

SEMESTER 10_5th YEAR ModIA

Introducing

Description

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en continu tout le long du semestre. En fonction des enseignements, elle peut prendre différentes formes : examen écrit, oral, compte-rendu, rapport écrit, évaluation par les pairs...

Practical info

Location(s)

 Toulouse

High Dimensional Statistics and Deep Learning



ECTS
3 crédits



Hourly volume
60h

Introducing

Description

The main theme of the course is learning methods, especially deep neural networks, for processing high dimensional media, such as images. We will cover the following topics:

* Neural networks and introduction to deep learning: definition of neural networks, activation functions, multilayer perceptron, backpropagation algorithms, optimization algorithms, regularization

* Convolutional neural networks: convolutional layer, pooling, dropout, deep convolutional network architectures (ResNet, Inception), transfer learning and fine tuning, applications to image classification, object detection, image segmentation, pose estimation, etc.

* Recurrent neural networks: sequence modelling, recurrent neuron, backpropagation through time, LSTM and GRU, applications to natural language processing and audio/video processing

* Neural networks for 3D processing: 3D convolutional networks for volumetric image processing (e.g. MRI), PointNet and PointNet++ networks for point cloud processing (e.g. LIDAR)

* Deep unsupervised learning and Generative Models : Autoencoders, Variational autoencoders (VAE), Generative Adversarial Networks (GAN)

* Functional decomposition on splines, Fourier or wavelets bases or Functional PCA : cubic splines, penalized least squares criterion, Fourier basis, wavelet bases, applications to nonparametric regression, linear estimators and nonlinear estimators by thresholding, links with the LASSO method. Functional PCA

* Anomaly detection: Main algorithms : One Class SVM, Random Forest, Isolation Forest, Local Outlier Factor. Applications to anomaly detection in functional data

Objectives

At the end of this module, the student will have understood and be able to explain (main concepts):

* How to use deep learning methods for classification in high dimension

* Classification of media or images

* Estimation of the prediction error

* Dimension reduction by projection onto orthonormal bases

* Anomaly detection

* Application of deep learning methods on real data set

At the end of this module, the student should be able to:

* Fit a deep neural network for media or image classification and regression

* Apply anomaly detection algorithms

* Implement deep learning methods in high dimension on real data sets with Python libraries.

Necessary prerequisites

Statistical modelling
Software for statistics

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en continu tout le long du semestre. En fonction des enseignements, elle peut prendre différentes formes : examen écrit, oral, compte-rendu, rapport écrit, évaluation par les pairs...

Practical info

Location(s)

 Toulouse

High Performance Scientific Computing



ECTS
3 crédits



Hourly volume
59h

Introducing

Description

- * Arnoldi's method for constructing Krylov's spaces, gmres solver, lanzcos solver, link with the conjugate gradient
- * deflation and extraction of spectral information (deflated CG, Ritz vectors), sequence of linear systems, techniques for computing eigenvalues/vectors
- * domain decomposition methods (Schwarz, P.-L. Lions), 1D and 2D theoretical study, application to parallel computing
- * multigrid geometric methods, smoothing property, V and FMG scheme
- * equivalence between sparse matrix and graph. Filling concept and dependencies between unknowns of a sparse problem (elimination tree). Matrix permutations to reduce filling. Operating complexity. Exploitation of parallelism in the factorization of sparse matrices,
- * Architecture of modern supercomputers (multi-core processors, multiprocessors or SMP and NUMA nodes, GPU accelerators, interconnection networks). Flynn's classification. Cache hierarchy. Principles of spatial and temporal locality. Parallel programming in OpenMP: parallel regions, synchronization directives, parallel loops, competition situations and tasks. Parallel programming in MPI: point-to-point and collective communications, reduction operations and possible optimizations. Performance analysis: Amdahl's law, high and low scalability, roofline model and critical path calculation.

Objectives

At the end of this module, the student will have understood and be able to explain (main concepts):

- * the principle of Krylov's methods to solve linear systems or compute eigenvalues and eigenvectors,
- * the concept of preconditioning, the construction and use of preconditioners,
- * theory and basic concepts of direct methods for sparse linear systems. Operating complexity and parallelism of direct methods,
- * basic notions of parallel computer architecture, programming models for shared memory (OpenMP) and distributed memory (MPI) systems,
- * basic concepts and methods for analyzing the performance of a parallel algorithm or code (Amdahl's law, cache hierarchy, principles of spatial and temporal locality, roofline model, critical path computation and high and low scalability).

At the end of this module, the student should be able to:

- * evaluate the costs (flops/memory) of the different methods,
- * analyze the influence of preconditioners,
- * use high-level languages for the discretization of partial differential equations,
- * program solvers, to parallelize simple codes according to the most adequate standard and to execute them on the appropriate resources,
- * to analyse the efficiency of a method with regard to the operational complexity, the computing time and the memory footprint used in a high-performance computing perspective.

Necessary prerequisites

- * Courses in Linear Algebra or Scientific Calculus, in particular the factorization methods LU or Cholesky
 - * Basics of computer architecture and imperative programming languages
-

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en continu tout le long du semestre. En fonction des enseignements, elle peut prendre différentes formes : examen écrit, oral, compte-rendu, rapport écrit, évaluation par les pairs...

Practical info

Location(s)

 Toulouse

Physics constrained machine learning



ECTS
3 crédits



Hourly volume
59h

Introducing

Description

Programme (detailed contents):

- Solution methods for PDEs based on ML: data centric approach, discretization based approach in large dimension, main NN architectures, stabilization of time-stepping methods, data augmentation, sampling of physics, methods based on energy conservation
- Design of data assimilation algorithm based on recurrent neural architectures. Variational Bayes methods based on KL, Jordan and Wasserstein divergences. Application to chaotic dynamical systems and comparison with state of the art variational and ensemble data assimilation techniques.
- Improvement (accuracy, speed) of existing discretization methods from numerical analysis with ML techniques. Model reduction with latent space techniques.

Objectives

At the end of this module, the student will have understood and be able to explain (main concepts):

- Main approaches for solving time dependent problem (EDP and Data assimilation) using ML

- Relevance of using physical constraints for solving problems with underlying physics (feature engineering), design of Neural networks
- Methods for handling nonlinearity and large scale (use of latent space, high performance computing)
- Performance of ML for solving problems with physical constraints.

At the end of this module, the student should be able to:

- Use ML for solving time dependent PDE and analysis the accuracy
- Analyse the HP performance of the solvers, and propose algorithmic enhancements
- Design a full data assimilation system based on ML, starting from a description of a system using partial differential equation and an observational system
- Assess the performance of a system, question the relevance of the mathematical assumptions

Necessary prerequisites

Numerical algebra for large scale, statistical estimation, non-convex smooth optimization, numerical solution of PDEs, data assimilation, machine learning

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en continu tout le long du semestre. En fonction des

enseignements, elle peut prendre différentes formes :
examen écrit, oral, compte-rendu, rapport écrit,
évaluation par les pairs...

Practical info

Location(s)

 Toulouse

Trusted Systems



ECTS
3 crédits



Hourly volume
60h

Introducing

Description

- Software and System Engineering
 - o Process, Methods and Tools
 - o Agile Methods
 - o Verification and Validation
 - o Certification and Qualification
 - o System Safety
 - o Test Driven Development
 - o Model Driven Development
- Modelling, Resolution and Proof
 - o Propositional and Predicate Logic
 - o Terms and Structural Induction
 - o Hoare Logic
 - o Program Proof
 - o Model Logic
 - o Fuzzy and Probabilistic Logic
 - o Constraint Logic Programming
 - o SAT/SMT solver

Objectives

Software play a key role in many industrial domains, including safety critical ones (transportation, health, business, ...) where defects can have a strong direct, or indirect, impact on human life.

This UE provides 2 courses that contribute to improving

the quality of software and the trust we can have in it.

- Software and System Engineering provides the core concepts needed to build trusted software intensive systems. Model Driven Engineering will be a core element as it allows to model application domain specific elements and to ease the building of domain specific tools.

- Modeling, Resolution and Proof provides the elements from discrete mathematics that allowing modeling formally the requirements for software systems and to carry formal proof of correctness about their behavior. These elements are also at the root of symbolic artificial intelligence in order to model knowledge, structured data and to explain the decision taking by systems. We will also illustrate how these tools can be used for discrete optimization.

This UE tackles both the theoretical knowledge and methods, and their use in representative tools.

Necessary prerequisites

Computer use
Programming
Basic general algebra

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en

continu tout le long du semestre. En fonction des enseignements, elle peut prendre différentes formes : examen écrit, oral, compte-rendu, rapport écrit, évaluation par les pairs...

Practical info

Location(s)

 Toulouse

IA Frameworks



ECTS
3 crédits



Hourly volume
24h

Introducing

Description

This course provides various application for data scientist.

* Data Labelling: Introduction to crowdsourcing platforms and their limitations (especially ethical limitations), aggregation of annotations (input/output agreement, find-fix-verify pattern, annotators evaluation through Expectation-Maximization), other annotation techniques, semi-automated annotation, active learning.

* Introduction to cloud computing with Google Cloud: Presentation of the main Google Cloud's tools (Compute Engine, IA instances, billing, etc) Writing script for terminal execution (no notebook). Send codes and data to a n instance, code execution, get the results back etc. OPTIONAL: Initiate instance with docker image to better reproducibility

* Natural language processing: Text cleaning (stemming, tokenisation). Vectorisation (one-hot encoder, tf-idf) Words Embedding (Word2vec, glove). Text classification, text generation (RNN). OPTIONAL: Text translation; chat bot.

* Reinforcement learning: Introduction to various types of algorithms. policy-based algorithms (Policy gradient) and Value-based algorithms (Q-Learning, DQN). Good practices to be used for real-case application OPTIONNEL: Actor-critic

* Recommendation system: user/user or item/item filters. Latent-based methods (NMF, SVD, SoftImpute), neural recommender system

Objectives

At the end of this module, the student will have understood and be able to explain (main concepts):

- Main concept of data labelisation and related tools.
- Main algorithms of natural language processing
- Main concepts of reinforcement learning.
- Main concepts of recommendation system.
- How to access tools to perform efficiently and with enough computation power those algorithms

The student will be able to:

- Organize en data labelisation strategy.
- Handle various types of complex datasets (Image, text, video, notations,...)
- Identify the correct algorithm to solve various problem on these data.
- run these algorithms on the appropriate ressource (cloud machine, container? GPU?)
- Share efficiently the results obtain

Necessary prerequisites

Exploratory Data Analysis

Machine Learning / Deep Learning (MLP, RNN, CNN)

R and Python languages

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en continu tout le long du semestre. En fonction des enseignements, elle peut prendre différentes formes : examen écrit, oral, compte-rendu, rapport écrit, évaluation par les pairs...

Practical info

Location(s)

 Toulouse

In-company training 4

 ECTS
15 crédits

 Hourly volume

Introducing

Description

Objectives

See Company training 1 information sheet

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en continu tout le long du semestre. En fonction des enseignements, elle peut prendre différentes formes : examen écrit, oral, compte-rendu, rapport écrit, évaluation par les pairs...

Practical info

Location(s)

 Toulouse