

# Formation des formateurs en Sciences des données et IA 2024-2026

## 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

- IA/ML Overview (12H)
- Outils statistiques et probabilistes (21H)
- Algorithmes et structures de données en Python (21H)
- Math & IA Overview (12H)
- Gestion des données et SQL (18H)
- 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 :** Frédéric Lechenault, Alexandre Vérine, 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 2024-2026

## OBJECTIVE

This course is open to all researchers, teacher-researchers, doctoral students and post-doctoral students at PSL University who undertaketo 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@psl.eu](mailto:justine.petit@psl.eu@psl.eu) for more information.

## PROGRAM

### Basic course

- IA/ML Overview (12H)
- Statistical and probabilistic tools(21H)
- Algorithms and data structures in Python(21H)
- Data management and SQL(18H)
- Math & IA Overview (12H)
- 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:** Frédéric Lechenault, Alexandre Vérine, Muni Sreenivas Pydi, Nicolas Schreuder, Kimia Nadjahi, Bruno Chaves, Ana Busic

## Target audience :

researchers, teacher-researchers, doctoral students and post-doctoral students at PSLUniversity

## 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.onportail.psl.eu/pegasus/index.php>

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

### Basic courses - from 5 to 8 pm, Estrapade

#### — IA/ML Overview (12h) with Frédéric Lechenault

All these lectures correspond to a “beginner” level.

4 lectures: from 14/10/2024 to 4/11/2024

- o Training for academics, overview 14/10/2024
- o History of AI 21/10/2025
- o Data science basics 28/10/2024 & 4/11/2024

#### — Statistics, probability, linear algebra (21h) with Alexandre Verine

7 lectures: from 12/11/2024 and 13/01/2025

- o Linear Algebra1 (*Beginner*) 12/11/2024
- o Linear Algebra2 (*Advanced*) 18/11/2024
- o Calculus 1 (*Beginner*) 25/11/2024
- o Calculus 2 (*Beginner*) 2/12/2024
- o Probability and statistics 1 (*Beginner*) 16/12/2024
- o Probability and statistics 2 (*Advanced*) 6/01/2025
- o Probability and statistics 3 (*Advanced*) 13/01/2025

#### — Data structures and algorithms Python (21h)

7 lectures: from 20/01/2025 to 3/03/2025

- o Introduction to Python Programming (*Beginner*) 20/01/2025
- o Introduction to Data Structures (*Beginner*) 27/01/2025
- o Algorithm Analysis and Complexity (*Advanced*) 3/02/2025
- o Sorting Algorithms (*Advanced*) 10/02/2025
- o Searching Algorithms (*Advanced*) 17/02/2025

- o Advanced Data Structures and Algorithms (*Advanced*) 24/02/2025
- o Applications of Data Structures and Algorithms (*Beginner*) 3/03/2025

#### — Foundations of supervised and unsupervised learning (30h)

10 lectures: from 10/03/2025 to 19/05/2025

- o Introduction (*Beginner*) 10/03/2025
- o Linear regression (*Beginner*) 17/03/2025
- o Binary classification (*Advanced*) 24/03/2025
- o Linear methods in high dimension 1 (*Beginner*) 31/03/2025
- o Linear methods in high dimension 2 (*Advanced*) 07/04/2025
- o Beyond linear methods 1 (*Beginner*) 14/04/2025

- o Beyond Linear methods 2 (*Advanced*) 28/04/2025
- o Beyond Linear methods 3 (*Expert*) 07/05/2025
- o Unsupervised learning 1 (*Basic*) 12/05/2025
- o Unsupervised learning 2 (*Advanced*) 19/05/2025

#### — Math and IA Overview (12h) with Frédéric Lechenault

4 lectures : from 26/05/2025 to 23/06/2025

All these lectures correspond to a “beginner” level.

- o Introduction to Machine Learning and Neural Networks 26/05/2025
- o artificial intelligence for vision 02/06/2025
- o artificial intelligence for time series and natural language processing 16/06/2025
- o Introduction to Reinforcement Learning and Advanced ML Techniques 23/06/2025

#### — Data management and SQL (18h) with Muni Sreenivas Pydi

6 lectures: 22/09/2025 to 27/10/2025

- o Introduction to Databases (*Beginner*) 22/09/2025
- o Introduction to SQL (*Beginner*) 29/09/2025
- o Joins and Relationships (*Beginner*) 6/10/2025
- o Creating a SQL database (*Beginner*) 13/10/2025
- o Advanced Queries (*Advanced*) 20/10/2025
- o Data Manipulation and Control (*Advanced*) 27/10/2025

### — Deep Learning (30h) with Nicolas Schreuder

10 lectures: from 3/11/2025 to 26/01/2026

- o Machine learning reminder (*Beginner*) 3/11/2025
- o Multi-Layer Perceptron : (*Beginner*) 17/11/2025
- o Introduction to Pytorch (*Beginner*) 24/11/2025
- o Convolutional neural networks (*Beginner*) 1/12/2025
- o Deep learning tricks (*Beginner*) 8/12/2025

- o Deep learning tricks (*Advanced*) 15/12/2025
- o Residual Networks (*Advanced*) 5/01/2026
- o The transformer model 1 (*Advanced*) 12/01/2026
- o The transformer model 2 (*Advanced*) 19/01/2026
- o Generative models (*Advanced*) 26/01/2026

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

6 lectures: from 2/02/2026 to 9/03/2026

- o Convexity (*Beginner*) 2/02/2026 & 9/02/2026
- o Standard algorithms for constrained or unconstrained problems (*Advanced*) 16/02/2026 & 23/02/2026
- o Large scale optimization (*Advanced*) 2/03/2026 & 9/03/2026

### Advanced courses - from 5 to 8 pm, Estrapade

#### — Web scraping (12h) with Bruno Chaves

4 lectures: from 12/02/2026 and 2/04/2026

This lecture only requires basic knowledge of Python & SQL.

- o The basics 1 12/02/2026
- o The basics 2 19/03/2026
- o Advanced Topics and legal aspects 1 26/03/2026
- o Advanced Topics and legal aspects 2 2/04/2026

#### — SQL in practice, NoSQL and Vector (9h) with Bruno Chaves

3 lectures: from 19/05/2026 to 10/06/2026

This lecture only requires basic knowledge of Python & SQL.

- o SQL in practice 19/05/2026
- o NoSQL databases 02/06/2026
- o Vector databases 10/06/2026

#### — Natural Language Processing (21h) to be announced

7 lectures: from 4/05/2026 to 22/06/2026

- o Background (refresh): basic concepts from Machine Learning and Neural Networks 4/05/2026
- o Word vectors: Latent Semantic Analysis (LSA), word2vec and GloVe 11/05/2026
- o Language Models: from N-gram models to Recurrent Neural Networks 18/05/2026
- o Sequence-to-sequence modelling and Machine Translation (MT) 1/06/2026
- o Attention, self-attention and Transformer-based language models 8/06/2026
- o Pretraining and Finetuning large language models 15/06/2026
- o Understanding ChatGPT 22/06/2026

#### — Reinforcement learning (18h) with Ana Basic

6 lectures: from 16/03/2026 to 27/04/2026

- o Introduction to reinforcement learning 16/03/2026
- o Markov decision processes and dynamic programming 23/03/2026
- o Bandit algorithms 30/03/2026
- o Reinforcement learning: stochastic approximation, algorithms, actor-critic, model-based 13/04/2026 & 20/04/2026
- o Reinforcement with approximation 27/04/2026



# SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

## IA/ML overview

12 hours, with Frédéric Lechenault

- **Lecture 1 (beginner): training for academics, overview**
- **Lecture 2 (beginner): history of AI**
- **Lectures 3 (beginner) and 4 (beginner): data science basics**

- Data types
- Dimensionality
- Visualization
- What does it mean? Neural networks, convolution, transformers, optimization ...
- Why we need matrices, vectors and calculus

## Statistics, probability, linear algebra

21 hours, with Alexandre Vérine

- **Lectures 1 (beginner) and 2 (advanced): linear algebra**

- Basic vector and matrix algebra
- Eigenvalues, eigenvectors
- Singular value decomposition
- Matrix inverse, orthogonal matrices
- Trace, determinant
- Matrix similarity, congruence
- Sylvester law of inertia

- **Lectures 3 (beginner) and 4 (beginner): calculus**

- Single variable calculus
- Integration
- Fundamental theorem of calculus
- Multivariate calculus
- Change of variables
- Jacobian, Hessian
- Taylor series
- Chain rule and backpropagation
- Optimization
- Conditions for local optimality
- Gradient descent
- Stochastic gradient descent
- Newton's method
- Constrained optimization, Lagrangian

- **Lectures 5 (beginner), 6 (advanced) and 7 (advanced):**

### probability and statistics

- Random variables, probability distributions
- Conditional distributions, sum and product rule
- Expectations, average, variance
- Important distributions (normal, multivariate, binomial, Poisson)
- Bayes theorem
- Conjugate priors
- Parameter estimation
- Maximum likelihood
- Regularization

- Examples: Laplace birth-rate problem, Luria-Delbrück, PCA
- Information theory
- Entropy
- Kullback-Leibler divergence
- Asymptotic inference
- Model selection
- Maximum entropy method
- Markov chain Monte Carlo, detailed balance, Metropolis rule

## Data structures and algorithms, Python

21 hours, with Muni Sreenivas Pydi

This class will include not only theoretical explanations, but also practical examples and hands-on exercises intended to improve algorithmic thinking.

- **Lecture 1 (beginner): introduction to Python programming**

- Basics of Python programming language
  - Data types, variables, and expressions
  - Control flow statements: if-else, for and while loops
  - Functions and modules

- **Lecture 2 (beginner): introduction to data structures**

- Lists, tuples, and sets
- Dictionaries and hash tables
- Stacks, queues, and linked lists
- Trees and graphs

- **Lecture 3 (advanced): algorithm analysis and complexity**

- Time and space complexity of algorithms
- Big-O notation and its significance
- Understanding algorithm efficiency and performance

- **Lecture 4 (advanced): sorting algorithms**

- Bubble sort, selection sort, and insertion sort
- Merge sort and quicksort
- Analysis of sorting algorithms

- **Lecture 5 (advanced): searching algorithms**

- Linear search and binary search
- Hashing and its applications
- Analysis of searching algorithms

- **Lecture 6 (advanced): advanced data structures and algorithms**

- Priority queues and heaps
- Advanced graph algorithms
- Dynamic programming
- Greedy algorithms

- **Lecture 7 (beginner): applications of data structures and algorithms**

- Algorithmic problem solving and puzzle solving
- Applications in real-world problems such as route planning and image processing



## SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

## Foundations of supervised and unsupervised learning

30 hours, with Muni Sreenivas Pydi

- **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

## Math and IA overview

12 hours, with Frédéric Lechenault

- **Lecture 1 (beginner):**
  - Overview of machine learning
  - Neural nets 101 (vocabulary, structure ...)
- **Lecture 2 (beginner): artificial intelligence for vision**
  - History
  - What is convolution? Why is it useful?
  - State-of-the-art in vision
- **Lecture 3 (beginner): artificial intelligence for time series and natural language processing**
  - Time series forecasting
  - Foundational models
  - ChatGPT & cie

- Lecture 4 (beginner)
  - Reinforcement learning
  - Other ML techniques (graph convolution ...)

## Data management and SQL

18 hours, with Muni Sreenivas Pydi

- **Lecture 1 (beginner): introduction to databases**
  - What is a database and why do we use them?
  - Types of databases (relational, NoSQL, etc.)
  - Common relational database management systems (RDBMS), e.g. MySQL
- **Lecture 2 (beginner): introduction to SQL**
  - Basics of SQL syntax: SELECT, FROM, WHERE, ORDER BY, LIMIT
  - Simple queries using one table
  - Data types and NULL values
  - Simple data manipulation: INSERT, UPDATE, DELETE
- **Lecture 3 (beginner): joins and relationships**
  - Inner join, left join, right join, full outer join
  - One-to-many, many-to-one, and many-to-many relationships
  - Foreign keys and primary keys
- **Lecture 4 (beginner): creating a SQL database**
  - Introduction to database creation
  - Database design considerations
  - Normalization, denormalization and trade-offs
- **Lecture 5 (advanced): advanced queries**
  - Aggregation and grouping: SUM, COUNT, AVG, MIN, MAX, GROUP BY, HAVING
  - Subqueries and nested queries
  - Combining queries with UNION, INTERSECT, EXCEPT
- **Lecture 6 (advanced): data manipulation and control**
  - Transactions and rollbacks
  - Constraints: NOT NULL, UNIQUE, PRIMARY KEY, FOREIGN KEY, CHECK
  - Views and indexes

## 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

# SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

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

- Tensors
- GPU computations
- Automatic differentiation

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

- Convolutional layers
- Image classification/segmentation

## — 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

## 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

# SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

— **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**

## 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**

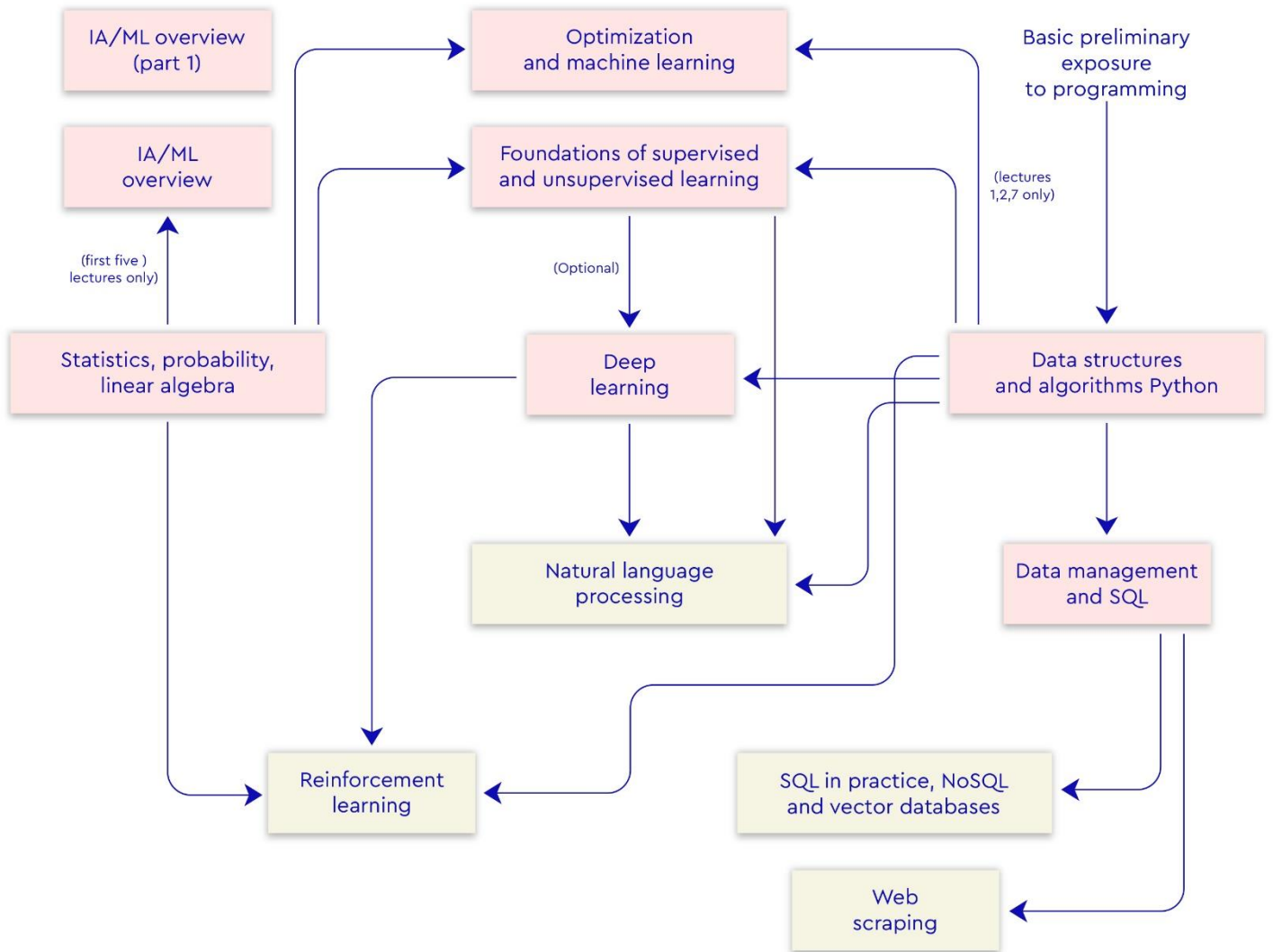
- Introduction to stochastic approximation
- Algorithms: TD-learning, SARSA, Q-learning
- Actor-critic
- Model-based reinforcement learning, upper confidence reinforcement learning (UCRL)

— **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** → **B** indicates that, to follow course B, it is necessary to be comfortable with the content of course A.