

## *Machine learning avancé/advanced topics in machine learning*

Florence d'Alche-Buc - Telecom

Cours : 20 heures

### *Objectif*

L'apprentissage statistique est généralement présenté sous la forme de trois problèmes: la régression, la classification binaire/multiclasse et la classification non supervisée. Dans les cas réels d'application de l'apprentissage, le scénario est souvent différent. Par exemple, pour résoudre une tâche de classification ou de régression, des données non étiquetées peuvent être disponibles et être utilisées pendant la phase d'apprentissage il s'agit alors de l'apprentissage semi-supervisé. Dans d'autres cas, on parle d'apprentissage sous faible supervision; par exemple, certaines classes sont peu représentées parmi les exemples d'apprentissage, il s'agit alors de "One-shot/few-shot learning", voire quand certaines classes ne sont pas du tout observées pendant l'apprentissage, de "Zero-shot learning".

De manière plus générale, l'hypothèse que la distribution des exemples de tests est la même que celle des exemples d'apprentissage se révèle souvent caduque, et il faut alors être capable de transférer ce qu'on a appris pour une distribution vers une autre distribution. Dans d'autres cas, une des classes à découvrir est extrêmement rare, typiquement un défaut, on parle d'apprentissage à partir de classes déséquilibrées qui peut également se transformer en apprentissage à une classe, si les exemples d'une seule classe sont disponibles. Dans d'autres cas, l'utilisateur souhaite ordonner et non classer une liste d'objets qui lui est présentée, il s'agit alors de "Ranking".

L'objectif de ce cours est d'initier les élèves à ces différents scénarios et de leur présenter un ensemble d'algorithmes pour traiter ces problèmes et le cas échéant les résultats théoriques associés quand ils sont disponibles. Le point de vue du cours est agnostique en terme de modèles d'apprentissage.

L'évaluation du cours se fera par un examen de questions de cours et exercices.