

**Heure et salle :**

- Heure : Mercredi 8h30
- Salle : 9.02

**Enseignants :** Nicolas Verzelen [nicolas.verzelen@inra.fr](mailto:nicolas.verzelen@inra.fr) et Joseph Salmon [joseph.salmon@umontpellier.fr](mailto:joseph.salmon@umontpellier.fr)

**Horaires de consultation (*office hours*) :**

- Mercredi 17h-19h, sur rendez-vous par mail uniquement.

**Prérequis**

Les étudiants sont supposés connaître les bases de la théorie des probabilités, du calcul différentiel, d'optimisation convexe et d'algèbre linéaire.

**Probabilités :** Pour le démarrage en probabilité les ouvrages suivant sont en français [CGDM01], [Ouv08], [Ouv07], [GS01]. Une introduction aux martingales et aux espérances conditionnelles est bien détaillée dans [Wil91] (avec une preuve de la loi des grands nombres suffisamment rigoureuse). Pour aller plus loin, en particulier sur la concentration de la mesure on pourra consulter [BLM13]

**Algèbre Linéaire :** Il faut être suffisamment à l'aise avec le calcul matriciel pour bien commencer avec les modèles linéaires et aller jusqu'aux modèles économétriques : [Sch05] et [Gv13] peuvent être un bon début.

Les classiques du genre sont [Gv13, HJ94] et [TB97] est aussi un bon point d'entrée. Pour aborder les chaînes de Markov il est bon de connaître le théorème de Perron-Frobenius et ses nombreuses conséquences [Sen06, LN12]. Pour des questions plus avancées, par exemple de majorization [MOA11] (utile pour les notions de répartitions de richesses, etc.) et les matrices bistochastiques [Bha97].

**Calcul différentiel, optimisation et convexité** Pour un bon départ en calcul de gradient et autres outils de calcul différentiel [Rou09]. Pour l'optimisation et la convexité un bon départ est [BV04], et pour les concepts plus avancés [HUL93a, HUL93b], pour les bornes inférieures et supérieures pour les algorithmes de premier ordre [Nes04], et enfin des classiques [RW98, Roc97, BL06]. Pour la partie algorithmique de l'optimisation on consultera plutôt [NW06].

**Algorithmique** Les étudiants doivent aussi être capables d'implémenter des méthodes numériques basiques en utilisant un langage de haut niveau (e.g. Python, Matlab, R/S-Plus). Pour commencer en Python, un fichier d'introduction est disponible ici : [http://josephsalmon.eu/index.php?page=teaching\\_18\\_19&lang=fr](http://josephsalmon.eu/index.php?page=teaching_18_19&lang=fr). Enfin une référence contenant la plupart des algorithmes usuels en informatique est le livre [CLRS01].

**Description du cours**

Ce cours traite du cadre de l'apprentissage automatique sous un angle statistique. Nous nous intéresserons principalement au cadre supervisé (régression et classification) et introduiront quelques éléments du cadre non-supervisé à travers les méthodes de clustering. Au

delà des aspects de modélisations et de théorie, le cours couvrira aussi quelques éléments d'optimisation et d'implémentation des méthodes introduites.

### Notation

**TP Noté :** Ces TPs seront à charger sur le Moodle du cours.

- 10% de la note finale pour le TP1 (3/10/18) et à rendre pour le 5/09/2018, 23h59.
- 10% de la note finale pour le TP2 (17/10/18) et à rendre pour le 19/10/18, 23h59.

**Projet :** 30% de la note finale, énoncé donné en séance du 17/10/18, et dont le rendu est prévu pour le dimanche 04/11/18, 23h59.

**Examen :** 50% de la note finale, 6/11/2018.

### Thèmes des séances :

- Séance 1 : Introduction à l'apprentissage supervisé ; modèles linéaires.
- Séance 2 : Validation croisée, régression logistique, analyse discriminante.
- Séance 3 : Sélection de modèle et méthodes de régularisation.
- Séance 4 : Régression Splines et GAM ; Arbres de décision.
- Séance 5 : Forêts aléatoires et Boosting.
- Séance 6 : SVM.
- Séance 7 : Apprentissage non-supervisé.
- Séance 8 : Examen

Pour la partie numérique, il est demandé de suivre les vidéos de Jake VanderPlas : <http://jakevdp.github.io/blog/2017/03/03/reproducible-data-analysis-in-jupyter/>, au rythme d'une part semaine. Les vidéos 1 ; 2 ; 7 ; 8 ; 9 et 10 sont obligatoires, les autres pouvant être consulter par curiosité.

### Livres pour l'apprentissage statistique

- **An introduction to statistical learning (With applications in R)**, G. James and D. Witten and T. J. Hastie and R. Tibshirani Springer, New York, 2013 ; [JWHT13]
- **Understanding machine learning : From theory to algorithms**, S. Shalev-Shwartz and S. Ben-David, Cambridge University Press, 2014 ; [SB14]
- **A probabilistic theory of pattern recognition**, L. Devroye and L. Györfi and G. Lugosi, Springer-Verlag, New York, 1996 ; [DGL96]

### Page du cours :

[http://josephsalmon.eu/index.php?page=teaching\\_18\\_19&lang=fr](http://josephsalmon.eu/index.php?page=teaching_18_19&lang=fr)

## Références

- [Bha97] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.
- [BL06] J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization. Theory and examples*. Springer, New York, 2nd ed. edition, 2006. Theory and examples.
- [BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities : A Nonasymptotic Theory of Independence*. Oxford University Press, second edition, 2013.
- [BV04] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
- [CGDM01] M. Cottrell, V. Genon-Catalot, C. Duhamel, and T. Meyre. *Exercices de probabilités, Licence - Master - Écoles d'ingénieur*. Cassini, 3<sup>e</sup> edition, 2001.
- [CLRS01] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 2001.
- [DGL96] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*, volume 31 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1996.
- [GS01] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. Oxford University Press, New York, third edition, 2001.
- [Gv13] G. H. Golub and C. F. van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.
- [HJ94] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.
- [HUL93a] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. I*, volume 305. Springer-Verlag, Berlin, 1993.
- [HUL93b] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. II*, volume 306. Springer-Verlag, Berlin, 1993.
- [JWHT13] G. James, D. Witten, T. J. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 103 of *Springer Texts in Statistics*. Springer, New York, 2013. With applications in R.
- [LN12] B. Lemmens and R. Nussbaum. *Nonlinear Perron-Frobenius theory*, volume 189 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012.
- [MOA11] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities : theory of majorization and its applications*. Springer Series in Statistics. Springer, New York, second edition, 2011.
- [Nes04] Y. Nesterov. *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004.

- 
- [NW06] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.
- [Ouv07] J-Y. Ouvrard. *Probabilités : Tome 2, Licence - CAPES*. Enseignement des mathématiques. Cassini, 2 edition, 2007.
- [Ouv08] J-Y. Ouvrard. *Probabilités : Tome 1, Licence - CAPES*. Enseignement des mathématiques. Cassini, 2 edition, 2008.
- [Roc97] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997.
- [Rou09] F. Rouvière. *Petit guide de calcul différentiel : à l'usage de la licence et de l'agrégation*. Enseignement des mathématiques. Cassini, 2009.
- [RW98] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.
- [SB14] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning : From theory to algorithms*. Cambridge University Press, 2014.
- [Sch05] J. R. Schott. *Matrix analysis for statistics*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2005.
- [Sen06] E. Seneta. *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York, 2006.
- [TB97] L. N. Trefethen and D. III Bau. *Numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
- [Wil91] D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.