



FORMATION PAR APPRENTISSAGE : MODIA

Version du 17/06/2021

Année universitaire : 2021-2022

Formation par apprentissage : ModIA 1

 Schéma de la formation, volumes horaires, crédits ECTS et International 2

 Volumes horaires, Crédits ECTS et Modalités d'évaluations 3

 Mobilité à l'internationale, TOEIC et Implication citoyenne 3

 Séjour à l'étranger 3

 TOEIC 3

 Implication citoyenne 3

 Semestre 7 3

 Semestre 8 3

 Semestre 9 4

 Semestre 10 4

 Semestre 11 4

 Calendrier d'alternance 4

 Syllabus des enseignements 5

 4ème année 5

 Semestre 7 5

 Modélisation & Calcul Scientifique 5

 Eléments de modélisation statistique 5

 Optimisation et optimisation stochastique 6

 Analyse des données 6

 Sciences Humaines 6

 Semestre 8 7

 Traitement du Signal 7

 Infrastructure pour le cloud et le big data 7

 Programmation fonctionnelle & théorie des graphes 7

 Machine Learning 7

 Sciences humaines 8

 5ème année 9

 Semestre 9 9

 Modélisation & Éléments Finis 9

 Assimilation de données 9

 Plans d'expériences et Méta-Modèles 9

 Durées de vie et Processus de Poisson appliqués à l'actuariat et la fiabilité 10

 Communiquer dans les organisations 10

 Semestre 10 11

 Statistique en grande dimension et Apprentissage profond 11

 HPSC 11

 Apprentissage sous contraintes physiques 11

 Systèmes de confiance 12

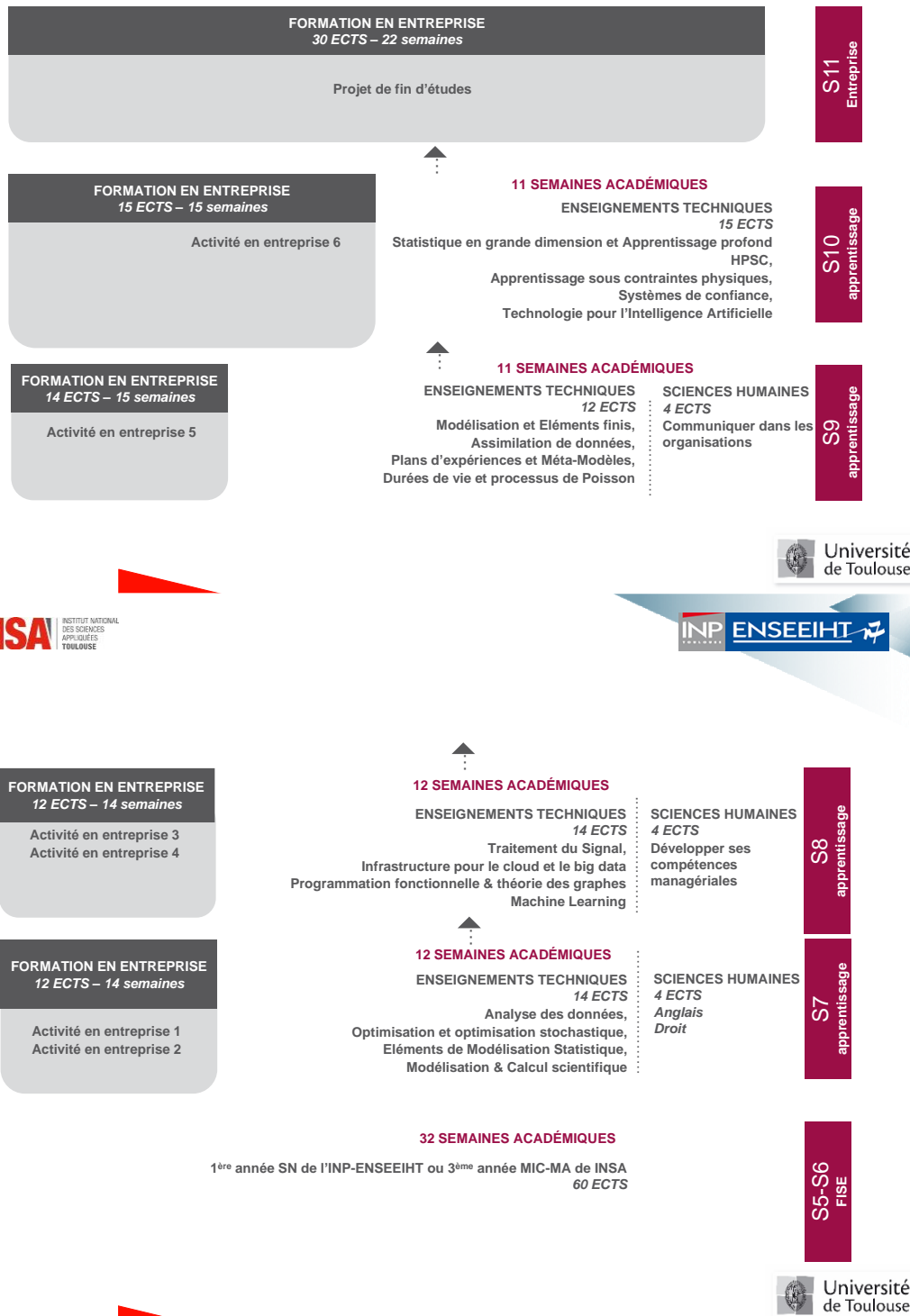
 Technologies pour l'Intelligence Artificielle 12



SCHEMA DE LA FORMATION, VOLUMES HORAIRES, CREDITS ECTS ET INTERNATIONAL



Schéma du cursus du Double Diplôme Mathématiques Appliquées INSA et Informatique & Télécommunication de l'INP-ENSEEIHHT sous statut Apprenti





VOLUMES HORAIRES, CREDITS ECTS ET MODALITES D'EVALUATIONS

La formation s'étale sur 5 semestres à partir du début de 4^{ème} année post-bac : année 4, année 5 et premier semestre de la 6^{ème} année.

Le volume total d'heures de formation en école est de : 1344h pour 66 crédits ECTS selon le découpage ci-dessous. Le reste de la formation est entièrement réalisé en entreprise.

Les évaluations sont effectuées selon une ou plusieurs modalités suivantes : Examen écrit, bureau d'études, compte-rendu de TP et projet (avec éventuelle soutenance).

MOBILITE A L'INTERNATIONALE, TOEIC ET IMPLICATION CITOYENNE

Pour obtenir les diplômes des deux écoles, les trois critères suivants doivent être validés.

SEJOUR A L'ETRANGER

A l'exception des étudiants étrangers, une mobilité à l'international d'au minimum 12 semaines est exigée. Toute mobilité antérieure à l'international, n'ayant pas déjà été prise en compte par un diplôme précédent, est acceptée dans ModIA. Cette mobilité peut être effectuée en une ou plusieurs périodes.

Il est fortement conseillé de réaliser ce ou ces séjours dans le cadre de mission(s) à l'international dans une filiale, chez un sous-traitant, chez un partenaire industriel, ou dans un laboratoire de recherche. Les apprentis qui ne pourraient pas disposer d'une telle opportunité pourront effectuer leur mobilité à l'international en suivant un semestre académique dans une université étrangère, par exemple au cours du semestre S9, sous réserve de compatibilité avec le cursus ModIA.

TOEIC

Un score minimum de 785 est nécessaire.

IMPLICATION CITOYENNE

L'UF implication citoyenne (appelée également engagement civique pour les étudiants de l'ENSEEIH) doit être validée pendant les années de la formation ModIA ou en amont pendant les années post-bac.

SEMESTRE 7

- ✓ Enseignements à l'INSA (12 semaines, 368h, 18 ECTS)
 - Modélisation & Calcul scientifique (72 h, 4 ECTS)
 - Eléments de modélisation statistique (72 h, 3 ECTS)
 - Optimisation et Optimisation stochastique (82 h, 4 ECTS)
 - Analyse des données (62 h, 3 ECTS)
 - SHS (40 heures, 4 ECTS)
- ✓ Activités en Entreprise (14 semaines, 12 ECTS)

SEMESTRE 8

- ✓ Enseignements à l'ENSEEIH (12 semaines, 368h, 18 ECTS)
 - Traitement du Signal (62 h, 3 ECTS)
 - Infrastructure pour le cloud et le big data (62 h, 3 ECTS)
 - Programmation fonctionnelle & Théorie des Graphes (82 h, 4 ECTS)
 - Machine Learning (82 h, 4 ECTS)
 - SHS (40 heures, 4 ECTS)
- ✓ Activités en Entreprise (14 semaines, 12 ECTS)



SEMESTRE 9

- ✓ Enseignements à l'INSA (11 semaines, 344h, 16 ECTS)
 - Modélisation et éléments finis (66h, 3 ECTS)
 - Assimilation de données (66 h, 3 ECTS)
 - Plans d'expériences et Méta-Modèles (66 h, 3 ECTS)
 - Durées de vie et Processus de Poisson appliqués à l'Actuariat et à la Fiabilité (66 h, 3 ECTS)
 - SHS (40 heures, 4 ECTS)
- ✓ Activités en Entreprise (15 semaines, 14 ECTS)

SEMESTRE 10

- ✓ Enseignements à l'ENSEEIH (11 semaines, 264h, 15 ECTS)
 - Statistique en grande dimension et Apprentissage profond (60 h, 3 ECTS)
 - High performance scientific computing (60 h, 3 ECTS)
 - Apprentissage sous contraintes physiques (60 h, 3 ECTS)
 - Systèmes de confiance (60 h, 3 ECTS)
 - Atelier de l'IA (24h, 3 ECTS)
- ✓ Activités en Entreprise (15 semaines, 15 ECTS)
 - Activités (15 ECTS)

SEMESTRE 11

- ✓ PFE en Entreprise (22 semaines, 30 ECTS)
- ✓ Soutenance en école (1 semaine)

CALENDRIER D'ALTERNANCE

		Semestre 1																						
		S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S01	S02	S03	S04	S05	
1ère Année			4					5				4					5					4		
2ème Année				6					4				5							7				
		Semestre 2																						
		S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26		
1ère Année			4					5					5						7					
2ème Année			3				6						7							5				
		Eté																						
		S27	S28	S29	S30	S31	S32	S33	S34	S35														
1ère Année						9																		
2ème Année						9																		
		Semestre 1																						
		S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S01	S02	S03	S04	S05	S06
3ème Année																								22

Légende

- Présence à l'école
- Présence en entreprise

Remarque

Les années à 53 semaines, la semaine supplémentaire est passée en entreprise



SYLLABUS DES ENSEIGNEMENTS

4EME ANNEE

SEMESTRE 7

MODELISATION & CALCUL SCIENTIFIQUE

Rappels de calcul différentiel :

- ✓ DL d'une fonction de plusieurs variables, gradient, matrice jacobienne, matrice hessienne, opérateur Laplacien... 4^{ème} année
- ✓ Dérivation des fonctions composées
- ✓ Formule de Green-Ostrogradsky, intégration par partie pour les fonctions de plusieurs variables

Equations différentielles (EDO)

- ✓ Exemples de problèmes de physique, biologie, économie, modélisés par des EDO ou des systèmes d'EDO
- ✓ Notions théoriques essentielles sur les EDO : existence et unicité locale et globale, stabilité
- ✓ Méthodes numériques pour les EDO : méthodes de Runge-Kutta, méthodes multipas, cas des systèmes raides

Equations aux dérivées partielles (EDP)

- ✓ Exemples de problèmes de physique, biologie, ou encore économie, modélisés par des EDP ou des systèmes d'EDP (linéaires et non-linéaires)
- ✓ Classification et notions théoriques essentielles sur quelques EDP modèles : existence, unicité, estimation d'énergie, principe du maximum, conditions aux limites, calcul de solutions exactes
- ✓ Introduction aux méthodes de Différences Finies et Volumes Finis sur quelques problèmes modèles. Applications à travers des TP

Problèmes aux valeurs

- ✓ Modes propres, valeurs propres d'un opérateur
- ✓ Décomposition modale, solutions à variables séparées
- ✓ Résolution numérique par méthode de Fourier

Projet sur la résolution numérique d'un modèle couplé

ELEMENTS DE MODELISATION STATISTIQUE

Tests non paramétriques : fonction de répartition empirique, test de Kolmogorov d'adéquation, tests de comparaison de deux échantillons (Kolmogorov-Smirnov et test de Wilcoxon), tests de normalité (Kolmogorov et Shapiro-Wilk)

Tests d'ajustement et d'indépendance : test du khi-deux d'ajustement, d'adéquation à une famille de loi, d'indépendance et d'homogénéité

Modèle linéaire : estimation des paramètres (modèle régulier et singulier sous contraintes d'identifiabilité), intervalle de confiance pour les paramètres, intervalle de confiance pour la réponse moyenne, intervalle de prédiction, test de Fisher de sous-modèle, sélection de variables. Régression linéaire, ANOVA à un et deux facteurs, ANCOVA

Modèle linéaire généralisé : inférence statistique, sélection de variables, Régression logistique, modèle loglinéaire

Les TP et le projet seront effectués avec le logiciel R.



OPTIMISATION ET OPTIMISATION STOCHASTIQUE

L'optimisation non-convexe et différentiable

- ✓ Globalisation des méthodes (Wolfe, Trust-Region)
- ✓ Équations du premier ordre : Théorie de KKT
- ✓ Méthodes d'ordre 2 : L-BFGS
- ✓ Méthodes de gradient projeté

Optimisation stochastique

- ✓ Optimisation par Batch / Epoch
- ✓ Momentum

Optimisation convexe non-lisse

- ✓ Théorie du sous-gradient, transformée Legendre
- ✓ Algorithmes proximaux (Splitting et Acceleration)

ANALYSE DES DONNEES

Introduction à Python

Visualisation de données. Conditional plots

Analyse en composantes principales (ACP)

Applications de l'ACP

- ✓ à la classification de données quantitatives : analyse linéaire discriminante. Lien avec l'approche probabiliste. Analyse quadratique discriminante.
- ✓ à l'exploration d'une matrice de dissimilarités : positionnement multi-dimensionnel.

Clustering : k-means, classification hiérarchique ascendante, DBSCAN.

Généralités de l'apprentissage statistique : compromis biais-variance, validation (externe, k-fold, courbes ROC), corrélation / causalité, risque de sur-apprentissage...

Arbres de régression et de classification (CART)

Bagging, forêts aléatoires

Imputation de données manquantes. missForest

SCIENCES HUMAINES

Anglais

Pour les publications scientifiques

- ✓ Compréhension écrite et expression orale à partir de documents techniques (contexte ; vocabulaire spécifique)
- ✓ Définir les parties d'un abstract et écrire un abstract selon les conventions en vigueur
- ✓ Concevoir un poster pour présenter des travaux scientifiques
- ✓ Comprendre les enjeux éthiques liés à son domaine.

Pour les étudiants les plus faibles :

- ✓ Cours spécifiques de préparation au TOEIC

Droit

- ✓ Les principes et les règles du droit dans des situations simples.
- ✓ Prise en compte des paramètres juridiques dans les fonctions d'ingénieur.



SEMESTRE 8

TRAITEMENT DU SIGNAL

Analyse Hilbertienne

- ✓ Espaces de Hilbert, Exemples d'espaces de Hilbert : L^2 , L^2 à poids
- ✓ Approximation : bases hilbertiennes, polynômes orthogonaux
- ✓ Projection sur un convexe, Convergence faible.
- ✓ Analyse de Fourier et aspects algorithmique (FFT, Fourier à fenêtre, échantillonnage) : application au traitement du son, traitement des images.

Ondelettes

- ✓ Ondelettes de Haar/Analyse MultiRésolution
- ✓ Autres types d'ondelettes (1d, 2d) : application au traitement du son et des images
- ✓ Transformée en ondelettes, lien coefficient/régularité

INFRASTRUCTURE POUR LE CLOUD ET LE BIG DATA

Partie cloud : C1) introduction au cloud ; TP1) utilisation de AWS/EC2 ; C2) virtualisation (KVM) ; TP2) KVM/Xen ; C3) conteneurs (Docker et outils) ; TP3) Docker ; C4) administration de cluster virtualisé (OpenStack, Kubertenes) ; TP4) Kubernetes ; C5) Services du cloud (AWS/Azure) ; TP5) AWS/scaling

Partie big data : C1) introduction au big data + Hadoop ; C2) Hadoop ; TP1) Hadoop ; C3) Spark ; TP2) Spark ; C4) Spark-streaming + Storm ; TP3) Spark-streaming ; TP4) passage à l'échelle ; TP5) passage à l'échelle + SparkML

PROGRAMMATION FONCTIONNELLE & THEORIE DES GRAPHES

Programmation fonctionnelle

- ✓ Fonction et composition de fonctions
- ✓ Types inductifs
- ✓ Filtrage
- ✓ Ordre supérieur
- ✓ Itérateur

Théorie des graphes

- ✓ Définitions et Concepts élémentaires
- ✓ Connexité
- ✓ Graphes Eulériens et Hamiltoniens
- ✓ Planarité et Coloration
- ✓ Structures de données
- ✓ Algorithmes de parcours

MACHINE LEARNING

Sélection de modèles : variables, pénalisation (ridge, lasso, etc.)

Clustering : modèles de mélange, algorithme EM

Méthodes à noyaux : SVM, ACP spectral, etc...

Boosting. Xgboost

Réseaux de neurones (introduction)

Apprentissage par renforcement

Biais dans les algorithmes et éthique des données



SCIENCES HUMAINES

Anglais

Développer son éloquence et ses compétences en communication écrites, orales, en interaction et médiation :

- ✓ Rédiger un *reaction paper* (commentaire argumenté), selon les conventions en vigueur, sur un sujet d'actualité
- ✓ Organiser et participer à des *debates* (joutes oratoires), selon les conventions en vigueur, sur des sujets d'actualité

Pour les étudiants les plus faibles :

- ✓ Cours spécifiques de préparation au TOEIC

Finance

- ✓ Comprendre le rôle de la finance dans les entreprises et dans la prise de décision
- ✓ Prendre en compte des paramètres financiers dans les entreprises
- ✓ Analyser le concept de l'actualisation.



5EME ANNEE

SEMESTRE 9

MODELISATION & ÉLEMENTS FINIS

Analyse mathématique et principes de la méthode EF

- ✓ Analyse (EDP elliptiques linéaires): solutions faibles, espaces de Sobolev H_m , théorie de Lax-Milgram.
- ✓ Principe des EF : discrétisation, approximation, implémentation, estimations d'erreur a-priori.
- ✓ Courbes de convergence, validation codes de calcul
- ✓ TP (Python–Fenics ou Julia) programmation algorithme d'assemblage

Modélisation et compléments EF

- ✓ Modélisation par EF (TP FreeFEM++ ou Python-Fenics). Ex : écoulements géophysiques – hydraulique spatiale (ondes diffusantes : convection – diffusion non linéaire).
- ✓ Compléments méthode EF : Terme de transport et stabilisation (ex : SUPG) ; Termes non linéaires et linéarisations ; Raffinement de maillage – concept de estimateurs a-posteriori. TP Python-Fenics.
- ✓ Modèles réduits POD : Stratégie offline – online. TP Python-Fenics.

Couplages de modèles et de codes de calcul

- ✓ Application de la modélisation EF au problème de l'élasticité
- ✓ Couplage faible de domaines élastiques (pénalisation, mortar, Nitsche)
- ✓ Notion d'interface non-conforme entre les domaines
- ✓ Résolution itérative non-intrusive du couplage.
- ✓ TP Python : calcul de la propagation d'une fissure avec utilisation de codes en boîtes noires.

ASSIMILATION DE DONNEES

Introduction. Préliminaires (statistiques, conditionnement), méthodes Bayésienne, modèles d'EDP à coefficients incertains, moindres carrés non linéaires, problèmes inverses mal posés, régularisation.

Assimilation de données et contrôle optimal. Contrôle optimal d'un modèle d'EDP, modèle linéaire tangent, système d'optimalité, calcul de gradient par méthode adjointe. Calcul d'exemples. Algorithmiques (3D/4D-var, incrémental). Stratégies de diagnostics. Validation de codes adjoints. Concept de différenciation automatique. Modélisation de matrices de variance. TP Python – Fenics eg en hydrologie spatiale (convection-diffusion non linéaire).

Estimation séquentielle et filtre de Kalman. Formalisation statistique pour un système dynamique observé, transfert d'optimalité et lien avec l'approche variationnelle, filtrage et lissage, filtre racine carrées et erreurs numériques, réduction de la dimension et filtres réduits.

Estimation ensembliste. Erreur d'échantillonnage, dérivation des principaux filtre/lisseurs stochastiques (EnKF) et déterministes (ETKF), hyperparamètres (inflation, localisation)

PLANS D'EXPERIENCES ET META-MODELES

Plans d'expériences

- ✓ Modèles linéaires de covariance, interactions multiples, modèles mixtes
- ✓ Principe de la randomisation des expériences et plans d'expériences classiques
- ✓ Plans factoriels, plans fractionnaires
- ✓ Exemples d'application

Metamodélisation

- ✓ Introduction : exemples d'applications



- ✓ Deux métamodèles célèbres : polynômes de chaos, processus gaussiens (krigeage)
- ✓ Simulation de processus gaussiens non conditionnels / conditionnels
- ✓ Prise en compte d'information métier et personnalisation de noyaux de covariance
- ✓ Optimisation assistée par métamodèle (optimisation bayésienne)
- ✓ Planification d'expériences numériques : focus sur les plans remplissant l'espace
- ✓ Analyse de sensibilité globale : focus sur la décomposition ANOVA (décomposition de Sobol)
- ✓ Application industrielle : quantification d'incertitudes.

DUREES DE VIE ET PROCESSUS DE POISSON APPLIQUES A L'ACTUARIAT ET LA FIABILITE

Durées de vie : censure, troncature, risque instantané, modèles de vieillissement, estimations paramétriques et non paramétriques (Kaplan-Meier et Nelson-Aalen) des fonctions d'intérêt, modèles avec covariables (modèles de régression de Cox et d'Aalen), approche bayésienne.

Structure d'un système et Fiabilité : diagramme de fiabilité, systèmes série, parallèle, k/n ou mixte, fonction de structure.

Modélisation de la récurrence des pannes en Fiabilité ou des sinistres en Actuariat : définition(s) et construction(s) d'un processus de Poisson (homogène ou non homogène) ou de renouvellement, statistique inférentielle pour les processus de Poisson homogènes (estimation ponctuelle, intervalles de confiance et tests sur l'intensité).

Mathématiques et Apprentissage pour l'actuariat : étude des processus de Poisson composés et du modèle de Cramér-Lundberg, mesures de risques, algorithmes d'apprentissage pour les problèmes de tarification et de provisionnement.

Approfondissement par projet (au choix) :

- ✓ Fiabilité des systèmes réparables : processus intensité de défaillance, processus de renouvellement, maintenance corrective et préventive, modèles de réparation imparfaite, modèles de dégradation et politique de maintenance.
- ✓ Statistique : test d'homogénéité, Statistique inférentielle pour le processus de Weibull.
- ✓ Actuariat : modèles de type Cramér-Lundberg, processus de Hawkes et produits dérivés en assurance.

COMMUNIQUER DANS LES ORGANISATIONS



SEMESTRE 10

STATISTIQUE EN GRANDE DIMENSION ET APPRENTISSAGE PROFOND

Réseaux de neurones et introduction à l'apprentissage profond : définition des réseaux de neurones, fonctions d'activation, perceptron multicouches, algorithme de rétropropagation du gradient, algorithmes d'optimisation, régularisation.

Réseaux de neurones convolutionnels : couche convolutionnelle, pooling, dropout, architecture des réseaux convolutionnels (ResNet, Inception), transfert d'apprentissage, applications à la classification d'images, la détection d'objet, la segmentation d'image, l'estimation de posture, etc.

Réseaux de neurones récurrents : modélisation de séquences, neurone récurrent, rétropropagation à travers le temps, LSTM et GRU, applications au traitement du langage naturel et au traitement des signaux audio et vidéo

Réseaux de neurones et 3D : réseaux convolutifs 3D pour le traitement des données volumétriques (ex: IRM), réseaux PointNet et PointNet++ pour le traitement des nuages de points 3D (ex: LIDAR). * Apprentissage profond non-supervisé et modèles génératifs : Autoencodeurs, Auto-encodeurs variationnels (VAE) , Réseaux Génératifs Antagonistes (GAN)

Décomposition fonctionnelle sur des bases de Spline, Fourier , ondelettes ou ACP fonctionnelle Functional decomposition on splines, Fourier or wavelets bases : splines cubiques, critère des moindres carrés pénalisés, bases de Fourier, bases d'ondelettes, application en régression non paramétrique, estimateurs linéaires et non linéaires par seuillage, liens avec la méthode LASSO. ACP fonctionnelle.

Détection d'anomalies : Principaux algorithmes : One Class SVM, Random Forest, Isolation Forest, Local Outlier Factor. Applications à la détection d'anomalies pour des données fonctionnelles.

HPSC

- ✓ Méthode d'arnoldi pour construire les espaces de Krylov, solveur gmres,
- ✓ Solveur de lanczos, lien avec le gradient conjugué
- ✓ Déflation et extraction d'informations spectrales (deflated CG, vecteurs de Ritz), séquence de systèmes linéaires, techniques pour le calcul des valeurs/vecteurs propres
- ✓ Méthodes de décomposition de domaine (de Schwarz, de P.-L. Lions), étude théorique 1D et 2D, application au calcul parallèle
- ✓ Méthodes multigrille géométrique, propriété de lissage, schéma V et FMG
- ✓ Equivalence entre matrice creuse et graphe. Concept de remplissage et dépendances entre inconnus d'un problème creux (arbre d'élimination). Permutations de matrices permettant de réduire le remplissage. Complexité opératoire. Exploitation du parallélisme dans la factorisation de matrices creuses
- ✓ Architecture des supercalculateurs modernes (processeurs multi-coeurs, multiprocesseurs ou noeuds SMP et NUMA, accélérateurs GPU, réseaux d'interconnexion). Classification de Flynn. Hiérarchie des mémoires caches. Principes de localité spatiale et temporelle. Programmation parallèle en OpenMP: régions parallèles, directives de synchronisation, boucles parallèles, situations de compétition et tâches. Programmation parallèle en MPI : communications point a point et collectives, opérations de réduction et éventuellement optimisations. Analyse des performances : loi d'Amdahl, scalabilité forte et faible, modèle roofline et calcul du chemin critique.

APPRENTISSAGE SOUS CONTRAINTES PHYSIQUES

- ✓ Méthodes de résolutions d'EDP à base de ML : approche data based, approche basée sur une discrétisation en grande dimension, principales structures de réseaux de Neurones, techniques de stabilisation en temps, augmentation de données, échantillonnage de la physique, méthodes basées sur la conservation de l'énergie totale
- ✓ Ré-écriture des méthodes pour l'assimilation de données en termes d'algorithmes de réseau récurrents à mémoire. Méthodes bayésiennes-variationnelles (BV) basées sur les divergences de KL, de Jordan, de Wassertein.



- ✓ Amélioration (précision, vitesse) de techniques de discrétisation usuelles par ML et réduction d'espace par espaces latents variationnels

SYSTEMES DE CONFIANCE

Génie du Logiciel et des Systèmes

- ✓ Processus, Méthodes et Outils
- ✓ Méthodes Agiles
- ✓ Vérification et Validation
- ✓ Certification et Qualification
- ✓ Sécurité de fonctionnement
- ✓ Développement Dirigé par les Tests
- ✓ Développement Dirigé par les Modèles

Modélisation, Résolution, Preuve

- ✓ Logique des propositions et des prédicats
- ✓ Termes et Induction structurelle
- ✓ Logique de Hoare
- ✓ Preuves de programme
- ✓ Logique modale
- ✓ Logique floue et probabiliste
- ✓ Programmation Logique Contrainte
- ✓ Résolution SAT/SMT

TECHNOLOGIES POUR L'INTELLIGENCE ARTIFICIELLE

Labellisation des données : Présentation des plateformes de crowdsourcing et de leurs limites (notamment éthiques), Aggrégation d'annotations (input/output agreement, Find Fix Verify, évaluation conjointe des annotateurs par Expectation-Maximization), autres méthodes d'annotation, techniques d'annotation semi-automatique, apprentissage actif.

Introduction au cloud computing avec Google Cloud : Présentation de l'outil google cloud (Compute Engine, IA instances, facturation, etc..) Ecriture de script python exécutable via un terminal (pas de notebook), envoi des codes et des données, exécution des codes, récupération des résultats. OPTIONNEL: Initialiser une instance à l'aide d'une image docker pour une meilleure reproductibilité des résultats.

Traitement naturel du langage : nettoyage de texte (racinisation, tokenisation). Vectorisation (one-hot encoder, tf-idf) Words Embedding (Word2vec, glove). Classification de texte. Génération de texte (RNN). OPTIONNEL : Traduction, élaboration d'un chatbot etc..

Apprentissage par renforcement : Introduction aux différentes familles d'algorithmes. Les algorithmes policy-based (Policy gradient) et les algorithmes Value-based (Q-Learning, DQN). Parcours des différentes techniques utilisées pour leur application en cas réel. OPTIONNEL : Actor-critic

Système de recommandation : filtre utilisateur/utilisateur ou produit/produit. Méthodes basées sur les facteurs agents (NMF, SVD, SoftImpute). Système de recommandation profond.