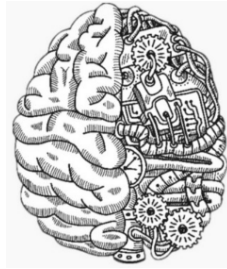


# Formation en **APPRENTISSAGE MACHINE** au **CRIUSMM**



**Automne 2019**

**Etienne Dumesnil & Pierre Orban (pierre.p.orban@gmail.com)**

## **Organisation**

### **Lieu et horaire**

- Centre d'enseignement Dr Frederic Grunberg, Pavillon Bédard 3e étage.
- Les lundis de 13h à 16h

### **Cours théoriques et démonstrations**

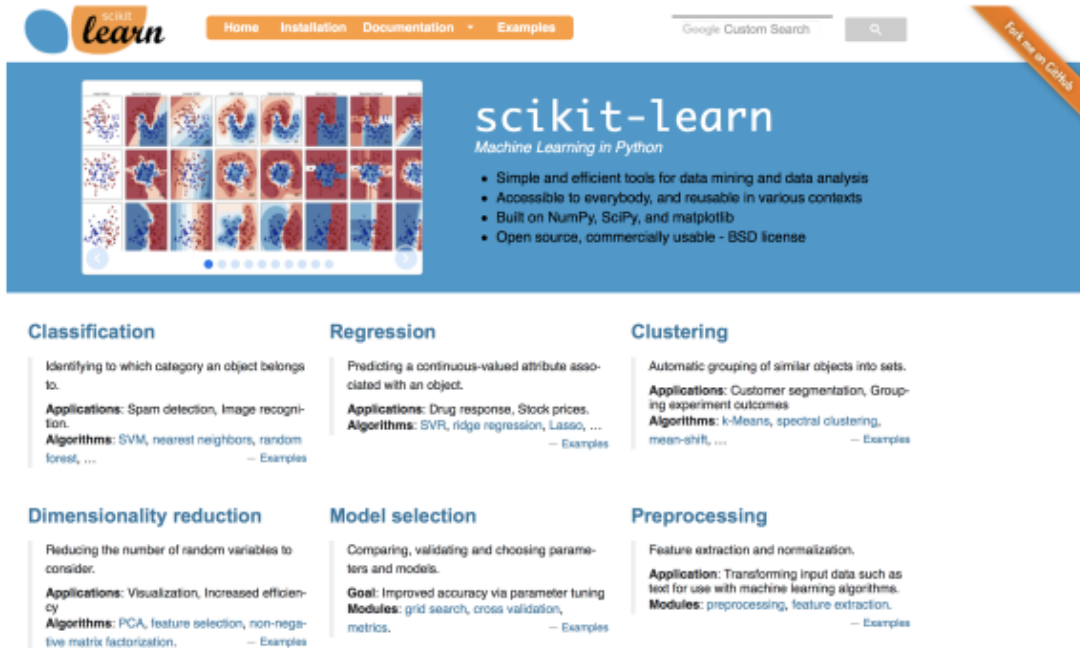
- 21 octobre 2019: BE-322-09 salle polyvalente
- 28 octobre 2019: BE-322-09 salle polyvalente
- 4 novembre 2019: BE-322-09 salle polyvalente
- 11 novembre 2019: BE-320-16
- 18 novembre 2019: BE-322-09 salle polyvalente

### **Travaux pratiques**

- 2 décembre 2019: BE-322-09 salle polyvalente
- 16 décembre 2019: BE-320-16

# Le pratico-pratique d'abord: Doit-on coder?

- Toutes les démonstrations reposeront sur *SciKit Learn* (<https://scikit-learn.org/>), une bibliothèque *Python* dédiée à l'apprentissage machine.



**scikit-learn**  
Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

**Classification**  
Identifying to which category an object belongs to.  
**Applications:** Spam detection, Image recognition.  
**Algorithms:** SVM, nearest neighbors, random forest, ... — Examples

**Regression**  
Predicting a continuous-valued attribute associated with an object.  
**Applications:** Drug response, Stock prices.  
**Algorithms:** SVR, ridge regression, Lasso, ... — Examples

**Clustering**  
Automatic grouping of similar objects into sets.  
**Applications:** Customer segmentation, Grouping experiment outcomes  
**Algorithms:** k-Means, spectral clustering, mean-shift, ... — Examples

**Dimensionality reduction**  
Reducing the number of random variables to consider.  
**Applications:** Visualization, Increased efficiency  
**Algorithms:** PCA, feature selection, non-negative matrix factorization. — Examples

**Model selection**  
Comparing, validating and choosing parameters and models.  
**Goal:** Improved accuracy via parameter tuning  
**Modules:** grid search, cross validation, metrics. — Examples

**Preprocessing**  
Feature extraction and normalization.  
**Application:** Transforming input data such as text for use with machine learning algorithms.  
**Modules:** preprocessing, feature extraction. — Examples

- Tout le matériel est contenu dans des "calepins" *Jupyter* (<https://jupyter.org> (<https://jupyter.org>)).



- Les "calepins" *Jupyter* se trouvent dans un "répo" *Github*, **connectez vous**:

[https://github.com/pnplab/ML\\_CRIUSMM](https://github.com/pnplab/ML_CRIUSMM)

pnplab / ML\_CRIUSMM

Unwatch 2 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

Formation en apprentissage machine au CRIUSMM - automne 2019 Edit

Manage topics

63 commits 3 branches 0 releases 3 contributors

Branch: master - New pull request Create new file Upload files Clone or download Find file

File	Commit	Time ago
porban Update README.md	Latest commit 8f2d4a9	2 hours ago
cours_1	sync	16 hours ago
cours_2	css files	17 hours ago
data	dataset	2 hours ago
img	logos	16 hours ago
intro	more material	2 hours ago
.gitignore	Update .gitignore	2 days ago
README.md	Update README.md	2 hours ago
environment.yml	Update environment.yml	2 days ago

README.md



**CENTRE DE RECHERCHE DE L'USMM**  
CENTRE VIRTUEL À CONSCIENCE DE MONDÉRIE

### Formation en apprentissage machine au CRIUSMM

Ce "répo" github contient le matériel utilisé dans la formation en apprentissage machine donnée au CRIUSMM à l'automne 2019.

Le matériel est disponible dans des dossiers séparés pour les 5 sessions théoriques et de démonstration ( intro/cours\_1, cours\_2, cours\_3, cours\_4 et cours\_5 ) et les 2 séances de travaux pratiques ( TP\_1 et TP\_2 ).

Chaque cours repose sur un notebook Jupyter (<https://jupyter.org/>), une application WEB qui permet de créer et de partager des documents contenant du code et d'en visualiser le résultat. Toutes les démonstrations reposent sur ScikitLearn (<https://scikit-learn.org/>), une bibliothèque libre Python dédiée à l'apprentissage machine.

Les données utilisées pour tous les exemples proviennent de données simulées (fichier /data/sim\_data\_signature.csv contenant des données transversales pour 543 patients et 90 contrôles, avec 222 variables) à partir de la banque de données Signature (<https://www.banquesignature.ca/>). *N'est strictement interdit d'utiliser ces données à d'autres fins que de formation. Toute demande d'accès doit être formulée en bonne et due forme: <https://www.banquesignature.ca/demande-acces/faire-une-demande-dacces/>*

- Les "calegins" *Jupyter* peuvent être exécutés de manière dynamique et fluide sur *Binder*, **démarrez le**:



Starting repository: pnplab/ML\_CRIUSMM/master

Your launch may take longer the first few times a repository is used. This is because our machine needs to create your environment.

```
Build logs
```

```
Step 36/51 : ENV CONDA_DEFAULT_ENV ${KERNEL_PYTHON_PREFIX}
--> Using cache
--> db21020523dc
Step 37/51 : COPY src/environment.yml ${REPO_DIR}/environment.yml
--> a7d3b90cd2b4
Step 38/51 : USER root
--> Running in 18ab284e100b
Removing intermediate container 18ab284e100b
--> 46da409d7681
Step 39/51 : RUN chown -R ${NB_USER}:${NB_USER} ${REPO_DIR}
--> Running in 62683b7474b2
Removing intermediate container 62683b7474b2
```

# Quelles données pour les démonstrations?

## Signature

- Toutes les démonstrations seront faites en utilisant des données **simulées** à partir de la banque de données *Signature*.
- Les données simulées (/data/sim\_data\_signature.csv) comprennent 222 variables démographiques et cliniques (P) pour 543 patients psychiatriques et 90 contrôles (N).

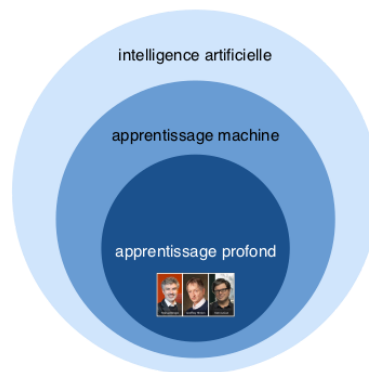
222 variables (prédicteurs)

633 sujets (exemples)

The image shows a screenshot of a CSV file named 'sim\_data\_signature.csv' opened in a spreadsheet application. The title bar indicates '222 variables (prédicteurs)'. The spreadsheet has 633 rows, each representing a subject, and 222 columns, each representing a variable. The first column contains subject IDs, with the first few being 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222. The subsequent columns contain numerical and categorical data for each subject. The first few rows show data for subjects 1 through 10, with values ranging from 0 to 1.0 for many variables. The spreadsheet application's interface is visible, including the menu bar (File, Edit, View, Format, Tools, Window, Help) and the toolbar.

# Apprentissage machine: le buzz!

- Intelligence artificielle, apprentissage machine (automatique ou statistique) et apprentissage profond (réseaux de neurones profonds) ne sont **pas** des synonymes.
- Boom de l'intelligence artificielle dû aux performances de l'apprentissage profond (prix Turing 2018 pour Bengio, Hinton et LeCun), permises par l'explosion des gros jeux de données (\*Big data\*) et de la puissance computationnelle (GPU).
- Cependant, la pertinence de l'apprentissage machine n'est **\*\*pas\*\*** unique à des données massives!



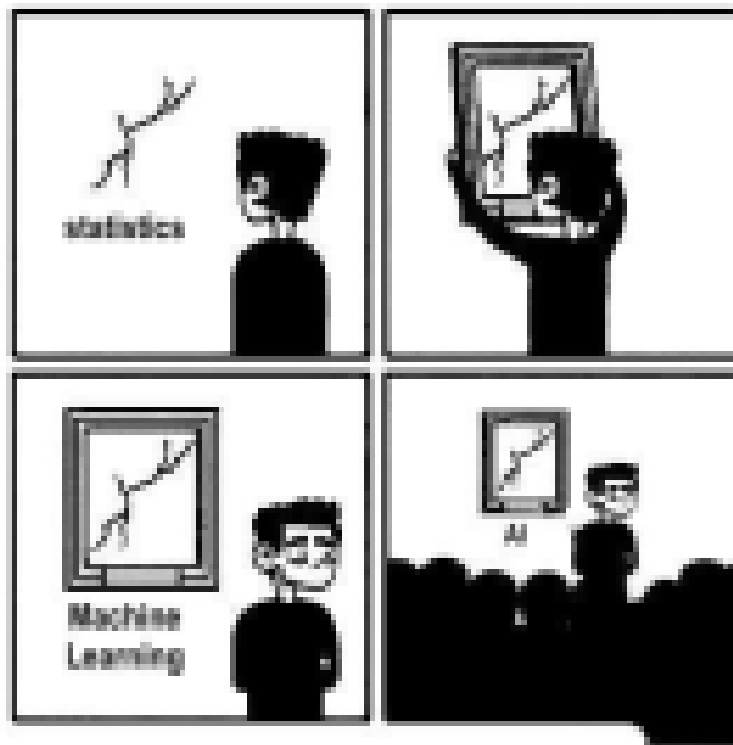
# **Le son de l'apprentissage machine (intelligence artificielle) quand il ne s'agit que d'une régression linéaire**

0:00





## Glorification?



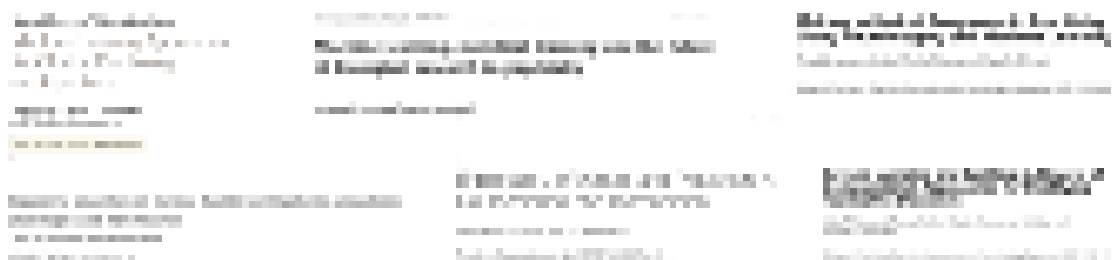
- Oui et non. L'apprentissage machine partage certaines techniques avec les statistiques inférentielles classiques.

- Mais l'usage de ces techniques et, surtout, leur finalité diffèrent de façon importante.



## Importance de l'apprentissage machine pour la psychiatrie

- **Psychiatrie de précision:** Apprentissage machine supervisé pour des prédictions individuelles dans des perspectives diagnostique, pronostique et théranostique.
- **Hétérogénéité des troubles psychiatriques:** Apprentissage machine non supervisé dans une perspective transdiagnostique.



# Plan de cours

## Cours théoriques et démonstrations

- **Cours 1:** Intro à l'apprentissage machine
  - Stats inférentielles vs apprentissage machine
  - Apprentissage machine supervisé vs non supervisé
- **Cours 2:** Apprentissage machine supervisé - régression
  - Hyperparamètres et régularisation (ridge, lasso, elastic-net)
  - Méthodes analytique vs itérative
  - Validation (croisée, croisée nichée)
- **Cours 3:** Apprentissage machine supervisé - classification
  - Évaluation: matrice de confusion et courbe "ROC"
  - Régression logistique
  - Machines à vecteurs de support
- **Cours 4:** Apprentissage machine supervisé - classification (suite)
  - Arbres de décision
  - Bagging: forêts aléatoires
  - Boosting de gradient
  - Classification naïve bayésienne
  - K plus proches voisins
  - Sélection de caractéristiques
  - Classification multi-classes
- **Cours 5:** Apprentissage machine non supervisé
  - Réduction de dimensions (PCA, t-SNE)
  - Regroupement: k-moyennes, décalage de moyenne, DBSCAN, Mélange gaussien, Hiérarchique
  - Validation interne vs externe
- **Cours 5:** Réseaux de neurones (profonds)

# Références

Une liste de ressources est proposée dans le ReadMe du "répo" *Github*:

[https://github.com/pnplab/ML\\_CRIUSMM/blob/master/README.md](https://github.com/pnplab/ML_CRIUSMM/blob/master/README.md)  
([https://github.com/pnplab/ML\\_CRIUSMM/blob/master/README.md](https://github.com/pnplab/ML_CRIUSMM/blob/master/README.md))

- **Livres**

