

Perturbation Analysis for Similarity Based on Entropy in a Linear Subspace Method

Kuniyoshi Hayashi*

Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan
k-hayashi@ems.okayama-u.ac.jp

Fumio Ishioka

School of Law, Okayama University, Okayama, Japan
fishioka@law.okayama-u.ac.jp

Hiroshi Suito

Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan
suito@ems.okayama-u.ac.jp

Koji Kurihara

Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan
kurihara@ems.okayama-u.ac.jp

Sensitivity analysis based on the influence function of the statistical discrimination method is important because it enables the detection of large influential training samples such as mislabeled data for prediction accuracy. In fact, by this analysis, we can enhance the performance of a target classifier in terms of the prediction accuracy. However, thus far, diagnostics has been limited to statistical methods in the field of statistics. We have therefore developed a sensitivity analysis method based on the influence function of Class-Featuring Information Compression (CLAFIC) method, which is a discriminant method in pattern recognition. In the CLAFIC method, the squared value of the norm calculated by projecting a test observation into the subspace in each class and the entropy value calculated by its squared value are used for the similarity. In our previous studies, we have mainly focused on the diagnostics for the discrimination based on the similarity of the projection norm and developed its sensitivity analysis based on the influence function. In this paper, to expand the possibility of the diagnostics for the CLAFIC method, we propose a sensitivity analysis method for the similarity based on the entropy value. We perturb the average discriminant score in each class calculated from the similarity based on the entropy value. Then, we derive the influence function and present a diagnostic procedure with the derived influence function.

Key words: CLAFIC method, diagnostics, influence function, pattern recognition