Tradeoff in Unsupervised Learning: Interplay of Model Error and Data Error

Gilhan Kim,* Hojun Lee, Junghyo Jo, and Yongjoo Baek

Seoul National University, Republic of Korea

(Dated: July 2, 2023)

In conventional supervised learning, the generalization error (GE) typically exhibits bias-variance tradeoff: lowering the bias by increasing he model complexity entails an increase of the variance. Meanwhile, little is known about whether the same tradeoff exists in unsupervised learning. In this study, we propose that unsupervised learning generally exhibits a different two-component tradeoff involving the model error (ME) and the data error (DE)—using a more complex model reduces the ME at the cost of the DE. This is corroborated by training the restricted Boltzmann machine (RBM) to generate the configurations of the two-dimensional Ising model at a given temperature T and the totally asymmetric simple exclusion process (TASEP) with given entry and exit rates α and β. Our results indicate that the optimal model tends to be more complex when the training data are more complex.

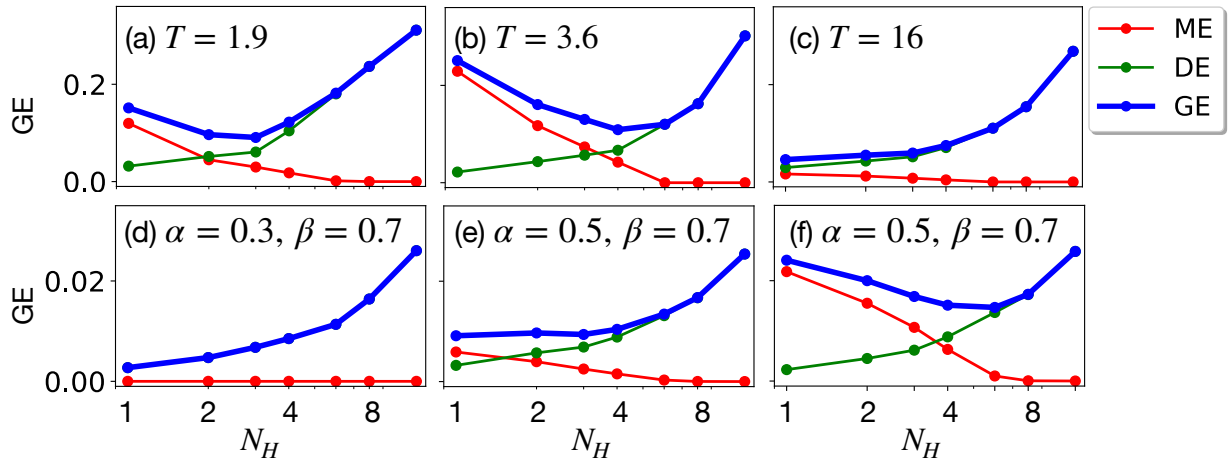

FIG. 1. Dependence of various error measures on the number of hidden nodes in the RBM. We show results for (a–c) the 2-d Ising model and (d–f) the TASEP.

* physicaone@snu.ac.kr