

A Feature selection method based on topological aspect of feature manifolds

Jea-Min PARK¹, JungHo PARK¹, Hyun-Young CHOI¹, Hyeon-Woo JEONG¹

- 1) *Department of Mathematical science, Seoul National University, Seoul 151-742, KOREA*
- 2) *Department of Mathematical science, Seoul National University, Seoul 151-742, KOREA*
- 3) *Department of Mathematics, Chungnam National University, Yuseong,Daejeon,305-764,Korea*
- 4) *Department of Electrical Engineering, Inha University, Incheon 402-751, KOREA*

ABSTRACT

Topological Data Analysis(TDA) is a well-known method in data analysis. We introduce a feature selection method in ML by TDA techniques[1,2]. In the ML process, feature selection is an important part of model training[3,4]. Much of feature selection methods are depending on the domain insight from a dataset. The insight contains the independency between features and the fitness of a feature to data. Our new method overcomes this issue by designing the combination method of max-min selection and a Morse theory. The max-min selection gives an independency between features and a proper Morse function enables the selection of features with desired properties. Moreover, the selection process is automatic after choosing a Morse function and cover of its range. Especially, it works well to data set with a large number of features. We will introduce a proper metric on feature manifold and an example of the Morse function that represents the purpose of analysis about a data set.

REFERENCES.

- (1) Paper in a journal
 1. Vin de Silva and Gunnar Carlsson “*Topological estimation using witness complexes,*” Eurographics Symposium on Point-Based Graphics (2004)
 2. Gurjeet Singh, Facundo Mémoli, “*Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition,*” Eurographics Symposium on Point-Based Graphics (2007)
- (2) Paper in Conference Proceedings
 3. X He, D Cai, P Niyogi, “*Laplacian score for feature selection,*” Advances in neural information processing systems, 507-514
 4. Quanquan Gu, Zhenhui Li, Jiawei Han, “*Generalized Fisher Score for Feature Selection,*” In Proc. of the 27th Conference on Uncertainty in Artificial Intelligence (UAI), Barcelona, Spain, 2011