

1EL2500 - Théorie de l'information

Responsables : **Richard COMBES**

Département de rattachement : **DÉPARTEMENT SIGNAL, INFORMATION, COMMUNICATION**

Langues d'enseignement : **ANGLAIS**

Campus où le cours est proposé : **CAMPUS DE PARIS - SACLAY**

Nombre d'heures d'études élèves (HEE) : **60**

Nombre d'heures présentielles d'enseignement (HPE) : **30**

Année académique : **2024-2025**

Niveau avancé : **non**

Présentation, objectifs généraux du cours :

- La théorie de l'information fournit un ensemble de concepts puissants permettant de comprendre et de résoudre des problèmes fondamentaux de notre monde moderne tels que les grands réseaux de communication, le stockage des grandes masses de données, la transmission fiable et efficiente d'information, la sécurité des données, la biologie, l'apprentissage automatique. La théorie de l'information, fruit des premiers travaux de scientifiques tels que Hartley, Shannon, Wiener et Kolmogorov, est à présent un domaine pluridisciplinaire au carrefour des probabilités, des statistiques, de la physique et de l'ingénierie. Elle a permis le développement des sciences et des technologies de l'information et de la communication qui ont été sources d'innovation et de transformations de nos modes de vie, de notre économie, et de nombreux secteurs industriels. Les enjeux actuels sont considérables pour de nombreux secteurs (spatial, énergie, santé, mobilité, services, industrie du futur...).
- Le but de ce cours est de présenter les outils fondamentaux de la théorie de l'information tels que l'entropie et l'information mutuelle, et de montrer comment ces outils s'appliquent à des problèmes pratiques : représentation et codage d'informations, compression de données, communication, correction d'erreur, cryptanalyse, inférence. Pour toutes ces grandes familles de problèmes applicatifs, le cours montrera comment des solutions élégantes et optimales peuvent être proposées en se basant sur les outils de la théorie de l'information. Les liens entre les concepts développés durant le cours et les ponts naturels avec d'autres cours sont nombreux: probabilités, statistiques et apprentissage, traitement du signal, communications numériques, réseaux et sécurité.

Période(s) du cours (n° de séquence ou hors séquence) :

SG3

Prérequis :

- Probabilités

Plan détaillé du cours (contenu) :

Les cours magistraux permettent d'inculquer les concepts fondamentaux de la théorie de l'information et de découvrir la diversité des applications. Les séances des TD/TP proposent des

résolutions concrètes de problématiques d'ingénieur pour modéliser, analyser et concevoir des architectures de traitement pour des champs applicatifs variés : cryptographie, compression d'image, communication, apprentissage, finance.

1 - Introduction à la théorie de l'information

- entropie et thermodynamique
- mesure de l'information d'un événement, information conditionnelle, entropie
- unités de mesure de l'information (bit)
- mesure de l'entropie du langage naturel
- exemples d'applications

2 Mesure de l'information

Outils mathématiques fondamentaux de la théorie de l'information

- entropie, entropie conditionnelle, entropie relative, information mutuelle

Propriétés et interprétations

- positivité de l'entropie, entropie maximale
- relations entre les différentes mesures d'information : règles de chaînes
- inégalités : log-somme, Fano-Shannon, systèmes traitement de données
- limites fondamentales de l'équipartition asymptotique et de la typicité
- conséquences applicatives

3 - Jeux et Information (exercices et travaux pratiques – 3h)

- expériences de Shannon
- application de l'entropie au jeux de type "20 questions"
- apprentissage de stratégies optimales à partir de données

4 - Représentation des données : limites fondamentales

- représentation binaire des données
- codage d'une suite de données dans un alphabet de code discret M-aire
- construction de codes à longueur variable à décodage sans ambiguïté
- inégalité de Kraft pour les codes à préfixe
- conséquence de l'inégalité de Kraft : limite fondamentale pour minimiser la longueur moyenne d'un message codé (entropie)

5 - Compression de données

Algorithmes de compression d'une suite de données discrètes sans perte

- algorithme de codage d'Huffman
- codage de sources de Markov
- compression universelle de données

Algorithmes de compression avec perte

- quantification vectorielle versus quantification scalaire
- théorie de la distorsion de taux
- fonction de distorsion de taux pour des vecteurs gaussiens

6 - Compression d'images (exercices et travaux pratiques – 3h)

- mesure du taux de compression d'une image
- techniques de compression d'images avec ou sans perte
- conception d'un algorithme de compression sur des images tests

7 - Information mutuelle et communication

Systèmes de traitement de données

- exemples de problématiques: stockage, exploration spatiale, IoT/WiFi/5G etc.
- information mutuelle sortie-entrée d'un système

Systèmes à entrées-sorties discrètes

- messages d'information erronés : fautes de frappe, données effacées, erreurs de transmission etc.
- information mutuelle sortie-entrée maximale d'un canal (capacité)
- capacité du canal binaire à erreurs symétriques indépendantes
- schéma de réalisation d'un codage optimal pour atteindre la capacité (rendement)
- exemples de mise en œuvre pratique (codage correcteur d'erreur)

Systèmes avec bruit additif gaussien

- calcul de l'information mutuelle maximale de canaux à entrées-sorties continues (capacité)
- canaux à bruit additif gaussien à temps discret
- canaux gaussiens parallèles et algorithme de « waterfilling »
- limites fondamentales du théorème de Shannon pour une transmission sans erreur
- compromis : débit d'information / bande passante / rapport signal à bruit
- exemples de mise en œuvre pratique (modulation et codage)

- contraintes et métriques de performance (complexité, fiabilité, secret, latence, énergie, efficacité spectrale...)

8 - Communication audio (exercices et travaux pratiques - 4h30)

- information mutuelle et conséquences applicatives
- conception et expérimentation d'une communication numérique audio
- évaluation des performances

9 - Économie et finance : théorie du portefeuille

- marchés boursiers et portefeuilles
- optimalité des portefeuilles log-optimaux
- portefeuilles universels

10 - Apprentissage automatique et statistiques : inférence

- théorie de l'information & inférence
- f-divergences et distances statistiques
- taux minimax pour les problèmes d'estimation

11 - Investissement en bourse (exercice et session pratique de simulation – 3h)

- théorie de l'information & stratégie de portefeuille
- optimisation d'une stratégie d'investissement avec de l'argent fictif

Déroulement, organisation du cours :

Le cours est proposé une seule fois (SG3). Le cours est programmé en 19 créneaux de 1h30 et un créneau d'examen de 2h.

- 10 créneaux de cours (15h00)

- 9 créneaux de TD/TP (13h30) organisées en 3 séances de 3h TD/TP avec exercices et mise en œuvre pratique (9h) et 1 séance TD de 1h30 avec évaluation de type QCM (1/2h)

Organisation de l'évaluation :

La note finale est calculée par la formule suivante :

$$\text{MAXIMUM}(A, 80\% \times A + 20\% \times B)$$

ou A désigne la note de l'examen écrit final (tous documents autorisés) et B la note de l'évaluation intermédiaire (contrôle continu CC) de type QCM (sans documents).

Moyens :

- - Equipe pédagogique (cours magistraux): Richard Combes
- - Taille des groupes de TD/TP : 25 élèves par groupe

Acquis d'apprentissage visés dans le cours :

- A l'issue de ce cours les élèves seront capables :
- - de modéliser des systèmes porteurs d'information
- - de manipuler les mesures d'informations et de comprendre leur sens physique
- - de comprendre les limites fondamentales pour la transmission de données, le stockage, la sécurité, le traitement de données
- - de développer des algorithmes pour des problèmes pratiques et de comprendre les contraintes applicatives et évaluer leur performance
- - de concevoir des architectures de traitement

Description des compétences acquises à l'issue du cours :

- The benchmark skills assessed are skills C1 (C1.1 C1.2 C1.4) and C2 (C2.1) . The intermediate exam is used to assess milestone 1 of skill C1. The final exam assesses milestone 1 of skills C1 and C2. Skills C1 and C2 are validated if the final mark is greater or equal to 10.

Bibliographie :

- Notes de cours fournies aux élèves
- Thomas M. Cover and Joy A. Thomas, Elements of Information Theory
- David J.C. MacKay, Information Theory, Inference, and Learning Algorithms
- Imre Csiszar, Information Theory and Statistics: A Tutorial
- Alexandre B. Tsybakov, Introduction to Nonparametric Estimation