

Share



Email destinataire :

Vous devez renseigner un email valide

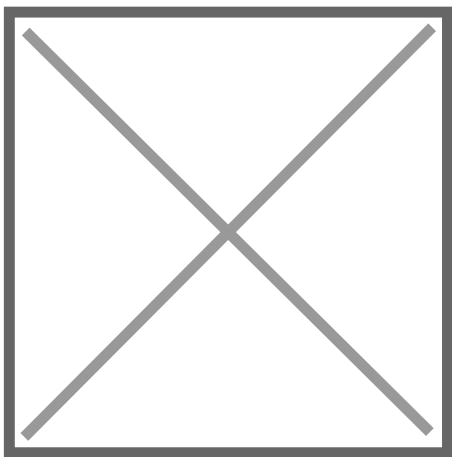
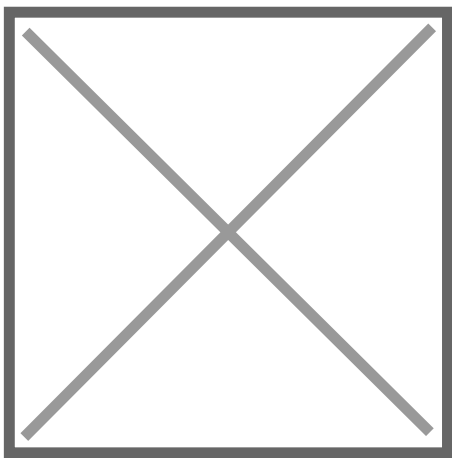
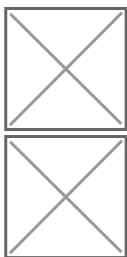
Envoyer

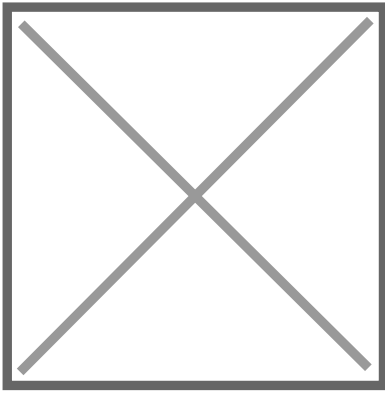
Une erreur est survenue. Merci de réessayer ultérieurement

Le mail de partage a bien été envoyé.

Add to Compare

M2 Bioinformatique, Biostatistiques et applications de l'intelligence artificielle





Master's degree

Specialisation Bioinformatics / Computational Biology

Language(s) of instruction

French

The **M2 BIBS-IA** program of the Master in Bioinformatics trains **research and development experts** at the intersection of biology, computer science, and mathematics. Students acquire in-demand skills in **artificial intelligence, data management and analysis, and modeling** across various areas of bioinformatics—genomics, structural biology, evolutionary biology, and digital health. These skills are applied in the life sciences, medicine, environmental sciences, and agronomy.

The **M2 BIBS-IA** program is built as a continuation of the **M1 BIBS-IA**. The “*Profile*” section specifies the expected prerequisites.

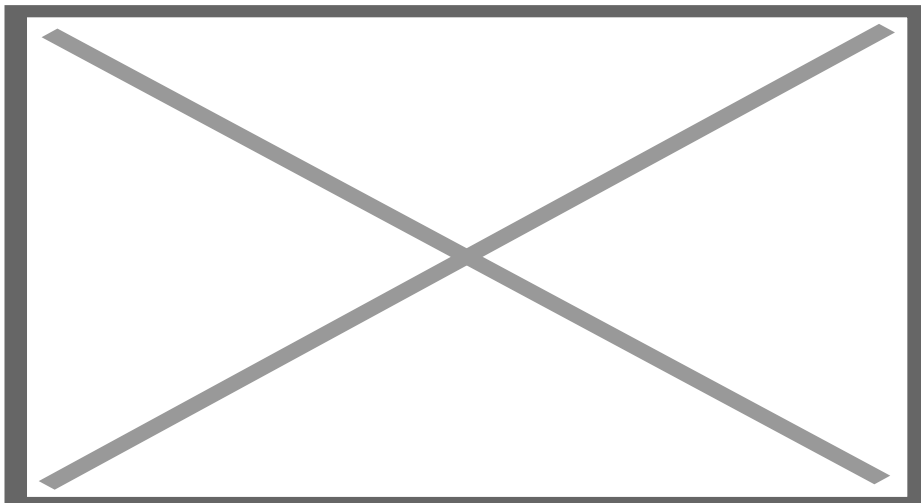
Within the BIBS-IA track, **artificial intelligence and data science** play a central role, approached both from a **theoretical perspective** (in mathematics and computer science) and from an **applied perspective**, closely linked to the specificities of biological and biomedical data.

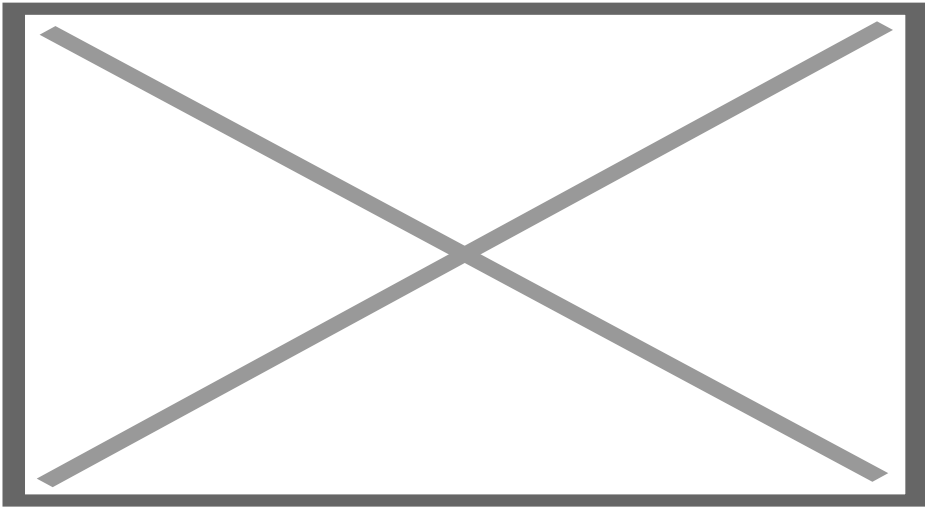
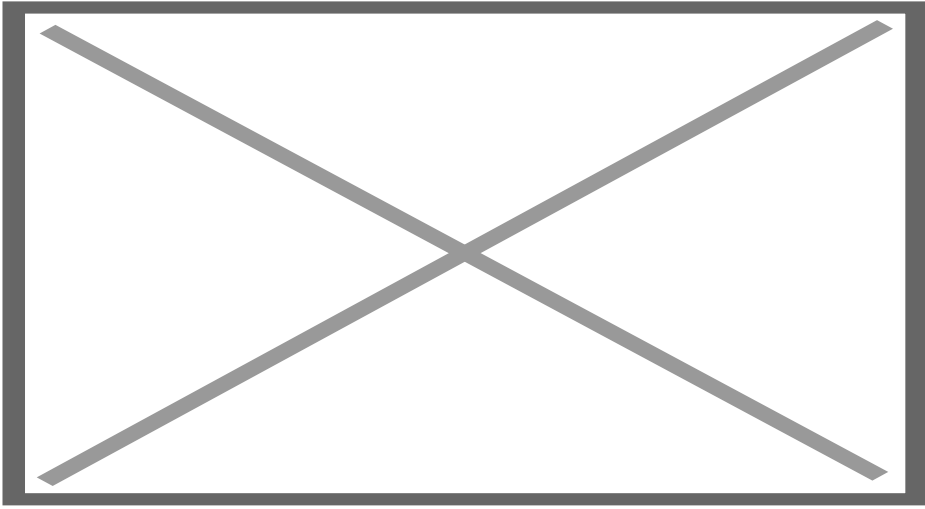
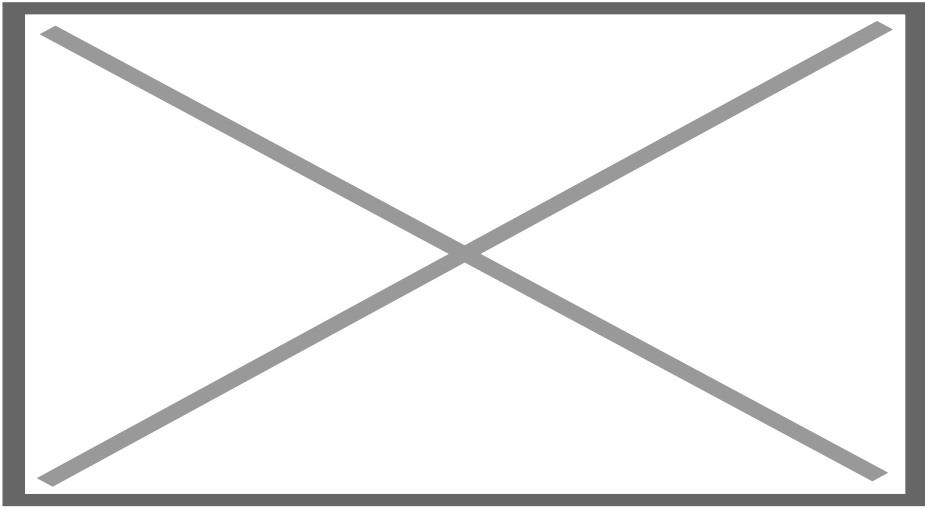
Multidisciplinary projects and hackathons are organized by experts from academia and industry. Students have access to a wide range of elective teaching modules, allowing them to specialize (applications of artificial intelligence in structural biology, genomics, clinical contexts, agronomy, etc.).

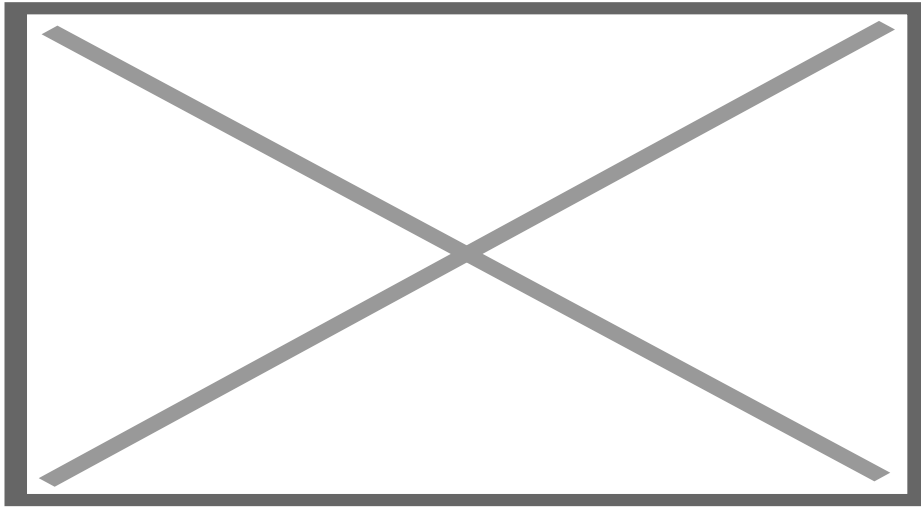
A **mandatory 4- to 6-month internship** ensures strong future career prospects for students, whether in research or in industry.

[Télécharger la fiche détaillée](#)

[Apply Now](#)







Information

- [Overview](#)
- [Application requirements](#)
- [Schedule](#)

Présentation

Skills

The skills acquired upon completion of the **M2 BIBS-IA** are aligned with the national framework of the Master in Bioinformatics ([RNCP 38964](#)).

- **Understand** current challenges related to complexity and diversity in biology and the life sciences in general, and **contribute actively** to research and innovation in the academic, pharmaceutical, biotechnological, agri-industrial, and healthcare sectors;
- **Address** the challenges arising from the evolution and development of high-throughput data production technologies, coming from multiple and heterogeneous sources;
- **Master** computational approaches and artificial intelligence to analyze, model, and interpret large-scale biomedical data;
- **Lead** bioinformatics projects independently, in multidisciplinary environments, by developing innovative solutions using appropriate computational tools and programming languages;
- **Experiment with, develop, and test** new methods to explore hypotheses in the life sciences, at the interface between fundamental and applied research.

Objectives

The **M2 BIBS-IA** program is dedicated to the advanced research and development needs of companies and research organizations in **bioinformatics** and **biostatistics**.

This **multidisciplinary training program** meets these needs by preparing specialists at the interface of biology, computer science, and mathematics.

Upon completion of the M2 BIBS-IA, students **master several programming languages**, a wide range of **artificial intelligence**, algorithmic, **statistical, and data science** methods, enabling them to store, model, and analyze large-scale genomic and biomedical data. They also deepen their knowledge in key areas of bioinformatics (systems biology, structural biology, functional genomics) and apply the methods learned to diverse projects in life sciences and medicine, as well as in environmental sciences and agronomy.

The curriculum of the M2 BIBS-IA is organized into six major knowledge and skills blocks:

Block 1 – Management of large-scale data and optimization.

This block includes three teaching units focused on optimizing the processing of various types of data (relational databases and graph data). A project based on massive real-world datasets allows students to put their acquired knowledge into practice.

Block 2 – High-throughput genomics.

This block comprises two teaching units: one focused on comparative genomics and another on population genomics, with strong emphasis on the application of statistical methods in these domains.

Block 3 – Data analysis and AI for biology.

This block includes three teaching units, covering topics ranging from fundamental AI concepts such as machine learning models, to applications of AI in integrative biology and image analysis.

Block 4 – Scientific integrity and reproducibility.

This block comprises two pre-professional teaching units: a *reprohackathon project* and a unit supporting professional integration. Students acquire best practices in open science, scientific integrity, ethics, and energy efficiency (ecological impact of digital practices, particularly artificial intelligence).

Block 5 – Advanced applications of Artificial Intelligence.

Students select four teaching units within this block, which includes six dedicated to the M2 BIBS-IA track and focused on diverse applications of AI. Examples include a *data camp* (a machine learning hackathon on real-world datasets, jointly organized with École Polytechnique and Télécom Paris) and data mining in agronomy (jointly with AgroParisTech). In addition, students may choose elective units from other Master's programs (Computer Science, Biology-Health, etc.).

Débouchés

Career prospects

After a Master's + PhD: researcher or lecturer

Bioinformatics Engineer (Research and Development)

Data Manager

Data scientist

Data Scientist

Further Study Opportunities

PhD in Bioinformatics

R&D researcher or data modelling and analysis expert in leading-edge companies or laboratories.

Data Scientist, Data Analyst, Ingénieur·e en Machine Learning dans des secteurs innovants (tech, finance, santé, énergie, etc.) ;

Calendar

Start of programme (indicative date)

03/10/2025

End of programme (indicative date)

04/09/2026

Admission

Capacity

Available Places

24

Target Audience and Entry Requirements

Most courses in the **M2 BIBS-IA** program are taught in **French**, therefore proficiency in the French language is required.

The M2 BIBS-IA is designed as a continuation of the **M1 BIBS-IA** and brings together three student profiles:

- Students who have **successfully completed the M1 BIBS-IA**, all of whom are admitted into the M2 BIBS-IA;
- Students from **AgroParisTech** enrolled in the IODAA track (*From Information to Decision through Analysis and Learning*), all of whom complete a double degree year with the M2 BIBS-IA;
- (Subject to availability) Students who have completed an M1 equivalent to the M1 BIBS-IA, i.e., who have acquired equivalent knowledge and skills in biology, mathematics, and computer science (see the M1 BIBS-IA program for further details).

Students who already hold a Master's degree in **Bioinformatics** are **not eligible**. Students holding a Master's degree in a field other than bioinformatics (biology, computer science, or mathematics) may apply for entry into the M1. However, their applications will be considered **less of a priority** than those from students applying directly after a Bachelor's degree.

Période(s) de candidature

Inception Platform

From 04/05/2026 to 13/07/2026

Supporting documents

Compulsory supporting documents

Motivation letter.

All transcripts of the years / semesters validated since the high school diploma at the date of application.

Curriculum Vitae.

Additional supporting documents

Certificate of French (compulsory for non-French speakers).

Certificate of English level (compulsory for non-English speakers) or GMAT / GRE test results.

Detailed description and hourly volume of courses taken since the beginning of the university program.

VAP file (obligatory for all persons requesting a valuation of the assets to enter the diploma).

Document indicating the list of local M2 choices available here : <https://urlz.fr/i3Lo>.

Supporting documents :

- Residence permit stating the country of residence of the first country
- Or receipt of request stating the country of first asylum
- Or document from the UNHCR granting refugee status
- Or receipt of refugee status request delivered in France
- Or residence permit stating the refugee status delivered in France
- Or document stating subsidiary protection in France or abroad
- Or document stating temporary protection in France or abroad.

Programme

Bloc 1 - Gestion de grands volumes de données et optimisation

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	practical class	directed study/practical class	discovery
----------	------	----------	---------	----------------	-----------------	------------------------	-----------------	--------------------------------	-----------

Projet données massives

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 2

Practical study : 18

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Savoir développer en Python, se servir de git

Programme/contents

- Architecture client/serveur
- Prise en main du framework Django
- Focus sur certaines fonctionnalités de Django
- Projet

Learning objectives

Concevoir une application web permettant à plusieurs profil d'utilisateurs de consulter et enrichir les données.

Avoir conscience de enjeux en terme de sécurité informatique.

Concevoir une base de donnée répondant aux besoins de l'application, et la volumétrie des données.

Savoir comprendre et se mettre à la place des utilisateurs, et anticipier les besoins.

Conduire et mener à terme un projet de développement informatique.

Apprendre à travailler en équipe.

Overall organisation

Une série de cours/TD initiaux permet aux étudiants de se familiariser avec le framework Django, l'architecture client/serveur et un framework css. Au cours de cette UE, les étudiants réalisent un projet en groupe de 3-4 où ils doivent concevoir et développer une application web interrogeant et enrichissant une base de données. Le suivi du projet se fait sous forme d'entretiens hebdomadaires avec un enseignant.

Bibliography

<https://bbrancot.pages.pasteur.fr/prog-web-client-server/>

<https://github.com/bbrancot/prog-web-client-server>

Bases de
Données
Avancées

2

Semestre
1

10

10



Bases de Données Avancées

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Directed study: 10

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Savoir implémenter en SQL la création de schémas, tables, clés primaires et étrangères, l'insertion/suppression/mise à jour de données et l'interrogation à l'aide des opérateurs Select-From-Where-Group by-Having.
Avoir des notions en Algèbre relationnelle (opérateurs, requêtes)

Programme/contents

- Stratégies de placement des données sur le disque (stockage des données)
- Index : types et structures
- Niveaux d'optimisation d'une BD : schéma, requête
- Plan de requêtes, algorithmes existants pour l'implementation des opérateurs
- Réécriture de requêtes

Learning objectives

L'objectif de cette UE est de présenter les techniques les plus avancées pour optimiser l'accès aux données d'une base de données :
Savoir optimiser une requête SQL ; savoir optimiser un schema de bases de données.

Overall organisation

Séances hebdomadaires de Cours-TD

Bibliography

Database Management Systems
de Raghu Ramakrishnan, Johannes Gehrke
McGraw Hill, 3ème édition

Database Tuning. Principles, Experiments, and Troubleshooting
de Dennis Shasha, Philippe Bonnet
Morgan Kaufmann

Type of assessment

Graphes biologiques et optimisation

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 12

Directed study: 6

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Programme d'algorithmique du M1 BIBS

Programme/contents

- Théorie de graphes. : bases, coloration, graphes planaires. - Algorithmique des graphes : parcours, plus courts chemins, arbres coouvrant de poids minimum, flots. Illustrations sur des problèmes en bioinformatique.
- Théorie de la complexité : problèmes P, NP, NP-complets; exemples dans les graphes. - Optimisation et Approximabilité.

Learning objectives

Acquérir les bases et les compétences en théorie et algorithmique des graphes et en complexité

Overall organisation

Séances hebdomadaires de Cours-TD

Bibliography

Claude Berge. Graphes. Dunod.

Type of assessment

Evaluation Continue non Intégrale

Bloc 2 - Génomique haut débit

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	practical class	directed study/practical class	distance learning
----------	------	----------	---------	----------------	-----------------	------------------------	-----------------	--------------------------------	-------------------

Génomique
des
populations

2

Semestre
1 10

10



Génomique des populations

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Practical study : 10

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Bases en analyse de séquences, notions de statistiques et probabilités, et familiarité avec Python.

Programme/contents

Cette UE propose une **introduction aux méthodes utilisés en génétique des populations à travers une application biologique concrète**. Après avoir présenté les concepts clés (mutation, recombinaison, généalogie des individus) les étudiantes et étudiants explorent un cas de métissage entre populations. Ce fil conducteur permet de découvrir différentes **approches d'inférence**, basées sur des simulations (Approximate Bayesian Computation, Machine Learning,...) et sur des **modèles probabilistes** (Hidden Markov Models). Les étudiants apprennent à manipuler un simulateur de séquences versatile et des bibliothèques Python pour analyser, visualiser et interpréter des données, souvent représentées sous forme de graphes de généalogie.

Learning objectives

Acquérir des compétences en **analyse et inférence à partir de données complexes de type graphe**, dans le contexte de la **génétique des populations**. Les étudiants utilisent des outils informatiques et des modèles probabilistes pour explorer et interpréter des scénarios issus de simulations.

Overall organisation

L'UE alterne cours magistraux et travaux pratiques. Les cours présentent les concepts fondamentaux de la génétique des populations et les approches d'inférence. Les TP permettent aux étudiantes et étudiants de manipuler des

Génomique
comparée

2

Semestre
1 10

10



Génomique comparée

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Practical study : 10

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Connaître un langage de script

Avoir des connaissances de base en génomique

Programme/contents

Concepts d'orthologie, core et pan génomes.

Détermination du contenu en gènes d'un génome

Application à la comparaison de génomes complets

Analyse des génomes : Duplications, Pertes et Gains de Gènes, Transferts horizontaux

Génomique comparée (Diagramme de Venn / Conservation de la synténie)

Learning objectives

Savoir manipuler, analyser et interpréter les données génomiques pour répondre à une question biologique

Cette UE vise à donner aux étudiants une vision synthétique de la constitution et de l'évolution des génomes procaryotes et eucaryotes.

Overall organisation

L'enseignement se fait en salle informatique et consiste à analyser des jeux de données extraits de la bibliographie ou produits dans les laboratoire du campus afin d'appréhender en situation réelle tous les aspects de l'analyse des données génomiques et de la comparaison des génomes.

Type of assessment

Evaluation Continue non Intégrale

Bloc 3 - Analyse de données et IA pour la biologie

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	distance learning
Analyse d'Images en Biologie	2	Semestre 1	8		12				

Analyse d'Images en Biologie

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 8

Practical study : 12

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Aucun

Programme/contents

Décrire et expliquer les bases théoriques de l'image numérique et de son interprétation :

:

1. Byte, bit, pixel, formats d'image, rang dynamique, histogramme, LUTs, opérations mathématiques sur les images.
2. Échantillonnage
3. Espace de Fourier, opérations dans l'espace des fréquences.
4. Segmentation, types de bruit et filtres (espace réel et Fourier)
5. Projection et rétroprojection
6. Visualisation 3D, utilisation d'ImageJ (Fiji), utilisation de plug-in, programmation Macro
7. Convolution et corrélation

Extraire des données quantitatives de l'image :

1. Segmentation par seuillage et introduction aux approches d'apprentissage profond pour la segmentation
2. Densitométrie
3. Granulométrie

Traiter et analyser les signaux :

1. Signaux périodiques (cristaux 2D et/ou structures empilées - hélices). Calcul de paramètres de signaux périodiques
2. Rapport signal sur bruit, contrastes, entropie
3. Débruitage

Réaliser des reconstructions 3D à partir de projections :

1. Algorithmes de reconstruction en imagerie : théorème de la section central, rétroprojection

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	dista learn cou
Analyse de données omiques et biologie intégrative	2.5	Semestre 1	15	5	20				

Analyse de données omiques et biologie intégrative

ECTS : 2.5

Semester: Semestre 1

Detail

Lecture: 15

Practical study : 20

Directed study: 5

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Bonne maîtrise du logiciel R; Principes des techniques d'analyses de données multivariées (ACP, MDS, etc.); Statistique inférentielle (estimation, tests d'hypothèses, etc.); Techniques de séquençage haut débit (RNAseq, ChIPseq, etc.)

Programme/contents

Différents types de données issues des approches expérimentales haut débit (omiques) seront traitées (par exemple RNAseq, ChIPseq, spectrométrie de masse). Une première partie de l'enseignement sera consacrée à l'analyse exploratoire des données multivariées, puis seront abordées différentes problématiques telles que l'analyse différentielle (RNAseq), la détection de pics (peak calling, ChIPseq) et l'inférence de réseaux de gènes.

Learning objectives

- Modéliser les données omics à l'aide de variables aléatoires adaptées et poser un cadre statistique inférentiel nécessaire à l'analyse de ces données
- Implémenter des méthodes d'estimations de paramètres dans les modèles
- Mettre en oeuvre les méthodes proposées sur des jeux de données réelles

Overall organisation

Les séances (3 h 00) se déroulent en salle

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	dista learn cou
Modèles pour le Machine Learning	2.5	Semestre 1	15		15				

Modèles pour le Machine Learning

ECTS : 2.5

Semester: Semestre 1

Detail

Lecture: 15

Practical study : 15

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Estimation dans le modèle linéaire, Estimation par maximum de vraisemblance, Programmation statistique avec R

Programme/contents

Modèles de régression et classification supervisée, paramétriques et non-paramétriques;
Ré-échantillonnage bootstrap, validation croisée et évaluation de performances (estimation de l'erreur de généralisation, courbes ROC);
Sélection de variables, choix de modèles;
Arbres de décision (CART, Forêts aléatoires);
Méthodes d'apprentissage en grande dimension (LASSO, SVM)

Learning objectives

Génération, évaluation et comparaison de modèles de machine learning:

Être capable d'utiliser des méthodes variées d'apprentissage statistique et de les interpréter sur des données, d'implémenter les méthodes étudiées avec R;
Savoir mettre en place une stratégie de choix de modèle(s); Savoir évaluer et comparer les performances des modèles d'apprentissage.

Overall organisation

Chaque séance inclut une étude de cas permettant la mise en oeuvre des méthodes enseignées et l'analyse statistique des données associées

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class
Reprohackathon	3	Semestre 1	11		19			
■								

Reprohackathon

ECTS : 3

Semester: Semestre 1

Detail

Lecture: 11

Practical study : 19

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Savoir développer en python.

Type of assessment

Evaluation Continue non Intégrale

Mise en situation professionnelle	2	Semestre 1	10	15
---	---	------------	----	----



Mise en situation professionnelle

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Directed study: 15

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Aucun

Programme/contents

L'enseignement combine :

- Des ateliers pratiques de rédaction de CV et de préparation aux entretiens ;
- Une introduction aux bonnes pratiques de recherche éthiques et à la responsabilité environnementale ;
- Une junior conférence organisée par les étudiants sur deux jours, leur permettant de se mettre en situation réelle de communication scientifique et de gestion d'événement académique.

Learning objectives

Cette UE vise à préparer les étudiants à leur insertion professionnelle et à les sensibiliser aux enjeux éthiques et environnementaux de la recherche en bioinformatique.

À l'issue de l'UE, les étudiants sauront :

- Identifier les principaux débouchés de la bioinformatique dans la recherche publique et privée ;
- Préparer leur orientation entre la poursuite en doctorat et une carrière d'ingénieur ;
- Concevoir un CV et se préparer à un entretien professionnel ;
- Adopter de bonnes pratiques de recherche et une réflexion éthique et responsable sur leur activité scientifique.

Overall organisation

L'UE alterne cours et exercices centrés sur la mise en pratique.

L'évaluation repose sur la participation active, les

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	di
Méthodologie en Recherche Clinique et Statistiques	2	Semestre 1	10	10					

Méthodologie en Recherche Clinique et Statistiques

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Directed study: 10

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Introduction à la recherche clinique et épidémiologique (UE de M1 BIBS)

Learning objectives

Comprendre le développement d'un médicament / dispositif médical dans son environnement technique et réglementaire

Connaitre les acteurs et les métiers de la recherche clinique

Voir comment un parcours BIBS-IA peut intéresser les entreprises des sciences de la vie

Comprendre les grandes fonctions statistiques liées à la recherche clinique.

Overall organisation

Cours magistral et ateliers interactifs.

Type of assessment

Evaluation Continue Intégrale

[LLM pour les séquences protéiques](#)

2	Semestre 1	10	10
---	------------	----	----

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	di
■									

LLM pour les séquences protéiques

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Practical study : 10

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Maitriser un langage de programmation. Bases de statistiques et de bioinformatique structurale.

Programme/contents

Introduction de la révolution de l'IA dans la prédiction des structures 3D des protéines, impact sur l'annotation grande échelle des métagénomes. Introduction aux LLM de protéines. Analyse manuelle de quelques prédictions structurales et classification des prédictions en fonction de leur score de confiance. Analyse à l'échelle d'un protéome. Utilisation des embeddings de LLMs (ESMfold et ProtTrans) et de méthodes de bioinformatique pour l'analyse de séquences et structures pour interpréter et classifier les cas prédits avec mauvais score de confiance (ie, utiliser d'autres approches pour donner de la valeur aux cas prédits avec faible confiance, typiquement écartés des approches classiques). Tirer partie de cette analyse pour discuter des limites des LLM pour explorer des territoires inconnus (souvent le cas avec les données de métagénomiques venant d'environnements extrêmes ou sous-échantillonnés). Cette analyse offrira un cas pratique idéal pour appliquer certaines méthodes statistiques vues en cours (lois de mélange, ...).

Learning objectives

- Comprendre les modèles LLM entraînés sur les séquences protéiques : architecture et entraînement

- Connaître les applications principales des modèles LLM des protéines (prédiction de

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	di
UE libre Catalogue S1	2.5	Semestre 1							c
■									

UE libre Catalogue S1

ECTS : 2.5

Semester: Semestre 1

Language(s) of instruction

Français

Remote teaching

non

Type of assessment

Evaluation Terminale

[Advanced](#)

[Machine](#)

[Learning](#)

2

Semestre
1

10

10

Advanced Machine Learning

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Directed study: 10

Language(s) of instruction

Anglais

Remote teaching

non

Prerequisites

Avoir suivi un cours d'apprentissage automatique couvrant en particulier les apprentissages supervisé et non supervisé.

Programme/contents

- L'induction classique avec données i.i.d. Problème. Quel type de garantie peut-on vouloir et obtenir ? Conditions pour les obtenir.
- Quand on change la distribution pour obtenir de meilleurs résultats : boosting, apprentissage actif, ré-échantillonnage, Learning Using Privileged Information (LUPI)
- Apprentissage semi-supervisé. Apprentissage avec des agents communiquant entre eux : co-training, blending, analogie, effet tunnel cognitif, distillation
- Quand la distribution P_X change entre apprentissage et test. Adaptation de domaine et co-variate shift.
- Quand la distribution conditionnelle $P_{Y|X}$ change. Tracking, apprentissage multi-tâches ; apprentissage par transfert, zero-shot learning, modèles de fondations, Invariant risk minimization
- Apprentissage continu avec curriculum. Comment changer la tâche, comment mesurer la difficulté d'un nouvel exemple, comment choisir une bonne séquence d'exemples ou de tâches d'apprentissage.
- Apprentissage en-ligne : approches théorique et pratique. Effets de séquence. Oubli catastrophique.

Learning objectives

L'objectif du cours est d'étudier l'apprentissage dit « hors distribution », c'est-à-dire quand les données d'apprentissage ne sont pas issues de la

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	di le c
Data Camp	2	Semestre 1	20		10				

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	di
■									

Data Camp

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 20

Practical study : 10

Language(s) of instruction

Anglais

Remote teaching

non

Prerequisites

Une bonne connaissance de python et de l'écosystème de calcul scientifique en python est requise (numpy, scipy, scikit-learn). Une connaissance des bases de l'apprentissage automatique et des algorithmes de classification est aussi nécessaire.

Programme/contents

Jour 1 : Préparation et manipulation des données

- * Introduction du workflow (VSCode, Python, git, GitHub, tests, ...)
- * Cours avancé sur Pandas
- * Exercices GitHub : NumPy et Pandas

Jour 2 : Pipelines de ML et évaluation de modèles

- * Scikit-learn avancé : ColumnTransformer et pipelines
- * Traitement parallèle avec joblib
- * Généralisation et validation croisée
- * Exercices sur scikit-learn
- * Introduction à Codalab et présentation des challenges

Jour 3 : Métriques et gestion des données déséquilibrées

- * Présentation des différentes métriques de ML
- * Problèmes liés aux métriques avec des données déséquilibrées
- * Approches d'apprentissage automatique pour traiter les jeux de données déséquilibrés
- * Travail sur les challenges de données

Jour 4 : Données complexes

- * Feature engineering et encodage avancé des

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	di le c
Fouille de données pour le vivant	2	Semestre 1	10		10				

Fouille de données pour le vivant

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Practical study : 10

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Des bases en algorithmique et programmation (idéalement en Python) sont nécessaires.

Programme/contents

Sont abordées dans ce module des éléments théoriques permettant de comprendre les questions fondamentales de l'apprentissage non supervisé. Y seront présentés et détaillés quelques modèles, ainsi que des éléments de méthodologie et de mise en pratique.

Learning objectives

La fouille de données (data mining) est un domaine de l'intelligence artificielle qui s'intéresse à l'émergence de connaissances nouvelles à partir de données. On parle également d'apprentissage automatique non supervisé (unsupervised learning). L'objectif de ce module est d'étudier les problématiques de l'apprentissage automatique non supervisé et les différentes catégories de méthodes qui ont été proposées pour y répondre avec des applications dans le domaine des sciences du vivant et l'environnement. On se place dans un cadre "non supervisé" lorsqu'on n'a pas d'exemples pour lesquelles on connaît ce que l'on cherche à prédire. Par exemple, on dispose d'une collection d'images. Certaines représentent des plantes et d'autres des animaux mais on ne sait pas à quelle "catégorie" appartient chacune de nos images. Parfois, il est possible et facile d'obtenir ce type d'informations et un apprentissage automatique dit "supervisé" (supervised learning) sera envisageable. Dans le cas contraire, comme souvent en sciences du vivant et de

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	di le c
Analyse d'images et intelligence artificielle	2	Semestre 1	10	10					

Analyse d'images et intelligence artificielle

ECTS : 2

Semester: Semestre 1

Detail

Lecture: 10

Directed study: 10

Language(s) of instruction

Français

Remote teaching

non

Prerequisites

Bases d'analyse d'images biologiques; bases d'apprentissage profond

Programme/contents

L'enseignement est constitué autour d'une activité pratique d'apprentissage profond supervisé sur des images biologiques qui met en jeu des concepts et des techniques pertinentes et d'utilisation courante, et permet aux apprenants d'effectuer une telle activité avec une compréhension profonde des enjeux, en donnant une base solide des disciplines concernées pour minimiser les "boîtes noires". L'enseignement prévoit donc une importante partie de "projet", où les apprenants travaillent en groupe avec la supervision ponctuelle de l'équipe enseignante. L'activité de projet inclut les tâches suivantes : 1) Nettoyage et préparation des données ; 2) Choix d'un modèle existant pertinent ; 3) "Refinement" du modèle pour la tâche de prédiction souhaitée ; 4) Évaluation de la performance du nouveau modèle.

Learning objectives

- Comprendre les modèles d'apprentissage profond d'utilisation courante en analyse d'images (CNN et en particulier UNET) et mesurer la qualité de leurs performances
- Connaître les applications principales des modèles d'apprentissage profond pour l'analyse d'images biologiques (*e.g.* segmentation, classification, ...)
- Utiliser des modèles d'apprentissage profond

Bloc 6 - Stage

Subjects	ECTS	Semester	Lecture	directed study	practical class	Lecture/directed study	Lecture/practical class	directed study/practical class	distance learning courses
Stage	30	Semestre 2							
■									

Stage

ECTS : 30

Semester: Semestre 2

Detail

Tutored project 45

Language(s) of instruction

Français

Remote teaching

non

Type of assessment

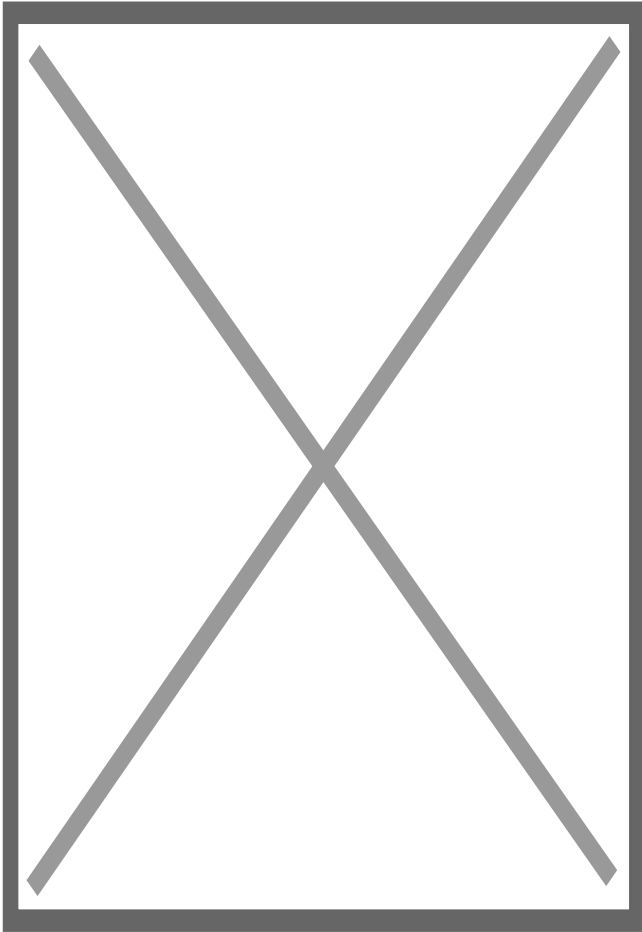
Evaluation Continue Intégrale

Teaching Location(s)

ORSAY

Training campus

Javascript est requis pour visualiser cette carte.



Saclay Moulon

Saclay / Gif-sur-Yvette / Orsay

Bus 4606 / 4609 / 4611 / 5154

RER B Le guichet

[Voir](#)

Student restaurant (CROUS)

Library

Sports facilities

Student residence

Contact

- **Fanny Pouyet**

MCF

<http://www.universite-paris-saclay.fr/formation/master/bio-informatique/m2-bioinformatique-biostatistiques-et-applications-de-lintelligence-artificielle-bibs-ia>