Machine Learning
For Natural Language Processing

Abdelhak Mahmoudi
abdelhak.mahmoudi@um5.ac.ma

2020

Content

1. Introduction
2. Machine Learning

   1. Supervised Learning, 2. Unsupervised Learning
3. Natural Language Processing

   1. Regular Expressions, 2. Tokenization, 3. Character Encoding, 4. Part-of-Speech Tagging, 5. Chunking, 6. Stemming and Lemmatization, 7. Parsing, 8. Named Entity Recognition, 9. Topic Segmentation
4. Introduction to Deep Learning for NLP

   1. Sequence models, 2. Embeddings, 3. BERT models

Introduction

- From Programming to Machine Learning!
- Definitions
- Terminologies
- How can I Apply?
- How can I Learn?

Motivation

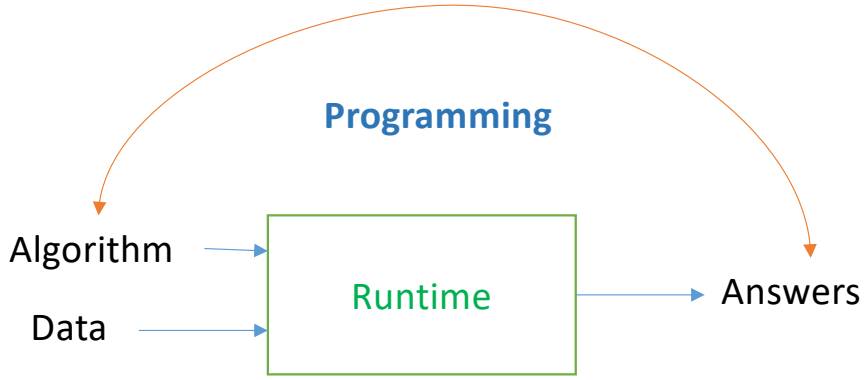
Forbes (2016): "The Top 10 AI And Machine Learning Use Cases Everyone Should Know About"

1. Data Security,
2. Personal Security,
3. Financial Trading,
4. Healthcare,
5. Marketing personalization,
6. Fraud Detection,
7. Recommendations,
8. Online Search,
9. **Natural Language Processing (NLP),**
10. Smart Cars

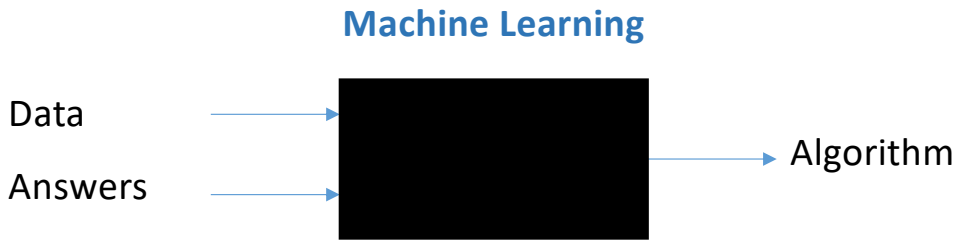
NLP Applications

From Programming to Machine Learning!

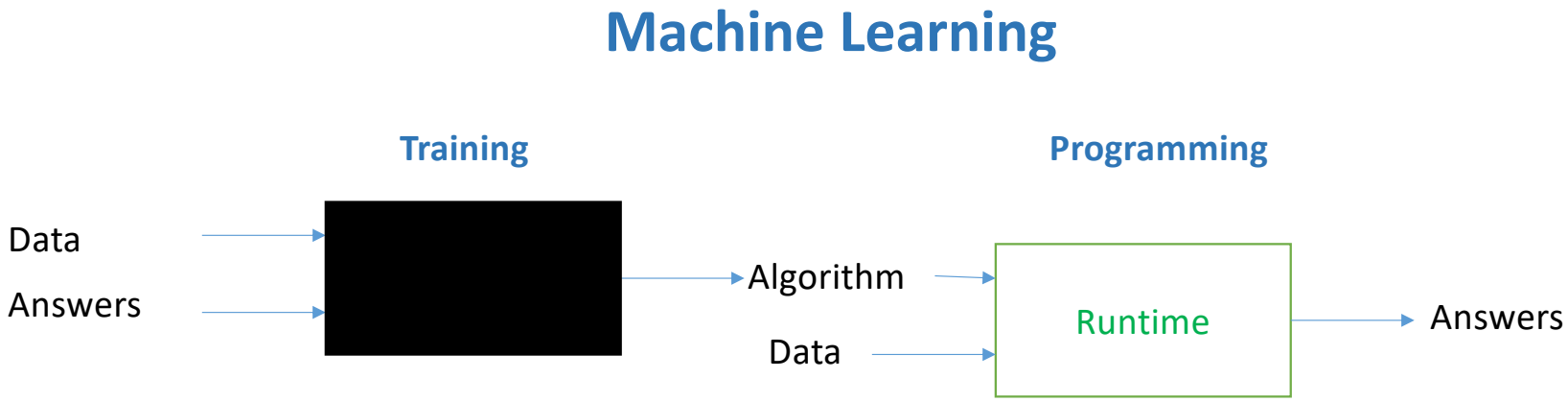
From Programming to Machine Learning!



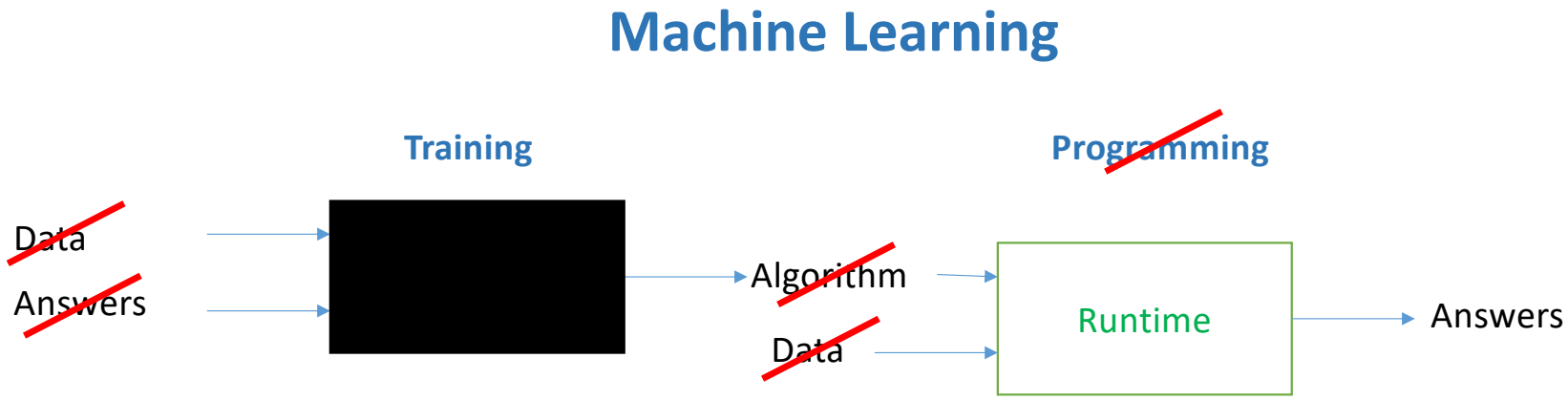
From Programming to Machine Learning!



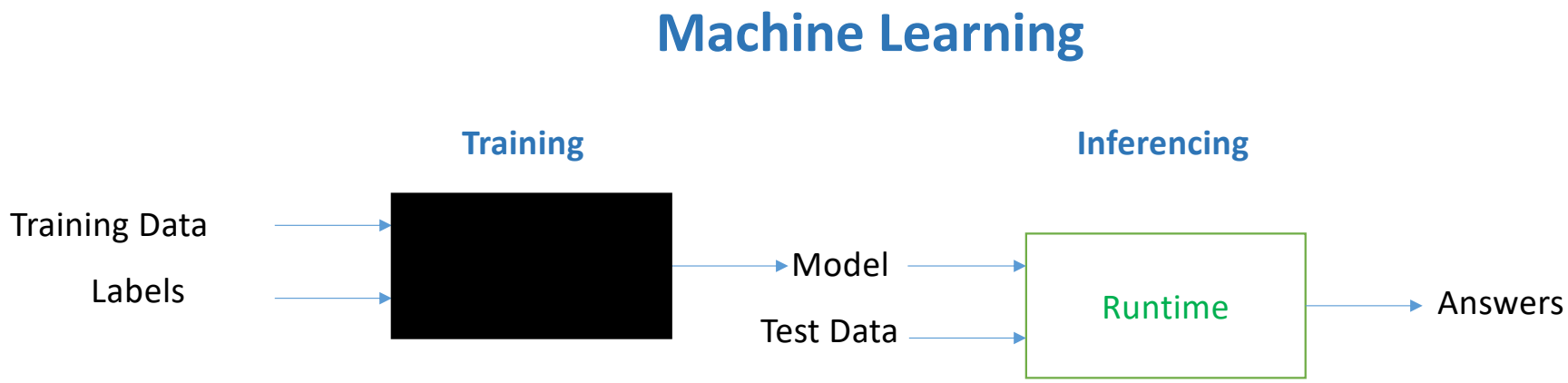
From Programming to Machine Learning!



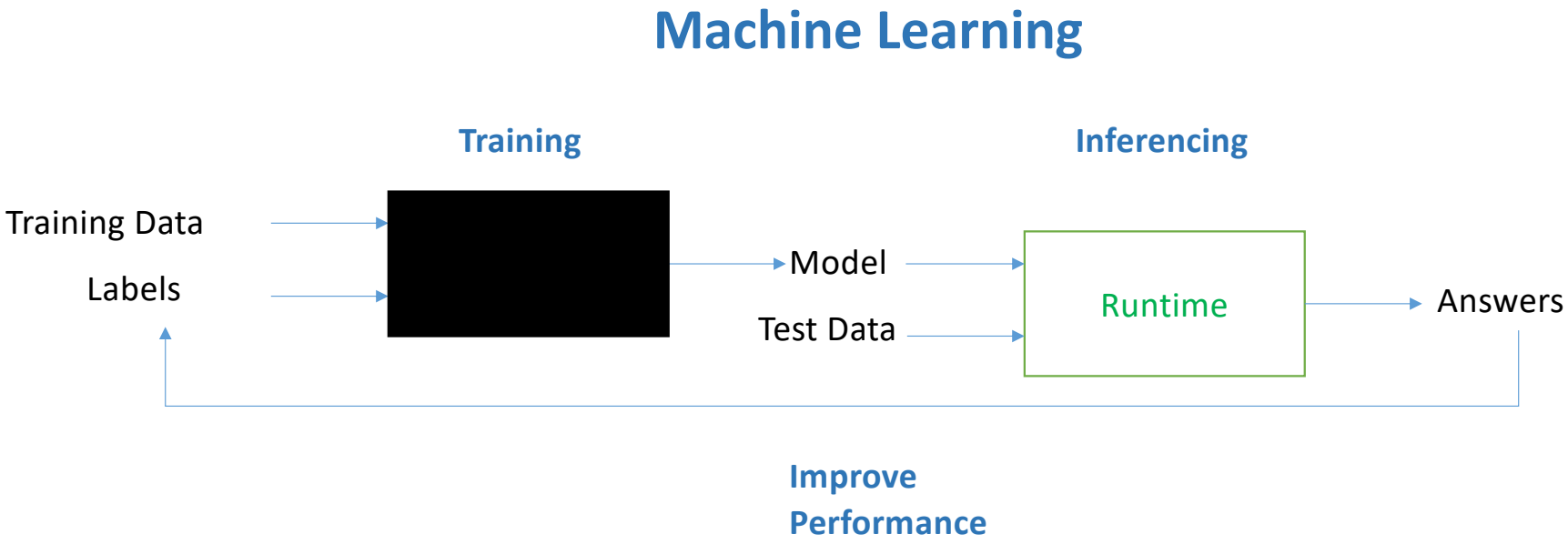
From Programming to Machine Learning!



From Programming to Machine Learning!

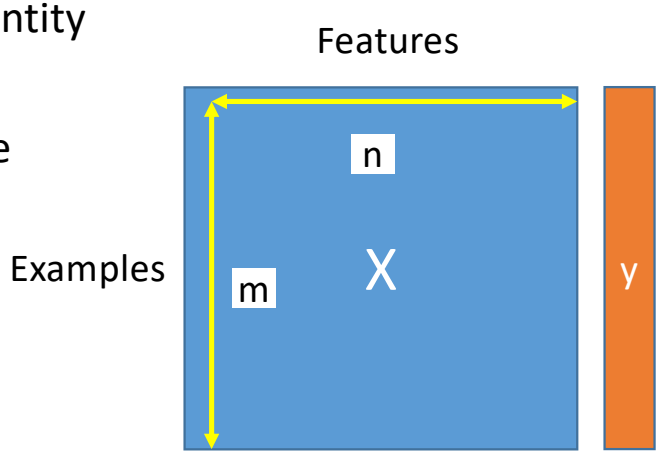


From Programming to Machine Learning!



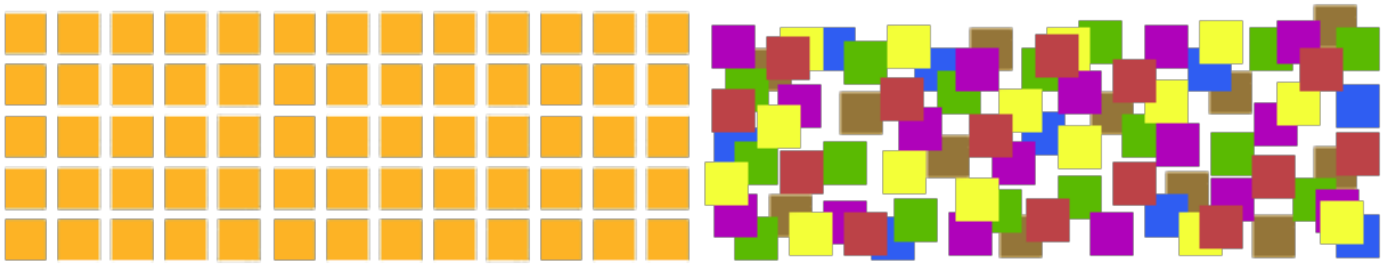
Definitions

- Data
  - Example $x^{(i)}$
    - Row/Instance/Input/Observation/Record/Point/Sample/Entity
  - Feature $x^{(i)}_j$
    - Columns/Variable/Predictor/Characteristic/Field/Attribute
    - Quantitative (numeric, continue)
    - Qualitative (textual, category)
  - Dimension, Visualization
    - m Examples: i = 1..m
    - n Features: j = 1..n
  - Output : $y_i = x^{(i)}_k$ (k in 1..n)
    - target/class/output
    - For each example (0/1)



Definitions

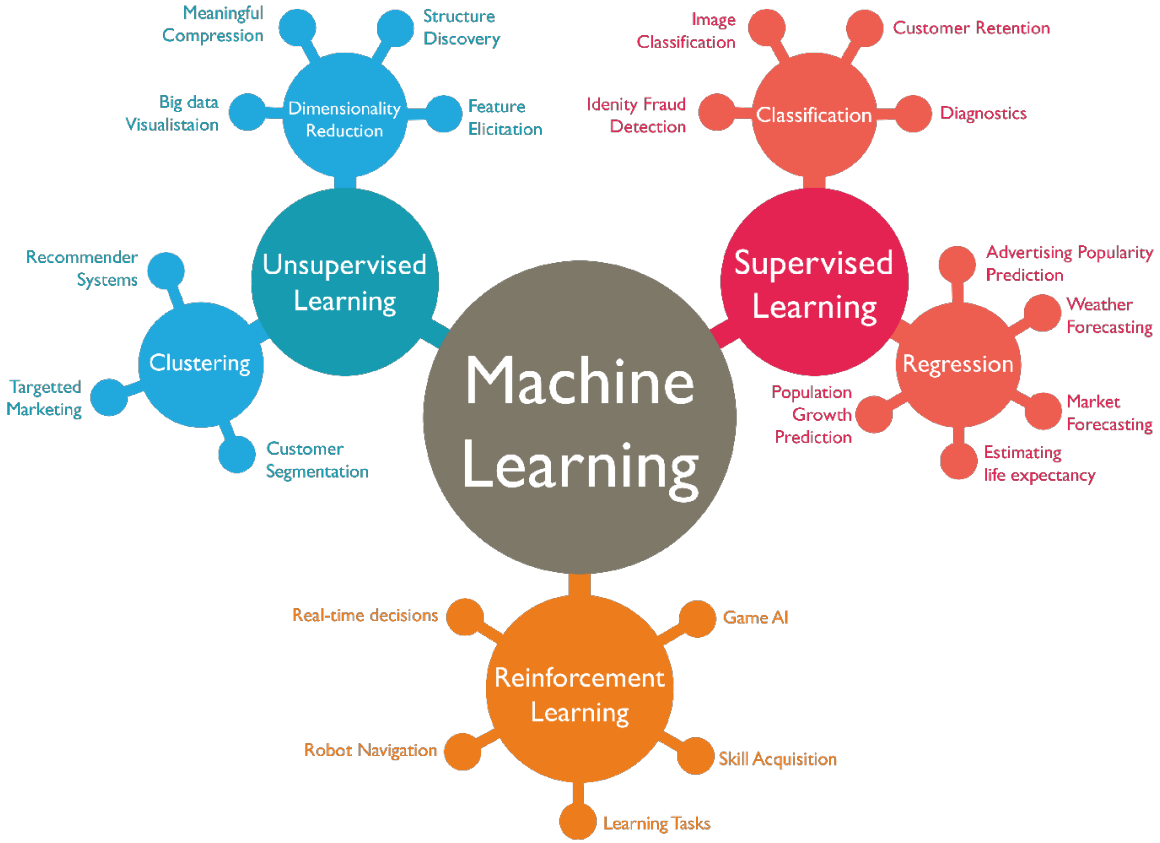
- Data
  - Structured
    - CSV, XML, JSON, XLSX, etc.
  - Unstructured
    - DOC, HTML, PDF, PNG, MP3, MP4, etc.



Text, Image, son

Definitions

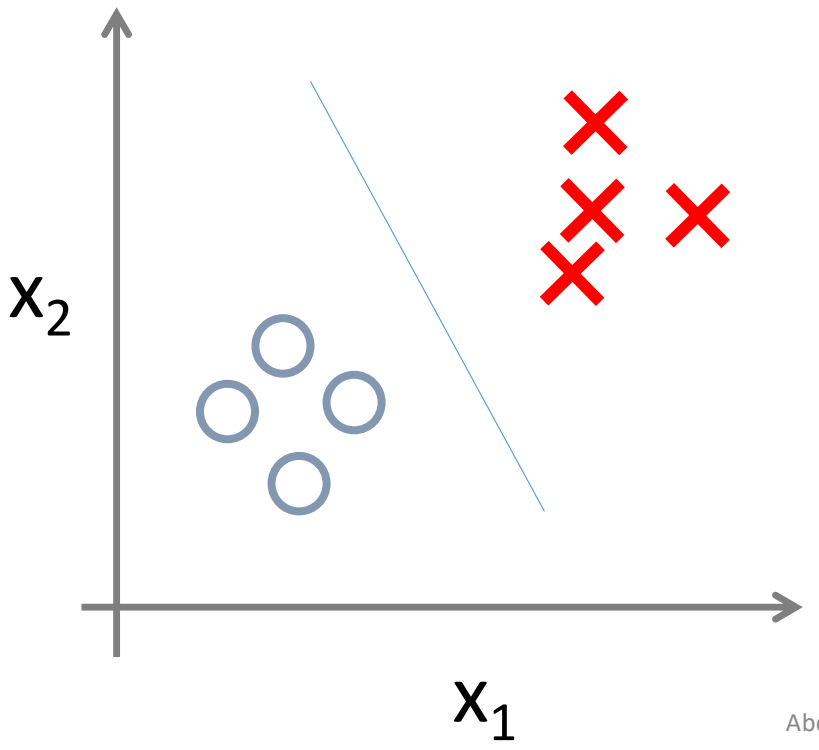
- Types of Learning
  - Supervised
    - Classification
    - Regression
  - Unsupervised
    - Dimensionality Reduction
    - Clustering
  - Semi-supervised
    - Little supervised data
  - Reinforcement



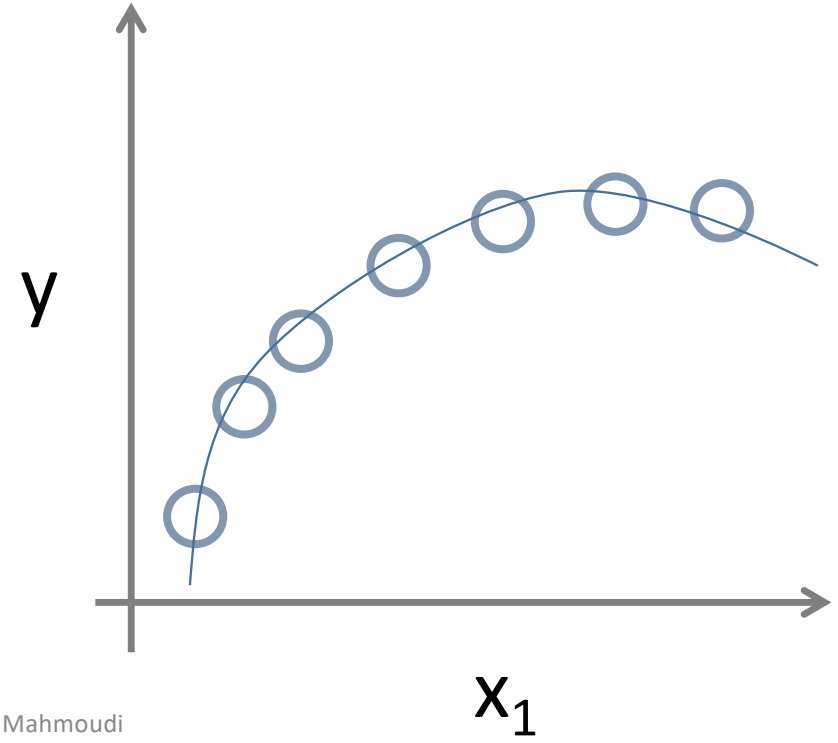
Definitions

Supervised Learning

Classification (y is discrete)



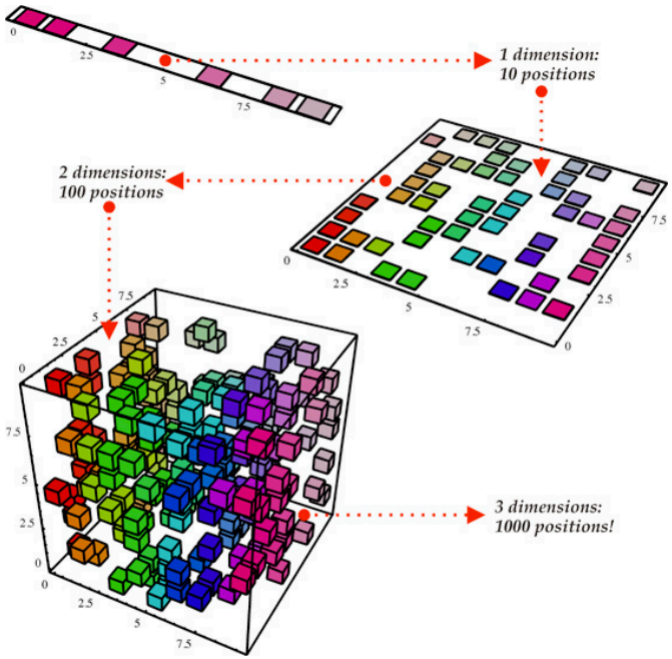
Regression (y is continuous)



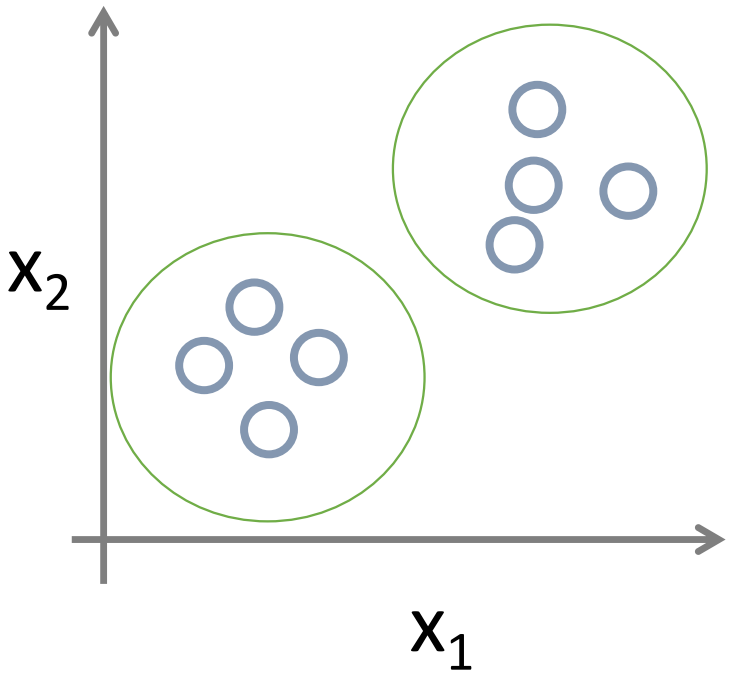
Definitions

Unsupervised Learning (y absent)

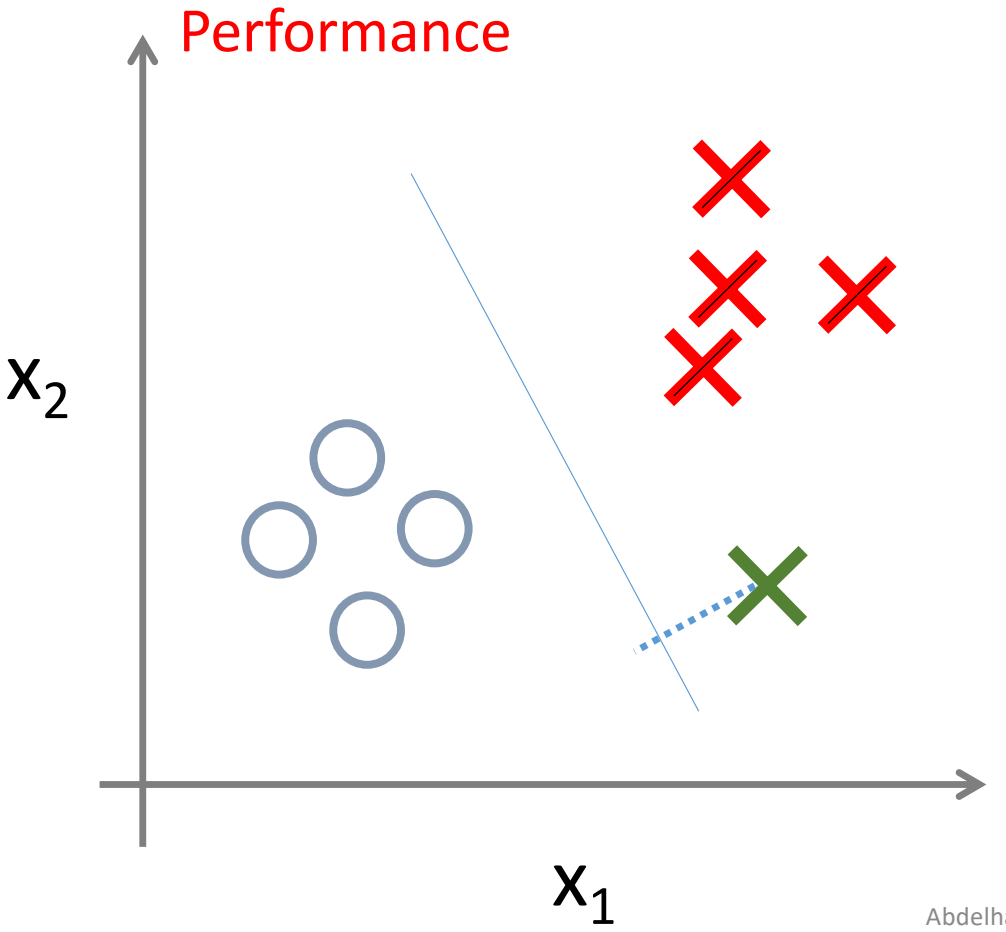
Dimensionality Reduction

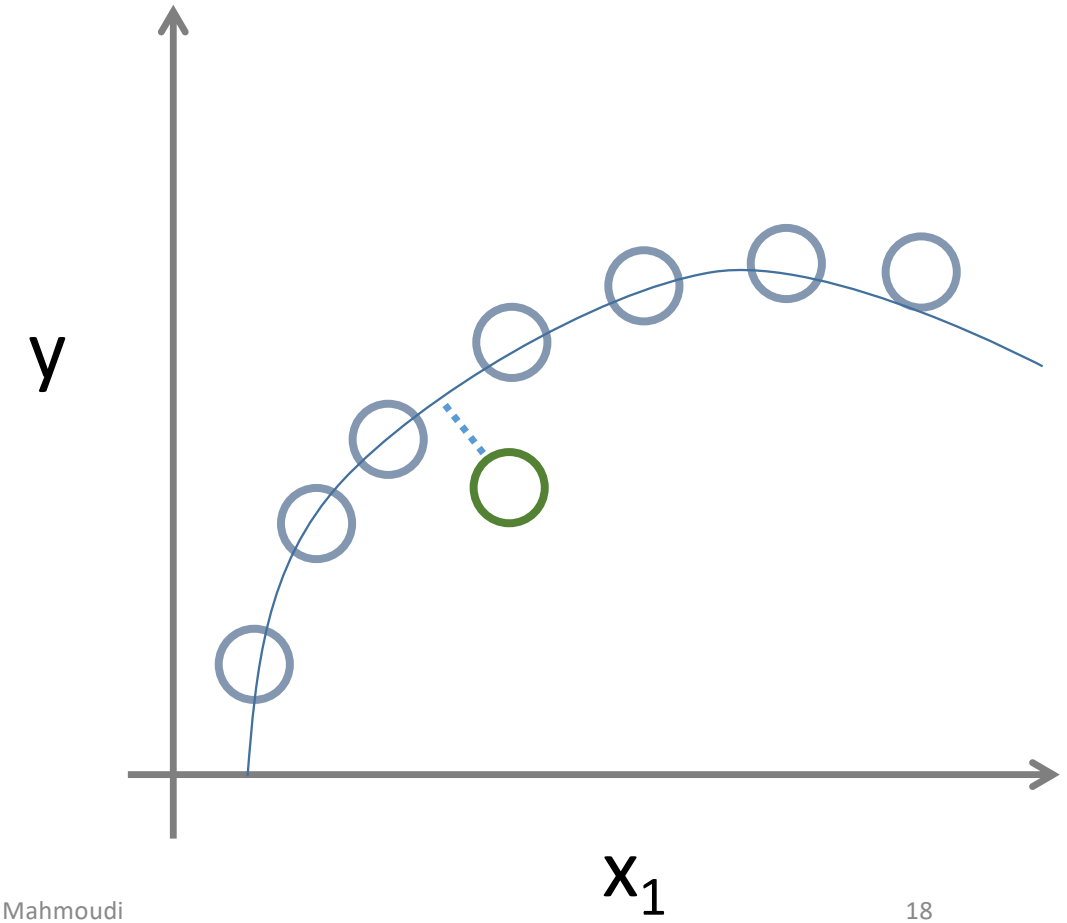


Clustering

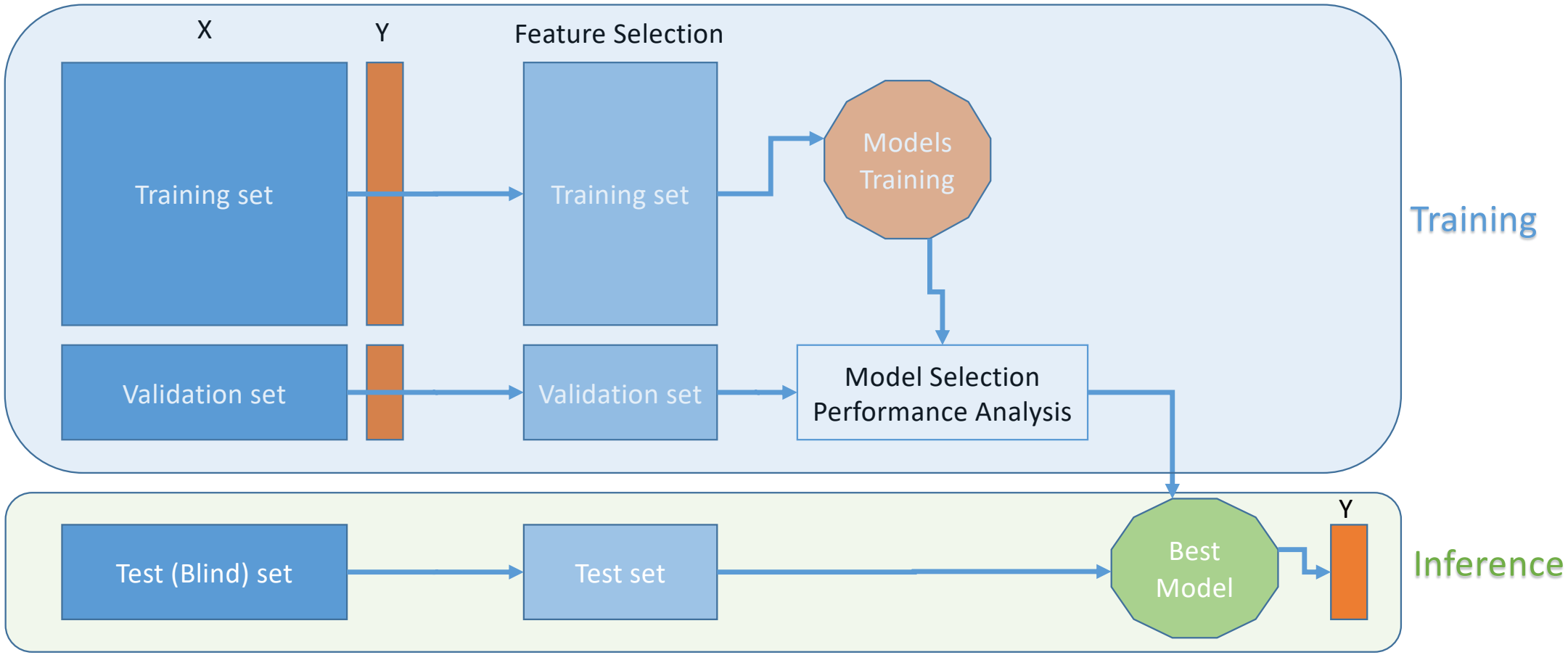


Definitions





All in one picture



Terminologies

- Artificial Intelligence
- Machine Learning, Deep Learning
- Statistical Learning
- Data Mining
- Deep Learning
- Natural Language Processing

Artificial Intelligence (1943)

- “The first work that is now generally recognized as AI was McCullouch and Pitts' 1943 formal design for Turing-complete "artificial neurons". Wikipedia
- Intelligent Machines mimics Natural Intelligence (NI)
- Natural Intelligence (General Intelligence)
  - Reasoning, Problem solving,
  - Knowledge representation, Learning,
  - Planning, Perception, Motion and manipulation, Natural Language
  - Etc.

Machine Learning (1959)

- "Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "Machine Learning" in 1959 while at IBM". Wikipedia
- A subfield of Computer Science and Artificial Intelligence which deals with building systems that can learn from data, instead of explicitly programmed instructions.
- Artificial Neural Networks (**1975**)
  - Begin in 1943, stagnated in 1969, relaunched in 1975 by the Backpropagation algorithm,
- Book: "Machine Learning". Tom M. Mitchell. 1997

Statistical Learning (1968)

- VC Theory. “On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities". Vapnik, V. N.; Chervonenkis, A. Ya, 1968
- A subfield of Mathematics which deals with finding relationship between variables to predict an outcome
- Support Vector Machines (**1995**)
  - Much simpler, overtook ANN, Vapnik V. N.
- Book
  - “An introduction to statistical learning with applications in R” (1st Edition 2013). Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani.

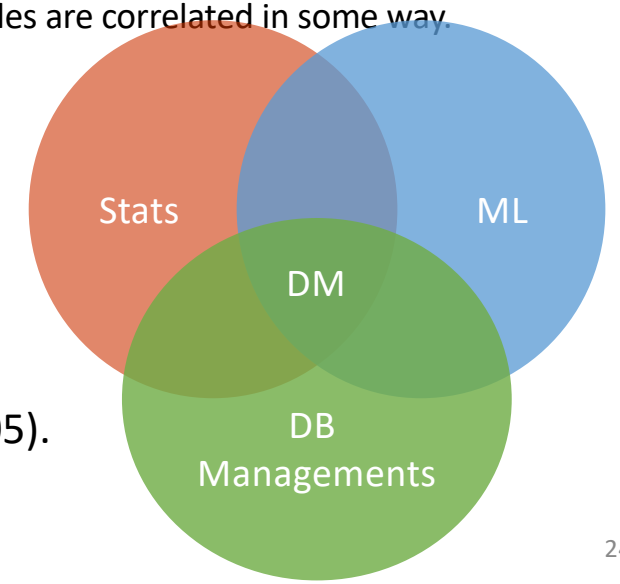
Data Mining (1990)

Appeared in the database and financial community to recognize customer and products trends

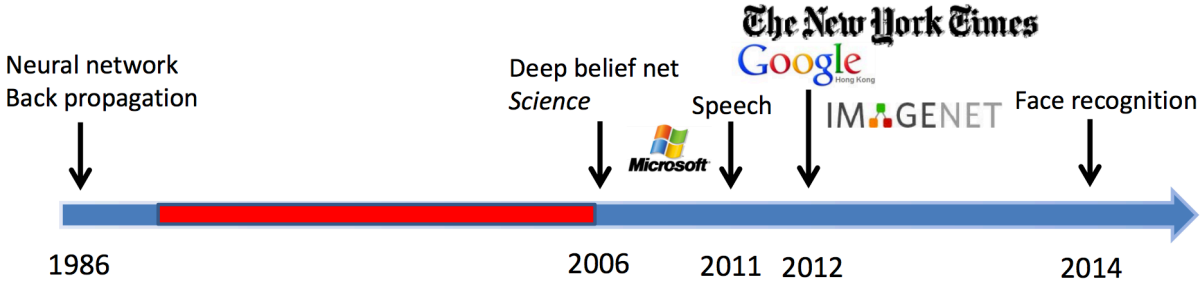
Definition : "The process of automatically discovering useful information in large repositories".

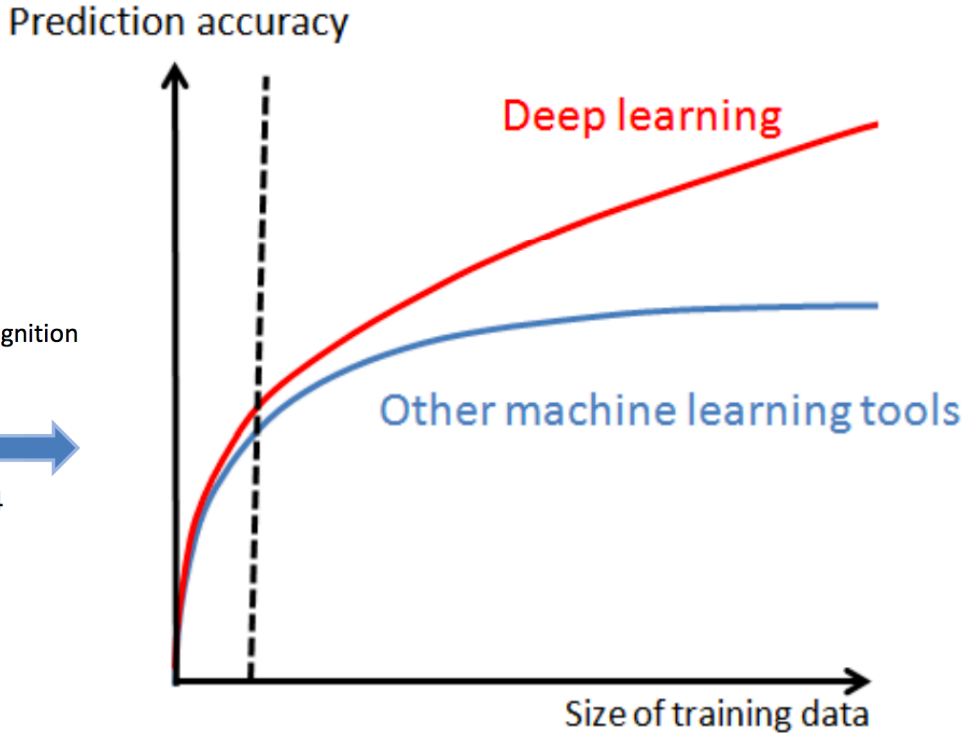
- Automatically
  - Stats: correlation between 2 variables, what is the problem?
  - DM: parallel correlation between 1000 variables, send and email if two variables are correlated in some way.
- Discovering useful information
  - Stats: answer a specific question
  - DM: look for any specific reason
- Large Repositories
  - Stats: Collect data to answer a specific question
  - DM: Collect all, you don't know the reason yet!

Book:Introduction to Data Mining (2nd edition 2018, 1st Edition in 2005). Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar



Deep Learning





Natural Language Processing

- **NLP** is a subfield of **linguistics**, **computer science**, **information engineering**, and **AI** concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.
- **1950**: **Turing** Test
- **1954**: Machine **Translation** (Russian -> English)
- **1966**: ELIZA computer psychotherapist
- **70s, 80s**: hand-written **rules** Chatbots (PARRY, Racter, and Jabberwacky)
- **Early 80s – 2010**:
  - From real based models (decision trees: complex if-then rules) to statistical models (probabilistic decisions)
  - IBM statistical machine translation (SMT) (Hidden Markov Models, language corpus, linguistics)
- **2010-2020**: Deep Learning models (Word embeddings, Neural Machine Translation, etc.)

How can I Learn?

- Math
  - Statistics, Probabilistic Graphical Models, Algebra, Optimization
- Programming Languages
  - Python, R, Julia !
- Books
  - Gilbert Strang, Linear Algebra and Learning from Data. 2018.
  - Ian Goodfellow et al. “Deep Learning”. 2016
  - Aurélien Géron. “Hands on ML with sklearn”. 2017
  - Gareth James et al., “An introduction to statistical learning with R”. 2013
  - Speech and Language Processing (Jurafsky and Martin)
  - Web...

How can I Learn?

- MOOCs
  - Coursera.org, Fast.ai, Udemy.com, ocw.mit.edu,, etc.
- StackOverflow
- Research Papers
  - Read and rewrite algorithms from scratch
- Follow People:
  - ML/ DL: Androw Ng, Yann LeCun, Jeff Hinton, Sebastian Thrun, Yoshua Bengio, etc.
  - NLP: Chris Manning, Dan Jurafsky, etc.

How can I Apply?

- Start small projects and use Framworks
  - Scikit-learn, TensorFlow, Keras, Pytorch, Caffe, Microsoft Cognitive Toolkit (CNTK), MXNet, Spark MLlib, etc.
- Challenge your self
  - Find data: Web, UCI Machine Learning Repo
  - Go for competitions: Kaggle, DrivenData, Zindi
- Github
  - Find codes
  - Share your code
- Softwares (for non-pro !)
  - Knime, IBM SPSS Modeler