Prérequis mathématiques, statistiques et informatique pour le MS Big Data

Avril 2022

Voici quelques points très importants à voir avant la rentrée (surtout les documents à télécharger)

1 Formation aux fondamentaux de l'informatique

- Structures de données, algorithmique et programmation (par exemple Java, Python ou C)

  Références à titre indicatif:

  - Cormen T., Leiserson C., Rovest R., Stein C. *Introduction to Algorithms (3rd ed.)*. MIT Press and McGraw-Hill; 2009.
  - Eckel B. *Thinking in Java (4th ed.)*. Prentice Hall; 2006.
  - Tutoriel Python: `https://docs.python.org/3/tutorial/index.html`

- Fondements des systèmes d'exploitation

  Références à titre indicatif:

  - Tanenbaum A., Bos H. *Modern Operating Systems 4th Edition*. Pearson; 2014. L'ouvrage est trop complet mais son survol permet d'avoir de bonnes notions du fonctionnement des systèmes d'exploitation.

- Principes de systèmes de gestion de bases de données (représentation des données, stockage, interrogation...)

  Références à titre indicatif:

  - Ullman J.D., Widom J. *A First Course in Database Systems (3rd ed.)*. Pearson education; 2007.
  - Hainaut J-L. *Bases de données - Concepts, utilisation et développement (4e édition)*. Dunod; 2018.

- Technologies web

  Références à titre indicatif:

  - Haverbeke M. *Eloquent JavaScript, A Modern Introduction to Programming (3rd Edition)*. No Starch Press; 2018.
    Disponible gratuitement en ligne: `https://eloquentjavascript.net/`

- Notions en compilation

Références à titre indicatif:

– Aho A., Lam M., Sethi R., Ullman J. *Compilers: Principles, Techniques, and Tools (2nd edition)*. Pearson; 2006. (trouvable en ligne)
À nouveau, l'ouvrage est très (trop) complet mais son survol permet d'avoir de bonnes notions du fonctionnement de la compilation.

2 Mathématiques

2.1 Algèbre linéaire

- Espace euclidiens, produit scalaire et formes hermitiennes
- Calcul matriciel, matrices semi-définies positives, diagonalisation de matrices

Références à titre indicatif:

- `https://www.edx.org/course/algebre-lineaire-partie-1`
- `https://www.edx.org/course/algebre-lineaire-partie-2`
- `https://www.edx.org/course/algebre-lineaire-partie-3`
- Cairoli R. *Algèbre linéaire (2nde édition)*. Presses polytechniques et universitaires romandes; 2004.
- Fabien Margairaz. Algèbre linéaire I & II: Notes de cours de l'EPFL.
`https://docplayer.fr/23918385-Algebre-lineaire-i-ii.html`
- Les notions de base d'algèbre sont aussi très bien décrites sur Wikipédia

2.2 Probabilités

- Espérance, variance
- Probabilités jointes et conditionnelles, formule de Bayes
- Lois usuelles (loi de Bernoulli, loi uniforme, loi normale)
- Estimation des paramètres d'une loi par maximisation de la vraisemblance

Références à titre indicatif:

- `https://fr.coursera.org/learn/probabilites-1`
- `https://fr.coursera.org/learn/probabilites-2`
- Ross SM. *A first course in probability*. Pearson; 2019.
(PDF trouvable en ligne, utile notamment pour le maximum de vraisemblance, les probabilités jointes/conditionnelles et les lois de probabilité)
- Olivier François. Notes de cours de Probabilités Appliquées. Les 40 premières pages. `http://membres-timc.imag.fr/Olivier.Francois/Poly_Cours_Proba.pdf`

2.3 Statistiques

- Statistiques descriptives : Population statistique, Estimateurs de tendance centrale et de dispersion, Représentations usuelles (histogramme, diagramme en bâtons, etc.)
- Notions élémentaires de test d'hypothèse : Echantillons, Hypothèse nulle, hypothèse alternative, risques de type I et II, Test de Student

Références à titre indicatif:

- Olivier Gaudoin. Principes et Méthodes Statistiques : Notes de cours, Ensimag 2A. Chapitres I,II, V (3 premières sections)
  `https://www-ljk.imag.fr/membres/Olivier.Gaudoin/PMS.pdf`