

Data Mining (6 ECTS)

Aurélie Fischer

2^e semestre

Présentation

Le cours de Data Mining a pour objectif de présenter différentes techniques d'apprentissage supervisé (en classification et en régression) et non supervisé. L'apprentissage statistique désigne un ensemble de méthodes et d'algorithmes permettant d'extraire des informations pertinentes d'un ensemble de données et d'apprendre des comportements à partir d'exemples.

Programme

Nous nous intéresserons à des approches telles que la classification par les centres mobiles, les modèles de mélange gaussien, l'analyse discriminante linéaire et quadratique, la régression logistique, les classifieurs des plus proches voisins et à noyau, les arbres de décision ainsi que les méthodes bagging, boosting et forêts aléatoires, les machines à vecteurs de support, la régression ridge ou lasso ou encore les réseaux de neurones. Nous définirons les courbes ROC, évoquerons les problèmes de sélection de modèle et les méthodes de validation croisée. Une introduction à la théorie de Vapnik-Chervonenkis complètera le cours.

Connaissances requises

Notions de base en probabilités, statistique, analyse des données, ou cours de la semaine de rentrée du M2 ISIFAR.