

IA et science des données

Cours 11 – mardi 8 avril 2025

Fin sur les ensembles. IA.

Christophe Marsala

Sorbonne Université

LU3IN026 - 2024-2025

Plan du cours

Le projet

Retour au supervisé : méthodes d'ensembles

L'intelligence artificielle

1 – Le projet –

Objectif du projet : analyser un jeu de données

- Résultat attendu
 - définition d'un certain nombre de problématiques et leur résolution
 - au moins un problème d'apprentissage supervisé et un autre non-supervisé
- Compte-rendu
 - un unique notebook complété par un package avec vos fonctions
 - un poster (électronique) expliquant de façon synthétique en une page les différentes expériences réalisées et leurs résultats
- Calendrier
 - séances 11 : prévue pour travailler sur le projet
 - lundi 19 mai, 11h : remise du compte-rendu sur Moodle
 - mardi 20 mai, 9h-13h : soutenances (10mn par binôme)

Marsala – 2025

LU3IN026 – cours 11 – 3

Plan du cours

Le projet

Retour au supervisé : méthodes d'ensembles

L'intelligence artificielle

1 – Le projet –

Les données 20-newsgroups

- Ensemble de textes
 - env. 18000 documents, répartis en 20 classes ("target") qui représentent des groupes de discussion
- Que peut-on apprendre de ces données ?
 - soyez imaginatifs !
- **IMPORTANT** :
 - citez vos sources éventuelles !
 - si vous utilisez un travail externe, ou même ChatGPT (ou autre), etc.
 - **dites-le !**
 - tout manquement à cette règle → note 0
 - tout plagiat sera sanctionné par un 0

Marsala – 2025

LU3IN026 – cours 11 – 4

2 – Retour au supervisé : méthodes d'ensembles –

Les forêts aléatoires (random forest)

- Idée : plus les arbres sont **diversifiés**, meilleur sera le score global
 - augmenter la diversité : plus d'aléatoire !
 - → **random forest**
- Soit \mathbf{X} une base d'apprentissage avec n exemples
- Soit B le nombre de classifieurs souhaités, $m < n$ le nombre d'exemples à choisir et $p \leq d$ variables de description à choisir
 1. Extraire B sous-bases de \mathbf{X} : $\mathbf{X}_1, \dots, \mathbf{X}_B$
 2. Construire un classifieur f_k pour chaque sous-base \mathbf{X}_k
- Remarque : B , m et p sont des **hyper-paramètres** de l'algorithme

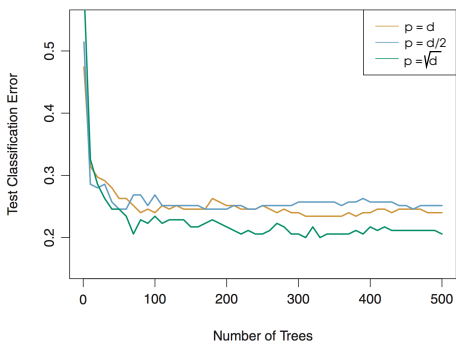
Marsala – 2025

LU3IN026 – cours 11 – 6

Evaluation d'ensembles

- Pour évaluer un ensemble construit par Bagging / random forest
- Validation croisée
 - très coûteuse pour évaluer un ensemble
 - il faut construire B classifieurs à chaque fois !
- Evaluation **Out Of the Bag** (OOB)
 - adaptée aux ensembles et suffisante pour les évaluer

Performances random forest : choix du nombre d'attributs



(source : "An introduction to statistical learning", Gareth et al.)

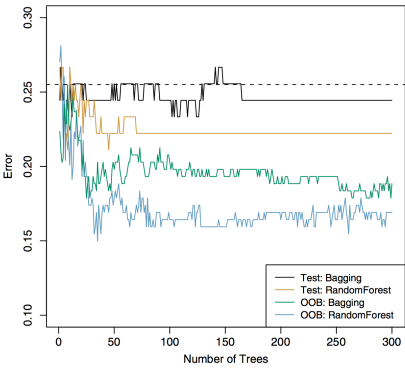
Plan du cours

Le projet

Retour au supervisé : méthodes d'ensembles

L'intelligence artificielle
une définition
un peu d'histoire...

Performances bagging vs random forest



(source : "An introduction to statistical learning", Gareth et al.)

Ensembles de classifieurs

- Approches de construction d'ensembles
 - bagging : bootstrap aggregating
 - random forests
- Evaluation :
 - validation croisée
 - approche Out Of Bag

Qu'est-ce que l'Intelligence artificielle ?



Turing



Shannon



Simon / Newell



McCarthy



Minsky

Une définition de l'IA

Science dont le but est de faire faire par la machine des activités que l'homme accomplit en utilisant son intelligence.

- Étude des activités intellectuelles pour lesquelles aucune méthode n'est, à priori, connue. (Laurière, 1986)
- \implies résoudre des problèmes pour lesquels il n'y a pas d'algorithme connu qui les résolve en temps raisonnable

Qu'est-ce que l'Intelligence artificielle ?

- ▶ Approches de résolution : informatique heuristique
 - choix guidés par des **heuristiques**
 - représentation et simplification de l'espace des solutions
- ▶ Quelques exemples
 - reconnaissance de forme
 - résumé de texte, traduction, reconnaissance de style d'auteur
 - prise de décision (diagnostic médical, résolution de pannes,...)
 - modélisation de systèmes complexes (systèmes biomimétiques,...)
 - jeux : échecs, go, gestion de PNJ, moteur de jeu (IA)

Quelques dates (1)

- ▶ 1949 : Claude Shannon propose un algorithme pour le jeu d'échecs
- ▶ 1950 : Alan Turing : "Computing machinery and intelligence"
 - le jeu de l'imitation (test de Turing)
- ▶ **1956** : **Conférence de Dartmouth College**
 - Marvin Minsky, John McCarthy, Herbert Simon,...
 - "Logic theorist" : démonstrateur de théorèmes
 - but : simuler le raisonnement humain pour résoudre des problèmes
 - approche multidisciplinaire

Quelques dates (2)

- ▶ ...
- ▶ 1996 : Deep Blue (IBM) bat Garri Kasparov (jeu d'échecs)
- ▶ 2005 : Univ. de Stanford (Sebastian Thrun) : voiture sans pilote
- ▶ 2011 : victoire de l'IA Watson (IBM) au jeu Jeopardy
- ▶ 2010 (env.) : émergence des modèles profonds (deep learning)
- ▶ 2016 : Alpha Go (Google Deepmind) bat Lee Sedol (4-1) au Go
- ▶ 2020 : Generative Pre-trained Transformer (GPT) d'OpenAI