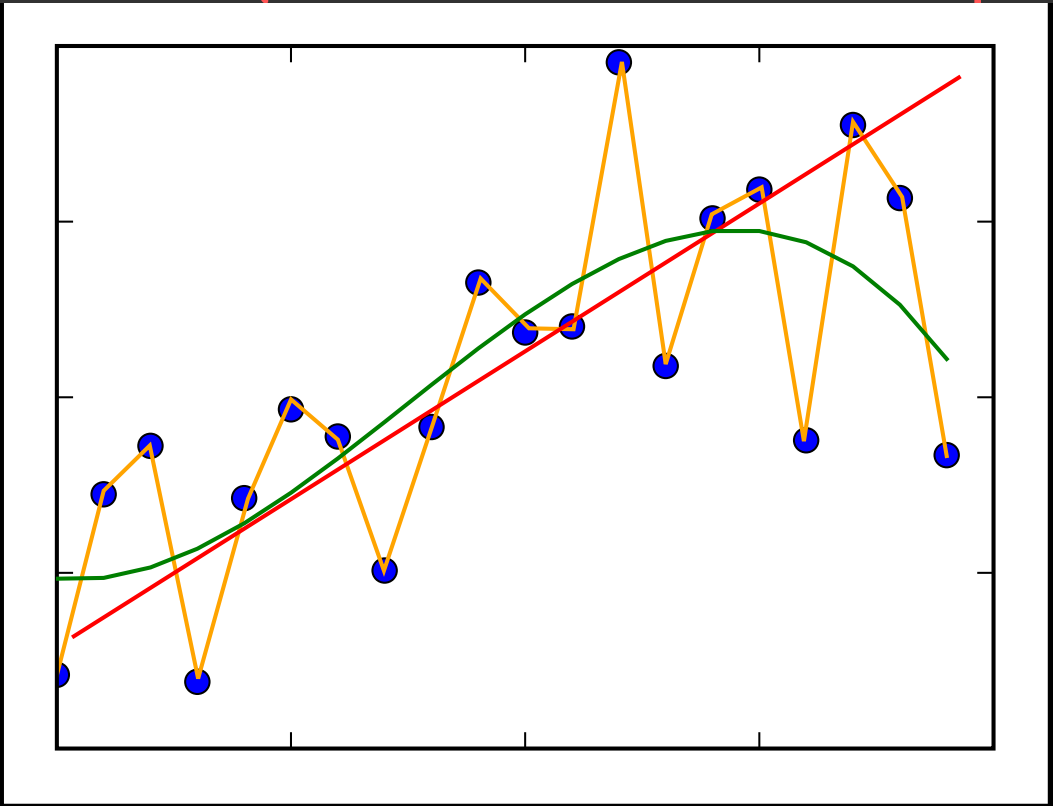
- Instead of writing a program that solves a task,
- We
  1. collect labeled data: input/output pairs
  2. automatically generate a program
     that computes an output for each new input
  3. profit!

The machine learns to generalize
from a limited number of examples,
like humans do.

Different Types of Tasks

- Supervised learning: some labels are known
  - classification: find the label of an example
  - regression: find the target value
- Unsupervised learning: no labels
  - clustering: group things together
  - pattern mining: find recurrent events
  - anomaly detection: find "outliers"

The Principle of "Overfitting"



A Lot of Different Methods

- Also called "models"
  - linear regression, logistic regression, SVM, kernel SVM, neural networks, k-means clustering, collaborative filtering, bayesian networks, expectation maximization, belief propagation, multiple kernel learning, metric learning, transfer learning, decision trees, gaussian processes, random forests, boosting, ...
- For different contexts
  - different tasks
  - different nature of data
  - different suppositions on the data
  - different amount of data

Different Ways to Start

- Use a product that uses ML
  - e.g. adwords, ibm bluemix, ...
- Use a prediction API
  - send your data to the service
  - get API to process new inputs
  - e.g., google pred. API, prediction.io, ...
- Dive into machine learning...

Into Machine Learning

- Using libraries
  - libraries exist in most languages
  - most models already implemented
  - test different methods with different parameters
- Learning machine learning
  - many online courses
  - get deeper understanding

Does Machine Learning Actually Matter?

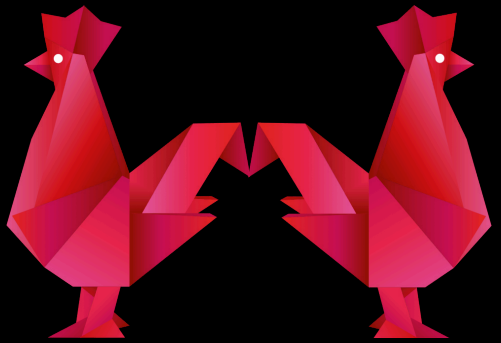
Example: The Netflix Challenge

NETFLIX

FAIR

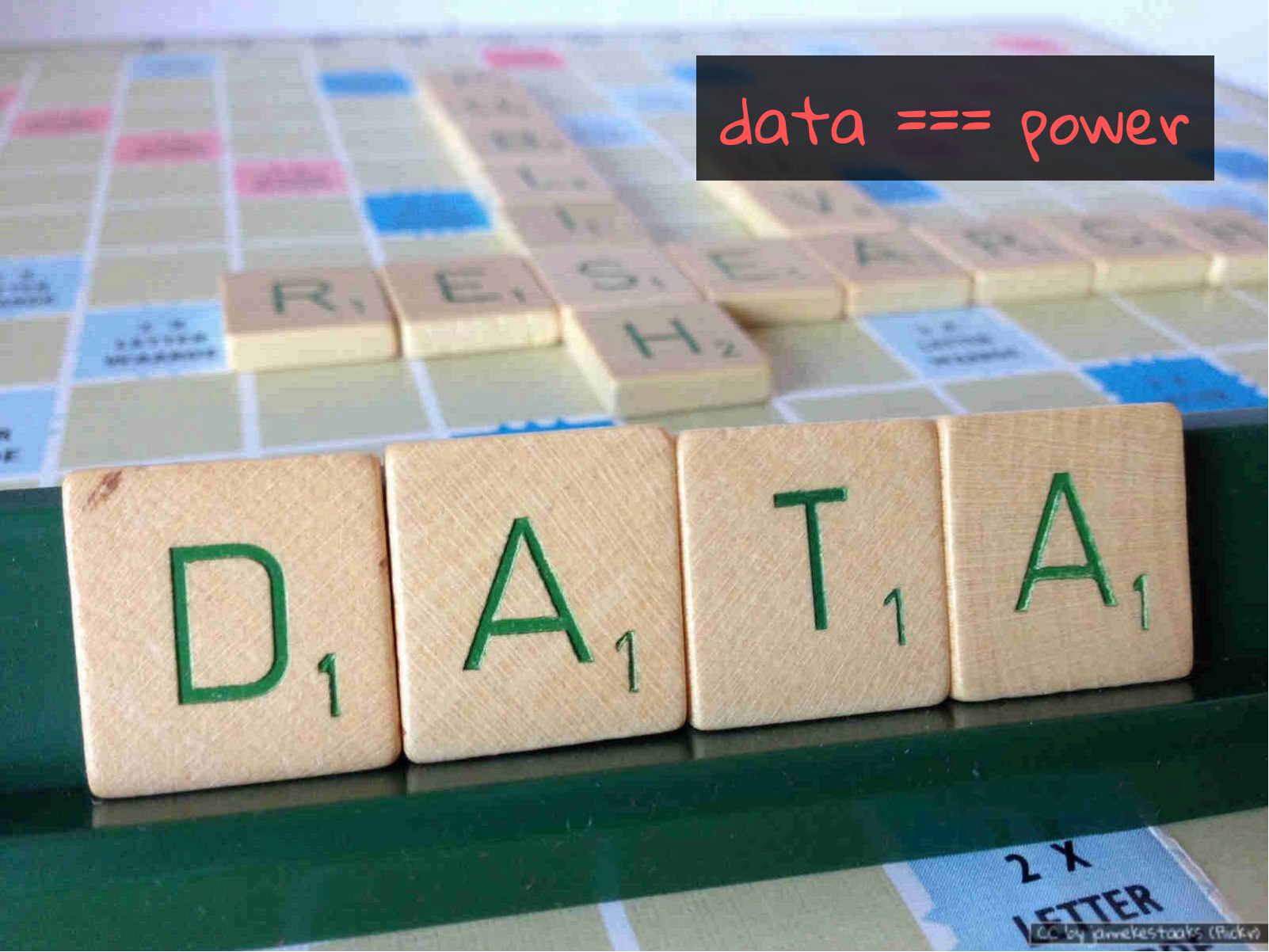
Example: Facebook AI Research

- Director: Yann LeCun
- Scientific Leads
  - Léon Bouttou
  - Rob Fergus
  - Florent Perronnin



Data, Data, Data

Data is Machine Learning's Fuel


data === power
D1 A1 T1 A1
2 X
LETTER
CC by jannekestaaks (Flickr)

Getting Data?
cc by francoisputfiney (flickr)

Getting Data?

- Collect from your services/applications
- Do it yourself
- Pay some people you know
- Use crowd-sourcing,
  e.g., Amazon Mechanical Turk (MTurk)
- Find existing datasets (open data, etc)
- Work for/with a "data rich" company
- Create your "intermediation" business

What Can It Do For Me

Search

Google Search, Bing, etc

Advertising

AdWords, etc

Recommendations

Netflix, Amazon, Youtube, app Stores, etc

Text Translation

Optical Character Recognition

(postcodes, checks, book scans, etc)

Visual Recognition (objects, plants, animals, etc)

Face Detection
Smile Detection
(embedded in Cameras)

Face Identification
(Picasa, Facebook, etc)

Kinect Controller

Self Driving Cars

Voice Recognition and Synthesis (GoogleNow, Siri, Cortana)

Sound Recognition
(birds, underwater sounds, safety, etc)

Fraud Detection
(Banking, Websites, etc)

Automated Trading

...

Customer/Person Profiling

BlueMix Watson, etc

Adaptive Websites
(automated A/B testing)

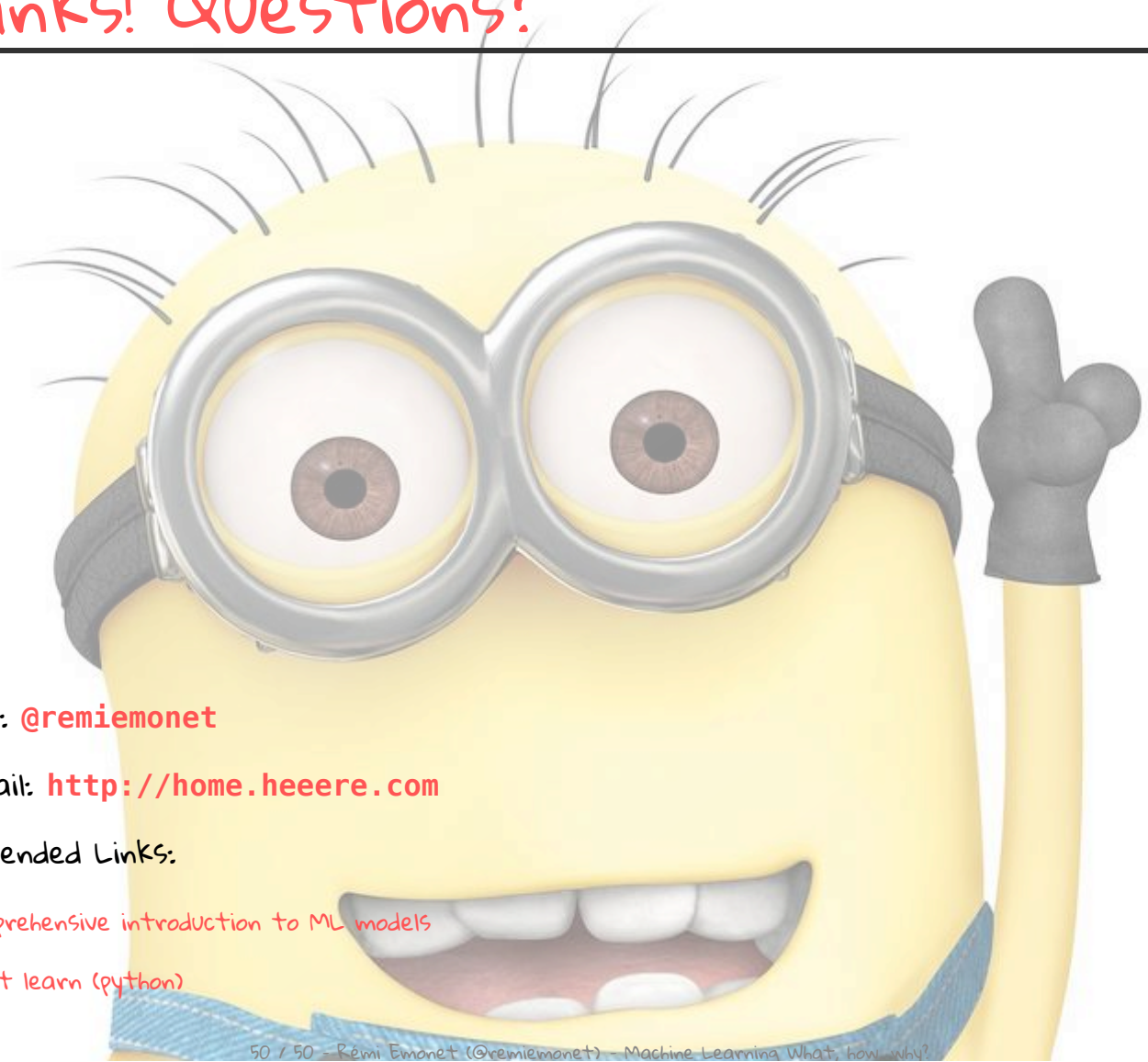
The "Big Data" Hype

and much more...
CC by eFlyera (flickr)

Where Will it Stop?

Singularity?
CC by JD Hancock (flickr)

Thanks! Questions?

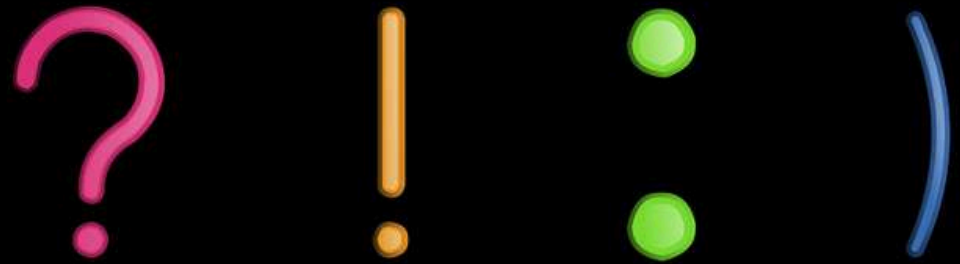


twitter: @remiemonet

web/email: http://home.heeere.com

Recommended Links:

- comprehensive introduction to ML models
- scikit learn (python)
- ...



tatadbb

FOSS

https://www.flickr.com/photos/tatadbb/6719845055/sizes/l

-Porsupah-
CC by -Porsupah- (Flickr)

Google
California
6YCT905
RX 450h
VISITOR
self-driving car
Google
ali eminov
CC by ali eminov (flickr)

efilpera
CC by efilpera (flickr)

DATA
jannekestaaks
2 X
CC by jannekestaaks (Flickr)

Asa-moya
CC by Asa-moya (flickr)

francoisjouffroy
cc by francoisjouffroy (flickr)

JD Hancock
CC by JD Hancock (flickr)