

Rudy SICARD ([LIP6](#))

Apprentissage automatique de modèles probabilistes (paramètres et structure) à l'aide du principe Bayesian Model Averaging (BMA)

L'apprentissage d'un modèle de forme paramétrique se limite souvent à l'optimisation de ses paramètres, p. ex. à maximiser la vraisemblance des paramètres en fonction de la base d'apprentissage. Cette approche a deux limitations. D'une part, elle ne tient pas compte de l'incertitude sur les valeurs des paramètres, et d'autre part elle n'offre pas de mécanisme inhérent pour choisir entre plusieurs formes paramétriques possibles. Autrement dit l'apprentissage concerne uniquement les paramètres du modèle mais pas sa structure. Pour cette raison, ce type d'approche est souvent utilisé avec un modèle très versatile, de structure complexe, pour le quel on limite le sur apprentissage à l'aide d'un critère de régularisation ad hoc.

Du point de vue théorique, il existe un meilleur principe d'apprentissage, nommé Bayesian Model Averaging (BMA), qui résout ces deux difficultés. Ce principe consiste simplement à appliquer la théorie des probabilités, c.-a.-d. à calculer des sommes, des intégrales, sans aucune optimisation. Cependant, du point de vue pratique ce principe est souvent plus difficile à appliquer qu'une optimisation.

Dans cette présentation, nous présentons brièvement le cadre probabiliste pour l'apprentissage automatique. Nous rappelons la définition du principe BMA et son caractère optimal en moyenne. Puis nous présentons quelques cas où ce principe est mis en oeuvre de manière exacte ou approchée. Ces cas illustrent le fonctionnement de BMA, et montrent comment ce principe gère la complexité du modèle. Notamment nous expliquons comment la sélection de variables et les modèles parcimonieux - en vogue depuis les Support Vector Machines (SVM)

- émergent, naturellement du principe BMA.