Table 1: Comparison of normalised training risks and computation times for rule ensembles, averaged over cognitive complexities between 1 and 50, using SIRUS(SRS), Gradient Sum(GS), Gradient boosting (GB), XGBoost (XGB) and FCOGB, for benchmark datasets

Dataset	d	n	Training Risks SRS	GS	GB	XGB	FCOGB	Testing Risks SRS	GS	GB	XGB	FCOGB
titanic	7	1043	.895	.653	.656	.646	**.621**	.894	.695	.707	.711	**.695**
tic-tac-toe	27	958	.892	.757	.641	.621	**.604**	.885	.809	.682	.605	**.586**
iris	4	150	.685	.239	.248	.329	**.232**	.745	.521	.440	.459	**.424**
breast	30	569	.569	.277	.310	.314	**.239**	.627	**.269**	.362	.338	.342
wine	13	178	.578	.198	.223	.192	**.183**	.621	.341	.443	.409	**.250**
ibm hr	32	1470	.980	.567	**.555**	.569	.564	.974	.640	.666	.630	**.617**
telco churn	18	7043	.944	.679	.683	.678	**.667**	.945	.663	.677	.665	**.653**
gender	20	3168	.566	.250	.254	.271	**.247**	.570	**.264**	.267	.285	.265
banknote	4	1372	.854	.305	.267	.290	**.228**	.858	.312	.268	.299	**.229**
liver	6	345	.908	.809	**.805**	.811	.830	.917	.909	.901	**.878**	.881
magic	10	19020	.906	.719	.709	.710	**.707**	.903	.700	.693	.693	**.687**
adult	11	30162	.804	.604	.609	.604	**.559**	.802	.612	.623	.615	**.573**
digits5	64	3915	**.248**	.355	.331	.354	.337	**.262**	.347	.328	.348	.331
insurance	6	1338	.169	.130	.142	.144	**.127**	.177	.132	.145	.147	**.131**
friedman1	10	2000	.180	.051	.052	.051	**.019**	.165	.054	.057	.059	**.022**
friedman2	4	10000	.082	.069	.064	.059	**.023**	.082	.071	.065	.060	**.023**
friedman3	4	5000	.093	.026	.022	.022	**.009**	.092	.029	.027	.027	**.011**
wage	5	1379	.427	.370	.361	.356	**.351**	**.341**	.358	.406	.367	.375
demographics	13	6876	.219	.214	.214	.214	**.212**	**.209**	.216	.217	.217	.215
gdp	1	35	.063	.020	.020	.020	**.020**	.059	.020	.020	.020	**.020**
used cars	4	1770	.373	**.130**	.205	.153	.132	.427	.101	.157	.100	**.082**
diabetes	10	442	.156	.138	.141	.139	**.132**	.188	**.140**	.142	.147	.148
boston	13	506	.101	.088	.086	.086	**.082**	.105	**.081**	.086	.087	.088
happiness	8	315	.109	.030	.031	.030	**.030**	.109	**.033**	.038	.038	.036
life expect.	21	1649	.109	.026	.026	.026	**.025**	.110	.027	.027	.027	**.026**
mobile prices	20	2000	.148	**.131**	.137	.137	.134	.140	**.134**	.143	.145	.139
suicide rate	5	27820	.547	.566	.561	.561	**.531**	.514	.544	.542	.542	**.509**
videogame	6	16327	**.895**	.953	.953	.953	.953	.850	**.720**	**.720**	**.720**	**.720**
red wine	11	1599	.072	.034	.035	.034	**.034**	.073	**.035**	.036	.036	**.035**
covid vic	4	85	NA	.153	.121	.132	**.105**	NA	.182	**.100**	.133	.104
covid	2	225	NA	.344	.371	.891	**.335**	NA	.459	.411	.741	**.407**
bicycle	4	122	NA	.317	.324	.337	**.295**	NA	.366	.478	.457	**.328**
ships	4	34	NA	**.174**	.181	1.927	.186	NA	.197	.203	4.156	**.172**
smoking	2	36	NA	.127	.128	.163	**.091**	NA	.136	.250	.322	**.103**

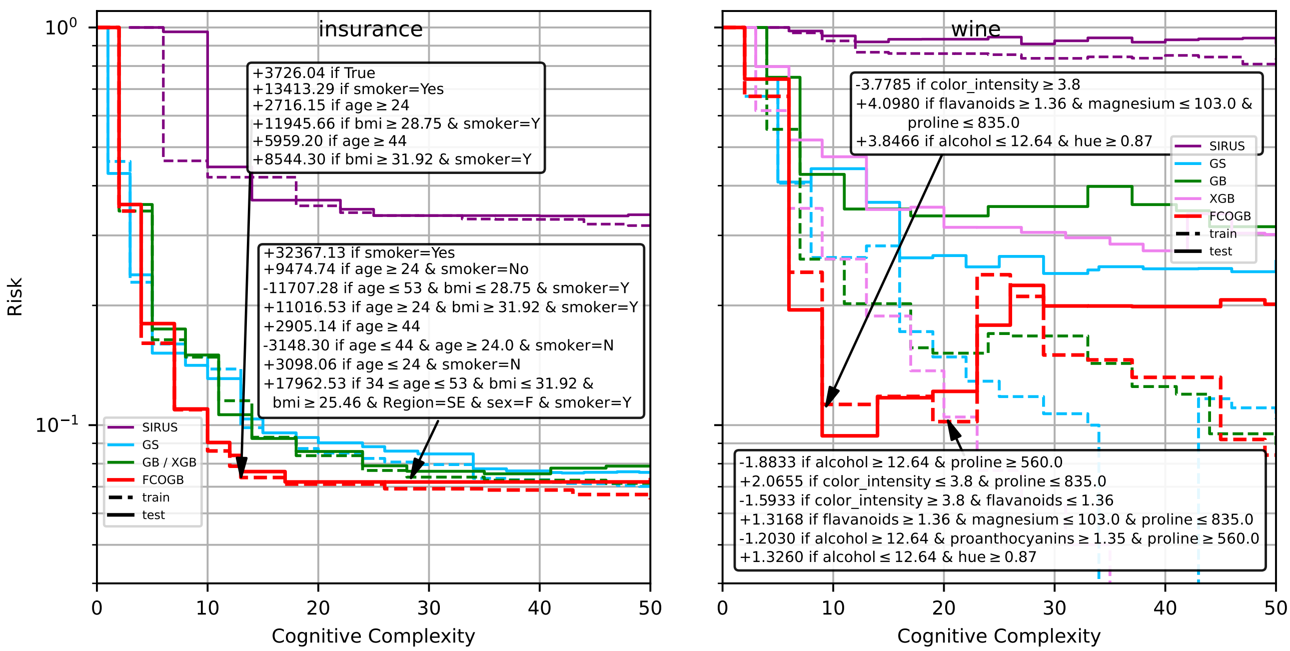

Figure 1: Risk/complexity curves for proposed orthogonalization approach (red) previous rule boosting variants (other colours) and for dataset `insurance` and `wine`. The two highlighted corners correspond to rule ensembles with roughly equivalent training risk but substantially reduced cognitive complexity for the proposed algorithm.