Demystifying ML, AI & Automation

Spyros Gkezerlis

24-25.10.2017

Intro

Spyros Gkezerlis - @sgez

About
BSc Automation Engineering – Telematic Applications
MSc Information Systems – Reinforcement Learning
Interests – Applied ML, RL especially in time domain

Experience
4 years Software Engineer in Greece
9 years Engineer and Manager in OTE
3 years Program Manager in DTAG

Ariadni Gkezerli
(c) 2017

TLAs

ML : Machine Learning
AI : Artificial Intelligence
BD : Big Data
NN : Neural Network
DL : Deep Learning
CNN: Convolutional Neural Network
NLP: Natural Language Processing

MATH
Mental Abuse To Humans

TLA: Two or Three Letter Acronym

Why Lectures?

- Get a better overview of the current landscape in ML, AI & Automation
- "Set the record straight" sessions
- Create a common understanding of what ML, AI & Automation
- “Start with the problem” philosophy

Pros cons, Tools, etc. should not dictate what we should use!

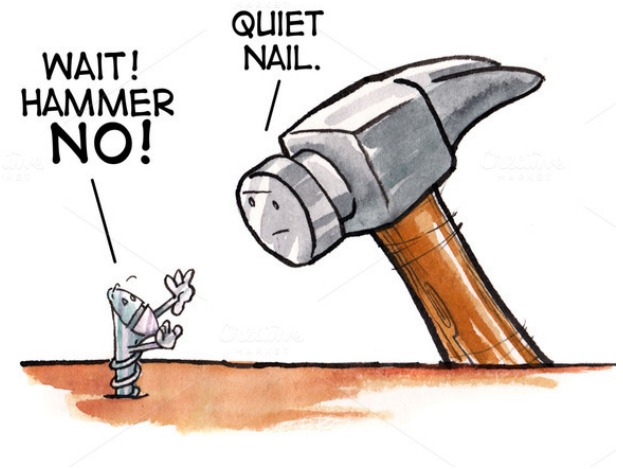
  - Identify what we want to solve
  - Work to the algorithms & models needed
  - Utilise best approach
- Assess ML potential

Do you want to become an ML Jedi?

Why ML, AI, Automation?

- Why use Automation, Adv.Stats, Probability Theory, ML, AI?

  Because it can help us:
  - Reduce complexity of network
  - Improve experience by Time-to-market, Time-to-repair
  - Repetitive caused costs can be targeted and reduced



- Also… AI is another fun way to solve problems ☺

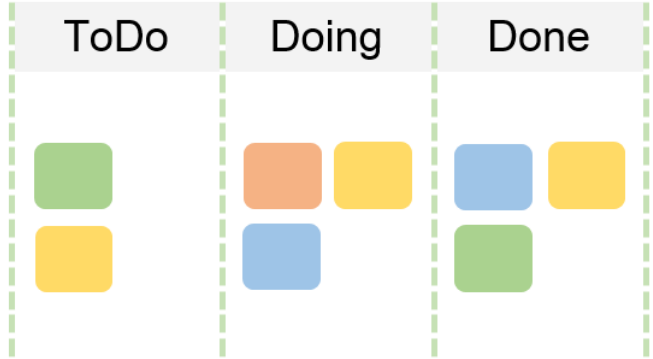
- Keep in mind it can solve particular types of problems!!!

Lecture Guidelines

- Stop for questions at any time! Be curious!!!
- Lunch all together for questions and NT
- Workflow to be used?
  Kanban board with ToDo, Doing, Done
- First iteration through everything is a must
  if more items:
  put in ToDo
- Set as ToDo from the beginning

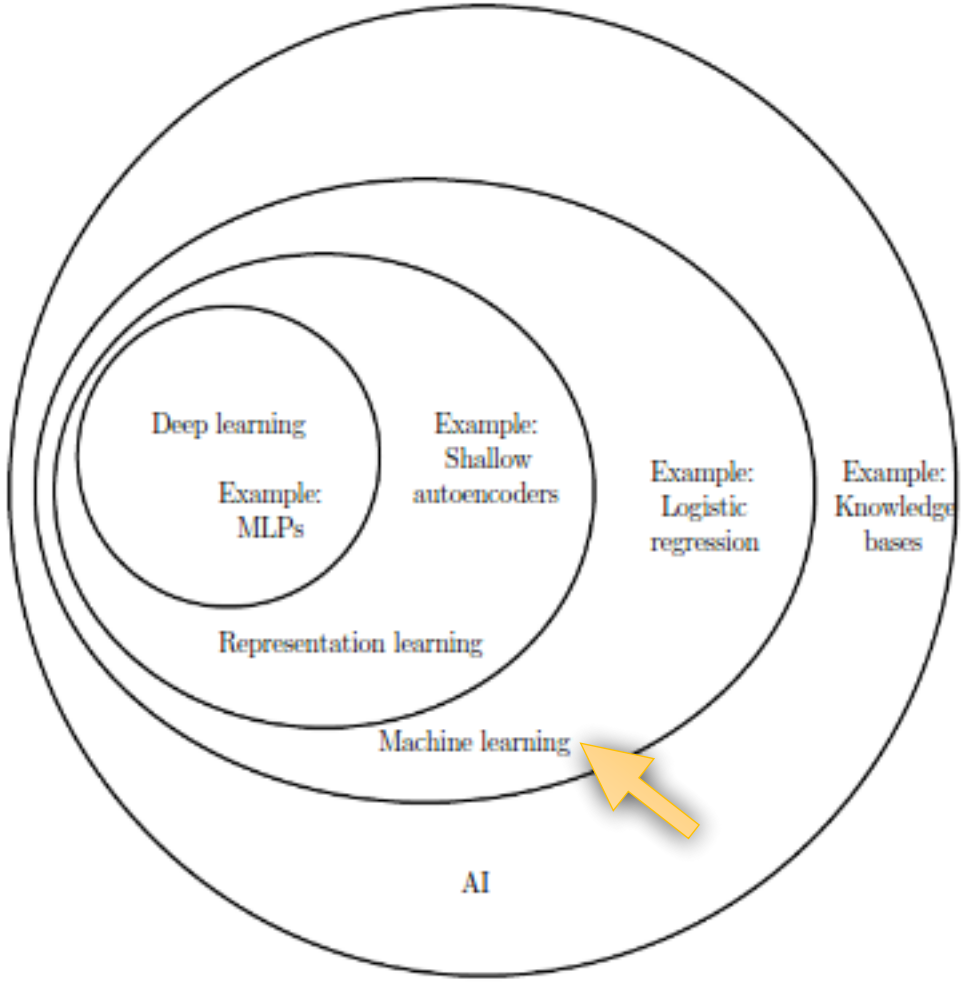
Basic understanding of the landscape

- If noted, slides have links with extra source material



Landscape overview

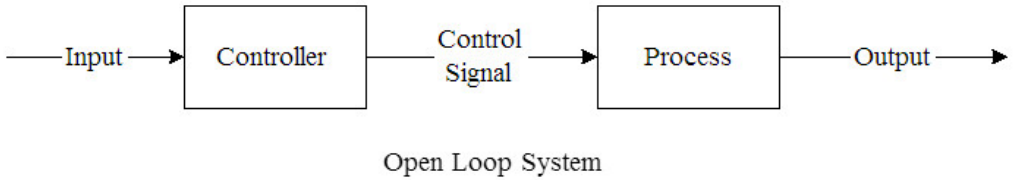
- Automation
- Statistics & Probability
- Data Mining
- Artificial Intelligence
- Machine Learning
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning
- Big Data Analytics



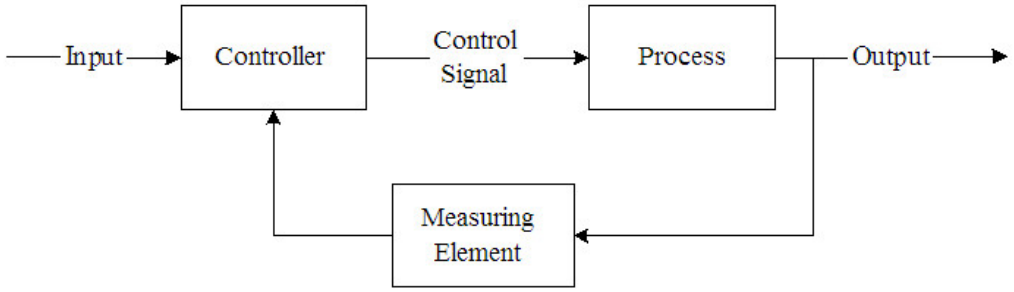
From Ian Goodfellow's Deep Learning Book: http://www.deeplearningbook.org

Automation

- Comes from ancient greek word which means the thing that wishes on its own or the has a will or fury by itself
- Describes the tasks that are or can be performed by machines autonomously
- It can be used to improve quality, accuracy, save costs and amplify precision



Open Loop System



Closed Loop System

Theory

- Probability Theory Bayes Theorem

  P(A|B) ==> Probability of A given that B

  - Exercise
    Out of 3000 emails received over a certain period, 2000 are spam and 1000 are not. The word “Rolex” appeared in 250 out of the 2000 which are spam and in 5 out of the 1000. So, if an email is received, and contains the word “Rolex”, what is the possibility that it is a spam?

Let S be the event that the message is spam, and E be the event that the message contains the word w. Under our assumption from before, we have that:

$$p(S|E) = \frac{p(E|S)}{p(E|S) + p(E|\overline{S})}$$

https://www.arl.wustl.edu/~todd/cse240/cse240_lecture17_3-24-16.pdf

Theory

- Example – Solution:
  Out of 3000 emails received over a certain period, 2000 are spam and 1000 are not. The word "Rolex" appeared in 250 out of the 2000 which are spam and in 5 out of the 1000.

  So, if an email is received, and contains the word "Rolex", what is the possibility that it is a spam?

Now p(w) and q(w) are empirical estimates of p(E|S) and p(E|S')

$$r(Rolex) = \frac{p(Rolex)}{p(Rolex) + q(Rolex)} =$$

$$\frac{0.125}{0.125 + .005} = \frac{0.125}{0.125 + .005} \approx 0.962$$

https://www.arl.wustl.edu/~todd/cse240/cse240_lecture17_3-24-16.pdf

Theory

- Statistics Markov Chains

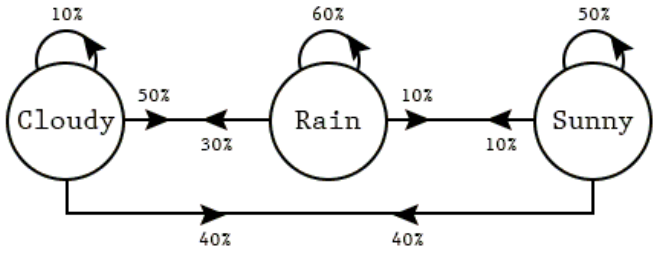
  Markov Chains is a probabilistic process, that relies on the current state to predict the next state. For Markov chains to be effective the current state has to be dependent on the previous state in some way; For instance, from experience we know that if it looks cloudy outside, the next state we expect is rain. We can also say that when the rain starts to subside into cloudiness, the next state will most likely be sunny.

  Not every process has the Markov Property, such as the Lottery, this weeks winning numbers have no dependence to the previous weeks winning numbers.

Markov Table of Probabilities

STATE	NEXT STATE	PROBABILITY	%
CLOUDY	CLOUDY	0.1	10%
CLOUDY	RAIN	0.5	50%
CLOUDY	SUNNY	0.4	40%
RAIN	CLOUDY	0.3	30%
RAIN	RAIN	0.6	60%
RAIN	SUNNY	0.1	10%
SUNNY	CLOUDY	0.4	40%
SUNNY	RAIN	0.1	10%
SUNNY	SUNNY	0.5	50%

Markov State Diagram



Current State Vector

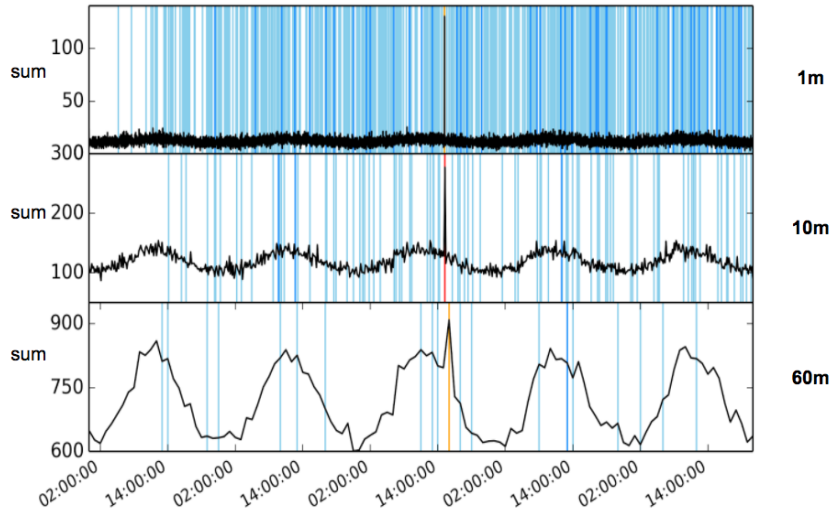
C	R	S
1	0	0

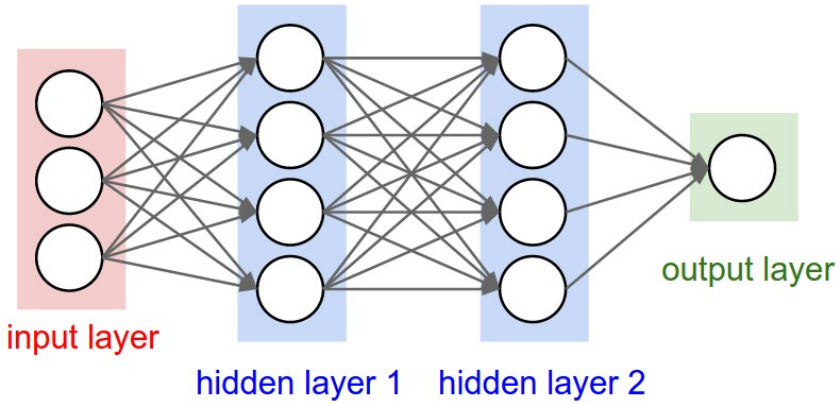
Figure 4

http://techeffigytutorials.blogspot.com/2015/01/markov-chains-explained.html

Theory

- Other Topics
  - Time Domain
  - Event & Anomaly Detection
  - Pattern recognition
- Structures
  - Trees
  - Graphs
  - Neural Networks



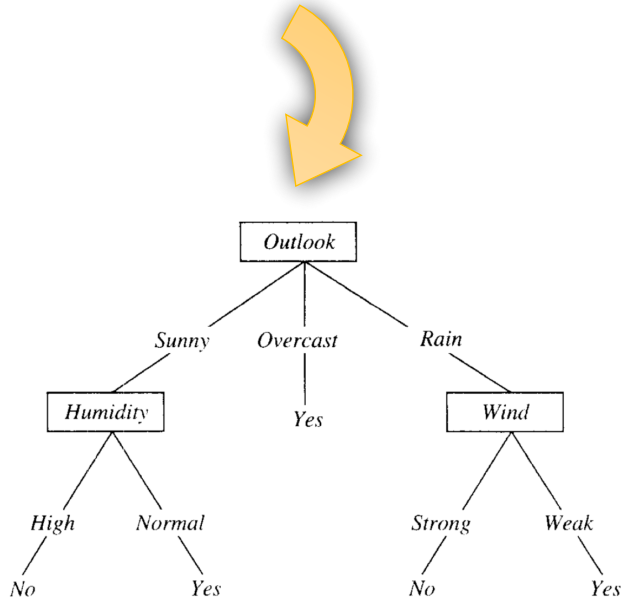


http://cs231n.github.io/neural-networks-1/

Theory

- Supervised Learning - Decision Tree Learning DTL is method for approximating discrete valued target functions, in which the learned function is represented by a decision tree. (Weka example will follow)

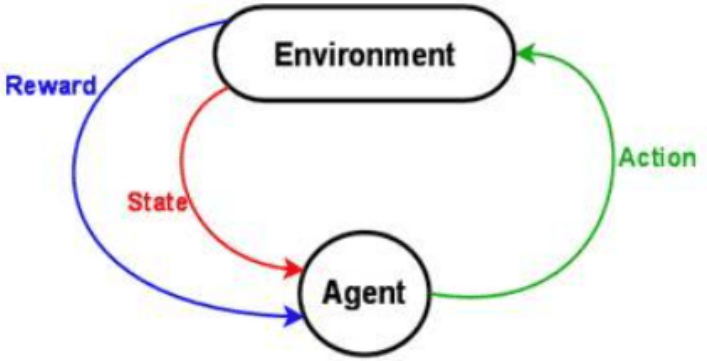
No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	70.0	FALSE	yes
...	rainy	75.0	80.0	FALSE	yes
...	sunny	75.0	70.0	TRUE	yes
...	overcast	72.0	90.0	TRUE	yes
...	overcast	81.0	75.0	FALSE	yes
...	rainy	71.0	91.0	TRUE	no

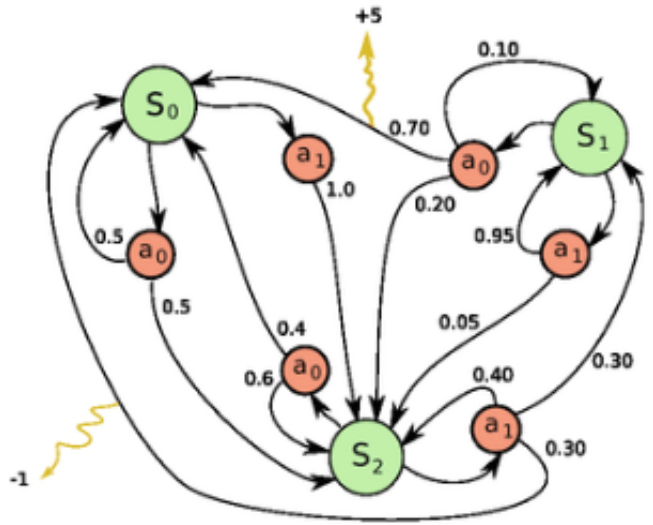


http://www.cs.princeton.edu/courses/archive/spr07/cos424/papers/mitchell-dectrees.pdf

Theory

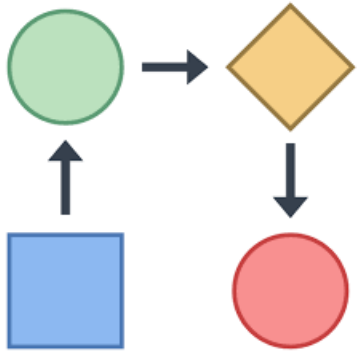
- Reinforcement Learning
  (inspired by behaviourist psychology)
  - RL is about how agents ought to take actions in an environment so as to maximise some notion of cumulative reward.
  - A state is the current status
  - An action is decided based on previous reward levels
  - A reward is given based on the next state
  - Usual structure that an agent can use to describe the environment can be a graph / NN which has starting states and end / target states.





Workflow – The ML Pipeline

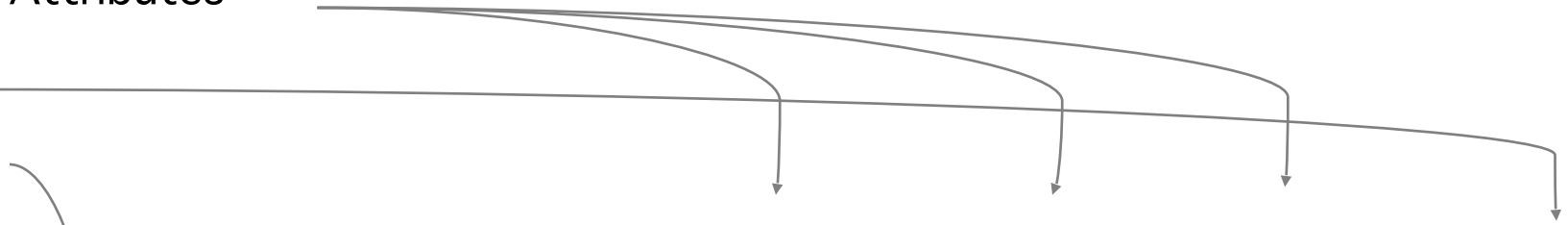
- Start with the question or problem we want to solve
- Find proper data and sources
- Prepare & create a data set
- Choose a model e.g. Decision Tree, J48 Algorithm
- Train system & classify
- Evaluate the system and fine-tune
- Predict / Forecast
- Apply to workflow
- Automate into workflow



Workflow loosely based on https://medium.com/towards-data-science/the-7-steps-of-machine-learning-2877d7e5548e

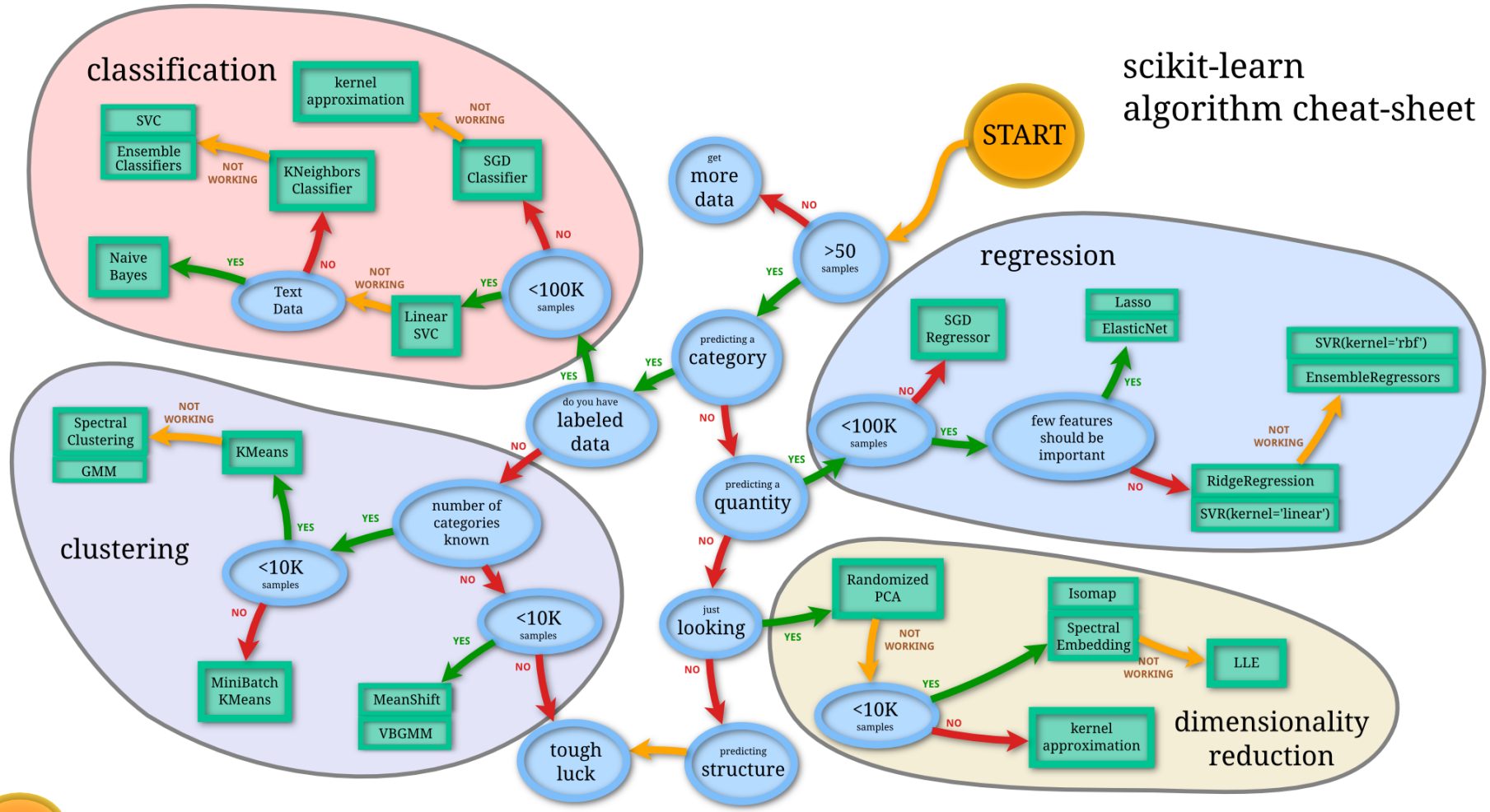
Prepare & create a dataset

- Dataset
  - Features or Attributes
  - Labels
  - Instances



Age	Nationality	Education	Job
25	Greek	MSc	Engineer
38	American	MSc	Analyst
35	Japanese	PhD	Assistant
…	…	…	…

Choose a model

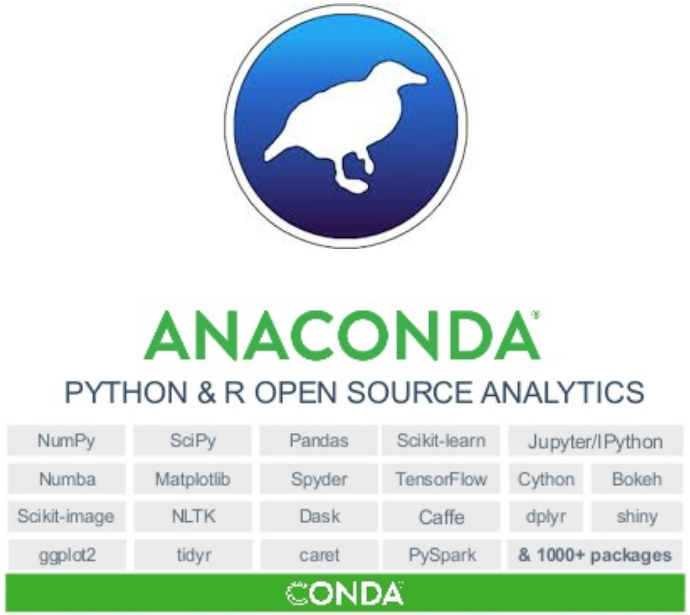


Back

http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

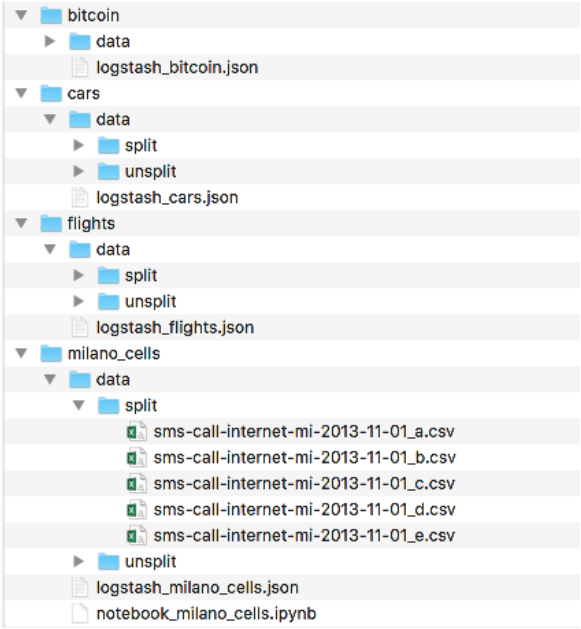
Tools – Overview

- Weka
  - Data Mining GUI / basic ML
- Anaconda
    - Python
      - Dask
      - Scikit-Learn for ML/DL
      - Tensorflow for DL
    - R Studio
- Elasticsearch – Logstash - Kibana



Example Datasets

- bitcoin
- cars
- flights
- milano_cells



Lectures Page

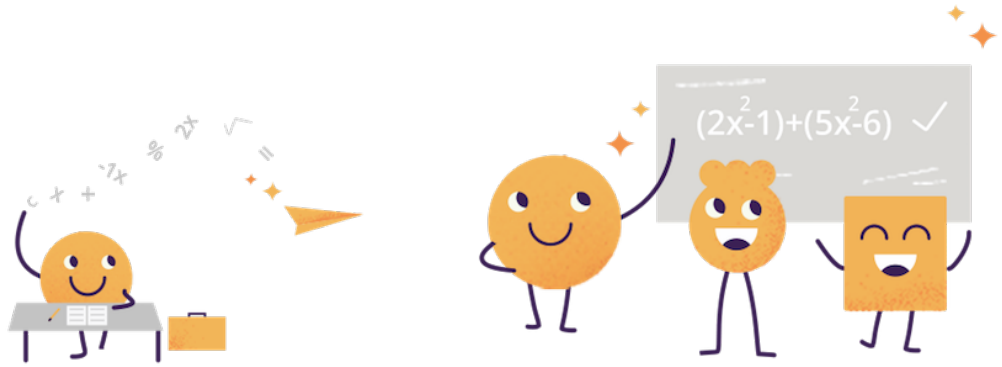
https://github.com/sgez/MLAI

Practice (fun part :-)

- Weka For Data Mining / basic ML
- Anaconda Python & R for ML
- Elasticsearch (ELK) for Visualisation & ML

on:

  - bitcoin
  - cars
  - flights
  - milano_cells



Special Thanks to Cosmote!

24.10.2017

- Diogenis Georgiadis
    - Zeta Papadopoulou
        - Christos Kavakopoulos

25.10.2017

- Christos Penitas
    - Antonis Galanakis
        - George Georgitselis