Stage : Modèles contrefactuels et déformations géodésiques.

L'utilisation massive de stratégies de prises de décisions automatiques de type 'Machine Learning' a ouvert de nouveaux problèmes éthiques dans la société [1]. Par exemple, il a été montré que l'algorithme COMPASS, utilisé aux Etats-Unis pour prédire la tendance à récidiver d'individus accusés de crime, avais tendance à être discriminant envers certaines populations[2].

Une approche pour caractériser l'impartialité d'un algorithme de prise de décision automatique est basée sur la notion d'appariement (mapping) entre les observations issues de deux sous-populations, une standard (S=1) et une potentiellement discriminée (S=0). Une décision peut par exemple être négative pour un individu du groupe S=0 mais positive pour l'individu qui lui est apparié le groupe S=1. L'appariement est dit contre-factuel si ce phénomène pénalise principalement un groupe plutôt que l'autre. Nous allons étudier ce type de modèles dans le cadre de ce stage.

La définition du mapping est alors essentielle ici et elle se fait classiquement à l'aide de modèles de transport optimal [3] qui calculent un appariement optimal de distributions de probabilités. Après une phase de prise en main de ces outils le ou la stagiaire étendra ces travaux en utilisant les méthodes d'appariement de nuages de points à l'aide de modèles de grandes déformations difféomorphique. Ces méthodes calculent une géodésique entre les nuages de points dans un espace Riemmanien [4n, 5n] et n'utilisent pas de distributions de probabilités de dimension p>1 calculées à partir des nuages de points. Un aspect clé de ces méthodes est qu'elles rendent la déformation optimale plus interprétable de par le choix explicite de noyaux de convolution, ce qui est un plus dans de nombreuses applications.

Le stage se déroulera à [IMT ou ANITI/B612] sous la direction de Jean-Michel Loubes et Laurent Risser. Le ou la candidat devra avoir la volonté forte d'utiliser de manière pertinente des outils avancés de mathématiques appliqués sur des cas d'applications concret. Il/Elle devra avoir un goût pour le monde de la recherche, un gout pour la géométrie serait un plus. Les développements se feront sous Python et une experience préalable en sciences des données sera appréciée.

[1] P. Besse, C. Castets-Renard, A. Garivier, and J.-M. Loubes. Can Everyday AI be Ethical? working paper or preprint, Oct. 2018.
[2] A. Chouldechova. Fair prediction with disparate impact : A study of bias in recidivism prediction instruments. Big Data, 5(2) :153–163, Jun 2017.
[3] E. Black, S. Yeom, and M. Fredrikson. Fliptest : Fairness auditing via optimal transport, 2019.
[4] Bigot J., Gadat S., Loubes J.M.: Statistical M-Estimation and Consistency in Large Deformable Models for Image Warping. Journal of Mathematical Imaging and Vision, 2009

[5] Younes L.: Diffeomorphic Learning. Arrive 2019

[2]