Demystifying ML, AI & Automation Part I

Spyros Gkezerlis

First version: Oct 2017

Current version: Apr 2018

Intro

Spyros Gkezerlis (Spi·ros Ge·zer·lis)

About
BSc Automation Engineering – Telematic Applications
MSc Information Systems – Reinforcement Learning
Current Research field – Applied ML, time series analysis

Experience
5 years Software Developer in Greece
8 years Engineer, Analyst & Manager in OTE
3 years Program Manager in DTAG

Ariadni Gkezerli (8 y.o), © 2017

github.com/sgez

kaggle.com/sgmtcl

twitter.com/sgez

Why Lectures?

- Get a better overview of the current landscape in ML, AI & Automation, because they can potentially can help us on:
  - Reducing complexity of network
  - Improving experience by Time-to-market, Time-to-repair
  - Repetitive caused costs can be targeted and reduced
  - Forecasting, Automating, Making predictions smarter
- “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
- Start changing mindset and attitude…

> “AI is the New Electricity”
>
> Andrew Ng

Simplistic Definitions

- Automation Comes from ancient compound greek word which means the thing that wishes on its own or the has a will or fury by itself
- Artificial Intelligence[1c] is Human Intelligence Exhibited by Machines
- Machine Learning[1a] is a field of computer science that gives computers the ability to learn without being explicitly programmed
- Deep learning[1b] is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms.

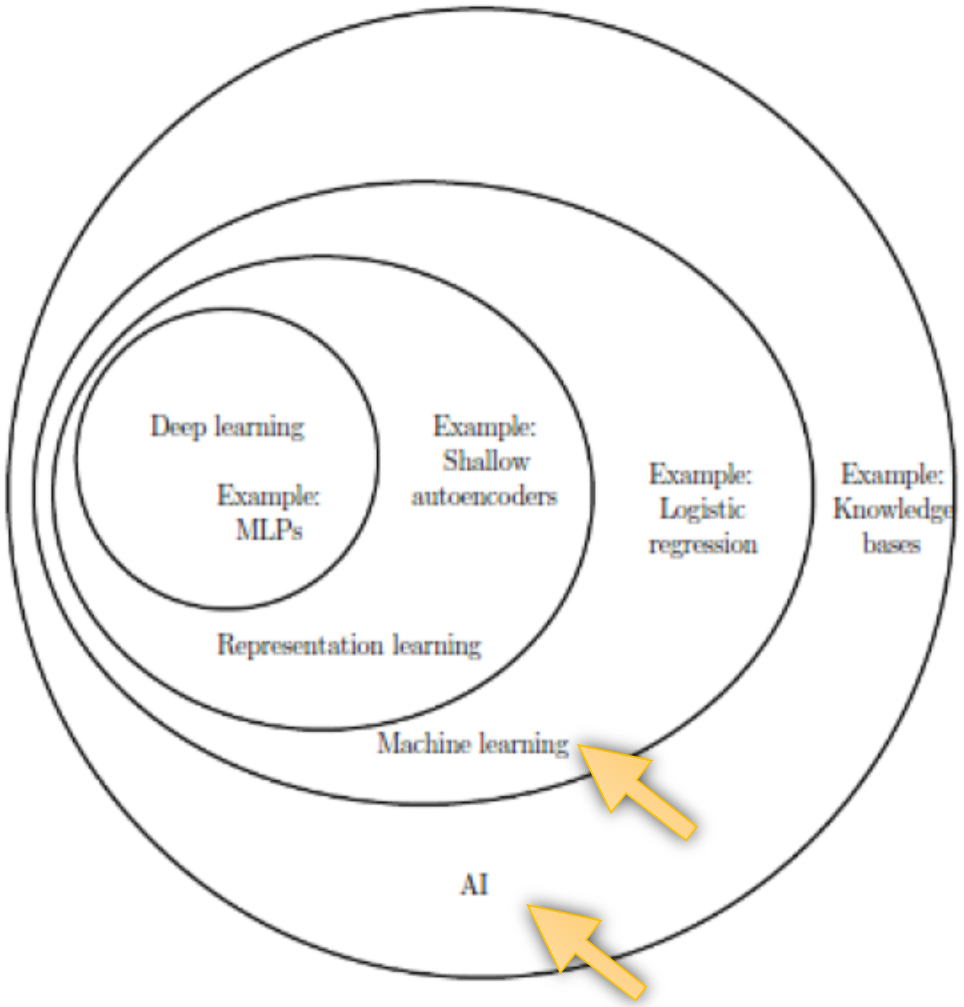
def·i·ni·tion
defəˈniSH(ə)n
noun
a statement of the exact meaning of a word, especially in a dictionary.

1a Machine Learning, Wikipedia Web page, 09 Nov 2017
https://en.wikipedia.org/wiki/Machine_learning
1b Deep Learning, Wikipedia Web page, 09 Nov 2017
https://en.wikipedia.org/wiki/Deep_learning#Definitions
1c What's the Difference Between Artificial Intelligence, Machine Learning, and Deep Learning?, Copeland, Nvidia Web page, 09 Nov 2017
https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

Landscape overview

Landscape and focus[2] areas vary:

- Automation
- Statistics & Probability
- Data Mining
- Artificial Intelligence
- Machine Learning
    - Supervised Learning
    - Unsupervised Learning
    - Reinforcement Learning
- Deep Learning
- Other Areas
    - Deep Reinforcement Learning

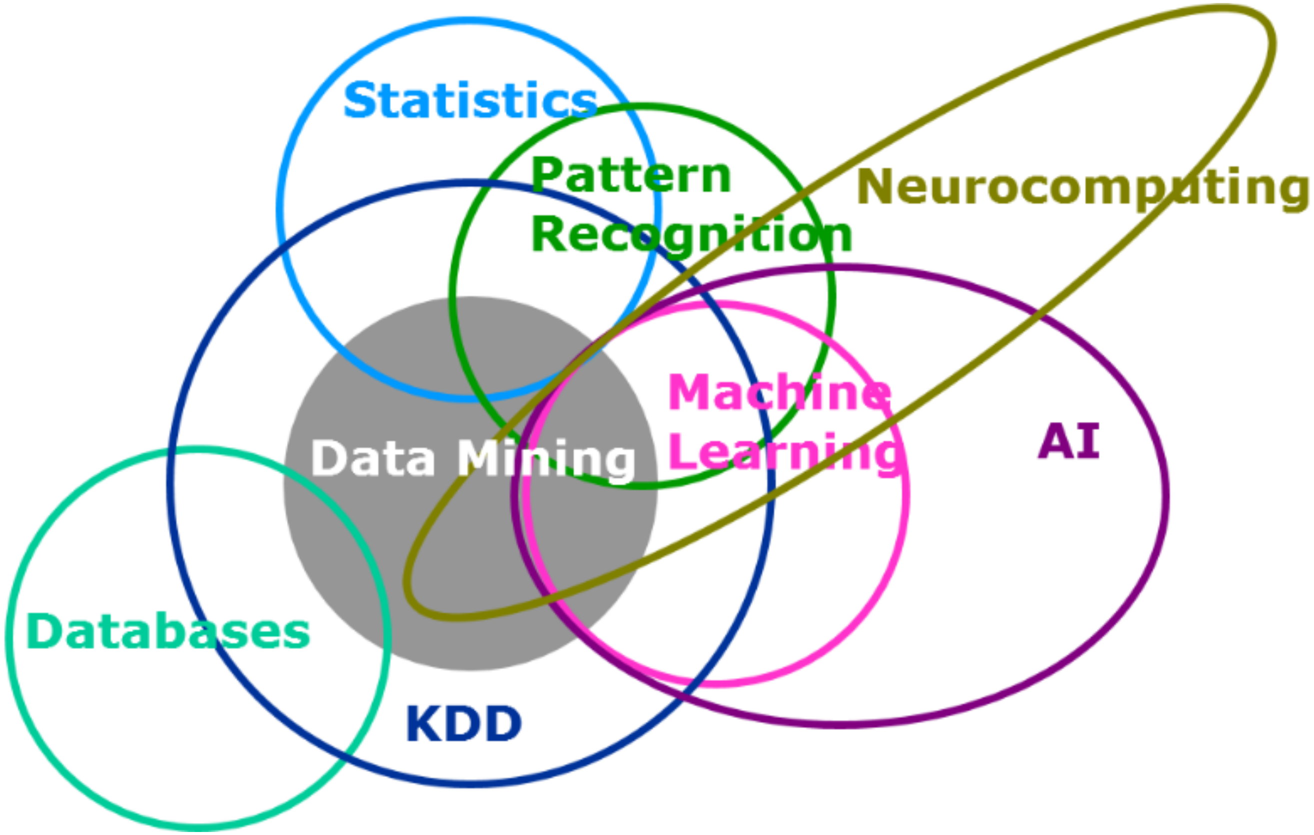


2 Deep Learning, Goodfellow; Bengio; Courville, 2016

http://www.deeplearningbook.org

Landscape overview

Landscape and focus[3] areas of ML also could have overlaps with other scientific disciples such as Data Mining, AI, Big Data an other.



3 Look at Popular Data mining Algorithms and Deep learning, Sharma, 26 Feb 2015 http://https://www.linkedin.com/pulse/look-popular-data-mining-algorithms-deep-learning-sandeep-sharma-me-

Bayes Theorem

- P(A|B) ==> Probability of A given that B
  - Exercise on Bayes[4]
    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, i.e. $email_{3001}$, 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:

$$P(S|E) = \frac{P(E|S)}{P(E|S) + P(E|\bar{S})}$$

4 CSE 240 Home Page, Sproull, Spring 2016
https://www.arl.wustl.edu/~todd/cse240/cse240_lecture17_3-24-16.pdf

Bayes Theorem exercise

- 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?

$$P(S|E) = \frac{P(E|S)}{P(E|S) + P(E|\bar{S})}$$

$$P(S|E) = \frac{\frac{250}{2000}}{\frac{250}{2000} + \frac{5}{1000}} =$$

$$\frac{0.125}{0.125 + 0.005} \approx 0.962$$

Markov Chains

- **Statistics & Probability - Markov Chains**[5]

  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.

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

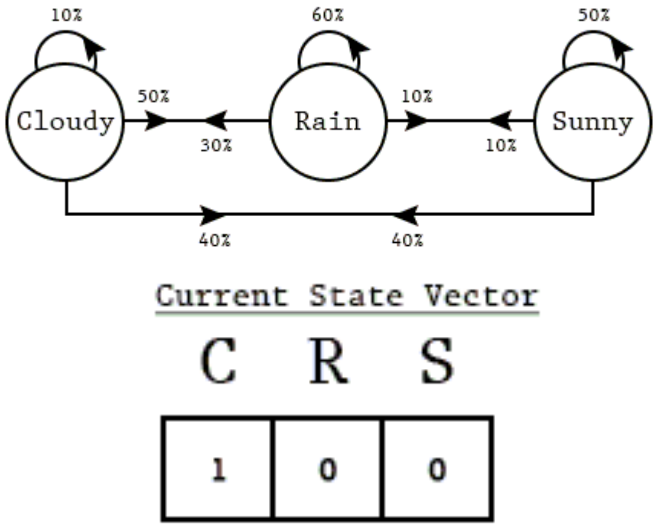


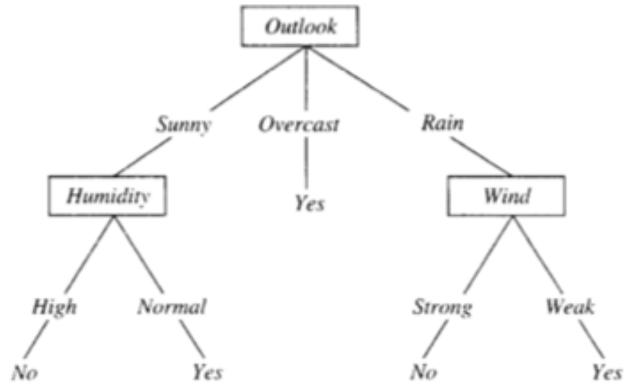
Figure 4

5 Markov Chains Explained, Armstrong, 2015
http://techeffigytutorials.blogspot.com/2015/01/markov-chains-explained.html

Supervised Learning
Decision Trees

- 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)
- Example[7a] dataset converted via algorithm to Decision tree
- Methodology[7b] of is whenever a feature is able to tell us more about our class, it is selected as a node

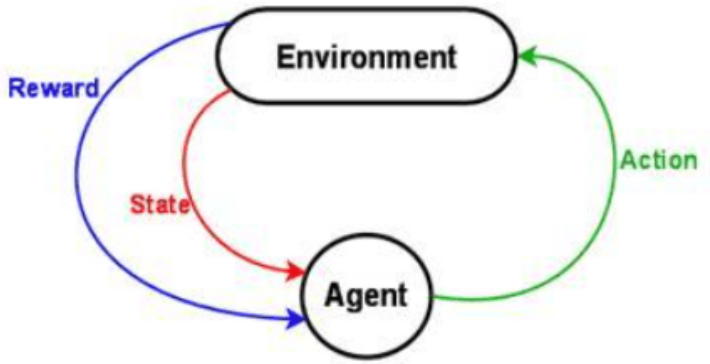
No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	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

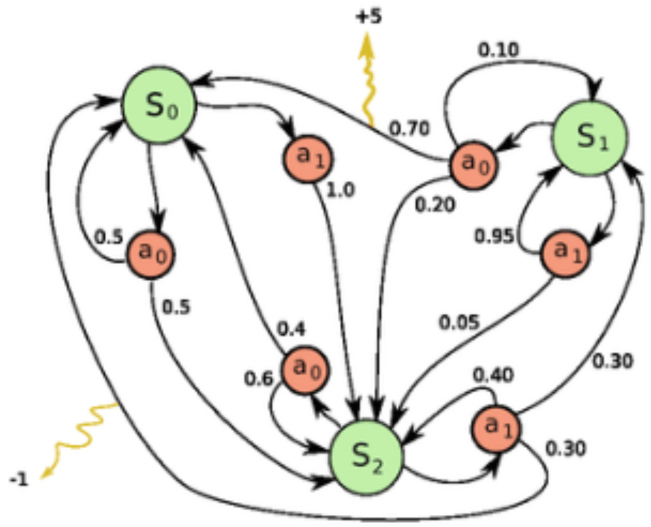


7a Machine Learning, Mitchell, McGraw, 1997.
http://www.cs.princeton.edu/courses/archive/spr07/cos424/papers/mitchell-dectrees.pdf
7b Classification Methods, Padhye, 2017.
http://www.d.umn.edu/~padhy005/Chapter5.html

Reinforcement Learning (RL)

- Reinforcement Learning[8] is learning what to do--how to map situations to actions--so as to maximise a numerical reward signal.
- Reinforcement learning is defined not by characterising learning methods, but by characterising a learning problem.





8 Reinforcement Learning: An Introduction, Sutton; Barto, 1998
https://mitpress.mit.edu/0262193981

Tools – Landscape[11]

Category	Sub-category	Description	Examples
Packages of ML Implementations	Statistical Software Packages	Software toolkits with a large set of implementations of ML algorithms, typically with visualization support	SAS, R, Matlab, SPSS
	Data Mining Toolkits	Software toolkits with a relatively limited set of ML algorithms, typically over a data platform, possibly with incremental maintenance	Weka, AzureML, ODM, MADlib, Mahout, Hazy-Classify
	Developability-oriented Frameworks	Software frameworks and systems that aim to improve developability, typically from academic research	GraphLab, Bismarck, MLBase
	SRL Frameworks	Implementations of statistical relational learning (SRL)	DeepDive
	Deep Learning Systems	Implementations of deep neural networks	Google Brain, Microsoft Adam
	Bayesian Inference Systems	Systems providing scalable inference for Bayesian ML models	SimSQL, Elementary, Tuffy
Linear Algebra- based Systems	Statistical Software Packages	Systems offering an interactive statistical programming environment	SAS, R, Matlab
	R-based Analytics Systems	Systems that provide R or an R-like language for analytics, typically over a data platform, possibly with incremental maintenance	RIOT, ORE, SystemML, LINVIEW
Model Management Systems		Systems that provide querying, versioning, and deployment support	SAS, LongView, Velox
Systems for Feature Engineering		Systems that provide abstractions to make feature engineering easier	Columbus , DeepDive
Systems for Algorithm Selection		Systems that provide abstractions to make algorithm selection easier	MLBase, AzureML
Systems for Parameter Tuning		Systems that provide abstractions to make parameter tuning easier	SAS, R, MLBase, AzureML

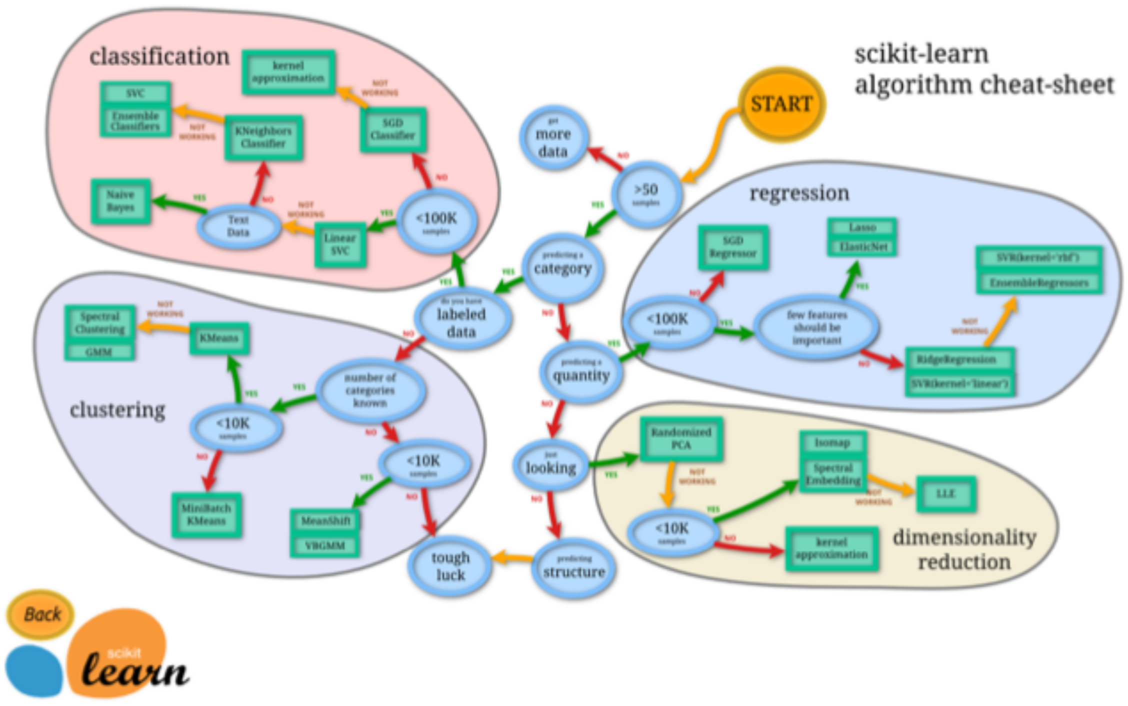
11 A Survey of the Existing Landscape of ML Systems, Kumar; McCann; Naughton; Patel, 27 Nov 2015
https://minds.wisconsin.edu/handle/1793/73834

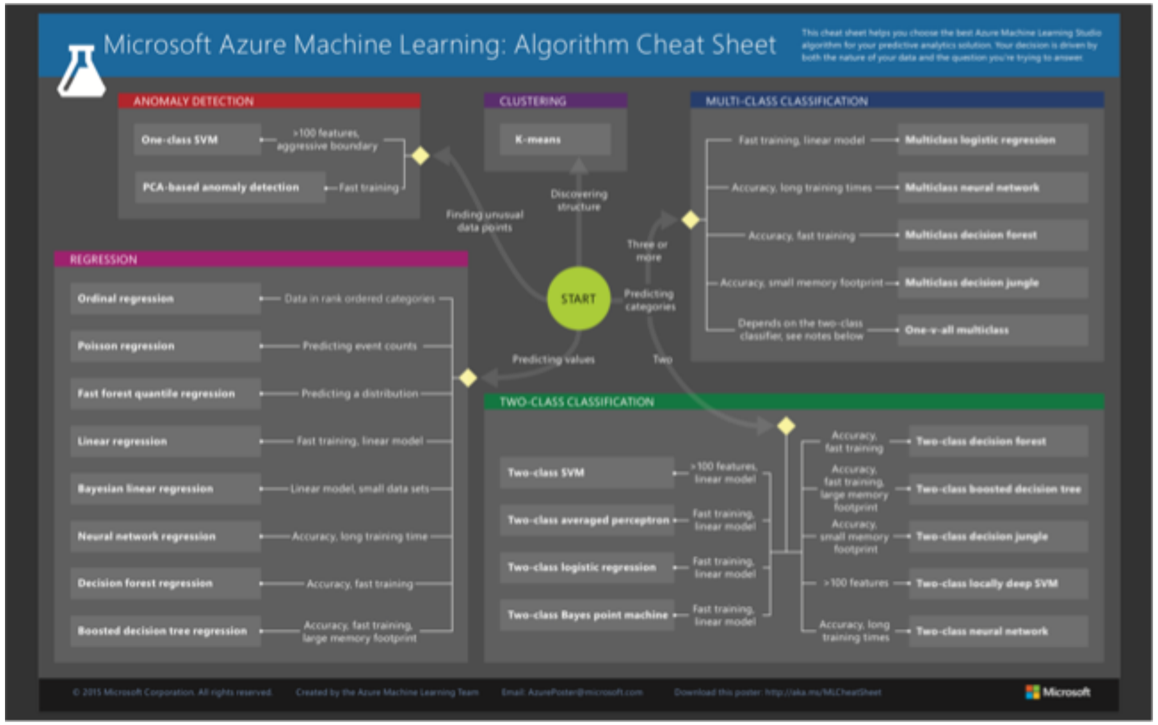
Deep Learning Toolkits comparison[12]

Toolkit	GPU Support	Other
Caffe	Yes	JSON-like text file to describe the network architecture
Deeplearning4j	Yes	Java on Scala API
Tensorflow	Yes	Google backing, high adoption - Python
Theano		Python
Keras		Python - uses Theano or Tensorflow as backend
MXNet	Yes	C++
Lasagne		Python - uses Then
CNTK		VS for ML - developed by Microsoft
DIGITS		Nvidia - web based tool
Torch		Written in C
PyTorch	Yes	Python frontend
Pylearn2		Python
Chainer		

12 Toolkits and Libraries for Deep Learning, Erickson; Korfiatis; Akkus; J Digit Imaging 30: 400., 2017
https://doi.org/10.1007/s10278-017-9965-6

Choose an algorithm[10a, 10b]





10a Microsoft Azure Machine Learning: Algorithm Cheat Sheet, Microsoft website, 09 Nov 2017

http://download.microsoft.com/download/A/6/1/A613E11E-8F9C-424A-B99D-65344785C288/microsoft-machine-learning-algorithm-cheat-sheet-v6.pdf

10b Scikit-Learn Algorithm selection Procedure, Scikit-learn website, 23 Oct 2017

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

Before-Selecting-a-Tool Checklist

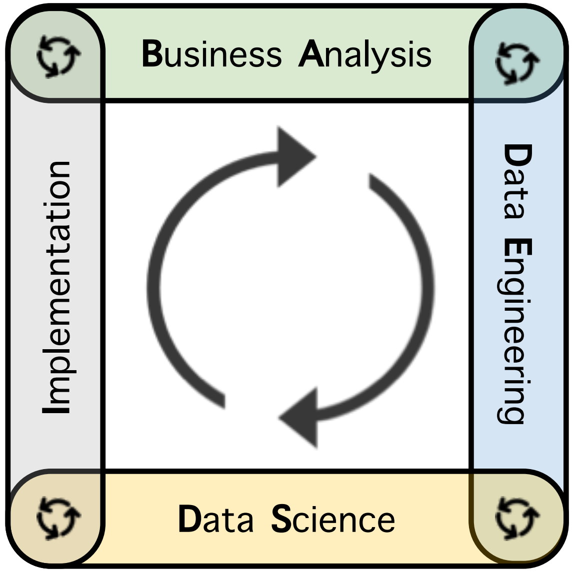
- Things to consider for a toolkit/tool/ecosystem
  - ✓ Environment ease of use
  - ✓ Dev & Exec speed
  - ✓ Training Speed
  - ✓ GPU Support
  - ✓ Community support & contributors
  - ✓ License contamination
  - ✓ Language to be used

Workflow – The ML Pipeline[9]

High Level Step	Short Description	Possible Skills *(actual skills & roles will vary, even from project to project!)*
Define Objective	Start with the question or problem we want to solve	*Analytics, Operational*
Allocate proper data	Find proper data and sources, prepare data set (train/dev/test)	*Data Engineering, Big Data*
Prepare & Evaluate Data	Identify features, flatten data in observations per row, clean, Exploratory Data Analysis	*Data Engineering, Analytics*
Develop Model	Select ML algorithm suitable for selected problem	*Data Science, ML*
Train Model	Train, classify dataset	*Data Science, ML*
Analysis and Testing	Test your model for performance - errors, correct classifications	*Data Science, ML, Analytics*
Deploy, Monitor & Operate	Publish model in live environment	*Development, Operational*
Accuracy Improvement	Evaluate accuracy of predictions/forecasts	*Operational, Analytics*

9 Deep Learning Dissected: The Role of DevOps Teams and Workflows, Adel El-Hallak, 08 Nov 2017
https://thenewstack.io/deep-learning-dissected-devops-teams-workflows/
9 End-to-End Predictive Model in AzureML using Linear Regression, Tejaswi, 15 Nov 2014
https://blogs.msdn.microsoft.com/continuous_learning/2014/11/15/end-to-end-predictive-model-in-azureml-using-linear-regression/
9 The 7 Steps of Machine Learning, Yufeng G, 31 Aug 2017
https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e

Workflow – More practical view on roles



- **Business Analysis** role is responsible to define the problem, verify added value of results.
- **Data Engineering** includes tasks such as collecting data, cleaning, sanitizing and creating working & qualitative data frames and data sets.
- **Data Science** includes the activities needed to derive correlations, information, analysis, forecasts and set an ML / AI agent to decide on its own.
- **Implementation** takes the outcomes from the Data Scientists and verifies it with the Business Analysts to deliver as expected. Also, within the role, activities may include O&M of the Automation & ML.

9 Deep Learning Dissected: The Role of DevOps Teams and Workflows, Adel El-Hallak, 08 Nov 2017
https://thenewstack.io/deep-learning-dissected-devops-teams-workflows/
9 End-to-End Predictive Model in AzureML using Linear Regression, Tejaswi, 15 Nov 2014
https://blogs.msdn.microsoft.com/continuous_learning/2014/11/15/end-to-end-predictive-model-in-azureml-using-linear-regression/
9 The 7 Steps of Machine Learning, Yufeng G, 31 Aug 2017
https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e

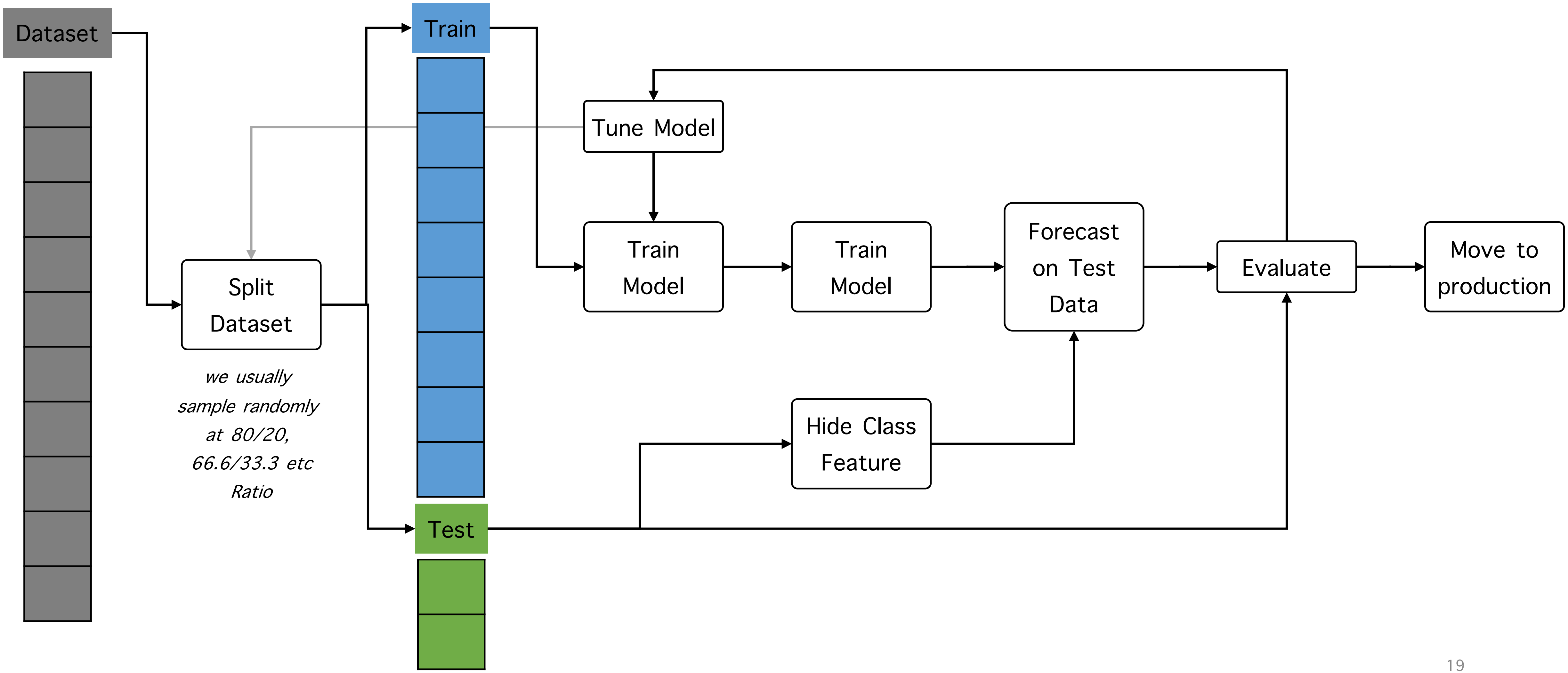
Dataset semantics for supervised learning

(we have a problem and we would like to do predictions!)

- Dataset
  - Features or Labels
  - Labels or Classes
  - Instances or Observations

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

Training/Testing Workflow – simplified view

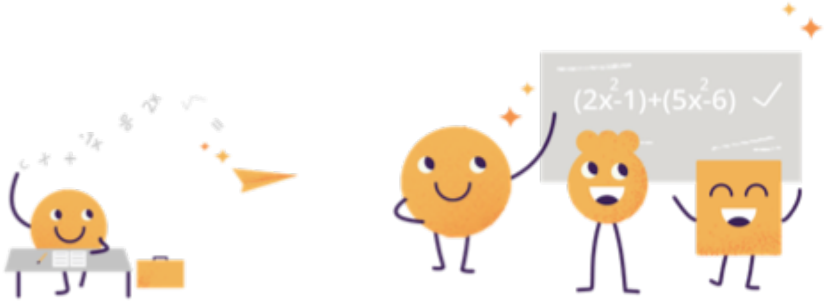


Practice (fun part :-)

- Weka - Supervised Learning - Decision Trees
- R Studio - Basic Statistics on Large Files
- Anaconda Python - Data Frames / Dask / Keras
- Orange Data Mining - SL Example - Predictions

on:

  - bitcoin (prices, open/close in time)
  - cars (values based on various features)
  - flights (features for flights in US 1989-2004)
  - milano_cells (Telecom Italia Milano area cell traffic)
  - Maintenance_data (machine break down prediction) <<<<



Lectures & Data Sources Page
https://github.com/sgez/MLAI

Important takeaways

- Start with the problem
  you are doing the analysis for a reason
- Be persistent
  you will have up's and down's
- Be methodical
  evaluate, be critical, do not be biased

Lectures & Data Sources Page
https://github.com/sgez/MLAI

Now what?
Resources to get you up to speed!

- Data Science Portals
    - Kaggle
    - KDnuggets
- Youtube channels
    - Siraj Raval
    - Stanford
- Online Lessons
    - Coursera
    - Udacity
    - Udemy
    - Datacamp
- Important DL/ML personalities

Thank you!


simply
skills
testing
being
argue
target
pessimists
best
health
optimists
psychology
people
positive
style
think
success
mindset
negative
attitude
optimism
research
thinking
life
mind
possible
logic
philosophy