USSI6H - Mathématiques S3

Présentation

Programme

Contenu

Positionnement : Semestre 3 **UE : Mathématiques** ECTS : 6

Nombre d'heures : 126h

Modalité : Présentiel (100%)

L'UE Mathématiques est constituée de 3 éléments constitutifs d'UE :

- Modélisation statistique
- Recherche opérationnelle avancée
- Apprentissage statistique supervisé : modélisation, sélection et évaluation en haute dimension

La note finale associée à cette UE est calculée en pondérant les notes des 3 ECUE avec les coefficients associés.

ECUE 1 : Modélisation statistique (Coef : 1) - 42h (Cours, TD) - 10h (Estimation temps de travail personnel)

Objectifs :

- Comprendre les concepts fondamentaux de la modélisation statistique.
- Acquérir des compétences pratiques pour résoudre des problèmes de modélisation statistique.

Contenu :

1. Introduction à la Modélisation statistique
2. Principes de l'inférence statistique
3. Intervalles de confiance
4. Vraisemblance
5. Test statistique
6. Modèle linéaire simple
7. Modèle linéaire multiple
8. Analyse de la variance
9. Analyse de la covariance
10. Modèle de mélange
11. Algorithme Espérance Maximisation (EM)
12. Quelques applications sous le logiciel R

Mis à jour le 09-04-2026

Code : USSI6H

Unité spécifique de type mixte

7 crédits

Responsabilité nationale : EPN06 - Mathématique et statistique / Giorgio RUSSOLILLO

Contact national :

Cnam - Nouvelle Aquitaine

2 Avenue Gustave Eiffel
Téléport 2
86960 Futuroscope Chasseneuil
05 49 49 61 20

na_info@lecnam.net

ECUE 2 : Recherche opérationnelle avancée (Coef : 1) - 42h (Cours, TP, Projet) - 10h (Estimation temps de travail personnel)

Objectifs : Ce cours est la suite du cours de recherche opérationnelle de première année. Il a pour but de présenter des notions de recherche opérationnelle avancée et d'aide à la décision indispensables pour de futurs ingénieurs, décideurs, responsables de projets.

Contenu :

- Modélisation sous forme de programme linéaire
- ALGORITHMES DU SIMPLEXE
- DUALITÉ (définition ; liens entre le primal et le dual, correspondance entre l'optimum du primal et du dual ; relations d'exclusion)
- PROGRAMMATION LINÉAIRE EN 0-1 et PROGRAMMATION LINÉAIRE EN NOMBRES ENTIERS (modélisation, algorithmes de résolution, saut de dualité)

Compétences visées :

- Aptitude à modéliser des problèmes issus de l'entreprise.
- Assimilation de méthodes et d'algorithmes fondamentaux en recherche opérationnelle et aide à la décision (en particulier pour l'optimisation de programmes linéaires).

Modalités d'évaluation : Examen final

ECUE 3 : Apprentissage statistique supervisé : modélisation, sélection et évaluation en haute dimension (Coef : 1) - 42h (Cours, TP, Projet) - 10h (Estimation temps de travail personnel)

Présentation du cours :

Ce cours propose une introduction approfondie à la modélisation statistique prédictive dans le cadre de l'apprentissage supervisé. Il combine les fondements théoriques avec une mise en œuvre pratique en langage R.

L'un des axes centraux du cours est la compréhension des limites des approches classiques face à la complexité des données : forte dimensionnalité, multicolinéarité, non-linéarités, faible rapport observation/variable. Ces situations nécessitent des approches adaptées.

Les apprentis seront ainsi initiés aux principes fondamentaux de l'apprentissage statistique :

- Comment évaluer la performance d'un modèle au-delà de l'ajustement aux données ?
- Comment gérer le compromis biais-variance et éviter le surapprentissage ?
- Quels outils permettent de maintenir des modèles robustes et interprétables en haute dimension ?
- Quelle stratégie adopter pour choisir un modèle parmi plusieurs alternatives possibles ?

Ce cours met également l'accent sur une démarche projet, en équipe, pour développer une réelle autonomie analytique et apprendre à communiquer efficacement les résultats statistiques.

Objectifs d'apprentissage :

1. Comprendre les fondements de l'apprentissage statistique supervisé
    1. Décomposer les composantes d'un modèle prédictif : variable cible, variables explicatives, fonction d'apprentissage, critère de performance.
    2. Identifier les différentes sources d'erreur de prédiction : biais, variance, bruit, et comprendre leur rôle dans la performance d'un modèle.
    3. Évaluer les effets de la complexité du modèle : analyser les conséquences du sous-apprentissage et du sur-apprentissage sur la capacité à généraliser.
2. Maîtriser les outils de modélisation statistique pour des données multivariées
    1. Maitriser et mettre en œuvre des techniques de régression adaptées à nature de la variable cible (numérique ou catégorielle) et de sa relation avec les prédicteurs (linéaire ou non linéaire) : régression des moindre carrés, régression polynomiale, splines, régression logistique (binaire et multinomiale).
    2. Utiliser des approches robustes pour l'estimation des paramètres (Bootstrap, Jackknife).
    3. Appliquer des méthodes non paramétriques comme les *k*-plus proches voisins, tout en comprenant les limites induites par la malédiction de la dimensionnalité.
    4. Justifier le choix du modèle en contexte appliqué, en tenant compte des enjeux pratiques et théoriques.
3. Comprendre les limites des méthodes classiques en haute dimension et les surmonter
    1. Identifier les contextes où les méthodes standards deviennent instables ou inefficaces.
    2. Expliquer pourquoi la forte dimension complexifie l'apprentissage.
4. Maitriser les techniques de sélection de variables (méthodes best subset, pas à pas), réduction de dimension (régression sur CP, PLS) ou de régularisation (Ridge, Lasso, Elastic Net) et les utiliser pour surmonter ces enjeux.
5. Évaluer la performance d'un modèle prédictif
    1. Comparer les modèles de complexité différent à l'aide de critères adaptés (AIC, BIC) et de mesure d'erreur prédictive issues de techniques de validation (séparation entraînement/validation/test, validation croisée)
    2. Évaluer la complexité du modèle
    3. Détecter le surapprentissage du modèle
6. Optimiser un modèle à l'aide des hyperparamètres
    1. Comprendre le rôle des hyperparamètres dans les différentes méthodes

    2. Mettre en œuvre des stratégies de recherche pour identifier les configurations optimales.
7. Comprendre la sélection de modèle comme une démarche globale visant à identifier, parmi plusieurs familles de modèles, celle qui offre le meilleur équilibre entre performance prédictive, complexité et interprétabilité.
8. Travailler en équipe sur un projet appliqué
    1. Concevoir et réaliser une analyse complète sur un jeu de données, depuis la phase exploratoire jusqu'à l'évaluation de la capacité prédictive du modèle choisi par suite d'une procédure d'apprentissage.
    2. Collaborer à la rédaction d'un rapport et à une présentation claire des résultats et des choix méthodologiques.

Modalités d'évaluation :

Les apprentis travailleront en groupe sur un projet appliqué, consistant à modéliser un jeu de données. Le projet donnera lieu à un rapport écrit et à une présentation orale. Le rapport devra être composé des sections suivantes :

- Présentation du jeu de données
- Description des données à l'aide d'analyses statistiques ; justification de pré traitements éventuels (re-calcul de données aberrantes, traitement de valeurs manquantes...)
- Objectif de l'étude en mettant en évidence la démarche choisie
- Présentation, interprétation des résultats de l'analyse et conclusions
- Source des données et références consultées

L'évaluation portera sur la pertinence des choix méthodologiques, la qualité de l'interprétation des résultats, la rigueur de l'évaluation du modèle, ainsi que sur la capacité à communiquer les résultats de manière claire et structurée.

Modalités de validation

- Contrôle continu
- Projet(s)
- Mémoire
- Examen final