Table 1: Results in the large-data regime: average test AUCs with different positive class-prior $\pi := \pi_{\mathrm{tr}} = \pi_{\mathrm{te}}$. Values in bold are not statistically different at the 5% level from the best performing method in each row according to a paired t-test. '# best' row represents the number of results of each method that are the best or comparable to it.

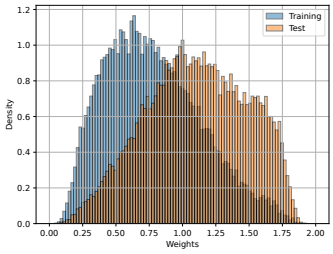
Data	π	Ours	trPU	tePU	trAUC	teAUC	trteAUC	PAUC	UDAUC	CPU	PURA
MNIST	0.01	**0.849**	0.719	0.690	0.795	0.823	**0.842**	0.417	0.823	0.704	0.808
	0.05	**0.829**	0.766	0.738	0.813	0.814	0.813	0.507	0.813	0.774	0.775
	0.1	**0.810**	0.779	0.736	**0.806**	**0.805**	**0.805**	0.524	**0.806**	0.757	0.776
Fashion	0.01	**0.942**	**0.928**	0.922	0.942	0.925	0.942	0.668	**0.945**	**0.916**	0.924
MNIST	0.05	**0.935**	0.864	0.750	**0.945**	0.912	0.912	0.728	**0.946**	**0.920**	0.925
	0.1	**0.913**	0.897	0.820	0.925	0.854	0.854	0.701	0.926	0.898	**0.946**
SVHN	0.01	0.763	0.504	0.504	0.628	**0.780**	**0.780**	0.252	0.628	0.504	0.503
	0.05	**0.757**	0.504	0.504	0.599	**0.756**	**0.765**	0.252	0.605	0.504	0.506
	0.1	**0.750**	0.504	0.504	0.587	**0.749**	**0.751**	0.304	0.586	0.504	0.503
CIFAR10	0.01	**0.920**	0.450	0.476	0.915	**0.922**	**0.922**	0.343	0.914	0.509	0.854
	0.05	**0.910**	0.464	0.520	**0.904**	**0.910**	**0.910**	0.487	**0.906**	0.521	0.849
	0.1	**0.900**	0.552	0.654	**0.885**	**0.877**	**0.879**	0.547	**0.900**	0.579	0.864
# best		11	1	0	4	7	8	0	5	2	1

Table 2: Results without covariate shift: average test AUCs over different positive class-prior $\pi := \pi_{\mathrm{tr}} = \pi_{\mathrm{te}}$ within $\{0.01, 0.05, 0.1\}$.

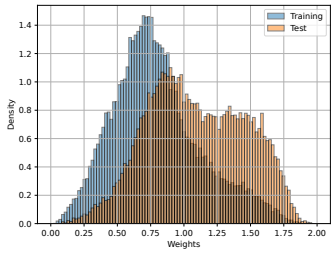
Data	Ours	trPU	trAUC
MNIST	0.927	0.814	0.927
FashionMNIST	0.982	0.916	0.982
SVHN	0.625	0.504	0.630
CIFAR10	0.885	0.436	0.875

Table 3: Results under class-prior shift: average test AUCs with different positive test class-prior π_{te} with postive training class-prior $\pi_{\mathrm{tr}} = 0.01$.

Data	π_{te}	Ours	trPU	tePU	trAUC	teAUC	trteAUC	PAUC	UDAUC	CPU
MNIST	0.01	**0.938**	0.729	0.729	**0.938**	**0.938**	**0.938**	0.562	**0.938**	0.704
	0.05	**0.929**	0.734	0.717	**0.930**	0.924	**0.930**	0.762	**0.930**	0.733
	0.1	**0.930**	0.736	0.709	**0.932**	0.921	**0.932**	0.866	**0.932**	0.731
Fashion	0.01	**0.983**	0.899	0.895	**0.982**	**0.984**	**0.984**	0.818	**0.982**	0.906
MNIST	0.05	**0.983**	0.900	0.770	**0.983**	0.980	**0.983**	0.929	**0.983**	0.897
	0.1	**0.988**	0.901	0.743	**0.988**	0.983	**0.988**	0.970	**0.988**	0.895
SVHN	0.01	**0.644**	0.504	0.504	**0.642**	**0.632**	**0.655**	0.499	**0.648**	0.504
	0.05	**0.648**	0.504	0.504	**0.625**	0.600	**0.635**	0.505	**0.624**	0.504
	0.1	**0.639**	0.504	0.504	**0.641**	0.619	**0.648**	0.507	**0.645**	0.504
CIFAR10	0.01	**0.884**	0.432	0.427	0.867	0.872	0.872	0.519	0.866	0.481
	0.05	**0.890**	0.433	0.424	0.878	0.872	0.878	0.680	0.878	0.480
	0.1	**0.891**	0.432	0.424	**0.885**	0.867	**0.886**	0.785	**0.885**	0.482
# best		12	0	0	10	3	10	0	10	0



$\pi = 0.01$



$\pi = 0.05$

Figure 1: Importance weight distributions on FashionMNIST.