Classificação ordinal baseada em um conjunto de classificadores

Luiz F. Bossa[*] **Maria E. Pinheiro**[†]
Universidade Federal de Santa Catarina – Departamento de Matemática
88040-900, Campus Trindade, Florianópolis, SC
{l.f.bossa, maria.eduarda.pinheiro}@posgrad.ufsc.br,

Luiz-Rafael Santos [‡]
Universidade Federal de Santa Catarina – Departamento de Matemática
89065-200, Campus Blumenau, Blumenau, SC
l.r.santos@ufsc.br.

RESUMO

Em muitos problemas de classificação em pesquisa e na indústria, os rótulos possuem uma ordenação natural. Os níveis de satisfação do cliente e as classificações *score* de crédito são exemplos típicos. Nesses casos, podemos falar em erros de previsão grandes e pequenos. No entanto, os métodos de classificação multiclasse padrão tratam esses rótulos ordinais como nominais, ignorando a relação de ordem inerente nos dados. Como consequência, o treinamento do modelo penaliza todas as classificações incorretas de forma igual, independentemente da sua gravidade. Para superar esta limitação, propomos um método de classificação multiclasse que tem explicitamente em conta a natureza ordinal da variável-objetivo, permitindo que o modelo aprenda não apenas a classificar corretamente, mas também a minimizar a gravidade dos seus erros, incorporando a "distância" entre os rótulos previstos e os verdadeiros no processo de treinamento.

Palavras-chave: *Classificação ordinal, Aprendizado de máquina, Conjunto de classificadores.*

Problemas de classificação ordinal são problemas de aprendizado de máquina nos quais o objetivo é classificar dados usando uma escala categórica, a qual possui uma ordem natural entre os rótulos [2]. Por exemplo, em um sistema de recomendação de filmes, as avaliações podem variar de 1 a 5 estrelas, classificação de risco de crédito podem ir do nível A até o nível H, conforme a capacidade de pagamento do cliente. Em ambos os casos, existe uma ordem natural entre as categorias, o que significa que uma classificação incorreta pode ter diferentes níveis de gravidade dependendo da distância entre o rótulo previsto e o verdadeiro [1].

Os algoritmos tradicionais de classificação multiclasse tratam os rótulos como categorias nominais, ignorando a ordem inerente entre eles. Isso pode levar a resultados subótimos, pois todos os erros de classificação são penalizados igualmente, independentemente da sua gravidade. Por exemplo, dar um rating de crédito A para um cliente D é um erro mais grave do que classificá-lo como C, mas os classificadores tradicionais não fazem essa distinção.

Um algoritmo de classificação ordinal deve ser capaz de reconhecer e explorar essa ordem natural entre os rótulos para melhorar a precisão das previsões. Além disso, deve ser capaz de minimizar a gravidade dos erros de classificação, penalizando mais severamente os erros que envolvem uma maior distância entre o rótulo previsto e o verdadeiro [3].

Considere a tarefa de classificação ordinal com K classes ordenadas $\{1, \ldots, K\} =: [1, K]$ e um conjunto de T classificadores. Seja $X = [x^1, \ldots, x^n] \subset \mathbb{R}^{p \times n}$ o conjunto de dados rotulados e $\bar{\imath} \in [1, K]$ a classificação do ponto x^i.

[*]Bolsista FAPESC (Chamada 62/2024)
[†]Pesquisadora Conhecimento Brasil do CNPq (314788/2025-5)
[‡]Bolsista de Produtividade em Pesquisa (CNPq 310571/2023-5, 407147/2023-3, FAPESC 2024TR002238)

Cada classificador $t \in [1, T]$ atribui cada ponto $x^i \in X$ a uma classe. Com base nisso, para cada ponto x^i e classe k, definimos o vetor $y_k(x^i) \in \{0, 1\}^T$. A t-ésima entrada de $y_k(x^i)$ é igual a 1 se o classificador t atribui x^i como k e 0 caso contrário.

O objetivo é encontrar uma combinação linear dos classificadores, ou seja, encontrar $\omega^* \in \mathbb{R}^T$ e $\xi^* \in \mathbb{R}^n$ que resolva o problema de otimização

$$\begin{aligned} \min_{\omega,\xi} \quad & \sum_{i=1}^{n} \xi_i \\ \text{s.a.} \quad & \omega^\top y_{\bar{\imath}}(x^i) \geq \omega^\top y_j(x^i) - \xi_i + \epsilon, && \forall j \in [1, K]\backslash\{\bar{\imath}\},\ i \in [1, n], \\ & \omega_t \in [0, 1] && t \in [1, T], \\ & \xi_i \geq 0, && i \in [1, n]. \end{aligned}$$

em que $\epsilon > 0$. Dado um novo ponto x com uma classe desconhecida, sua classificação é determinada por

$$\arg\max_{k \in [1,K]} \omega^\top y_k(x).$$

Note que para cada $i \in [1, n]$, ξ_i denota a penalidade por classificação errônea para o ponto x^i. Além disso, considere $\bar{\imath} > 2$, e os casos em que o ponto x^i é classificado errôneamente como $\bar{\imath} - 1$ ou $\bar{\imath} - 2$, com a mesma penalidade ξ_i. Ou seja,

$$\omega^\top y_{\bar{\imath}-1}(x^i) - \omega^\top y_{\bar{\imath}}(x^i) = \xi_i \text{ ou } \omega^\top y_{\bar{\imath}-2}(x^i) - \omega^\top y_{\bar{\imath}}(x^i) = \xi_i.$$

O modelo acima não leva em conta o fato de que classificar erroneamente x^i como $\bar{\imath} - 2$ é pior do que classificá-lo como $\bar{\imath} - 1$.

No modelo que propomos, a ideia é impor uma penalidade maior de classificação errônea à medida que a classificação prevista se desvia mais da classificação verdadeira:

$$\begin{aligned} \min_{\omega,\xi} \quad & \sum_{i=1}^{n} \xi_i \\ \text{s.a.} \quad & \omega^\top y_{\bar{\imath}}(x^i) \geq \omega^\top y_j(x^i) - \frac{\xi_i}{|j - \bar{\imath}|} + \epsilon, && \forall j \in [1, K]\backslash\{\bar{\imath}\},\ i \in [1, n], \\ & \omega_t \in [0, 1] && t \in [1, T], \\ & \xi_i \geq 0, && i \in [1, n]. \end{aligned}$$

Nese novo modelo, se tivermos os casos

$$\omega^\top y_{\bar{\imath}-1}(x^i) - \omega^\top y_{\bar{\imath}}(x^i) = A \text{ ou } \omega^\top y_{\bar{\imath}-2}(x^i) - \omega^\top y_{\bar{\imath}}(x^i) = A.$$

com $A > 0$, no primeiro caso obtemos $\xi_i = A$ e no segundo $\xi_i = 2A$. Ou seja, o modelo penaliza mais quando o ponto é atribuído a uma classe mais distante da verdadeira.

Referências

[1] Eibe Frank and Mark Hall. A simple approach to ordinal classification. In Luc De Raedt and Peter Flach, editors, *Machine Learning: ECML 2001*, volume 2167 of *Lecture Notes in Computer Science*, pages 145–156, Berlin, Heidelberg, 2001. Springer.

[2] Pedro Antonio Gutiérrez, María Pérez-Ortiz, Javier Sánchez-Monedero, Francisco Fernández-Navarro, and César Hervás-Martínez. Ordinal regression methods: Survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):127–146, 2016.

[3] R. Herbrich, T. Graepel, and K. Obermayer. Support vector learning for ordinal regression. In *9th International Conference on Artificial Neural Networks (ICANN 1999)*, volume 1, pages 97–102. IEE, 1999.