Regression on a manifold with a Laplace eigenbasis and topological penalty

Olympio Hacquard, Gilles Blanchard, Clément Levrard
Université Paris-Saclay, Université Paris-Saclay, Université Paris Diderot

Email : ohacquar@universite-paris-saclay.fr

Mots Clés : Régression statistique, Laplacien de graphe, Analyse topologique de données, Persistance topologique.

Biographie – Ancien élève de l'Ecole Normale Supérieure Paris-Saclay (2016-2020), Olympio Hacquard a débuté en septembre 2020 une thèse à l'Institut Mathématiques d'Orsay au sein de l'équipe Datashape, spécialisée dans l'analyse topologique de données, sous la direction de Gilles Blanchard et Clément Levrard sur des modèles de réduction de dimension pour l'homologie persistante. Il bénéficie à ce titre d'un contrat CDSN pour les trois années de sa thèse.

Resumé :

On considère un problème de regression où l'on observe des données X sur une variété différentielle et des étiquettes réelles Y. Sous l'hypothèse que chaque donnée est telle que $Y_i = f(X_i) + \varepsilon_i$ où ε_i est un terme de bruit, on cherche à retrouver la fonction de régression f. Pour ce faire, on tente d'approcher f par une combinaison linéaire des premières fonctions propres de l'opérateur de Laplace-Beltrami. Puisque ces fonctions sont très compliquées à estimer (en particulier lorsque la variété n'est pas connue ce qui est courant en pratique), on construit un graphe sur les données et on remplace les fonctions propres de l'opérateur de Laplace-Beltrami par les vecteurs propres de la matrice Laplacienne du graphe. Il est prouvé dans [5] que sous de bonnes hypothèses, les vecteurs propres ainsi construits convergent vers les fonctions propres de l'opérateur de Laplace-Beltrami. Généralisation naturelle de la base de Fourier à une variété générale, les bases Laplaciennes sont composées de fonctions oscillantes et sont particulièrement propices aux phénomènes de surinterprétation. On discutera deux type de pénalités topologiques construites sur l'homologie persistante des sous-ensembles de niveaux de fonctions (voir [1]) permettant de proposer une certaines robustesse au bruit et de pallier aux problèmes de surinterprétation. S'inspirant du débruitage par pénalisation de la variation totale, les modèles de pénalité topologique sont néanmoins beaucoup plus généraux et offrent une bien meilleure robustesse, notamment en grande dimension. Des expériences concluantes et aux résultats explicables, à la fois en reconstruction et en prédiction montrant l'intérêt de telles pénalités topologiques seront complétées par la présentation de résultats théoriques inédits garantissant une vitesse rapide de convergence. L'utilisation d'information topologique pour des problèmes d'apprentissage statistique a donné lieu à des développements très variés et très prometteurs ces dernières années, comme par exemple dans [2], [3]. ou encore [4]. Néanmoins, de nombreuses interrogations subsistent concernant les garanties théoriques de ces modèles et les résultats développés ici constituent un premier pas vers une meilleure compréhension de l'intérêt des méthodes topologiques.

Références

[1] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and topological inference*, volume 57. Cambridge University Press, 2018.

[2] Rickard Brüel-Gabrielsson, Bradley J Nelson, Anjan Dwaraknath, Primoz Skraba, Leonidas J Guibas, and Gunnar Carlsson. A topology layer for machine learning. *arXiv preprint arXiv:1905.12200*, 2019.

[3] Mathieu Carriere, Frédéric Chazal, Marc Glisse, Yuichi Ike, and Hariprasad Kannan. A note on stochastic subgradient descent for persistence-based functionals: convergence and practical aspects. *arXiv preprint arXiv:2010.08356*, 2020.

[4] Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang. A topological regularizer for classifiers via persistent homology. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2573–2582, 2019.

[5] Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error estimates for spectral convergence of the graph laplacian on random geometric graphs toward the laplace–beltrami operator. *Foundations of Computational Mathematics*, pages 1–61, 2019.