

Programme de colle - Semaine 15 (12 février)

La démonstration des énoncés marqués d'une étoile est exigible

1 Apprentissage supervisé

- Principe de l'apprentissage supervisé
- Arbres binaires de décision : définition et écriture de programmes simples sur les arbres de décision.
- Éléments de théorie de l'information : définition de l'entropie, cas du pile ou face traité en cours, calcul de l'entropie dans le cas d'un ensemble d'exemples d'apprentissage, calcul du gain d'information pour un attribut binaire uniquement **Savoir étudier $H(X)$ pour X suivant une loi de Bernoulli de paramètre p en fonction de p (*)**.
- Algorithme ID3 : uniquement dans le cas d'apprentissage d'arbres binaires de décision.
- Algorithme du plus proche voisin. Algorithme des k plus proches voisins. Il faut savoir calculer les k plus proches voisins d'un ensemble de n points en $O(nk)$ c'est-à-dire sans trier la liste des points.
- Arbres k -dimensionnels : définition et exemples. Représentation du partitionnement de l'espace obtenu à l'aide d'un k -arbre dimensionnel. Insertion dans un arbre k -dimensionnel. Algorithme de recherche du plus proche voisin à l'aide d'un arbre k -dimensionnel.
- Révisions : le cas de la dimension 1 est correspond à celui de l'arbre binaire de recherche vu en MP2I. On pourra poser toute question de révision sur ce point (recherche, insertion, suppression).

2 Complexité et décidabilité

- Notion de problème de décision.
- Problèmes décidables. La notion de machine de Turing est hors programme, une machine est un algorithme ou un programme écrit en C ou en OCaml s'exécutant sur une machine à mémoire infinie.
- Classe **P**.
- Principe de réduction d'un problème A à un problème B noté $A \leq B$.
- Principe de réduction polynomiale d'un problème A à un problème B noté $A \leq_P B$.
- Si $B \in \mathbf{P}$ et $A \leq_P B$ alors $A \in \mathbf{P}$ (*)
- Classe **NP** : définie comme la classe des problèmes admettant dont les instances positives admettent des certificats qui peuvent être vérifiés en temps polynomial. Notion de certificat pour un problème de décision.
- $\mathbf{P} \subset \mathbf{NP}$ (*)
- Problèmes **NP-complets** : définis comme les problèmes de **NP** les plus difficiles.

- Si A est **NP-complet**, $B \in \mathbf{NP}$, et $A \leq_p B$ alors B est NP-complet (*)
- Théorème de Cook-Levin.
- Exemples de problèmes NP-complets vus en cours : CNF-SAT, 3-SAT (réduction faite en cours), CLIQUE (réduction faite en cours), INDEPENDANT, SUBSET-SUM
- Notion de machine universelle : les machines M sont représentables sous forme d'une donnée $\langle M \rangle$ et une machine dont l'entrée est $\langle M \rangle$ peut simuler le comportement de la machine M sur toute entrée.
- Problèmes indécidables. Problème de l'arrêt HALT prenant en entrée le code d'une machine $\langle M \rangle$ une entrée I pour M et décidant si la machine M termine sur l'entrée I . **HALT est indécidable (*)**.