

**CLASSIFICATION PAR PLUS PROCHES VOISINS,
CONSISTANCE EN DIMENSION FINIE ET CADRE
FONCTIONNEL**

Sébastien Gadat

**UMR 5219- Institut Mathématiques de Toulouse
Université Paul Sabatier**

L' algorithme des k plus proches voisins est une des plus anciennes méthodes d'apprentissage statistique et reste encore très populaires dans de nombreuses applications de classification supervisée ou de régression. Les premières études statistiques remontent à Györfi [1], Devroye & Wagner [2], Cover & Hart [3], ou Stone [4]. Nous étudions la consistance de ces méthodes dans un contexte de classification supervisée et sous des conditions de marge introduites par Audibert & Tsybakov [5], nous établissons un résultat nouveau d'optimalité asymptotique de classification par plus proches voisins en dimension finie d . Nous montrons que la vitesse est fortement influencée par la dimension ambiante et la taille de n , effectif de l'échantillon, et de k le nombre de voisins utilisés par l'algorithme.

Nous étudions ensuite le cas fonctionnel où les entrées sont perturbées par un modèle de bruit blanc, puis un cas particulier où un opérateur aléatoire vient perturber les observations. Nous concluons par quelques illustrations numériques et des perspectives statistiques, numériques et de modélisation.

Bibliographie

- [1] Györfi, Lazlo. On the rate of convergence of nearest neighbor rules., IEEE Trans. Inform. Theory 24 (1978), no. 4, 509-512.
- [2] Devroye, Luc P. ; Wagner, T. J. The strong uniform consistency of nearest neighbor density estimates. Ann. Statist. 5 (1977), no. 3, 536-540.
- [3] Cover Thomas and Hart, Peter. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1) :2127, 1967.
- [4] Stone, Charles J. Consistent nonparametric regression. Ann. Statist, 5(4) : 595-620, 1977.
- [5] Audibert, Jean-Yves ; Tsybakov, Alexandre B. Fast learning rates for plug-in classifiers. Ann. Statist. 35 (2007), no. 2, 608-633.
- [6] Gadat, S. & Klein, T. & Marteau, C. Optimal nearest neighbour classification through margin assumption. (2013), Preprint.