Introduction

- If you use or reuse this document or part of it; you need to give the credit to
  « J. Mothe (2018) » or

- Objectives:
  - What machine learning covers?
  - Basics in machine learning

- Plan
  - Introduction
  - Data representation for ML
  - ML approaches
  - Applications
  - Books and MOOCs

Your turn!

I. INTRODUCTION

Machine Learning

- What does Machine Learning cover?
- What is it linked to?
Machine Learning

- Volume
- Variety
- Velocity
- Veracity

Data Mining

- Un-supervised
- Descriptive analysis

Machine Learning

- Data features
- Matrices

Data representation

Visualisation

Machine Learning

- Supervised: Train / test
- Predictive analysis
- Interpretation
- Graphs
- Synthetic views

Is part of a more general process

Information extraction

Model extraction

Data representation

Data mining

Data visualisation

Interpretation

Results

Machine Learning

- Machine learning designs and studies algorithms that can learn from data and make predictions on data

Training / Learning

- Annotated data: Examples for which we know the decision

Testing / Predicting

- Non-annotated data: for which we want the decision
### Machine Learning

#### Training / Learning
- Annotated data: Examples for which we know the decision

<table>
<thead>
<tr>
<th>Ind</th>
<th>Weight</th>
<th>Height</th>
<th>Hair</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>20</td>
<td>1.12</td>
<td>Blond</td>
<td>Child</td>
</tr>
<tr>
<td>I2</td>
<td>75</td>
<td>1.80</td>
<td>Brown</td>
<td>Adult</td>
</tr>
<tr>
<td>I3</td>
<td>80</td>
<td>1.74</td>
<td>Brown</td>
<td>Adult</td>
</tr>
<tr>
<td>I4</td>
<td>18</td>
<td>0.80</td>
<td>Brown</td>
<td>Child</td>
</tr>
</tbody>
</table>

#### Testing / Predicting
- Non-annotated data: for which we want the decision

<table>
<thead>
<tr>
<th>Ind</th>
<th>Weight</th>
<th>Height</th>
<th>Hair</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>I8</td>
<td>15</td>
<td>0.87</td>
<td>Blond</td>
<td>?</td>
</tr>
</tbody>
</table>

### II. DATA REPRESENTATION FOR MACHINE LEARNING

- Indexing
- Information Extraction

- Quality
- Cleaning, sampling, completeness

- Individuals / Features - Variables
- Matrices or Vectors

### Data representation

- Depends on the data type
- Text – non structured
  - Indexing = term extraction
  - Vectors
- Images
  - Histograms of colors
  - CNN
- Structured data
  - Vectors: characteristics of each individuals
  - Matrices
Forms of data representation

• Vectors or matrices
  • For each text document the terms it is composed of
  • For each person her characteristics or feature values
  • For each individual the value of each variable

Variable & information representation

Types of variables / Features

• Quantitative Variable
  Numerical feature
  Arithmetic's can be applied with meaningful results.

• Examples
  Age
  Temperature

• Continuous vs discrete
  – Discrete: variable with a finite number of values
  – Continuous: variable with an infinite number of values
Types of variables / Features

• Qualitative Variable
  Non-numerical feature
  Arithmetic's is not meaningful

• Examples
  Color of eyes
  Level of satisfaction
  – Nominal: categories with no order
  – Ordinal: categories can be ordered/ranked

Type of variables

• Number of failures of a system
• Time needed to process a task
• Color of eyes
• Answer to the question: « do you think the system answers correctly »?

Variables

• Number of failures of a system
• Time needed to process a task
• Color of eyes
• Quantitative & discrete
• Quantitative & continuous
• Qualitative & nominal
• Qualitative & ordinal

Inspired from ebrunelle.ep.profweb.qc.ca/MD/Chapitre2.pdf
## III. MACHINE LEARNING APPROACHES

### Machine learning approaches

<table>
<thead>
<tr>
<th>Descriptive analysis</th>
<th>Predictive analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Analysis of the data</td>
<td>• Analysis of the training set</td>
</tr>
<tr>
<td>• No train/test</td>
<td>• Extract the model</td>
</tr>
<tr>
<td>• Extract the model from the data</td>
<td>• Use the model on the test or data for which we need a prediction</td>
</tr>
</tbody>
</table>

#### Data quality

- Completeness of the training data set
- Lack of examples

- SUN
- BALLON
Data quality

- Completeness of the features
- Lack of values in training or testing

<table>
<thead>
<tr>
<th>Ind</th>
<th>Weight</th>
<th>Height</th>
<th>Hair</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.12</td>
<td>Blond</td>
<td>Child</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>1.80</td>
<td>Brown</td>
<td>Adult</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
<td>1.74</td>
<td>Adult</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>Brown</td>
<td>Child</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ind</th>
<th>Weight</th>
<th>Height</th>
<th>Hair</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1.30</td>
<td>Blond</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

Over-fitting

Simple/Multiple linear regression

- Simple: Relationships between two continuous (quantitative) variables
  - One (simple) variable is regarded as the predictor/explanatory/independent variable
  - The other variable is regarded as the response/outcome/dependent variable
  - Extract the trend / relationship that may exist between the predictor and the response
  - Linear
Simple/Multiple linear regression

• fitted using the least squares (minimizes the sum of squared residuals) approach (or other)

[Image of regression plot]

https://onlinecourses.science.psu.edu/stat501/node/252

Decision Tree

• The model predicts the value of a target (response) variable by learning simple decision rules inferred from the data features.

• Both categorical and continuous input/predictor and output/response variables

[Image of decision tree]

https://onlinecourses.science.psu.edu/stat501

https://r-tutor.com/elementary-statistics

Decision Trees

• CART (Classification And Regression Tree)

• Binary tree

• A greedy approach is used to divide the space: recursive binary splitting.

• All the values are lined up and different split points are tried and tested using a cost function. The split with the lowest cost is selected.

• Regression predictive modeling: the cost function is the sum squared error across all training samples

https://medium.com/@haydar_ai/learning-data-science-day-12-decision-tree-fa83499f89fd

Decision Tree

- More?

---

SVM

- Support Vector Machine
  - Binary classification (multi-classes too)

[www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf](www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

---

SVM

![SVM](www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

- More?
  - [www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf](www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf) (slides)
  - [cs229.stanford.edu/notes/cs229-notes3.pdf](cs229.stanford.edu/notes/cs229-notes3.pdf) (textual)
From NN to Deep Learning

- Perceptron
  - Unique neuron
  - Learns the connection weights
  - Using examples
  - Input
  - Expected output

From NN to Deep Learning

- Multi layer NN

From NN to Deep Learning

- Convolutional NN

Want to know more?
- Udacity (Google) https://www.udacity.com/course/deep-learning--ud730
- Coding http://course.fast.ai/
- Tensorflow library: https://www.tensorflow.org/
- Keras: https://elitedatascience.com/keras-tutorial-deep-learning-in-python
Evaluation in supervised methods

- Cross validation
  - A training set (to calculate the model)
  - A testing set (to check the results)
  - [A validation set]
- How to choose the training/testing sets?
  - 80 / 20

Evaluation in supervised methods

- 10-folds cross validation

Evaluation in supervised methods

- Leave-one out

A few non-supervised methods

http://www.nosimpler.me/machine-learning/
Agglomerative clustering

- Needs a distance

<table>
<thead>
<tr>
<th>Nom</th>
<th>Paramètre</th>
<th>Fonction</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance de Manhattan</td>
<td>1-distance</td>
<td>$\sum_{i=1}^{n}</td>
</tr>
<tr>
<td>distance euclidienne</td>
<td>2-distance</td>
<td>$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$</td>
</tr>
<tr>
<td>distance de Minkowski</td>
<td>$p$-distance</td>
<td>$\sum_{i=1}^{n}</td>
</tr>
</tbody>
</table>

- Needs an aggregation criteria

- Simple linkage
- Average linkage
- Complete linkage
Agglomerative clustering

- Ward criteria
  - minimum of the intra-cluster inertia or
  - maximum of the inter-group inertia

K-means and variants

- Initialisation:
  - Choose Q group centroids
  - Cluster the objects (closest centroid)
  - First partition
  - Repeat
    - Calculate the new centroids considering the objects in the cluster
    - Reclude the objects
  - Until either no changes, or a certain number of iterations, or …

K-means and variants

- Parameters - variants
  - Choice of the initial Q centroids
  - Distance « closest »
  - New centroids

Others

- Factorial analysis
  - Principal Component Analysis
  - Correspondance Analysis
  - Search for the best axis in order to visualize a N dimensional space into a P dimensional space P << N
  - Based on singular value decomposition
  - Graphical visualization of the results
IV. WANT TO KNOW MORE?

MOOCs
Books
Practical works

MOOCs
- Machine Learning:
  - Andrew Ng, Co-founder, Coursera; Adjunct Professor, Stanford University; formerly head of Baidu AI Group/Google Brain

MOOCs
- Machine Learning:
  - https://lagunita.stanford.edu/courses/HumanitiesSciences/StatLearning/Winter2016/about
  - This is an introductory-level course in supervised learning, with a focus on regression and classification methods. The syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis; cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines. Some unsupervised learning methods are discussed: principal components and clustering (k-means and hierarchical).
MOOCs

- Deep learning: https://www.udacity.com/course/deep-learning--ud730
  We’ll show you how to train and optimize basic neural networks, convolutional neural networks, and long short term memory networks. Complete learning systems in TensorFlow will be introduced via projects and assignments. You will learn to solve new classes of problems that were once thought prohibitively challenging, and come to better appreciate the complex nature of human intelligence as you solve these same problems effortlessly using deep learning methods.

  Vincent Vanhoucke, Principal Scientist at Google

Practice ML with R

  This small tutorial is meant to introduce you to the basics of machine learning in R; it will show you how to use R to work with KNN

  In this step-by-step tutorial you will:
  1. Download and install R and get the most useful package for machine learning in R.
  2. Load a dataset and understand it's structure using statistical summaries and data visualization.
  3. Create 6 machine learning models, pick the best and build confidence that the accuracy is reliable.

Practice ML with Python

  In this step-by-step tutorial you will:
  1. Download and install Python SciPy and get the most useful package for machine learning in Python.
  2. Load a dataset and understand it's structure using statistical summaries and data visualization.
  3. Create 6 machine learning models, pick the best and build confidence that the accuracy is reliable.

Books
Books

https://github.com/josephmisiti/awesome-machine-learning/blob/master/books.md