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 linar 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 transofmr)
·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

— **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

— **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)