Differentially Private Small Dataset Release Using Random Projections

Lovedeep Gondara **Ke Wang**
Department of Computing Science
Simon Fraser University

Supplementary Material

Experiments

Privacy Budget Allocation

For this comparison, we keep the core setup the same as the main results. With the total privacy budget kept constant at $\epsilon = 4, \delta = 0.0001$, and the values for k_2 kept constant at $0.6d$. Table 1 shows the results. We see that the best results are obtained as we increase the privacy budget allocation for the random projection, especially $\geq 40\%$, leading to a less noisy random projection. Signaling that random projection plays a "larger" role in the reconstruction compared to the right singular vector.

Computational Complexity

We compare the average run time of DPRP and GANs[1] over the 50 runs per dataset using percent reduction in computational time as a comparison metric, calculated

[1]Both GANs have similar run times, hence for this comparison, we take the average of both.

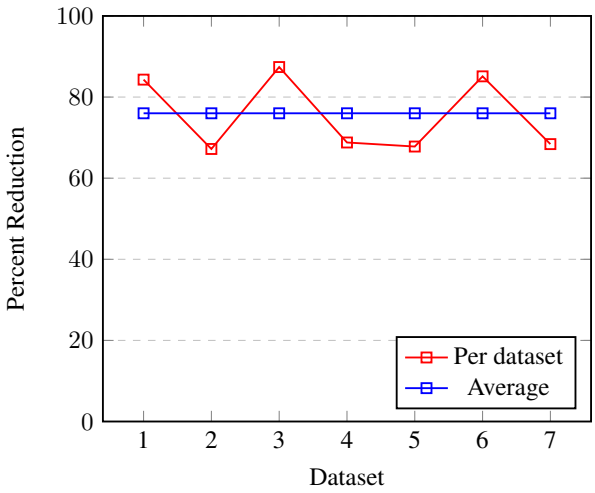

Figure 1: Percent reduction in computational time (DPRP vs GANs), the red line is per-dataset and the blue line is the average over all datasets. Datasets are enumerated according to their position in Table 1. We observe a significant decrease in computational time when using DPRP compared to GANs.

Data	RP:20%, SVD:75% (Acc,AUPRC)	RP:40%, SVD:55% (Acc,AUPRC)	RP:60%, SVD:35% (Acc,AUPRC)	RP:80%, SVD:15% (Acc,AUPRC)
Coimbra BC	0.48, 0.43	0.49, 0.50	0.55, 0.51	0.52, 0.59
Wisconsin BC	0.58, 0.59	0.70, 0.61	0.68, 0.62	0.68, 0.65
Indian Liver	0.68, 0.61	0.75, 0.65	0.77, 0.65	0.77, 0.66
Dermatology	0.29, 0.28	0.32, 0.30	0.38, 0.34	0.32, 0.30
Cervical Cancer	0.95, 0.90	0.96, 0.92	0.96, 0.92	0.96, 0.92
Caesarian	0.51, 0.55	0.54, 0.59	0.53, 0.65	0.55, 0.66
HCC	0.64, 0.56	0.68, 0.59	0.72, 0.61	0.72, 0.62

Table 1: Evaluating the effect of privacy budget allocation on the outcome. Results are shown using AUPRC and Classification Accuracy (Acc). Privacy budget is kept fixed at $\epsilon = 4, \delta = 0.0001$ with allocation among differentially private random projection and differentially private SVD varied as shown in the table. We observe better outcomes as the privacy budget for random projection is increased.

as

$$\%\text{Reduction} = \frac{T_{GAN} - T_{DPRP}}{T_{GAN}} \times 100 \qquad (1)$$

where T_{GAN} is the average time taken by GANs and T_{DPRP} is the time taken by DPRP.

Figure 1 shows the results. The red line is the percent reduction in computational time using DPRP compared to GANs per dataset and the blue line is the average reduction across all datasets. Datasets are enumerated according to the listing in Table 1 (that is, Coimbra BC is 1, Wisconsin BC is 2, and so on). We observe that DPRP offers a reduction in computational time greater than 65% on all datasets, with gains close to 90% on some datasets. The average reduction in computational time across all datasets is close to 80%, a significant decrease compared to GANs.