Orthogonal Gradient Boosting for Interpretable Additive Rule Ensembles Supplementary Information

Anonymous Author(s)

ACM Reference Format:
Anonymous Author(s). 2023. Orthogonal Gradient Boosting for Interpretable Additive Rule Ensembles Supplementary Information. In *Proceedings of (KDD 2023).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 PROOF OF THEOREM 4.3

The condition of Theorem 4.3 states that:

Let $\mathbf{Q} \in \mathbb{R}^{n\times(t-1)}$ be the selected query matrix, $\mathbf{g}$ the corresponding gradient vector after a full weight correction, and $\mathbf{q}^*$ be a maximizer of the **orthogonal gradient boosting objective** function defined by

$$\mathrm{obj}_{\mathrm{ogb}}(q) = \frac{|\mathbf{g}^T\mathbf{q}|}{\|\mathbf{q}_\perp\| + \epsilon} \tag{1}$$

where $\mathbf{q}_\perp$ is the projection of q onto the orthogonal complement of range $\mathbf{Q}$.

1.1 Property a

Proposition 1.1. *For $\epsilon \to 0$, $[\mathbf{q}_1, \ldots, \mathbf{q}_{t-1}, \mathbf{q}^*]$ is the best approximation to $[\mathbf{q}_1, \ldots, \mathbf{q}_{t-1}, \mathbf{g}]$.*

Proof. If $\epsilon \to 0$, then $\mathrm{obj}_{\mathrm{ogb}}(q) \to \dfrac{|g^T q|}{\|q_\perp\|}$. If $\mathbf{q}^*$ is a maximizer of $\mathrm{obj}_{\mathrm{ogb}}$, then as shown in Lemma 4.1, $\mathbf{q}^*$ minimises the minimum distance from all

$$\mathbf{f} \in \mathrm{span}\{\mathbf{q}_1, \cdots, \mathbf{q}_{t-1}, \mathbf{g}\}$$

to the subspace of

$$\mathrm{span}\{\mathbf{q}_1, \cdots, \mathbf{q}_{t-1}, \mathbf{q}^*\}.$$

Therefore, the subspace spanned by $[\mathbf{q}_1, \cdots, \mathbf{q}_{t-1}, \mathbf{q}^*]$ is the best approximation to the subspace spanned by $[\mathbf{q}_1, \cdots, \mathbf{q}_{t-1}, \mathbf{g}]$. □

1.2 Property b

Proposition 1.2. *For $\epsilon \to \infty$, $\mathbf{q}^*$ is also a maximizer of $\mathrm{obj}_{\mathrm{gs}}$ and any maximizer of $\mathrm{obj}_{\mathrm{gs}}$ is also a maximizer of $\mathrm{obj}_{\mathrm{ogb}}$.*

Proof. Let q_1 and q_2 be any two queries and denote by $\mathrm{obj}^{(\epsilon)}_{\mathrm{ogb}}(q)$ the $\mathrm{obj}_{\mathrm{ogb}}$-value of q for a specific ϵ. Then

$$\begin{aligned}
&\lim_{\epsilon\to\infty} \epsilon \left(\mathrm{obj}^{(\epsilon)}_{\mathrm{ogb}}(q_1) - \mathrm{obj}^{(\epsilon)}_{\mathrm{ogb}}(q_2)\right) \\
=& \lim_{\epsilon\to\infty} \epsilon \left(\frac{|g^T\mathbf{q}_1|}{\|\mathbf{q}_1^\perp\| + \epsilon} - \frac{|g^T\mathbf{q}_2|}{\|\mathbf{q}_2^\perp\| + \epsilon}\right) \\
=& \lim_{\epsilon\to\infty} \left(\frac{|g^T\mathbf{q}_1|}{\|\mathbf{q}_1^\perp\|/\epsilon + 1} - \frac{|g^T\mathbf{q}_2|}{\|\mathbf{q}_2^\perp\|/\epsilon + 1}\right) \\
=& |g^T\mathbf{q}_1| - |g^T\mathbf{q}_2| \\
=& \mathrm{obj}_{\mathrm{gs}}(q_1) - \mathrm{obj}_{\mathrm{gs}}(q_2)
\end{aligned}$$

Thus for large enough ϵ, the signs of $\mathrm{obj}^{(\epsilon)}_{\mathrm{ogb}}(q_1) - \mathrm{obj}^{(\epsilon)}_{\mathrm{ogb}}(q_2)$ and $\mathrm{obj}_{\mathrm{gs}}(q_1) - \mathrm{obj}_{\mathrm{gs}}(q_2)$ agree. Therefore, a query q is a $\mathrm{obj}_{\mathrm{gs}}$-maximizer, i.e., $\mathrm{obj}_{\mathrm{gs}}(q) \geq \mathrm{obj}_{\mathrm{gs}}(q')$ for all $q' \in Q$, if and only if q is a $\mathrm{obj}_{\mathrm{ogb}}$-maximizer, i.e., $\mathrm{obj}_{\mathrm{ogb}}(q) \geq \mathrm{obj}_{\mathrm{ogb}}(q')$ for all $q' \in Q$. □

1.3 Property c

Proposition 1.3. *For $\epsilon = 0$ and $\|\mathbf{q}_\perp\| > 0$, the ratio $(\mathrm{obj}_{\mathrm{ogb}}(q)/\mathrm{obj}_{\mathrm{gb}}(q))^2$ is equal to $1 + (\|\mathbf{q}_\|\|/\|\mathbf{q}_\perp\|)^2$.*

Proof. If $\epsilon = 0$ and $\|q_\perp\| > 0$, then

$$\begin{aligned}
\left(\frac{\mathrm{obj}_{\mathrm{ogb}}(q)}{\mathrm{obj}_{\mathrm{gb}}(q)}\right)^2 &= \frac{\dfrac{|\mathbf{g}^T\mathbf{q}|^2}{\|\mathbf{q}_\perp\|^2}}{\dfrac{|\mathbf{g}^T\mathbf{q}|^2}{\|\mathbf{q}\|^2}} = \frac{\|\mathbf{q}\|^2}{\|\mathbf{q}_\perp\|^2} \\
&= \frac{\|\mathbf{q}_\|\|^2 + \|\mathbf{q}_\perp\|^2}{\|\mathbf{q}_\perp\|^2} \\
&= 1 + \left(\frac{\|\mathbf{q}_\|\|}{\|\mathbf{q}_\perp\|}\right)^2
\end{aligned}$$

□

1.4 Property d

Proposition 1.4. *The objective value $\mathrm{obj}_{\mathrm{ogb}}(q)$ is upper bounded by $\|\mathbf{g}\|/(1+\epsilon)$.*

Proof. If we divide the numerator and denominator of $\mathrm{obj}_{\mathrm{ogb}}(\mathbf{q})$ with $\|\mathbf{q}_b ot\|$, then we can get

$$\begin{aligned}
\mathrm{obj}_{\mathrm{ogb}}(\mathbf{q}) &= \frac{|\mathbf{g}^T\mathbf{q}|}{\|q_\perp\| + \epsilon} \\
&= \frac{\dfrac{|\mathbf{g}^T\mathbf{q}_\perp|}{\|\mathbf{q}_\perp\|}}{1 + \dfrac{\epsilon}{\|\mathbf{q}_\perp\|}}
\end{aligned}$$

KDD 2023, August 6-10, 2023, Long Beach, CA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

because $\frac{|\mathbf{g}^T\mathbf{q}|}{\|\mathbf{q}_\perp\|} = \frac{\|\mathbf{g}\|\|\mathbf{q}_\perp\|| \cos\langle \mathbf{g}, \mathbf{q}_\perp\rangle |}{\|\mathbf{q}_\perp\|} = \|\mathbf{g}\|| \cos\langle \mathbf{g}, \mathbf{q}_\perp\rangle |$ is always less than or equal to $\|\mathbf{g}\|$,

$$\mathrm{obj}_{\mathrm{ogb}}(\mathbf{q}) \leq \frac{\|\mathbf{g}\|}{1 + \frac{\epsilon}{\|\mathbf{q}_\perp\|}}$$

as $\|\mathbf{q}_\perp\|$ is upper bounded by the number of data points n,

$$\mathrm{obj}_{\mathrm{ogb}}(\mathbf{q}) \leq \frac{\|\mathbf{g}\|}{1 + \frac{\epsilon}{n}}.$$

Let $\epsilon' = \frac{\epsilon}{n}$, then

$$\mathrm{obj}_{\mathrm{ogb}}(\mathbf{q}) \leq \frac{\|\mathbf{g}\|}{1 + \epsilon'}$$

□

2 GREEDY APPROXIMATION TO BOUNDARY FUNCTION

Solving the prefix optimization problem for a query described in Section 4.3 results in an approximation to the optimal objective function value for the best objective value which can be obtained by adding the query into the model.

We test 2000 instances for different gradient values to see the difference between the greedy approach, the pre-sorting greedy approach, and the actual optimal objective values.

Figure 1 shows the number of instances of ratios between the best objective values and the true optimal objective value. It shows that more than 80% of the best objective function values of prefixes have the same value with the optimal objective value, and more than 95% of the instances have an objective value more than 95% of the optimal objective. The smallest ratio between the result of greedy approach and the optimal solution is about 75%.

3 COMPARISON OF TESTING RISKS

Table 1 compares the testing empirical risks of the same benchmark datasets in Section 5. Table 1 has the same format with Table **??** except that it does not contain the comparison of running times.

As discussed in Section 5, the test performance of these algorithms follows similar trends to the training performance.

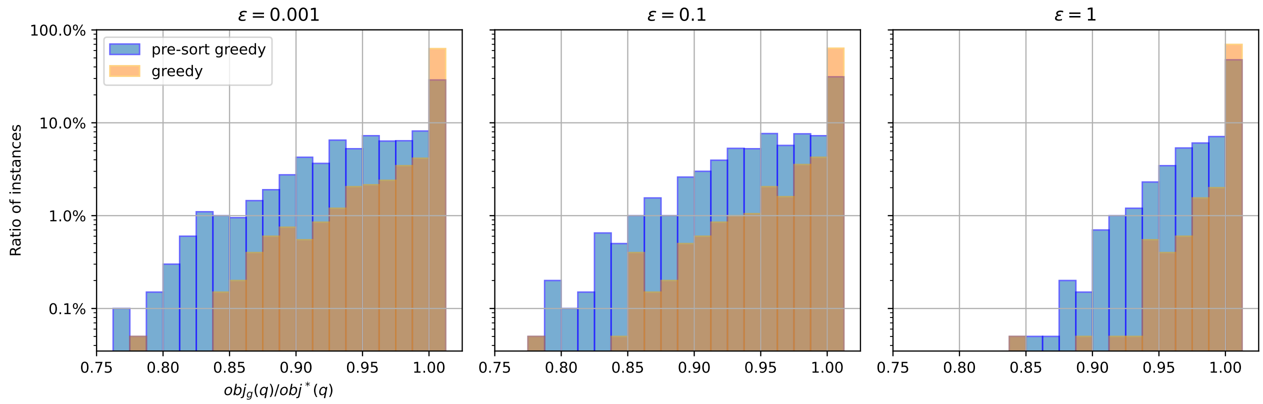

Figure 1: The number of instances of ratios between the best objective values obtained from the greedy search and the true optimal objective value.

Table 1: Comparison of Test Risks of Gradient Sum(S), Gradient boosting (G), XGBoost (X) and FCOGB (O) for benchmark datasets of classification (upper), regression (middle) and Poisson regression problems (lower).

DATASET	FEAT	ROW	$\tilde{\bar{R}}_{\mathrm{O}}$	FCOGB vs. GS					FCOGB vs. GB					FCOGB vs XGBoost				
				$\Delta^{\mathrm{bc}}_{\mathrm{SO}}$		$\bar{\Delta}_{\mathrm{SO}}$	$\Delta^{\mathrm{wc}}_{\mathrm{SO}}$		$\Delta^{\mathrm{bc}}_{\mathrm{GO}}$		$\bar{\Delta}_{\mathrm{GO}}$	$\Delta^{\mathrm{wc}}_{\mathrm{GO}}$		$\Delta^{\mathrm{bc}}_{\mathrm{XO}}$		$\bar{\Delta}_{\mathrm{XO}}$	$\Delta^{\mathrm{wc}}_{\mathrm{XO}}$	
TITANIC	7	1043	.712	.074	(17.6)	.035	.000	(2.4)	.147	(2.4)	.025	-.025	(28.2)	.147	(2.4)	.022	-.015	(4.4)
TIC-TAC-TOE	27	958	.751	.174	(23.6)	.101	-.058	(1.4)	.111	(6.2)	.060	.000	(2.4)	.089	(16.8)	.030	-.013	(28.4)
IRIS	4	150	.552	.141	(5.8)	-.089	-.180	(7.2)	.149	(2.4)	-.083	-.294	(5)	.149	(2.4)	-.058	-.147	(25.8)
BREAST	30	569	.352	.024	(7.8)	-.055	-.229	(7)	.009	(7.8)	-.011	-.026	(26)	.100	(7.8)	.031	-.006	(20.2)
WINE	13	178	.368	.314	(2.4)	.020	-.270	(5.8)	.409	(6)	.135	.003	(4)	.433	(6)	.090	-.043	(27)
IBM HR	32	1470	.217	.019	(6.2)	.003	-.005	(8.2)	.034	(15.4)	.004	.000	(1.4)	.065	(4)	.012	.004	(1.4)
TELCO CHURN	18	7043	.688	.058	(2.4)	.005	-.159	(1.4)	.037	(23.4)	.019	-.032	(8.2)	.017	(9.4)	.006	-.009	(12.4)
GENDER	20	3168	.999	.003	(2.4)	.001	.002	(3.2)	.000	(4.2)	.000	.000	(2.4)	.003	(11.4)	.001	.000	(2.4)
BANKNOTE	4	1372	.355	.142	(19.6)	.055	-.075	(8.2)	.133	(19.6)	.024	-.079	(9.4)	.120	(19.6)	.049	-.049	(9.4)
LIVER	6	345	.999	-.012	(3.8)	-.093	-.195	(29.8)	.057	(3.8)	-.024	-.067	(15.4)	.057	(3.8)	-.066	-.164	(29.4)
MAGIC	10	19020	.710	.056	(8.2)	.018	-.037	(4.2)	.018	(15.4)	.007	.000	(5.2)	.017	(18.4)	.007	-.003	(25.4)
ADULT	11	30162	.619	.146	(2.4)	.007	-.191	(1.4)	.059	(10.4)	.018	.000	(2.4)	.058	(4.2)	.011	.004	(20.4)
DIGITS5	64	3915	.381	.030	(4.2)	.014	-.008	(3.2)	.009	(4.2)	-.031	-.058	(19.4)	.070	(4.2)	-.004	-.034	(19.4)
INSURANCE	6	1338	.163	.104	(7.2)	.017	-.567	(1.4)	.172	(4.2)	.018	-.011	(5.2)	.172	(4.2)	.020	-.011	(5.2)
FRIEDMAN1	10	2000	.083	.013	(4)	-.002	-.012	(3.2)	.025	(4)	.006	.000	(1.4)	.025	(4)	.005	.000	(1.4)
FRIEDMAN2	4	10000	.149	.165	(3.2)	-.021	-.612	(1.4)	.084	(10.4)	.019	-.068	(5.2)	.080	(8.2)	.016	-.068	(5.2)
FRIEDMAN3	4	5000	.060	.009	(4.6)	.003	-.012	(4)	.021	(4.6)	.002	.000	(1.4)	.021	(4.6)	.002	.000	(1.4)
WAGE	5	1379	.419	.017	(6.6)	-.021	-.048	(4.6)	.065	(6.6)	.000	-.027	(18.2)	.065	(6.6)	-.002	-.014	(21.4)
DEMOGRAPHICS	13	6876	.229	.011	(5.2)	.003	.000	(1.4)	.007	(3.2)	.002	.000	(1.4)	.007	(3.2)	.002	.000	(1.4)
GDP	1	35	.038	.003	(3.2)	.000	-.001	(5.2)	.003	(3.2)	.000	-.001	(5.2)	.003	(3.2)	.000	-.001	(5.2)
USED CARS	4	1770	.198	.178	(3.2)	-.019	-.549	(1.4)	.113	(14)	.055	.000	(3.2)	.089	(8.8)	.035	.000	(3.2)
DIABETES	10	442	.169	.026	(4.4)	-.004	-.033	(3.4)	.058	(4.4)	-.002	-.011	(29.8)	.058	(4.4)	.002	-.008	(29.4)
BOSTON	13	506	.097	.019	(4.4)	-.001	-.025	(3.8)	.044	(4.4)	.006	-.011	(8.6)	.044	(4.4)	.006	-.011	(9.2)
WORLD HAPPINESS	8	315	.051	.010	(5.2)	-.003	-.012	(3.8)	.013	(5.2)	.002	-.010	(4.4)	.023	(5.2)	.002	-.001	(23.6)
LIFE EXPECTANCY	21	1649	.041	.003	(4.2)	.000	-.001	(8.2)	.007	(4.2)	.001	.000	(20.4)	.007	(4.2)	.001	-.001	(26.4)
MOBILE PRICES	20	2000	.168	.168	(2.4)	-.008	-.648	(1.4)	.058	(3.2)	.002	-.004	(8.2)	.058	(3.2)	.004	.000	(4.2)
SUICIDE RATE	5	27820	.534	.081	(2.4)	.016	-.333	(1.4)	.018	(13.4)	.008	.000	(2.4)	.024	(11.4)	.009	.000	(2.4)
VIDEO GAMES	6	16327	.723	.000	(4.2)	.000	.000	(3.2)	.000	(10.4)	.000	.000	(3.2)	.000	(10.4)	.000	.000	(3.2)
RED WINE	11	1599	.048	.003	(4.2)	.000	-.002	(3.2)	.004	(4.2)	.001	.000	(1.4)	.004	(4.2)	.001	.000	(7.4)
COVID VIC	4	85	.185	.068	(3)	.030	-.062	(10)	.103	(10.6)	-.005	-.692	(2.8)	3.448	(3)	.418	.002	(25.8)
COVID	2	225	.515	.071	(6.8)	-.037	-.399	(2)	.279	(3.8)	.025	-.003	(27)	25.66	(6.8)	5.388	.290	(26.8)
BICYCLE	4	122	.505	.213	(7.8)	-.021	-.333	(3.4)	.054	(4.6)	-.035	-.337	(4)	.011	(7.8)	-.025	-.061	(22.2)
SHIPS	4	34	.338	.076	(3.2)	-.089	-.498	(2.4)	.160	(11)	.015	-.066	(23.2)	993.4	(3.2)	181.3	.605	(29.6)
SMOKING	2	36	.216	-.017	(16.6)	-.046	-.189	(6.6)	.075	(10.2)	.029	-.043	(4.2)	.950	(2.4)	.124	.035	(14.2)