

Formation des formateurs en Sciences des données et IA 2023-2025

OBJECTIFS

Cette formation est ouverte à tout chercheur, enseignant-chercheur, doctorant et post-doctorant de l'Université PSL qui s'engage à créer au moins un cours reprenant les enseignements reçus :

- Proposer des modules adaptés aux différentes disciplines liées aux sciences des données et à l'intelligence artificielle.
- Les **participants peuvent choisir les modules** les plus adaptés à leurs besoins et à ceux de leur discipline.
- Chaque séance de 3H est qualifiée par niveau : débutant, intermédiaire, avancé
- Renforcement de la capacité des formateurs à dispenser une formation de qualité dans leur domaine respectif
- **Chaque participant bénéficiera d'un soutien à la création de nouveaux modules de formation** au sein de PSL à la suite du cours de formation

PROGRAMME

Cours de base

- Fondements de l'apprentissage supervisé et non supervisé (30H)
- Deep learning (30H)
- Optimisation et apprentissage (18H)

Cours avancés

- SQL in practice, NoSQL and vector databases (9H)
- Web scraping (12H)
- Natural Language Processing (21H)
- Apprentissage par renforcement (18H)

[Retrouver le détail du programme](#)



Date et horaires :

Retrouver le détail sur [la page web du programme DATA](#)

Durée Totale de l'ensemble des modules: 234H

Lieu : Estrapade

Formateurs :

Vincent Divol, Muni Sreenivas Pydi, Nicolas Schreuder, Kimia Nadjahi, Bruno Chaves, Ana Busic

Public concerné

chercheur, enseignant-chercheur, doctorant et post-doctorant de l'Université PSL

POUR VOUS INSCRIRE



1 créez votre compte - <https://psl-ecoleinterne-inscription.helvetius.net/sinscrire.php>

2 connectez-vous au portail de l'Ecole Interne PSL - <https://psl-ecoleinterne.monportail.psl.eu/pegasus/index.php>

3 faites votre demande en « ça m'intéresse » pour être recontacté à l'ouverture de la prochaine session ou inscrivez-vous à la date indiquée.

IA & Data Training for Academics 2023-2025

OBJECTIFS

This course is open to all researchers, teacher-researchers, doctoral students and post-doctoral students at PSL University who undertake to create at least one course incorporating the lessons received :

- Main objective: to offer modules adapted to the various disciplines linked to data sciences and artificial intelligence; to strengthen the capacity of trainers to provide high-quality training in their respective fields.
- **Participants can choose the modules** best suited to their needs and those of their discipline.
- Each 3H session is qualified by level: beginner, intermediate, advanced
- **Researchers and teacher-researchers** who commit **to creating a new training module** within PSL at the end of their training will be eligible for **support**. Contact justine.petit@psl.eu for more information.

PROGRAM

Basic course

- Fundamentals of supervised and unsupervised learning (30H)
- Deep learning (30H)
- Optimization and machine learning (18H)

Advanced courses

- SQL in practice, NoSQL and vector databases (9H)
- Web scraping (12H)
- Natural Language Processing (21H)
- Reinforcement learning (18H)

[Find the details of the program](#)



Date and times :

Find the details on the [DATA webpage](#)

Total duration of all modules: 234h

Place: Estrapade

Trainers: Vincent Divol, Muni Sreenivas Pydi, Nicolas Schreuder, Kimia Nadjahi, Bruno Chaves, Ana Busic

Target audience :

researchers, teacher-researchers, doctoral students and post-doctoral students at PSL University

TO REGISTER



1 Create your account - <https://psl-ecoleinterne-inscription.helvetius.net/sinscrire.php>

2 Log in to the PSL Internal School portal - <https://psl-ecoleinterne.monportail.psl.eu/pegasus/index.php>

3 Apply in "I'm interested" to be contacted again at the opening of the next session or register on the date indicated.

DATA SCIENCE & AI

| TRAINING FOR ACADEMICS 2023/25 |

Basic courses - Mondays from 5 to 8 pm, Estrapade

Advanced courses - from 5 to 8 pm, Estrapade

— Foundations of supervised and unsupervised learning (30h) with Vincent Divol

10 lectures: from 1/7/24 to 4/11/24

- o Introduction (Beginner) 1/7/24
- o Linear regression (Beginner) 8/7/24
- o Classification and logistic regression (Advanced) 15/7/24
- o Regularization 1 (Advanced) 23/9/24
- o Regularization 2 (Expert) 30/9/24
- o Beyond linear methods 1 (Beginner) 7/10/24
- o Beyond linear methods 2 (Advanced) 14/10/24
- o Beyond linear methods 3 (Expert) 21/10/24
- o Unsupervised learning 1 (Beginner) 28/10/24
- o Unsupervised learning 2 (Advanced) 4/11/24

— Deep Learning (30h) with Nicolas Schreuder

10 lectures: from 12-11-24 to 3/2/25

- o Machine learning reminder : linear model, loss minimization (Beginner) 18/11/24
- o Multi-Layer Perceptron : expressivity, back-propagation (Beginner) 25/11/24
- o Deep-Learning Tricks : activation function, batch-norm, drop-out (Beginner) 02/12/24
- o Introduction to Pytorch, MLP (Beginner) 9/12/24
- o Convolution Neural Network (Beginner) 6/01/25
- o Recurrent Network (Advanced) 13/01/25,
- o Transformers (Advanced) 20/01/25, 27/01/25
- o Generative models (Advanced) 03/02/25, 06/02/25

— Optimization & machine learning (18h) with Kimia Nadjahi

6 lectures: from 10/2/25 to 17/3/25

- o Introduction (Beginner) 10/02/2025
- o Convexity (Beginner) 17/02/2025
- o Gradient descent and Newton's method (Advanced) 24/02/2025
- o Linear and nonlinear constrained optimization (Advanced) 3/03/2025
- o Advanced deterministic methods (Expert) 10/03/2025
- o Stochastic optimization (Advanced) 17/03/2025

— Web scraping (12h) with Bruno Chaves

4 lectures on Thursdays: from 20/03/25 to 10/04/25

This lecture only requires basic knowledge of Python & SQL.

- o Lectures 1 and 2 (total 6 hours): the basics 20/03/2025 and 27/03/2025
- o Lectures 3 and 4 (total 6 hours): advanced topics and legal aspects 3/04/2025 and 10/04/2025

— SQL in practice, NoSQL and vector databases (9h) with Bruno Chaves

3 lectures: on Mondays from 19/05/25 to 10/6/25

This lecture only requires basic knowledge of Python & SQL.

- o SQL in practice 19/05/2025
- o No SQL databases 2/06/2025
- o Vector databases 10/06/2025

— Natural Language Processing (21h)

7 lectures: on Mondays from 24/03/25 to 12/05/25

- o Background (refresh): basic concepts from Machine Learning and Neural Networks 24/03/2025
- o Word vectors: Latent Semantic Analysis (LSA), word2vec and GloVe 31/03/2025
- o Language Models: from N-gram models to Recurrent Neural Networks 7/04/2025
- o Sequence-to-sequence modelling and Machine Translation (MT) 14/04/2025
- o Attention, self-attention and Transformer-based language models 28/04/2025
- o Pretraining and Finetuning large language models 5/05/2025
- o Understanding ChatGPT 12/05/2025

— Reinforcement learning (18h) with Ana Busic

6 lectures on Thursdays

- o Introduction to reinforcement learning 15/05/25
- o Markov decision processes and dynamic programming 22/05/25
- o Bandit algorithms 05/06/25
- o From dynamic programming to reinforcement learning 12/06/25 & 19/06/25
- o Reinforcement learning with approximation 26/06/25

SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

Foundations of supervised and unsupervised learning

30 hours, with Vincent Divol

— **Lecture 1 (beginner):**

- Challenges of supervised and unsupervised learning
- Module presentation

— **Lecture 2 (beginner): linear regression**

- Univariate linear regression
- R-squared coefficient
- Reminders on linear algebra
- Multidimensional linear regression

— **Lecture 3 (advanced): binary classification**

- Logistic regression
- SVM
- Convexification of binary loss

— **Lectures 4 (advanced) and 5 (expert): linear methods in high dimension**

- Why regularization is necessary
- Ridge regularization
- Sparsity hypothesis
- Lasso regularization

— **Lectures 6 (beginner), 7 (advanced) and 8 (expert): beyond linear methods**

- Using feature maps in non-parametric regression (e.g. Fourier transform)
- Underfitting and overfitting (analysis using VC dimension)
- Kernel methods

— **Lectures 9 (basic) and 10 (advanced): unsupervised learning**

- Clustering (k-means method and spectral clustering)
- Dimensionality reduction (PCA, MDS, Isomap)
- Kernel density estimation
- Nearest neighbors estimation

Deep learning

30 hours, with Nicolas Schreuder

— **Lecture 1 (beginner): machine learning reminders**

- Linear models
- Loss minimisation

— **Lecture 2 (beginner): multi-layer perceptron**

- Expressivity
- Back-propagation
- Activation function

— **Lecture 3 (beginner): introduction to PyTorch**

- Tensors
- GPU computations
- Automatic differentiation

— **Lecture 4 (beginner): convolutional neural networks**

- Convolutional layers
- Image classification/segmentation

SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

— Lectures 5 (beginner) and 6 (beginner): deep learning tricks

- Batch-normalisation
- Dropout
- Feature extraction
- Fine-tuning of pre-trained models

— Lecture 7 (advanced): residual Networks

- Depth
- Residual connections

— Lectures 8 (advanced) and 9 (advanced): the Transformer model

- Attention mechanism
- Natural Language Processing (NLP)
- Tokenisation
- Vision transformer

— Lecture 10 (advanced): generative models

- Generative Adversarial Networks (GANs)
- Diffusion models

Optimization and machine learning

18 hours, with Kimia Nadjahi

The objective of this module is to give an overview of the main optimization algorithms, with a focus on their application to statistical learning.

— Lectures 1 (beginner) and 2 (beginner): convexity

- Introduction and convexity
- Convex optimization
- Duality

— Lectures 3 (advanced) and 4 (advanced): standard algorithms for constrained or unconstrained problems

- Unconstrained optimization: gradient descent, Newton's method
- Constrained optimization: interior-point methods

— Lectures 5 (advanced) and 6 (advanced): large scale optimization

- Advanced first-order methods
- Stochastic algorithms

Web scraping

12 hours, with Bruno Chaves

— Lectures 1 and 2: the basics

- Python, HTML and CSS: the basics
- XPath and CSS selectors
- Scraping with http queries
- Scraping with browser emulation
- Practice

— Lectures 3 and 4: advanced topics and legal aspects

- Public and private APIs (Application Programming Interface)
- Hidden APIs
- Identifying/bypassing blocking points
- Recurring and/or large-scale scraping
- The legal context of research data: discussion with a data officer
- Practice

SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

SQL in practice, NoSQL and vector databases

9 hours, with Bruno Chaves

— Lecture 1: SQL in practice

- Understanding SQL in practice: SQL structures and research questions
- SQL in the era of ChatGPT
- Hands-on lab with SQLite (using Python)

— Lecture 2: NoSQL databases

- Overview of NoSQL databases
- Types of NoSQL databases and use cases
- Hands-on lab with MongoDB (using Python)

— Lecture 3: Vector databases

- Overview of vector databases and their use cases
- Basics of vector similarity search
- Hands-on lab: Implementing a vector database for a machine learning application

Natural language processing

21 hours

Natural Language Processing (NLP) is a subfield of artificial intelligence, which studies principles and algorithms that allow an artificial system to understand and process human natural language. In this course, students will learn about the latest advancements in using deep learning for NLP. They will gain both conceptual understanding and practical skills to build and use their own models for various purposes.

— Lecture 1: Background (refresh)

- Basic concepts from machine learning and neural networks
- Exercise: Basics of Scikit-Learn and PyTorch

— Lecture 2: Word vectors

- Latent Semantic Analysis (LSA)
- word2vec and GloVe
- Exercise: GloVe implementation

— Lecture 3: language models: from N-gram models to Recurrent Neural Networks

- Exercise: LSTM-based Language Models

— Lecture 4: sequence-to-sequence modelling and Machine Translation (MT)

- Exercise: English-French Translation

— Lecture 5:

- Attention
- Self-attention
- Transformer-based language models
- Exercise: natural-language generation with transformers

— Lecture 6: pretraining and finetuning large language models

- Exercise: tutorial on Hugging Face

— Lecture 7: Understanding chatGPT

SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

Reinforcement learning

18 hours, with Ana Busic

— **Lecture 1: introduction to reinforcement learning**

— **Lecture 2: Markov decision processes and dynamic programming**

· Finite horizon, infinite horizon (discounted, optimal stopping, stochastic shortest path)

· Algorithms: value iteration, policy iteration, linear programming

— **Lecture 3: bandit algorithms**

· Epsilon greedy and the EXP3 algorithms

· Contextual bandits, bandits with expert advice and the EXP4 algorithm

· Stochastic bandits, Upper Confidence Bound (UCB) algorithm

· Thompson sampling

— **Lectures 4 and 5: from dynamic programming to reinforcement learning**

· Introduction to stochastic approximation

· Algorithms: TD-learning, SARSA, Q-learning

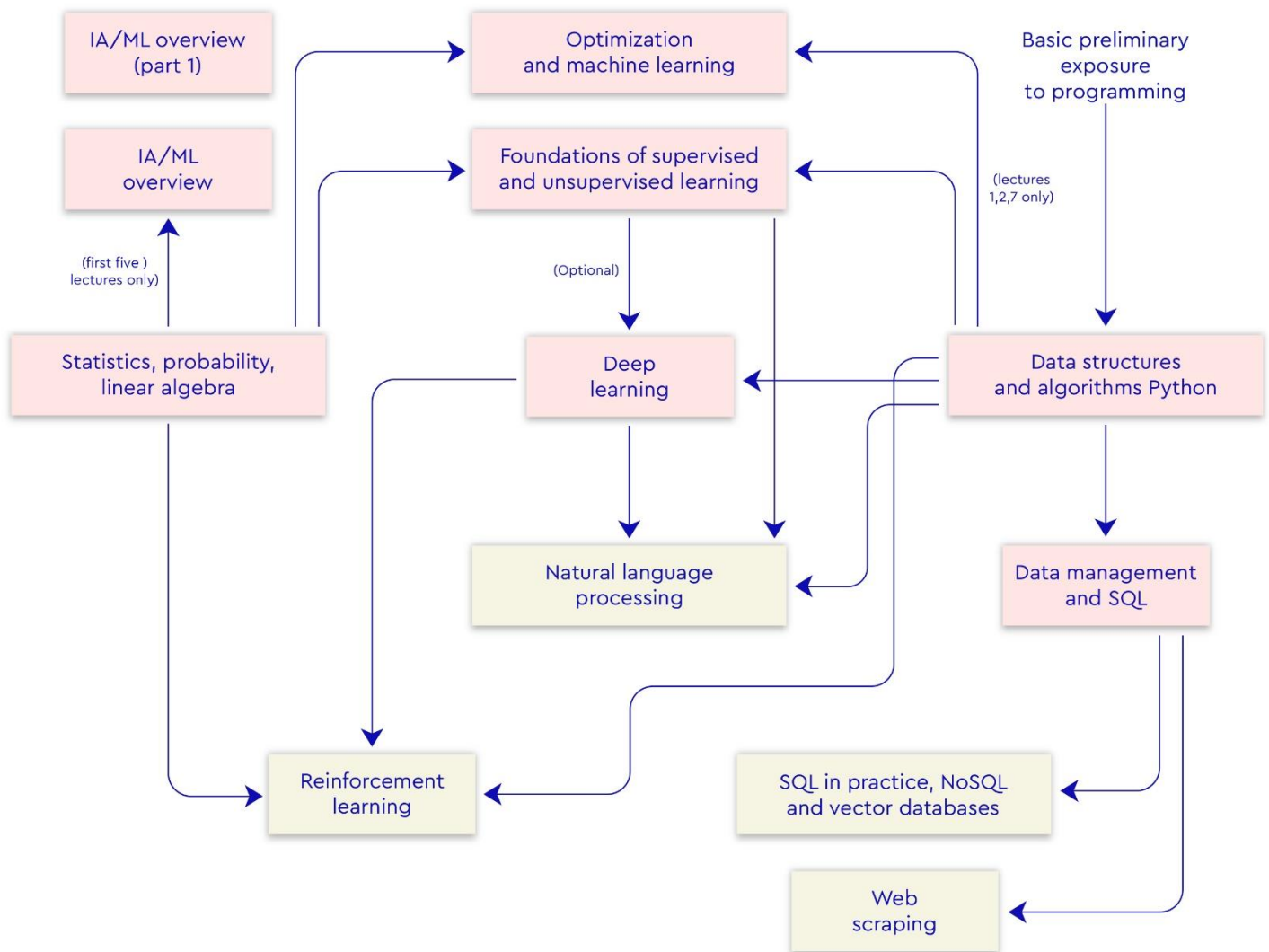
· Actor-critic

· Model-based reinforcement learning, upper confidence reinforcement learning (UCRL)

— **Lecture 6: reinforcement learning with approximation**

· From linear to Deep Neural Networks (DNN)

Prerequisite



How to read this picture ?

Basic courses are in pink. Advanced courses are in light green.

An arrow of the form $A \rightarrow B$ indicates that, to follow course B, it is necessary to be comfortable with the content of course A.