23 – 25 July 2018

Lviv Data Science Summer School 2018

Machine Learning for Medical Applications:

Algorithms

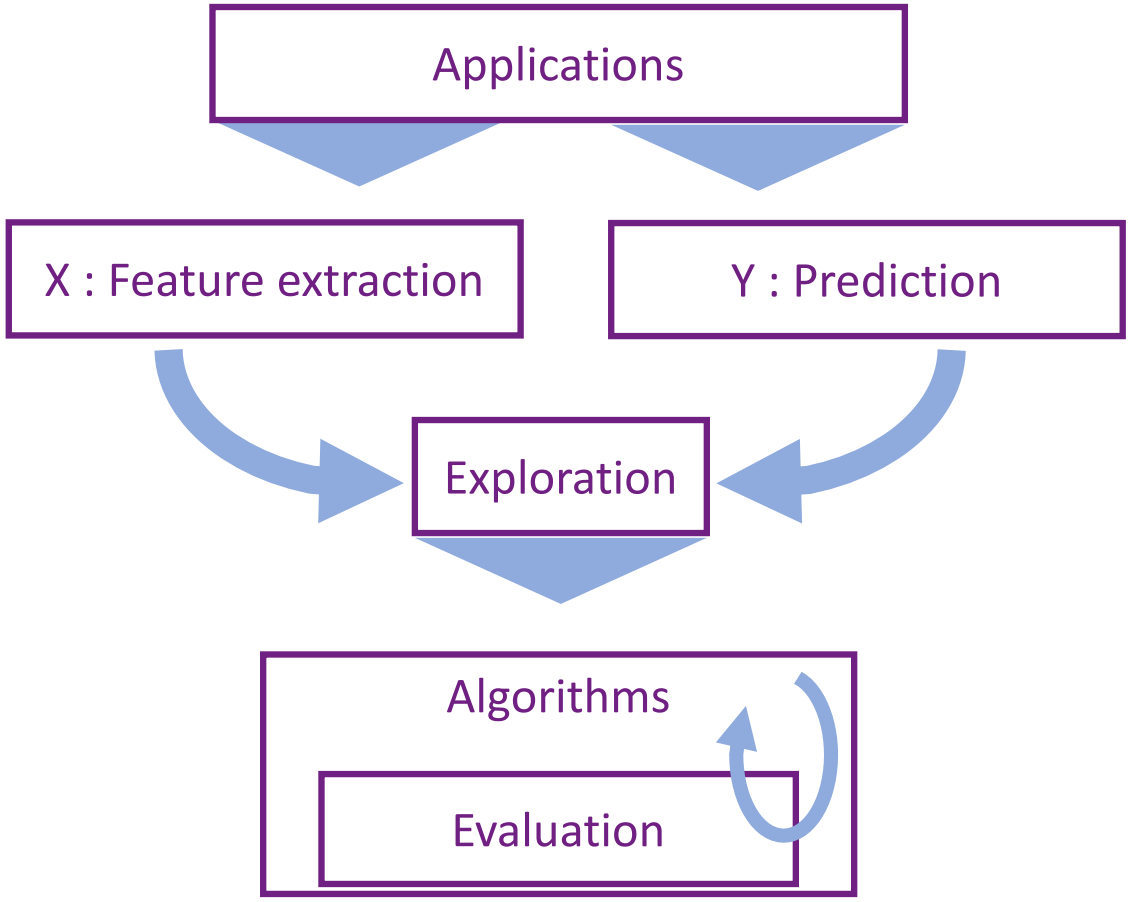
Igor Koval

PhD Student in Applied Mathematics

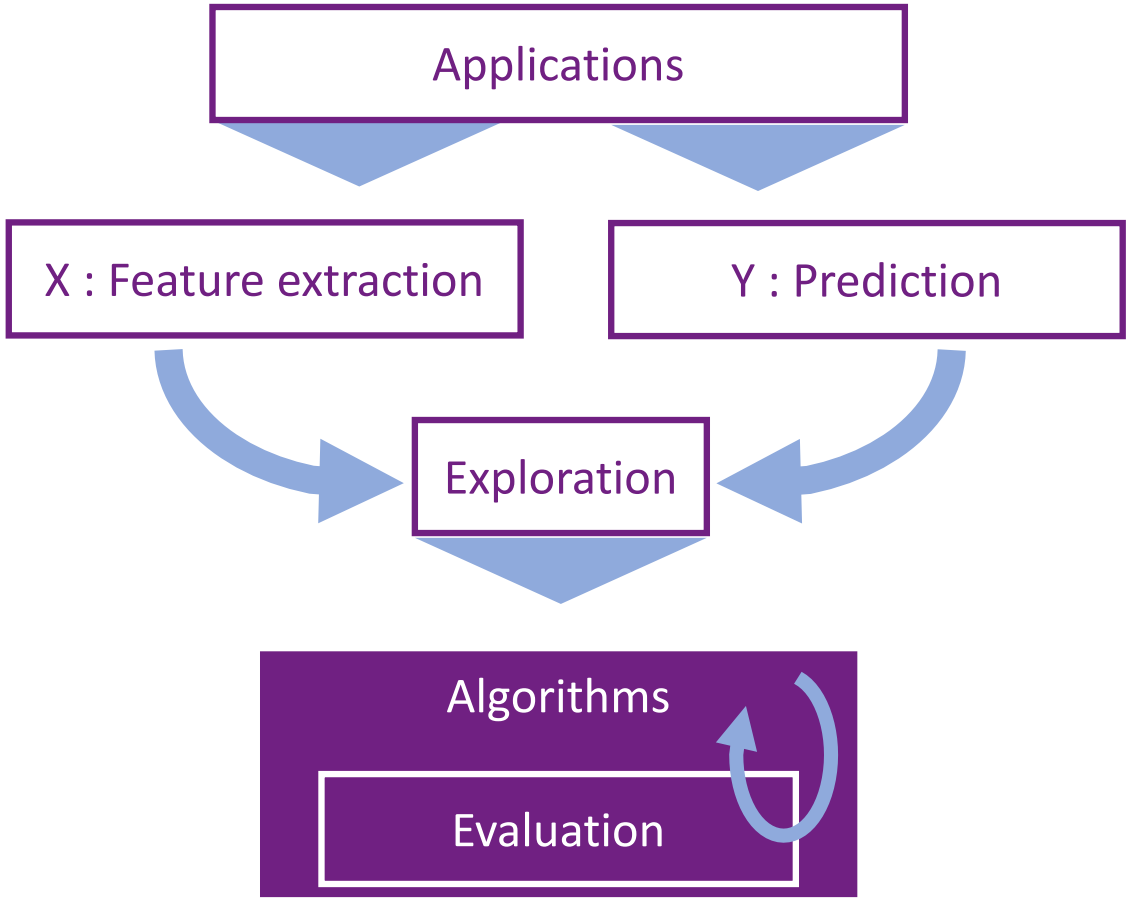
Brain and Spine Institute, Pitié Salpétrière Hospital, Paris, France
& Mathematical Laboratory of Ecole Polytechnique

igor.koval@icm-institute.org

Plan


Applications
X : Feature extraction
Y : Prediction
Exploration
Algorithms
Evaluation

Plan


Applications
X : Feature extraction
Y : Prediction
Exploration
Algorithms
Evaluation

Algorithms

ARAMIS LAB BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

Unsupervised

Clustering

- K-Means
- DBSCAN
- Spectral clustering
- Density models
- ...

Supervised

Classification

- SVM & Kernels
- Logistic Regression
- Decision Trees:
  - RandomForest
  - AdaBoost
  - ...
- ...

Regression

- Regression
  - Linear
  - Ridge penalty
  - Lasso penalty
  - ElasticNet penalty
- Decision Trees
- ...

Will be deepen during the practical session

sklearn is your best friend
http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

Deep Learning in Lecture 5.

Semi-Supervised

Part of the data is not labeled but used to determine some "density" around labeled data

Weakly-Supervised

The data are not labeled with the good labels

Metrics & Loss function

ARAMIS LAB BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

The algorithm optimizes a loss function

≠

Algorithms are compared according to a given metrix

Unsupervised

Clustering

- The key challenge is to know if a metric fits the task of comparing clustering algorithms

Supervised

Classification

- ROC Curve & AUC
- True/False Positive/Negative https://en.wikipedia.org/wiki/Precision_and_recall

Regression

- Mean absolute error L1-norm
- Mean squared error L2-norm
- Median absolute error
- R^2 score
- ...

Too vast to cover all the metrics

http://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics

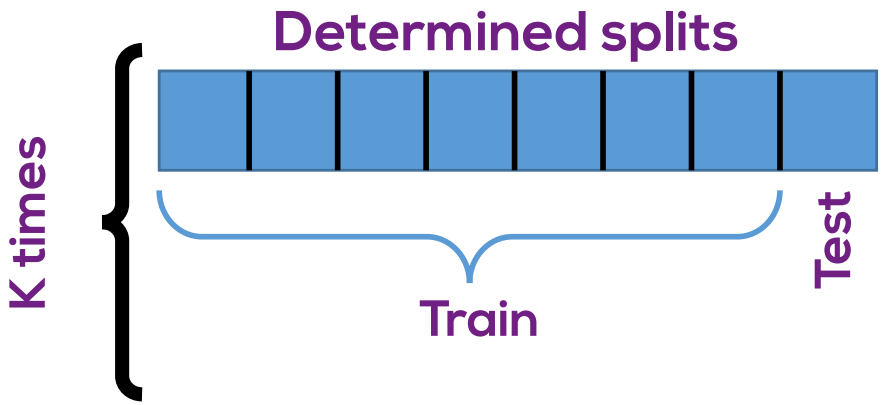
Be careful for unbalanced classes !

Validation

ARAMIS LAB BRAIN DATA SCIENCE

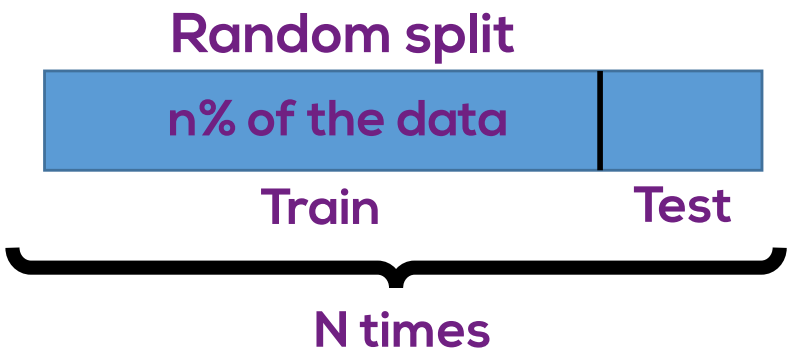
UKRAINIAN CATHOLIC UNIVERSITY FACULTY OF APPLIED SCIENCES

K-Fold



K-scores from which you compute the mean and standard deviation

Holdout Method



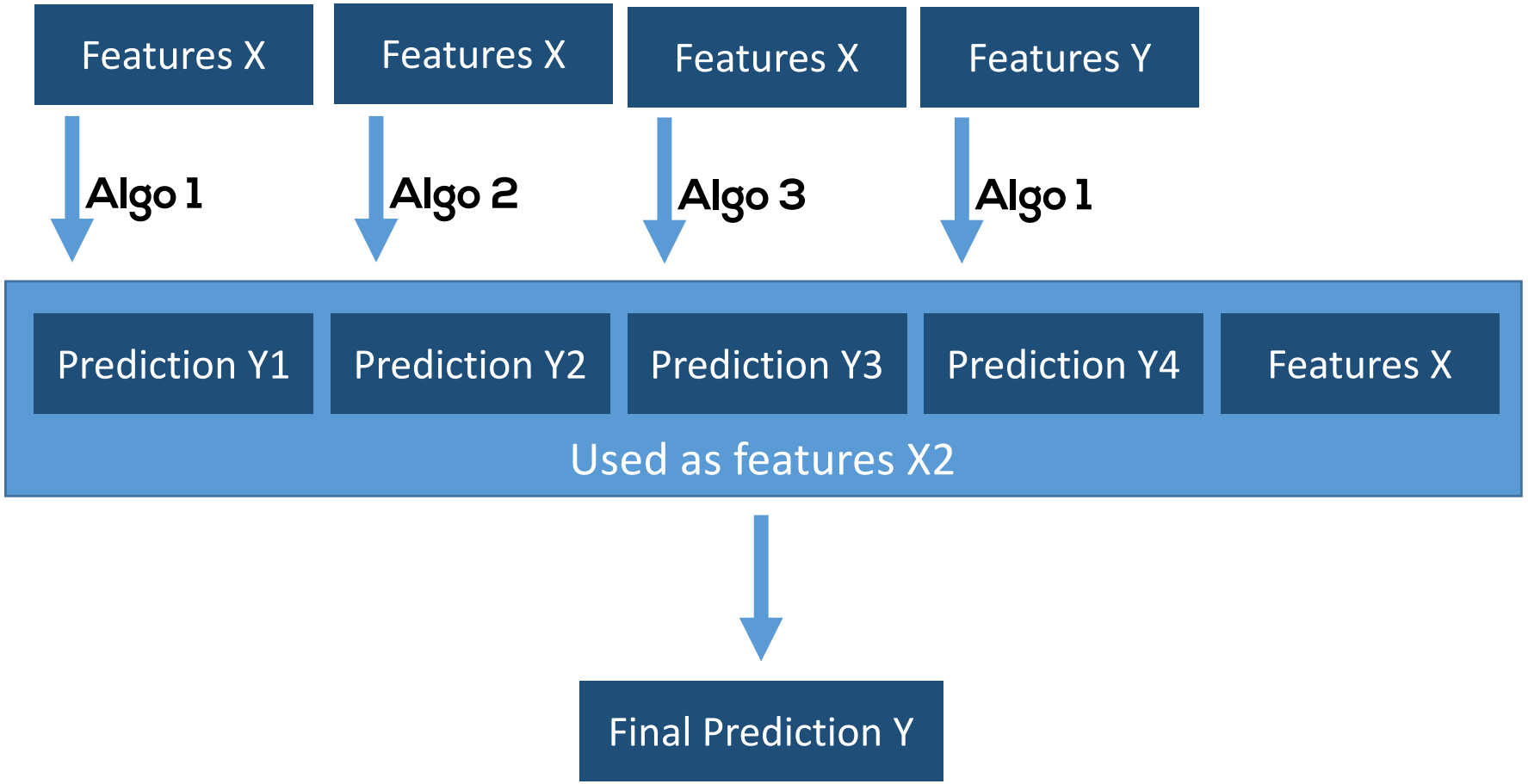
N-scores from which you compute the mean and standard deviation

It is possible to mix : N k-Fold

Hyperparameters learning

But you need part of the data to learn the hyperparameters

Ensembling methods



Practical session

ARAMIS LAB
BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY
FACULTY OF APPLIED SCIENCES

Objective :

1. Prediction of current stage of Alzheimer's Disease

2. Prediction of future stages

3. Trajectory patterns with longitudinal data

Data :

TADPOLE

- Different algorithms
- Hyperparameters tuning
- Cross-validation