
Clustering des observations des modèles de Markov cachées

Ibrahim Kaddouri^{*1}, Élisabeth Gassiat^{*†1}, and Zacharie Naulet^{*‡1}

¹Laboratoire de Mathématiques d'Orsay – Université Paris-Saclay, Centre National de la Recherche Scientifique, Centre National de la Recherche Scientifique : UMR8628 – France

Résumé

Grâce à leur structure de dépendance, les modèles de Markov cachés (HMM) non paramétriques permettent d'effectuer le clustering sans spécifier les distributions de groupes. L'objectif de ce travail est d'étudier le risque de Bayes de clustering lors de l'utilisation des HMM et de proposer des procédures de clustering associées. Nous donnons d'abord un résultat liant le risque de Bayes de classification et le risque de Bayes de clustering, que nous utilisons pour identifier la quantité clé déterminant la difficulté du clustering. Nous donnons également une preuve de ce résultat dans le cadre de données i.i.d., qui a un intérêt indépendant. Nous étudions ensuite l'excès de risque du classificateur de plugin. On montre que tous ces résultats restent valables dans le contexte en ligne où les observations sont regroupées de manière séquentielle. Des simulations illustrent nos résultats.

Mots-Clés: Chaînes de Markov cachées, Clustering, Modèles de mélange, statistiques non, paramétriques

*Intervenant

†Auteur correspondant: elisabeth.gassiat@universite-paris-saclay.fr

‡Auteur correspondant: zacharie.naulet@universite-paris-saclay.fr