

Figure 1: Comparison of testing accuracies from logistic regression, BSVRB^{v1} with $m = 1$ lower-level problem, and BSVRB^{v1} with $m = 100$ lower-level problems on various noise level of dataset *a8a*

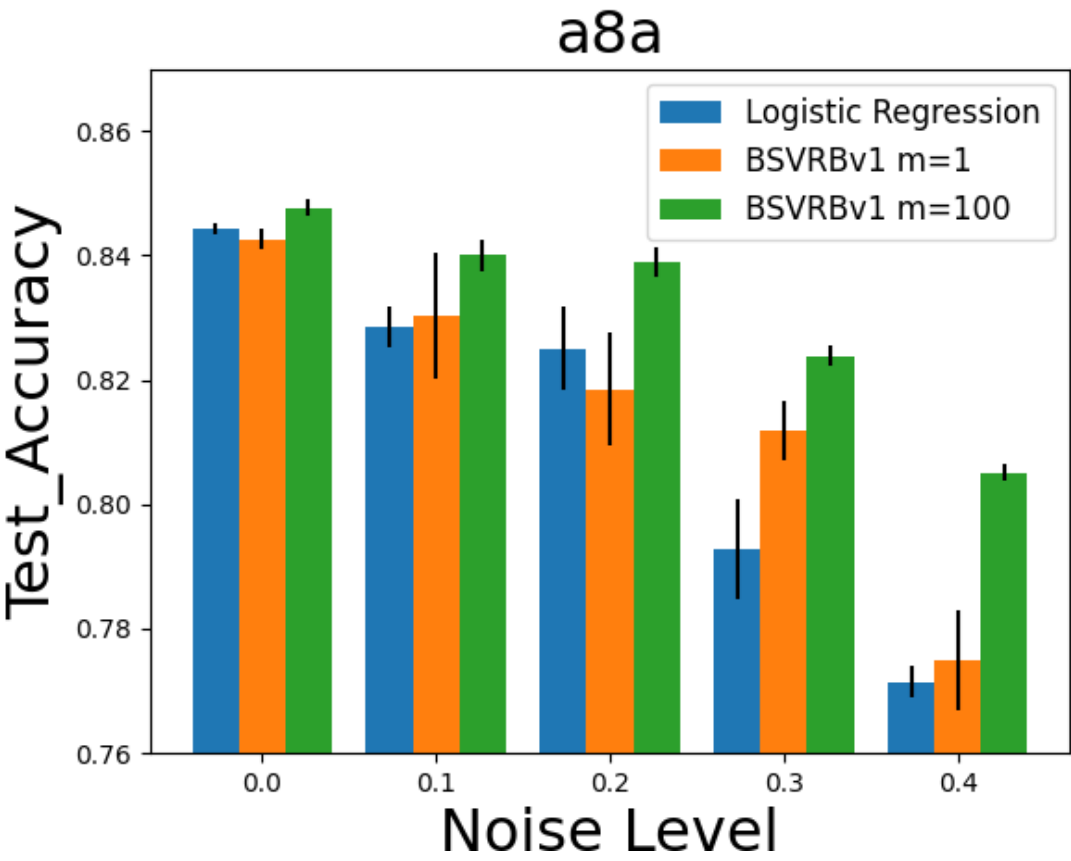


Table 1: Testing accuracies and standard deviation over 3 runs with different random seeds from logistic regression, BSVRB with $m = 1$ lower-level problem, and BSVRB with $m = 100$ lower-level problems on various noise level of dataset *a8a*. Noise level represents the proportion of training sample labels that are flipped. 70% of the positive samples are removed from training data except for noise level 0*, which means no label noise and no data imbalance.

Noise Level	Logistic Regression	BSVRB ^{v1} ($m = 1$)	BSVRB ^{v1} ($m = 100$)
0*	0.8528 \pm 0.0005	0.8526 \pm 0.0002	0.8509 \pm 0.0011
0	0.8442 \pm 0.0009	0.8426 \pm 0.0016	0.8477 \pm 0.0013
0.1	0.8285 \pm 0.0034	0.8303 \pm 0.0100	0.8400 \pm 0.0025
0.2	0.8250 \pm 0.0066	0.8185 \pm 0.0090	0.8388 \pm 0.0024
0.3	0.7929 \pm 0.0081	0.8118 \pm 0.0047	0.8239 \pm 0.0015
0.4	0.7715 \pm 0.0025	0.7749 \pm 0.0079	0.8051 \pm 0.0013