DFORD: Directional Feedback based Online Ordinal Regression Learning

Extended Abstract

Naresh Manwani
IIIT Hyderabad
Hyderabad, India
naresh.manwani@iiit.ac.in

M. Elamparithy
IIIT Hyderabad
Hyderabad, India
elamparithy.m@research.iiit.ac.in

Tanish Taneja
IIIT Hyderabad
Hyderabad, India
tanish.taneja@research.iiit.ac.in

ABSTRACT

In this paper, we introduce directional feedback in the ordinal regression setting, in which the learner receives feedback on whether the predicted label lies to the left or right of the actual label. We propose an online algorithm for ordinal regression, DFORD, that uses directional feedback. It preserves threshold ordering in the expected sense and achieves the expected regret of $O(\log T)$. We present experimental results to show the efficiency of DFORD.

KEYWORDS

Ordinal Regression; Online Learning; Directional Feedback

ACM Reference Format:

Naresh Manwani, M. Elamparithy, and Tanish Taneja. 2026. DFORD: Directional Feedback based Online Ordinal Regression Learning: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/QKDH8961

1 INTRODUCTION

Ordinal Regression (OR) predicts labels of data points that belong to ordered categories [2–8, 10, 11, 15, 18]. Label availability and reliability are major challenges in OR, especially where many labels result from human feedback and preferences. In such situations, using directional feedback can reduce ambiguity in human feedback: instead of an exact label, the feedback is at or below a certain ordinal level. For example, consider the hospital anxiety and depression scale [19] where the scores (ranging between 0 to 21) obtained from a questionnaire are divided into an ordinal scale: Normal '(< 7)'; 'Borderline (8 – 10)' and 'Clinical depression (or anxiety) (11+).

This paper proposes a new learning approach for OR using directional feedback. The main contributions are as follows.

- We propose a new weak supervision setting for OR called **directional feedback**. We propose an online learning algorithm for OR using directional feedback, which we call **DFORD**. We use an exploration-exploitation scheme [1] to handle the label uncertainty in this paper.
- DFORD maintains the ordering of thresholds in the expected sense. It achieves expected regret of $O(\ln T)$.
- Experimental results to show the effectiveness of DFORD.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/QKDH8961

2 PRELIMINARIES

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be the instance space and $\mathcal{Y} = [K]$ be the label space. Modeling OR requires a function $f : \mathcal{X} \to \mathbb{R}$ and thresholds $\boldsymbol{\theta} = [\theta_1 \ \cdots \ \theta_{K-1}]^\top$. To maintain the class order, we must ensure $\theta_1 \leq \ldots \leq \theta_K$. We assume $\theta_K = \infty$. OR predicts the class label $h : \mathcal{X} \to \mathcal{Y}$ as $h(\mathbf{x}) = 1 + \sum_{k=1}^{K} \mathbb{I}[f(\mathbf{x}) > \theta_k] = \min_{i \in [K]}\{i : f(\mathbf{x}) - \theta_i \leq 0\}$, where $\mathbb{I}[A]$ is an indicator function which take value 1 if the event A is true and 0 otherwise. A linear OR model assumes $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$, ($\mathbf{w} \in \mathbb{R}^d$). In Kernel OR, f is represented as $f = \sum_{\mathbf{x}_i \in \mathcal{X}} \alpha_i k(\mathbf{x}, .)$ where $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive definite kernel [16]. $k(.,.)$ has reproducing property (i.e., $\langle f, k(\mathbf{x}, .)\rangle = f(\mathbf{x})$ [16]. The mean absolute loss (MAE) used in OR is described below.

$$L^A(f, \boldsymbol{\theta}, \mathbf{x}, y) = \sum_{i=1}^{y-1} \mathbb{I}[f(\mathbf{x}) < \theta_i] + \sum_{i=y}^{K} \mathbb{I}[f(\mathbf{x}) \geq \theta_i] \qquad (1)$$

A convex surrogate of this loss is $L^H(f, \boldsymbol{\theta}, \mathbf{x}, y) = \sum_{i=1}^{K} [-z_i(f(\mathbf{x}) - \theta_i)]_+$, where $[a]_+ = \max(0, a)$, H stands for the hinge and $z_i,\ i \in [K]$ are defined as $z_i = \mathbb{I}[i \in \{1, \ldots, y-1\}] - \mathbb{I}[i \in \{y, \ldots, K\}]$. The regularized loss for an example $\mathbf{x}$ is defined as follows.

$$L^H_{reg}(f, \boldsymbol{\theta}, \mathbf{x}, y) = \frac{\lambda(\|f\|^2 + \|\boldsymbol{\theta}\|^2)}{2} + \sum_{i=1}^{K} L^H(f, \boldsymbol{\theta}, \mathbf{x}_i, y_i) \qquad (2)$$

Here, the objective is to minimize L^H_{reg} using an online algorithm.

3 DIRECTIONAL FEEDBACK AND LABEL UNCERTAINTY

Let $\tilde{y}$ be the label predicted. In this setting, we only get to know $\tilde{y} < y$. If $\tilde{y} < y$, then $y \in \{\tilde{y} + 1, \ldots, K\}$. Similarly, if $\tilde{y} \geq y$, then $y \in \{1, \ldots, \hat{y}\}$. Thus, label uncertainty remains even after feedback.
Efficient Exploration of Labels Under Directional Feedback: Let the model output the label $\hat{y}$, then asking directional feedback for $\hat{y}$ seems the best choice. However, in this way, we do not explore other labels different from $\hat{y}$. Thus, we assign nonzero probabilities for exploring all labels. At each round t, we use a mixture of two probability distributions P_1^t and P_2^t as the label distribution. Thus,

$$P^t = (1 - \gamma)P_1^t + \gamma P_2^t, \qquad (3)$$

where $P_1^t(i) = \mathbb{I}[i = \hat{y}^t], i \in [K]$. Distribution P_2^t should assign maximum probability to $\hat{y}^t$ and symmetrically decreases on both sides of $\hat{y}^t$. We use $P_2^t(i) = (Z^t)^{-1}\left(1 + d^t_{\max} - |i - \hat{y}^t|\right),\ i \in [K]$, where $d^t_{\max} := \max\{\hat{y}^t,\ K - \hat{y}^t\}$. $Z^t = (2K+1)\hat{y}^t - (\hat{y}^t)^2 - 0.5K(K-1)$ if $2\hat{y}^t \geq K$ and $Z^t = 0.5K(K+1) - \hat{y}^t(\hat{y}^t - 1)$ if $2\hat{y}^t < K$.

Algorithm 1 DFORD-Linear

Input: Training Dataset $\mathcal{S}$, λ
Initialize Set $t = 2$, $\mathbf{w}^2 = \mathbf{0}$, $\theta_1^2 = \ldots = \theta_{K-1}^2 = 0$, $\theta_K^2 = \infty$
for $i \leftarrow 2$ to T **do**
  Get example $\mathbf{x}^t$
  Find $\hat{y}^t$ as $\hat{y}^t = \min_{i \in [K]} \{i : \mathbf{w}^t \cdot \mathbf{x}^t - \theta_i^t \leq 0\}$
  Define $P^t(r)$, $r = 1 \ldots K$ using eq.(3)
  Sample $\tilde{y}^t$ from distribution P^t. Predict $\tilde{y}^t$
  Observe the directional feedback $\mathbb{I}[\tilde{y}^t < y^t]$.
  Define $\tilde{z}_i^t = \frac{1}{P(i)}(2d^t - 1)\mathbb{I}[i = \tilde{y}^t],\ i \in [K]$
  Initialize $\tilde{\tau}_i^t = 0,\ i \in [K]$
  Define $\tilde{\tau}_i^t = \tilde{z}_i^t \mathbb{I}[\tilde{z}_i^t(f^t(\mathbf{x}^t) - \theta_i^t) \leq 0],\ i \in [K]$
  $\mathbf{w}^{t+1} = (1 - \frac{1}{t})\mathbf{w}^t + \eta_t \tilde{\tau}_{\tilde{y}^t}^t \mathbf{x}^t$
  $\theta_i^{t+1} = (1 - \frac{1}{t})\theta_i^t - \mathbb{I}[i = \tilde{y}^t]\eta_t \tilde{\tau}_{\tilde{y}^t}^t,\ i = 1 \ldots K$
end for
Output: $h(\mathbf{x}) = \min_{i \in [K]} \left\{i\ :\ \mathbf{w}^{T+1}.\mathbf{x} - \theta_i^{T+1} < 0\right\}$

Unbiased Estimators of Labels z_i^t: At round t, let the algorithm samples a label $\tilde{y}^t$ using $P^t(i)$. Algorithm predicts $\tilde{y}^t$ and receives feedback $d_t = \mathbb{I}[\tilde{y}^t < y^t]$. For all $i \in [K]$, we define unbiased estimator $\tilde{z}_i^t$ of z_i^t as follows.

$$\tilde{z}_i^t = \frac{(2d_t - 1)\mathbb{I}[i = \tilde{y}^t]}{P^t(i)} = \begin{cases} \frac{1}{P^t(\tilde{y}^t)}(2d_t - 1), & i = \tilde{y}^t \\ 0, & i \neq \tilde{y}^t \end{cases} \quad (4)$$

We can show that $\mathbb{E}_{P^t}[\tilde{z}_i^t] = z_i^t$, i.e., $\tilde{z}_i^t$ an unbiased estimator of z_i^t. We are also interested in quantity $\tau_i^t = z_i^t \mathbb{I}[z_i^t(\mathbf{w}^t.\mathbf{x}^t - \theta_i^t) < 0]$. We use $\tilde{\tau}_i^t = \tilde{z}_i^t \mathbb{I}[\tilde{z}_i^t(f^t(\mathbf{x}^t) - \theta_i^t) \leq 0]$ as an unbiased estimator of τ_i^t.

4 DFORD

DFORD-Linear: We initialize with $\mathbf{w}^2 = \mathbf{0}$ and $\boldsymbol{\theta}^2 = \mathbf{0}$. Let $\mathbf{w}^t, \boldsymbol{\theta}^t$ be the estimates of the parameters at the beginning of trial t. At t, let $\mathbf{x}^t$ be the example observed. We find $\hat{y}^t = \min_{i \in [K]}\{i : \mathbf{w}^t.\mathbf{x}^t - \theta_i^t \leq 0\}$. We sample $\tilde{y}^t$ using P^t and receive $d_t = \mathbb{I}[\tilde{y}^t < y^t]$. Using d_t, we define $(\tilde{z}_i^t, \tilde{\tau}_i^t),\ i \in [K]$. We update the parameters as:

$$\mathbf{w}^{t+1} = (1 - \eta_t\lambda)\mathbf{w}^t + \eta_t \mathbf{x}^t \tilde{\tau}_{\tilde{y}^t}^t$$
$$\theta_i^{t+1} = (1 - \eta_t\lambda)\theta_i^t - \mathbb{I}[i = \tilde{y}^t]\eta_t \tilde{\tau}_{\tilde{y}^t}^t;\ \ i = 1 \ldots K - 1.$$

where $\eta_t = \frac{1}{\lambda t}$. We repeat this process for T rounds. Complete details of DFORD-Linear are given in Algorithm 1. We now show that DFORD-Linear preserves the threshold orderings.

Lemma 4.1 (Order Preservation). *For $t \geq 2$, if $\mathbb{E}[\theta_{i+1}^t - \theta_i^t] \geq \frac{K\eta_t}{1-\eta_t\lambda},\ \forall i \in [K-1]$, then DFORD-Linear Algorithm ensures that $\mathbb{E}[\theta_{i+1}^{t+1} - \theta_i^{t+1}] \geq 0,\ \forall i \in [K-1]$.*

Theorem 4.2 (Regret Bound). *Let $\mathbf{x}^1, \ldots, \mathbf{x}^T$ be the sequence of examples presented to DFORD-Linear. Let $(\mathbf{u}^1), \ldots, (\mathbf{u}^{T+1})$ be the parameter vectors generated by the DFORD-Linear, where $\mathbf{u}^t = (\mathbf{w}^t, \boldsymbol{\theta}^t)$. Let $\|\mathbf{x}^t\| \leq R,\ \forall t \in [T]$. Then, for any $(\mathbf{w}, \boldsymbol{\theta})$, we have,*

$$\mathbb{E}\left[\sum_{t=1}^{T} L_{Reg}^H(\mathbf{u}^t, \mathbf{x}^t, y^t) - \sum_{t=1}^{T} L_{Reg}^H(\mathbf{u}, \mathbf{x}^t, y^t)\right] \leq \frac{16K^2(R^2+1)\ln K \ln T}{\lambda\gamma}.$$

See [13] for the proofs of Lemma 4.1 and Theorem 4.2.

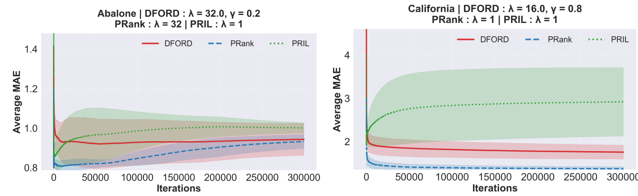

Figure 1: Average Mean Absolute Error (MAE) results.

DFORD-Kernel: In DFORD-kernel, threshold parameters $\boldsymbol{\theta}$ are updated similarly to the DFORD- Linear. f is updated as follows.

$$f^{t+1} = f^t - \eta_t[\lambda f^t - \tilde{\tau}_{\tilde{y}^t}^t k(\mathbf{x}^t, .)] = (1 - \eta_t\lambda)f^t + \eta_t \tilde{\tau}_{\tilde{y}^t}^t k(\mathbf{x}^t, .)$$

Lemma 4.3. *Let $\eta_t = \frac{1}{\lambda t}$, $f^1 = 0$. Then, the update rule for f^t given in Algorithm 3 reduces to $f^{t+1}(.) = \frac{1}{\lambda t}\sum_{i=1}^{t} \tilde{\tau}_{\tilde{y}^i}^i k(\mathbf{x}^i, .)$.*

See the proof in the full version [13]. To compute $f^{t+1}(\mathbf{x})$, we will need all $\mathbf{x}^i, i \in [t]$. We also have to do a linearly increasing number of kernel computations in each iteration. This makes DFORD-Kernel memory and compute-intensive. To tackle these issues, we use *truncation* [9], where we omit the sum from $i = 1$ up to a certain $i = t - \delta$ only (δ is the truncation parameter) while computing f^{t+1}.

$$f^{t+1}(.) = \frac{1}{\lambda t}\sum_{i=\max(1,t-\delta)}^{t} \tilde{\tau}_{\tilde{y}^i}^i k(\mathbf{x}^i, .)$$

Complete details of DFORD-Kernel are given in the full version [13]. DFORD-Kernel also maintains the orders of thresholds and achieves a regret bound of $O(\log T)$. See [13] for more details.

5 EXPERIMENTS

Datasets Used: We used Abalone [14] and California Housing (CH) [17] datasets. For Abalone, we used DFORD-Kernel with ($k(\mathbf{x}_1, \mathbf{x}_2) = (1 + \mathbf{x}_1^\top \mathbf{x}_2)^3$). For CH, we used DFORD-Linear. Target values for the Abalone dataset are discretized by mapping interval $[1, 7]$ to 1, $(7, 9]$ to 2, $(9, 12]$ to 3 and 12 onwards to 4. Target values in the CH dataset were divided into 10 equi-frequent categories [4].
Baselines: We compare DFORD with two baselines. (a) PRank [5], which is a full information based online OR approach, (b) PRIL [12] which is an interval label based online OR approach.
Experimental Results: We use average MAE loss (eq. (1)) over T iterations. We do average over 10 independent runs. Figure 1 shows that DFORD outperforms PRIL (partial label based approach). We see that DFORD performs comparably to PRank (a full information approach). These observations support the argument that directional feedback can lead to efficient ordinal regression models. Please refer [13] for detailed experimental results.

6 CONCLUSION

In this paper, we motivate the importance of directional feedback in OR tasks. We proposed an algorithm for online OR with directional feedback. We provide regret bounds for DFORD. DFORD performed comparably to the full information baseline and outperformed the weakly supervised baseline. Our findings suggest that directional feedback based learning is equally efficient for OR models.

REFERENCES

[1] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. 2002. The nonstochastic multiarmed bandit problem. *SIAM journal on computing* 32, 1 (2002), 48–77.

[2] Paul-Christian Bürkner and Matti Vuorre. 2019. Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science* 2, 1 (2019), 77–101.

[3] Wenzhi Cao, Vahid Mirjalili, and Sebastian Raschka. 2020. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters* 140 (2020), 325–331.

[4] Wei Chu and S. Sathiya Keerthi. 2005. New Approaches to Support Vector Ordinal Regression. In *Proceedings of the 22Nd International Conference on Machine Learning* (Bonn, Germany) *(ICML '05)*. 145–152.

[5] Koby Crammer and Yoram Singer. 2001. Pranking with Ranking. In *NIPS* (Vancouver, British Columbia, Canada). 641–647.

[6] Orla M Doyle, Eric Westman, Andre F Marquand, Patrizia Mecocci, Bruno Vellas, Magda Tsolaki, Iwona Kłoszewska, Hilkka Soininen, Simon Lovestone, Steve CR Williams, et al. 2014. Predicting progression of Alzheimer's disease using ordinal regression. *PloS one* 9, 8 (2014), e105542.

[7] Bhanu Garg and Naresh Manwani. 2020. Robust Deep Ordinal Regression under Label Noise. In *ACML*, Vol. 129. 782–796.

[8] R. Herbrich, T. Graepel, and K. Obermayer. 1999. Support vector learning for ordinal regression. In *ICANN*, Vol. 1. 97–102 vol.1.

[9] Jyrki Kivinen, Alexander J Smola, and Robert C Williamson. 2004. Online learning with kernels. *IEEE transactions on signal processing* 52, 8 (2004), 2165–2176.

[10] Yehuda Koren and Joseph Sill. 2013. Collaborative filtering on ordinal user feedback. In *IJCAI*.

[11] Ranjit Lall, Michael Campbell, S Walters, and Kevin Morgan. 2002. A review of ordinal regression models applied on Health related Quality of Life Assessments. *Statistical methods in medical research* 11 (03 2002), 49–67. https://doi.org/10.1191/0962280202sm271ra

[12] Naresh Manwani. 2019. PRIL: Perceptron Ranking Using Interval Labels. In *CODS-COMAD* (Kolkata, India). 78–85.

[13] Naresh Manwani, M Elamparithy, and Tanish Taneja. 2025. DFORD: Directional Feedback based Online Ordinal Regression Learning. arXiv:2512.19550 [cs.LG] https://arxiv.org/abs/2512.19550

[14] Kolby Nottingham Markelle Kelly, Rachel Longjohn. 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml

[15] Mariá Perez-Ortiź, Pedro Antonio Gutierrez, and César Hervás-Martínez. 2014. Projection-Based Ensemble Learning for Ordinal Regression. *IEEE Transactions on Cybernetics* 44, 5 (2014), 681–694.

[16] Bernhard Scholkopf and Alexander J. Smola. 2001. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, Cambridge, MA, USA.

[17] Luis Torgo. 2005. Regression datasets. *www. liaad. up. pt/˜ ltorgo/Regression/DataSets. html* (2005).

[18] Hong Wu, Hanqing Lu, and Songde Ma. 2003. A practical SVM-based algorithm for ordinal regression in image retrieval. In *ACM Multimedia* (Berkeley, CA, USA). 612–621.

[19] A. S. Zigmond and R. P. Snaith. 1983. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* 67, 6 (1983), 361–370.