



Yoann Offret

## Projets MIGS2 2016

**Sujet 1** *"Recherche de zones homogènes dans l'ADN : chaînes de Markov cachées et algorithme EM."*

L'objectif de ce projet est de présenter une modélisation probabiliste et un algorithme stochastique permettant de déterminer les parties codantes d'un brin d'ADN – séquence de paires de nucléotides Adénine, Cytosine, Guanine et Thymine. Pour cela on a envie de représenter une séquence d'ADN comme une réalisation partielle d'une chaîne de Markov (CM) à valeur dans l'ensemble des nucléotides. Cependant, selon que la partie de l'ADN soit codante ou non, on constate que la fréquence des nucléotides n'est pas la même. On modélise donc un brin d'ADN comme une **CM cachée** : la séquence formée par le brin d'ADN et l'information indiquant si on se trouve dans une partie codante ou non est alors markovienne. Toutefois de nombreux paramètres du modèle restent inconnus. Pour les retrouver à partir d'un échantillon expérimental on peut utiliser l'**algorithme EM** (Espérance-Maximisation). Cette procédure a été introduit dans [DLR77] et a de nombreuses applications aux méthodes de mélange, aux données censurées, *e.t.c.*. La principale référence sera [DJ06, Chap. 5].

**Sujet 2** *"Réseaux de neurones artificiels : le modèle de Hopfield"*

Les **réseaux de neurones** artificiels décrivent une familles d'algorithmes historiquement inspirés des réseaux de neurones humains, mais qui se sont ensuite éloignées des modèles du cerveau et rapprochées des méthodes statistiques. Parmi leurs applications on peut citer la reconnaissance de caractères ou encore l'optimisation. Un des modèles de référence, le modèle de Hopfield introduit dans [Hop82], traite de la mémoire associative. Ce réseau consiste en un ensemble de neurones interconnectés les uns les autres et dans différents états. Ceux-ci se mettent à jour au cours du temps – de manière synchrone ou asynchrone – suivant une certaine interaction. Le but de ce projet est de comprendre le fonctionnement de ce modèle et d'en illustrer quelques applications. Les principales références sont [FB96, Chap. XII.], [Roj96, Chap. 13.] et [WGP14, Chap. 17].

## Références

- [DJ06] Jean-François Delmas and Benjamin Jourdain. *Modèles aléatoires*, volume 57 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag,

- Berlin, 2006. Applications aux sciences de l'ingénieur et du vivant. [Applications to engineering and the life sciences].
- [DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1) :1–38, 1977. With discussion.
- [FB96] Michel Verleysen François Blayo. *Les réseaux de neurones artificiels*. Que sais-je? Presses Universitaires de France - PUF, 1996.
- [Hop82] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci. U.S.A.*, 79(8) :2554–2558, 1982.
- [Roj96] Raül Rojas. *Neural Networks : A Systematic Introduction*. Springer-Verlag Berlin Heidelberg, 1 edition, 1996.
- [WGP14] Richard Naud Wulfram Gerstner, Werner M. Kistler and Liam Paninski. *Neuronal Dynamics : From Single Neurons to Networks and Models of Cognition*. Cambridge University Press, 2014.