

- Innovateurs, hackers et startupper voulant tirer profit des possibilités offertes par exploitation de leurs données

| Pré-requis

- Ordinateur à apporter
- Connaître l'algorithmie
- Avoir une première expérience du langage Python
- Connaître quelques rudiments mathématiques, trigonométrie (sinus, cosinus) et algèbre linéaire (calcul matriciel)

| Plan de formation

Partie 1 : Présentation de l'écosystème scientifique "The Big Picture"

Ce chapitre vous offre une vision d'ensemble de l'environnement scientifique de Python afin que vous ne restiez pas devant l'arbre qui cache la forêt, et que vous sachiez trouver les chemins balisés.

- Présentation de quelques librairies scientifiques populaires :

- Les socle commun : La Scipy Stack

- L'astronomie

- La programmation concurrente

- La physique

- La chimie

- Le traitement d'images, la vision

- L'apprentissage automatique

- Les statistiques

- La biologie

- Choisir sa distribution

- Les distributions cross-platform

- Les distributions en ligne

- Les distributions spécifiques pour Windows

- Les outils de développement

Un bon ouvrier a toujours de bons outils

- L'interpréteur iPython

- Les éditeurs PyCharm et Spyder

- Les notebooks Jupyter

Environnements virtuels

Partie 2 : Calcul scientifique

Dans ce chapitre nous aborderons les librairies de la Scipy Stack.

- Ce que tout scientifique et informaticien devrait savoir sur la représentation des nombres en informatique

- Les normes (virgule flottante et fixe)
- Les erreurs d'arrondis
- Contourner et gérer ces erreurs

- **Numpy**

Numpy est une librairie Python de calcul numérique (algèbre linéaire) permettant de manipuler des vecteurs et matrices à plusieurs dimensions. Elle offre des types de données avancés pouvant être gérés très simplement. Implémentée en langage C elle mixe le meilleur des 2 mondes: la souplesse du langage Python avec les performances du C.

- Les types de données et de tableaux gérés par Numpy
- Indexation et manipulation de tableaux
- Les fonctions utiles: création, association, manipulation, extraction de données
- Entrées/Sorties et formats de fichiers : natif Numpy, NetCDF, **HDF5**, Matlab
- Spicy

SciPy fournit un ensemble d'algorithmes mathématiques et de fonctions utilitaires qui s'appuient sur la librairie Numpy.

- Algèbre linéaire (vecteurs, matrices, résolution d'équations linéaires)
- Analyse numérique (intégration, équations différentielles, interpolation, regression, optimisation)
- **Entrées/Sorties**
- Traitement du signal(transformées de Fourier, filtres)
- Statistiques et data mining (correlation, distribution)

- **Pandas**

Pandas est une librairie permettant d'analyser avec efficacité des structures de données complexes. Pandas permet de manipuler des matrices dont chaque colonne peut être d'un type différent et offrant de nombreuses et puissantes fonctionnalités de manipulation, réduction, recherche et statistiques.

- Séries
- Matrices riches (DataFrame)
- Sélection et indexation des données
- Manipulations de données (groupement, ajout, suppression, redimensionnement, pivots, ...)
- Opérations et statistiques
- Entrées/Sorties

Partie 3 : Visualisation de données

Maintenant que nous savons calculer et analyser nos données, il est temps d'éditer de beaux graphiques mettant en valeur les informations importantes que nous avons su y déceler.

- **Présentation de l'écosystème des librairies de visualisation pour Python**
- Les bases avec matplotlib
- **Les librairies orientées Web**
- Les librairies orientées big data et performantes

- Les bibliothèques spécialisées pour la 3D
- Les concepts de base avec Matplotlib
- Les concepts de Matplotlib
- Titres, légendes, annotations
- Axes, échelles de représentations
- Couleurs et styles
- 2D/3D
- Travailler avec plusieurs graphiques
- Interaction avec l'utilisateur
- L'art de faire de beaux graphiques
- Graphic Grammar
- La simplicité c'est mieux
- Publier ses graphiques sur Internet
- Les bibliothèques disponibles
- Cas d'application

1/2 journée?

Partie 4 : Le Calcul distribué

- Présentation
- Qu'est-ce que le parallélisme, la programmation concurrente ?
- Différences entre multithreading et multiprocessing et intérêts respectifs
- Parallélisme, Cluster computing, grid computing et cloud computing
- Multithreading
- Le module threading
- Démarrer, arrêter un thread
- La synchronisation
- Pourquoi le module multithreading de Python n'est pas vraiment une solution professionnelle viable : le GIL
- Les bibliothèques ou versions de Python contournant ce problème: Stackless Python, Jython, IronPython, Greenlet
- Multiprocessing
- Le module multiprocessing
- Le même exercice que pour les threads revisité avec les processus : une ligne de code à changer, ou presque...
- Faire communiquer des processus : Queues, Pipes, mémoire partagée, Proxy Objects, listeners et clients
- Les managers et pool of workers
- Calcul distribué avec Celery ou PySpark
- Présentation des 2 solutions
- Installation d'une solution
- Première application
- Réaliser différents types de traitement de tâches
- Visualiser des jeux de données volumineux (big data)

- Identifier les limites de votre librairie graphique
- Discrétisation des données
- Paralléliser le calcul de vos graphiques
- Quelques outils qui vous faciliteront la vie

Partie 5 : Apprentissage automatique - Machine Learning

Dans cette partie nous utiliserons la librairie Scikit-Learn pour mettre en oeuvre un algorithme d'apprentissage automatique.

- Concepts de base: présentation du Machine Learning
- Les types d'apprentissage
- Les algorithmes
- Savoir choisir son algorithme
- Mise en oeuvre

Dans cette partie du cours nous mettrons en oeuvre un cas concret d'apprentissage automatique basé sur l'algorithme des plus proches voisins

Partie 6 : Accélérer Python/Numpy

Cette section présente différentes approches disponibles pour accélérer vos projets Python, en utilisant différentes techniques de compilation.

- Compiler la scipy stack spécifiquement pour votre architecture
- Avec les librairies vectorielles open source
- Avec les librairies spécifiques par constructeur
- Utiliser les librairies de compilation à la volée (Just In Time)
- Présentation des différentes solutions
- Mise en oeuvre d'une solution et mesure des gains

| Suivi de formation en option

A l'issue de la formation, nos formateurs peuvent aussi intervenir pour vous accompagner dans la mise en application des compétences acquises :

- en répondant à vos questions lors de rendez-vous téléphoniques réguliers
- en étant présents physiquement à l'amorce du projet
- en réalisant un audit de vos pratiques quelques semaines/mois après la formation

Cette idée vous intéresse ? Faites-le nous savoir à l'acceptation du devis pour que nous trouvions la formule adaptée à votre situation.