

APPRENTISSAGE PAC-BAYÉSIEEN ET RÉSEAUX DE NEURONES

Pascal Germain ¹

¹ *Équipe-projet Modal, Inria Lille - Nord Europe,
40, avenue Halley 59650 Villeneuve d'Ascq, France*

Résumé. En apprentissage automatique, l'approche PAC-bayésienne permet d'obtenir des garanties statistiques sur la perte d'agrégation de plusieurs prédicteurs (par exemple, cette agrégation peut prendre la forme d'un vote de majorité pondéré). La théorie PAC-bayésienne «classique», initiée par McAllester (1999), étudie le cadre d'apprentissage inductif, sous l'hypothèse que les exemples d'apprentissage sont générés de manière indépendante et qu'ils sont identiquement distribués (i.i.d.) selon une distribution de probabilité inconnue, mais fixe.

Dans ce travail, nous explorons la possibilité d'utiliser les outils de la théorie PAC-bayésienne pour obtenir des garanties statistiques sur la capacité de généralisation de prédicteurs de type réseau de neurones. Cette recherche nous mène à suggérer une stratégie d'optimisation originale pour des réseaux à fonctions d'activations binaires.

Mots-clés. Apprentissage statistique, théorie PAC-bayésienne, réseaux de neurones.

Abstract. In machine learning, the PAC-Bayesian theory provides statistical guarantees on the loss of aggregation of several predictors (for example, this aggregation may take the form of a weighted majority vote). The “classical” PAC-Bayesian theory, initiated by McAllester (1999), studies the inductive learning framework, under the assumption that learning examples are generated independently and are identically distributed (i.i.d.) by an unknown but fixed probability distribution.

In this work, we explore the possibility of using the PAC-Bayesian framework to obtain statistical guarantees on the generalization properties of neural network predictors. This research leads us to an original optimization strategy for neural networks with binary activation functions.

Keywords. Statistical learning, PAC-Bayesian theory, neural networks.