We conducted new simulations in the same setting but with a different benchmark. We compare our PMI score with the following benchmark to avoid ties between candidate datasets: we first train a classifier on the evaluated candidate dataset using logistic regression, and then use its cross entropy on the small test set T as its score.

The simulation results are shown as follows. We tried three different Gaussian priors with covariance matrices $\sigma^2 \cdot \mathbf{I}$ with $\sigma^2 = 1, 9, 25$. The generalization errors of the two methods are shown in the tables and the win rates of the two methods are shown in the figures. The orange bars represent the win rates of our PMI score, the blue bars represent the win rates of the cross entropy benchmark, and the gray bars represent the rates of ties. The label under the bar indicates K, N_T, M.

K, N_T, M	PMI score	Cross entropy	Difference
$1, 20, 200$	0.7699	0.7677	0.0022
$1, 20, 400$	0.7696	0.7670	0.0026
$1, 20, 600$	0.7677	0.7647	0.0030
$1, 30, 200$	0.7757	0.7738	0.0019
$1, 30, 400$	0.7761	0.7743	0.0018
$1, 30, 600$	0.7737	0.7713	0.0024
$1, 40, 200$	0.7784	0.7773	0.0011
$1, 40, 400$	0.7753	0.7740	0.0014
$1, 40, 600$	0.7773	0.7758	0.0015
$3, 20, 200$	0.7735	0.7712	0.0022
$3, 20, 400$	0.7747	0.7714	0.0033
$3, 20, 600$	0.7719	0.7682	0.0038
$3, 30, 200$	0.7781	0.7761	0.0020
$3, 30, 400$	0.7791	0.7763	0.0028
$3, 30, 600$	0.7758	0.7727	0.0032
$3, 40, 200$	0.7793	0.7774	0.0018
$3, 40, 400$	0.7794	0.7771	0.0023
$3, 40, 600$	0.7790	0.7762	0.0028
$5, 20, 200$	0.7749	0.7733	0.0016
$5, 20, 400$	0.7750	0.7724	0.0027
$5, 20, 600$	0.7721	0.7690	0.0031
$5, 30, 200$	0.7764	0.7745	0.0019
$5, 30, 400$	0.7782	0.7756	0.0026
$5, 30, 600$	0.7786	0.7755	0.0031
$5, 40, 200$	0.7804	0.7785	0.0018
$5, 40, 400$	0.7826	0.7804	0.0022
$5, 40, 600$	0.7782	0.7752	0.0030

Table 1: Generalization errors of the two methods when $\sigma^2 = 1$.

K, N_T, M	PMI score	Cross entropy	Difference
$1, 20, 200$	0.7686	0.7671	0.0015
$1, 20, 400$	0.7685	0.7661	0.0024
$1, 20, 600$	0.7669	0.7642	0.0027
$1, 30, 200$	0.7766	0.7754	0.0012
$1, 30, 400$	0.7747	0.7729	0.0018
$1, 30, 600$	0.7740	0.7719	0.0022
$1, 40, 200$	0.7754	0.7747	0.0007
$1, 40, 400$	0.7763	0.7752	0.0011
$1, 40, 600$	0.7766	0.7752	0.0014
$3, 20, 200$	0.7732	0.7713	0.0019
$3, 20, 400$	0.7739	0.7713	0.0026
$3, 20, 600$	0.7715	0.7678	0.0037
$3, 30, 200$	0.7783	0.7763	0.0020
$3, 30, 400$	0.7797	0.7773	0.0024
$3, 30, 600$	0.7777	0.7745	0.0032
$3, 40, 200$	0.7807	0.7792	0.0015
$3, 40, 400$	0.7795	0.7771	0.0024
$3, 40, 600$	0.7797	0.7769	0.0028
$5, 20, 200$	0.7738	0.7724	0.0014
$5, 20, 400$	0.7730	0.7704	0.0026
$5, 20, 600$	0.7735	0.7703	0.0031
$5, 30, 200$	0.7780	0.7764	0.0016
$5, 30, 400$	0.7777	0.7753	0.0023
$5, 30, 600$	0.7771	0.7743	0.0028
$5, 40, 200$	0.7826	0.7810	0.0016
$5, 40, 400$	0.7794	0.7773	0.0022
$5, 40, 600$	0.7812	0.7785	0.0027

Table 2: Generalization errors of the two methods when $\sigma^2 = 9$.

K, N_T, M	PMI score	Cross entropy	Difference
$1, 20, 200$	0.7684	0.7671	0.0013
$1, 20, 400$	0.7691	0.7671	0.0020
$1, 20, 600$	0.7703	0.7677	0.0025
$1, 30, 200$	0.7726	0.7713	0.0013
$1, 30, 400$	0.7741	0.7722	0.0019
$1, 30, 600$	0.7764	0.7745	0.0019
$1, 40, 200$	0.7785	0.7775	0.0010
$1, 40, 400$	0.7748	0.7738	0.0010
$1, 40, 600$	0.7754	0.7741	0.0013
$3, 20, 200$	0.7721	0.7704	0.0017
$3, 20, 400$	0.7731	0.7706	0.0024
$3, 20, 600$	0.7713	0.7685	0.0028
$3, 30, 200$	0.7765	0.7750	0.0015
$3, 30, 400$	0.7773	0.7746	0.0027
$3, 30, 600$	0.7773	0.7739	0.0034
$3, 40, 200$	0.7806	0.7788	0.0018
$3, 40, 400$	0.7801	0.7777	0.0024
$3, 40, 600$	0.7793	0.7767	0.0026
$5, 20, 200$	0.7727	0.7715	0.0013
$5, 20, 400$	0.7726	0.7704	0.0022
$5, 20, 600$	0.7731	0.7704	0.0027
$5, 30, 200$	0.7775	0.7760	0.0014
$5, 30, 400$	0.7776	0.7749	0.0026
$5, 30, 600$	0.7799	0.7770	0.0030
$5, 40, 200$	0.7804	0.7788	0.0015
$5, 40, 400$	0.7796	0.7773	0.0024
$5, 40, 600$	0.7798	0.7769	0.0028

Table 3: Generalization errors of the two methods when $\sigma^2 = 25$.

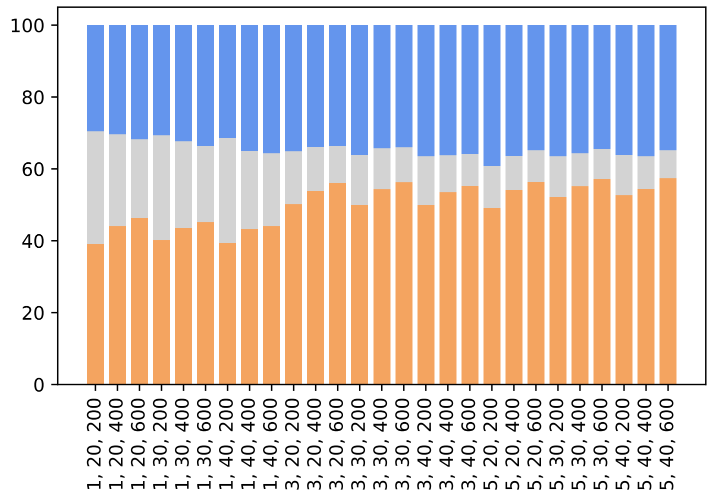

Figure 1: The win rates of the two methods when $\sigma^2 = 1$.

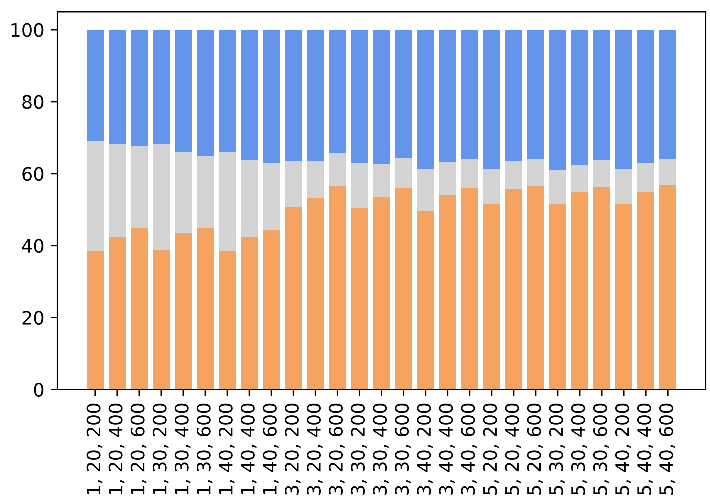

Figure 2: The win rates of the two methods when $\sigma^2 = 9$.

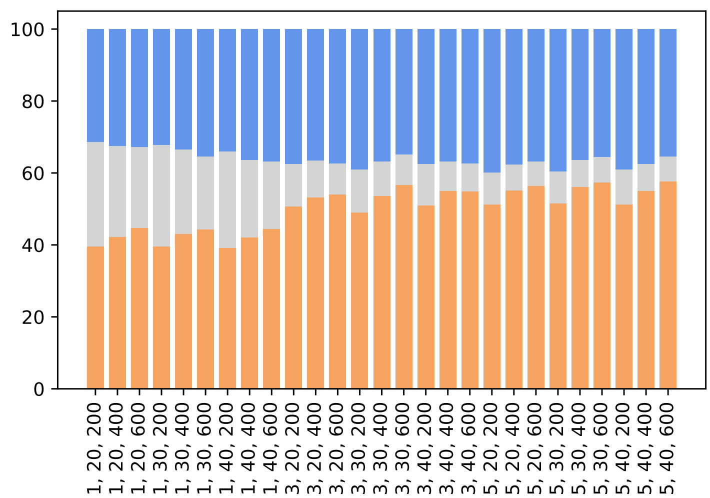

Figure 3: The win rates of the two methods when $\sigma^2 = 25$.