

Stage : Apprentissage machine supervisé et non supervisé, ENSTA, 20 et 21 octobre 2020

Effectif : environ 20

Auteur du CR : Jean-Julien Fleck

/ Déroulé du stage

L'ENSTA nous a accueillis pour un stage concernant l'apprentissage machine sur deux jours, avec exposés théoriques puis parties pratiques. Le stage s'est déroulé via Zoom du fait de la situation sanitaire, mais c'était plutôt un avantage car nombre de stagiaires n'auraient peut-être pas pu faire le déplacement physique même en conditions normales¹.

Le stage s'est conduit en deux journées complètes. La première était menée par Zacharie Alès, enseignant-chercheur de l'ENSTA, au sujet de l'apprentissage non supervisé. Le cours du matin portait sur une étude de la méthode d'analyse en composantes principales (ACP) et une présentation du module scikit-learn de Python qui permet de la mettre en œuvre. Un TP avec un fichier-squelette JuPyter fourni par l'enseignant a permis une mise en application rapide et efficace des concepts développés dans le cours.

Certes, le caractère distant est moins convivial pour discuter d'un détail de code qui ne fonctionne pas mais le très bon niveau du fichier et toutes les explications qui y ont été mises ont permis une utilisation optimale via la plateforme colab.research.google.com, ce qui a garanti un environnement de développement unifié sans avoir à se soucier de problèmes d'installation de bibliothèques potentiellement présentes sur les machines des uns mais pas des autres.

Le cours de l'après-midi portait quant à lui sur les méthodes de partitionnement (« clustering ») dans l'objectif de classifier plusieurs éléments d'un ensemble de données en divers groupes de manière automatique.

La suite du cours théorique a plus particulièrement porté sur la méthode de partition des k-moyennes, ainsi que sur la méthode par propagation d'affinité. La dernière partie du cours a, comme le matin, présenté les fonctions du module scikit-learn permettant de mettre en application les algorithmes décrits pour les appliquer sur des problèmes réels. Un TP a conclu cette séquence pour prendre en main de manière pratique les concepts développés durant la partie théorique.

La seconde journée a été conduite sous la direction de David Filliat, directeur de l'U2IS (Unité d'Informatique et d'Ingénierie des Systèmes). Il nous a présenté les méthodes d'apprentissage supervisé (c'est-à-dire, en se basant sur des données déjà étiquetées, être capable d'étiqueter correctement de nouvelles données, comme reconnaître des oiseaux, des voitures ou des chats dans des images). La matinée a été consacrée aux principes sous-jacents aux réseaux de neurones avec l'explication des différentes manipulations nécessaires à leur bon fonctionnement. Un TP a permis la mise en pratique avec utilisation du module PyTorch de Python via une prise en main de la bibliothèque puis une utilisation pour faire des reconnaissances de formes à partir d'une base de données d'images déjà traitées. La notion de matrice de confusion a permis de visualiser directement les identifications qui étaient bien ou mal faites selon les paramètres utilisés pour construire le réseau de neurones.

L'après-midi, la notion d'apprentissage supervisé a été creusée avec notamment les principes des arbres de décision puis des forêts aléatoires d'arbres de décision. Nous avons terminé sur une intéressante comparaison des différentes méthodes d'apprentissage (machines à vecteur support, arbres de décision, forêts aléatoires, réseaux de neurones) sur des modèles-jouets de données à deux dimensions. Un TP plus particulièrement centré sur les arbres de décision est venu compléter cette journée bien remplie.

1 C'était en tout cas mon cas

/ Conclusion et remerciements

En résumé, ce furent deux journées de stage bien denses et très intéressantes qui ont permis de présenter des méthodes que nos étudiants sont susceptibles de rencontrer ou d'utiliser notamment dans le cadre des TIPE, et ont ouvert des perspectives d'applications notamment dans le traitement d'images. Un grand merci aux organisateurs et intervenants.