Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250380

XGBOrdinal: An XGBoost Extension for Ordinal Data

Fabian KAHL[a,1], Iris KAHL[b], and Stephan M. JONAS[a]
[a] *Institute for Digital Medicine, University Hospital Bonn, Bonn*
[b] *Department of Mathematics and Technology, University of Applied Sciences Koblenz, Remagen*
ORCiD ID: Fabian Kahl https://orcid.org/0009-0003-2741-2611, Stephan M. Jonas https://orcid.org/0000-0002-3687-6165

Abstract. We propose XGBOrdinal, an extension of XGBoost designed for ordinal classification problems commonly found in fields like medicine, where outcomes are often represented as scores, scales, stages, or grades. The proposed approach builds on the theoretical method introduced by Frank and Hall (2001) to transform an ordinal classification problem into a series of binary classification problems. Evaluated on multiple datasets, XGBOrdinal outperformed XGBClassifier and XGBRegressor, as well as existing ordinal methods. The implementation is fully compatible with GridSearchCV and RandomizedSearchCV, making it a scalable and efficient solution for handling ordinal data in machine learning pipelines. The used code is available open source (https://github.com/digital-medicine/XGBOrdinal).

Keywords. Ordinal Regression, Gradient Boost, Ordinal Classifier, Decision Tree, Scale, Score, XGBoost

1. Introduction

Ordinal data, where classes follow a natural order but lack equidistance, are common in areas such as medicine. For example, the Glasgow Coma Scale (GCS) [1] is used to assess a patient's level of consciousness after a head injury. Such data have a natural order but lack equidistant classes. In machine learning (ML), theoretical approaches for this so-called ordinal classification have been proposed [2,3,4,5], but the number of available implementations remains limited. Notable approaches include ordinal regression [6,7] and LGBMOrdinal [8], which builds upon LightGBM but lacks published research.

Statistical supervised ML methods, such as XGBoost [9], have been designed primarily for regression and classification tasks. However, a lot of these methods do not support ordinal classification, presenting a challenge for practitioners. Neural networks can handle ordinal data but lack explainability, which can be significant drawbacks in many practical applications, like medicine. To address this in practice, one can transform the ordinal classification problem into a regression task by imposing fixed distances between classes. Alternatively, one can treat the problem as a standard classification task, thereby ignoring the ordinal nature of the data. Another strategy is to use specialized

[1] Corresponding Author: Fabian Kahl; E-mail: fabian.kahl@ukbonn.de.

implementations tailored to ordinal data [7,8], though these are often limited by availability and implementation issues.

In response to these challenges, this paper proposes an ordinal version of XGBoost. By integrating the theoretical approach outlined by Frank and Hall (2001), this method aims to incorporate ordinal information directly into the learning process, enhancing the ability to capture and utilize the inherent order in the data [2]. To the best of our knowledge, no prior implementation of this method has been developed for XGBoost. Our implementation furthermore provides evaluation error and feature importance aggregated over all sub-models per epoch, and compatibility with tools such as sklearn.model selection.GridSearchCV [10] and RandomizedSearchCV [10].

2. Methods

2.1. Cumulative Link Ordinal Classification

Frank and Hall (2001) proposed a method to address ordinal classification by transforming the target feature into multiple binary classification problems. Instead of predicting the k ordinal classes directly, their approach involves creating $k-1$ binary classifiers between adjacent classes. For example, with four ordinal classes C_1 through C_4, the method would generate three binary classifiers: one to distinguish C_1 from C_2 through C_4; another to differentiate C_1 and C_2 from C_3 and C_4; and a third classifier to separate C_1 through C_3 from C_4 [2] (see Figure 1).

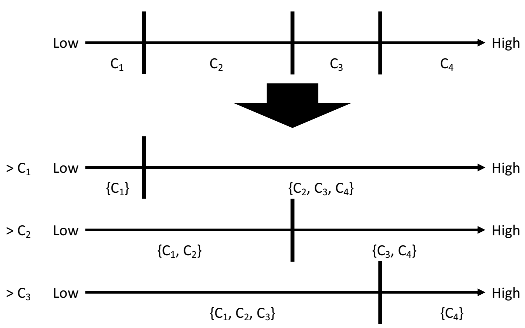

Figure 1. Transformation of four ordinal classes C_1 through C_4 into three binary classification problems.

During prediction, an element a with an unknown ordinal class is evaluated by each of the $k-1$ classifiers. The outcome $L(Ci)$ for a for each of the k ordinal classes $C_{i \in \{1,\ldots,k\}}$ is computed using the probabilities of the up to two classifiers surrounding it[2] [2]:

$$L(a \in C_1) = 1 - P(a > C_1) \quad (1)$$

$$L(a \in C_i) = P(a > C_{i-1}) - P(a > C_i),\ 1 < i < k \quad (2)$$

$$L(a \in C_k) = P(a > C_{k-1}) \quad (3)$$

Please note that the Kolmogorov axioms are violated, as negative and nonnormalized values are possible since the sub-models are independent. Therefore, we replaced all

[2]Please note that the original publication by Frank and Hall (2001) presented a different formula for Equation 2. That was later corrected and can be accessed on ResearchGate https://www.researchgate.net/publication/226877154_A_Simple_Approach_to_Ordinal_Classification

negative values with zero and normalized the outcomes to ensure a probability distribution (Equation 4):

$$P(a \in C_i) = \frac{\max(0, L(a \in C_i))}{\sum_{j=1}^{k} \max(0, L(a \in C_j))} \quad \text{such that} \quad \sum_{i=1}^{k} P(a \in C_i) = 1 \tag{4}$$

2.2. XGBoost Integration

We implemented the theoretical approach using the binary XGBClassifier [9]. During initialization, users can specify the preferred aggregation method for the evaluation error and feature importance. All other parameters are identical to those of the standard binary XGBClassifier and are passed to all binary sub-models.
To predict the probabilities $P(C_i)$, we use the equations described in the previous subsection (see Equations 1, 2, 3, and 4). The final predicted ordinal class per sample is the one corresponding to the highest probability.

2.3. Data Analysis and Datasets

We conducted two experiments comparing XGBOrdinal with other models. During the experiments, all methods ran 100 epochs with a random train-test split. In the first experiment, XGBOrdinal was evaluated against XGBClassifier, ignoring the ordinal nature of the target variable, and XGBRegressor, assuming equal intervals between ordinal classes and rounding predictions to the nearest class. This comparison was carried out across five datasets: the Car Evaluation dataset to predict quality class [11], the Cleveland subset of the Heart Disease dataset to predict the diagnosis [12], the White Wine subset of the Wine Quality dataset to predict wine quality [13], and a self-generated subset of the MIMIC-III database to predict the Glasgow Coma Scale (GCS) [14] both as a score (3 – 15) and as severity classes (3–8: severe, 9–12: moderate, 13–15: minor). The second experiment compared XGBOrdinal with ordinal regression [7] and LGBMOrdinal [8] on the same datasets. The models were evaluated using the mean absolute error (MAE) with standard deviation (SD) and the mean squared error (MSE) with SD.

3. Results

3.1. XGBOrdinal versus XGBClassifier and XGBRegressor

In comparison with XGBClassifier and XGBRegressor, XGBOrdinal achieved the lowest MAEs ± SDs on every tested dataset and the lowest MSE ± SD for the Car Evaluation dataset and the Wine Quality dataset. For the Heart Disease dataset and both versions of the MIMIC-III database, XGBRegressor received the best MSEs ± SDs (see Table 1).

3.2. XGBOrdinal versus Other Ordinal Methods

In comparison with ordinal regression and LGBMOrdinal, XGBOrdinal received the best MAE ± SD and MSE ± SD on the Car Evaluation, Wine Quality, and both versions of the MIMIC-III database. Ordinal Regression received the best MAE and MSE ± SD on the Heart Disease dataset (see Table 2).

Table 1. MAE ± SD and MSE ± SD for XGBOrdinal, XGBClassifier, and XGBRegressor on five datasets. The lowest MAE and MSE for each dataset are highlighted in bold.

		XGBOrdinal	**XGBClassifier**	**XGBRegressor**
Car Evaluation	MAE	**0.006 ± 0.005**	0.010 ± 0.006	0.027 ± 0.010
	MSE	**0.006 ± 0.005**	0.011 ± 0.008	0.029 ± 0.011
Heart Disease	MAE	**0.629 ± 0.076**	0.683 ± 0.068	0.665 ± 0.080
	MSE	1.106 ± 0.210	1.288 ± 0.203	1.083 ± 0.188
Wine Quality	MAE	**0.369 ± 0.016**	0.371 ± 0.017	0.381 ± 0.018
	MSE	**0.453 ± 0.026**	0.459 ± 0.027	0.461 ± 0.026
MIMIC-III	MAE	**0.383 ± 0.007**	0.384 ± 0.007	0.394 ± 0.006
(minor – severe)	MSE	0.559 ± 0.012	0.577 ± 0.013	**0.472 ± 0.009**
MIMIC-III	MAE	**1.985 ± 0.028**	2.062 ± 0.029	2.166 ± 0.023
(3 – 15)	MSE	13.100 ± 0.308	14.671 ± 0.323	**9.491 ± 0.188**

Table 2. MAE ± SD and MSE ± SD for XGBOrdinal, ordinal regression, and LGBMOrdinal. The lowest MAE and MSE for each dataset are highlighted in bold.

		XGBOrdinal	**Ordinal Regression**	**LGBMOrdinal**
Car Evaluation	MAE	**0.006 ± 0.005**	0.083 ± 0.012	0.084 ± 0.011
	MSE	**0.006 ± 0.005**	0.086 ± 0.014	0.093 ± 0.017
Heart Disease	MAE	0.629 ± 0.076	**0.595 ± 0.059**	0.730 ± 0.080
	MSE	1.106 ± 0.210	**1.090 ± 0.172**	1.591 ± 0.289
Wine Quality	MAE	**0.369 ± 0.016**	0.533 ± 0.012	0.415 ± 0.013
	MSE	**0.453 ± 0.026**	0.653 ± 0.017	0.490 ± 0.022
MIMIC-III	MAE	**0.383 ± 0.007**	0.472 ± 0.004	0.400 ± 0.006
(minor – severe)	MSE	**0.559 ± 0.012**	0.774 ± 0.008	0.630 ± 0.012
MIMIC-III	MAE	**1.985 ± 0.028**	2.414 ± 0.007	3.315 ± 0.373
(3 – 15)	MSE	**13.100 ± 0.308**	19.138 ± 0.085	29.951 ± 4.470

4. Discussion

The XGBOrdinal models outperformed both XGBClassifier and XGBRegressor on all five datasets based on MAE and on two out of five datasets based on MSE. Additionally, they performed better than existing ordinal methods on four out of five datasets across both MAE and MSE evaluations.

Regression models are well-suited for datasets with equidistant target features. However, XGBOrdinal is designed for target features with unknown equidistance, making it particularly valuable for tasks where this equidistance is not fulfilled.

The relatively lower performance of XGBOrdinal in MSE compared to MAE can be attributed to XGBOrdinal's limitation in penalizing "more severe" classification errors not more heavily than "less severe" ones (i.e., class "low" instead of "high" is as severe as "medium" instead of "high"), a distinction that regression methods can capture. Currently, XGBOrdinal lacks GPU support, which could enhance training efficiency on larger datasets.

An additional observation: The underlying approach comes with an inherent class imbalance in the binary classifiers. Specifically, in the edge cases, the classifier for the lowest and highest classes is basically a one-versus-all classifier. Therefore, exploring strategies to account for class imbalance could further improve performance. Future work should also investigate how this affects a general class imbalance within a dataset.

5. Conclusions

We introduced XGBOrdinal, an ordinal classification extension for XGBoost, designed to capture and utilize the inherent order in ordinal data. XGBOrdinal mostly outperformed XGBClassifier, XGBRegressor, and existing ordinal classification methods on several datasets, demonstrating its effectiveness in reducing MAE and MSE. By extending a widely-used ML framework with an ordinal-specific technique, XGBOrdinal provides a practical, scalable solution for tasks involving ordered data.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, we used generative AI to proofread the text and eliminate typos and grammatical flaws. After that, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Teasdale G, Jennett B. Assessment of coma and impaired consciousness: a practical scale. The Lancet. 1974;304(7872):81-4.
[2] Frank E, Hall M. A simple approach to ordinal classification. In: Machine Learning: ECML 2001: 12th European Conference on Machine Learning Freiburg, Germany, September 5–7, 2001 Proceedings 12. Springer; 2001. p. 145-56.
[3] Chu W, Keerthi SS. New approaches to support vector ordinal regression. In: Proceedings of the 22nd international conference on Machine learning; 2005. p. 145-52.
[4] Cardoso JS, da Costa JFP, Cardoso MJ. Modelling ordinal relations with SVMs: An application to objective aesthetic evaluation of breast cancer conservative treatment. Neural Networks. 2005;18(56):808-17.
[5] Yu S, Yu K, Tresp V, Kriegel HP. Collaborative ordinal regression. In: Proceedings of the 23rd international conference on Machine learning; 2006. p. 1089-96.
[6] Winship C, Mare RD. Regression models with ordinal variables. American sociological review. 1984:512-25.
[7] Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. SciPy. 2010;7(1).
[8] OrdinalGBT;. Accessed: 2024-09-18. https://ordinalgbt.readthedocs.io/en/latest/.
[9] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016. p. 785-94.
[10] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825-30.
[11] Bohanec M, Rajkovic V. Knowledge acquisition and explanation for multi-attribute decision making. In: 8th intl workshop on expert systems and their applications. Avignon France; 1988. p. 59-78.
[12] Detrano R, Janosi A, Steinbrunn W, Pfisterer M, Schmid JJ, Sandhu S, et al. International application of a new probability algorithm for the diagnosis of coronary artery disease. The American journal of cardiology. 1989;64(5):304-10.
[13] Cortez P, Cerdeira A, Almeida F, Matos T, Reis J. Modeling wine preferences by data mining from physicochemical properties. Decision support systems. 2009;47(4):547-53.
[14] Johnson AE, Pollard TJ, Shen L, Lehman LwH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016;3(1):1-9.