

Journée SciDoLySE 2

ENS Lyon

Anne-Laure, Aurélien, Paulo, Julien, Franck, Céline, Marc

Mardi 7 mai 2019

- 77 inscrits (dont certains tardifs...)
- 65 plateaux-repas à midi
- Prochaine réunion : date / lieu ?
 - Tutorial: optimal transport?
 - Créneau récurrent l'an prochain ?
- Forum <https://framateam.org/scidolyse/>
- Site web: <http://scidolyse.ens-lyon.fr>
 - Formations <http://scidolyse.ens-lyon.fr/training>

Programme

9h00-9h30 : accueil, présentation de la journée

9h30-10h30 : résultat récent: **Romain Azais** – De l'importance de la fonction de poids dans le noyau des sous-arbres

11h00-11h30 : **Débriefing de conf**: AISTAT (levgen) ALT (Aurélien)

11h30-12h30 : **Tutorial**: Generative adversarial network (Mathieu Serrurier)

12h30-13h30 : Pause déjeuner

13h30-15h00 : **Expression des besoins**: Groupe d'Analyse et de Théorie Economique

15h00-16h00 : **Tutorial**: Generative adversarial network (Liming Chen)

16h15-17h00 : **Groupes de lecture**

- Mathématiques du deep learning
- Interprétabilité des méthodes neuronales
- Quantum ML
- Apprentissage et jeux
- Hardware + neuromorphic engineering

De l'importance de la fonction de poids dans le noyau des sous-arbres

En classification supervisée, on souhaite construire une règle de prédiction d'une variable qualitative Y en fonction d'un prédicteur X (généralement à valeurs dans \mathbb{R}^d) à partir d'un ensemble d'exemples $(X_i, Y_i)_{1 \leq i \leq n}$. On suppose ici que X_i est un arbre, ordonné ou non, avec ou sans étiquette sur ses noeuds. Ce cadre est un peu particulier puisque l'espace d'état des prédicteurs n'est pas muni d'un produit scalaire canonique. Dans ce contexte, les méthodes à noyau permettent de transporter les données dans un espace de Hilbert rendant le problème plus facile à traiter. Se pose alors la question de la construction d'un bon noyau.

Le noyau des sous-arbres - introduit par Vishwanathan et Smola (NIPS '02) - a deux bonnes propriétés : une combinatoire réduite le rendant facile à calculer et une grande efficacité en pratique. Dans le cadre d'un modèle probabiliste, nous obtenons des garanties théoriques sur ce noyau et en déduisons une propriété importante de la fonction de poids. Cela nous amène à définir une nouvelle fonction de poids, apprise sur les données, qui permet dans des cas difficiles de déceler une information qui semblait indisponible. Nous illustrons ces résultats par une application à la prédiction de la langue d'articles Wikipedia à partir de leur structure. (Travail avec Florian Ingels)

Pierre-Philippe Combes:

Machine-learning coding of land use in France since the mid-XVIIIth c. from historical maps.

Sophie Hatte: "Twitter Revolution"

(with E. Madinier and E. Zhuravskaya) We examine how Twitter affects the likelihood and extent to which US TV channels report information. To this aim, we use the Israeli-Palestinian conflict as an ideal framework and study whether and how Twitter has affected the coverage of the conflict by US TV channels. Exploiting machine learning algorithms, we characterize the type of information released on Twitter. This procedure allows us to investigate potential heterogenous effects of the tweets on the TV coverage of the conflict.

Mathieu Couttenier: The impact of small scale mines on economic development in Africa.