

2020-10-04

Discretization and Feature Selection Based on Bias Corrected Mutual Information Considering High-Order Dependencie

Sharmin, Sadia

[Springer Link](#)

<http://ar.iub.edu.bd/handle/11348/494>

Downloaded from IUB Academic Repository

Mutual Information (MI) based feature selection methods are popular due to their ability to capture the nonlinear relationship among variables. However, existing works rarely address the error (bias) that occurs due to the use of finite samples during the estimation of MI. To the best of our knowledge, none of the existing methods address the bias issue for the high-order interaction term which is essential for better approximation of joint MI. In this paper, we first calculate the amount of bias of this term. Moreover, to select features using χ^2 based search, we also show that this term follows χ^2 distribution. Based on these two theoretical results, we propose Discretization and feature Selection based on bias corrected Mutual information (DSbM). DSbM is extended by adding simultaneous forward selection and backward elimination (DSbM_{fbfb}). We demonstrate the superiority of DSbM over four state-of-the-art methods in terms of accuracy and the number of selected features on twenty benchmark datasets. Experimental results also demonstrate that DSbM outperforms the existing methods in terms of accuracy, Pareto Optimality and Friedman test. We also observe that compared to DSbM, in some dataset DSbM_{fbfb} selects fewer features and increases accuracy.