单变量线性损失下的顺序线性对数时间最优单峰拟合

Kaan Gokcesu、Hakan Gokcesu

arXiv:2304.02141v1 [cs.LG] 4 Apr 2023

摘要——本文研究在线性损失函数下，对单变量学习模型的分数输出进行最优单峰变换。我们证明，分数值与目标区域之间的最优映射是一个矩形函数。为了对观测样本进行最优矩形拟合，我们提出一种顺序方法，可随着每个新样本的到来更新估计。该方法每次迭代的时间复杂度为对数级，且效率最优。

I. 引言

在检测、估计、预测和学习问题中[1], [2]，智能体常常需要在面对随机性、噪声和不完整数据等显著不确定性时作出决策。在这种情况下，智能体会结合各种特征，以确定能够最大化某种效用的行动[3]。这类问题广泛存在于多个领域，包括决策理论 [4], 控制理论 [5], 博弈论 [6], [7], 优化 [8]–[10], 分布估计 [11]–[14], 异常检测 [15], 信号处理 [16], 预测[17], [18] 和多臂老虎机 [19]。这些学习模型的输出旨在区分数据模式，并提供具有实际用途的准确估计。尽管大多数学习方法都会生成分数输出，但许多应用需要准确的目标估计。因此，开发用于对学习模型输出进行后处理以生成准确估计的方法，已变得至关重要。

一个例子是分类器校准 [20]。最终应用需要的是概率估计，而不是得分。生成校准良好的概率在许多领域都至关重要，包括科学（例如，确定开展哪项实验）、医学（例如，选择采用哪种疗法）和商业（例如，决定进行哪项投资）。在学习问题中，获得校准良好的分类器不仅对决策至关重要，对于组合[21] 或比较 [22]–[24] 不同分类器也必不可少。在分类问题中，概率估计发挥着重要作用 [25]。获得校准良好的分类模型有两种方法。第一种方法是通过修改目标函数，构建内在校准良好的模型，但这可能会增加计算成本 [20]。另一方面，第二种方法是对判别式分类模型的输出进行后处理以实现校准，这种方法灵活且通用 [20]。然而，如果操作不当，它可能会在提高校准程度的同时降低区分能力。校准方法主要有两种用途：可用于将缺乏明显概率解释的判别式模型输出转换为类别后验概率，或改善校准不良模型的校准效果 [26], [27]。

文献中有多种将学习模型输出映射为准确估计值的技术。一种方法是采用参数化形式，例如 Platt 方法[26], ，该方法使用 Sigmoid 函数将分数输出映射为概率，然后使用模型信赖最小化算法最大化参数的似然 [28]。该方法最初是为支持向量机模型开发的，但也已应用于其他分类器，包括朴素贝叶斯 [27]。然而，对于某些数据集，Sigmoid 函数可能并不适合拟合朴素贝叶斯分数 [29]，[30]。尽管 Platt 方法计算效率高且能防止过拟合，但它具有局限性，因为无论分数输出中存在何种噪声或误差，都会直接使用这些输出 [31]。

近年来，为解决参数学习方法 [32]的问题，限制较少的非参数学习方法日益受到关注。其中一种方法是等频直方图分箱，也称为分位数分箱或简称分箱 [25], [33]。在这种方法中，分数输出会先排序，再划分到各个箱中，并对每个箱分别进行优化 [29]。对于小型或不平衡的数据集，任意设定箱的数量和边界可能导致估计不准确。为解决这些问题，人们提出了多种扩展和改进方法，包括 ACP [31]，，它通过为每个预测构建置信区间来划分箱；以及 BBQ[20]，，它会考察多个分箱模型及其基于贝叶斯评分函数 [34]的组合。ABB [35]则通过对所有分箱模型进行贝叶斯平均，进一步改进了这些方法。然而，这些方法都没有利用输入估计器的判别能力。

保序回归已成为机器学习 [36]–[45]中最常用的非参数拟合技术之一。它介于参数拟合与分箱之间，根据分数输出的排序对映射进行建模 [29]。保序回归是一种非参数回归，采用相邻对违背者算法（PAV）等算法 [38], [39]。一些旨在解决分箱问题的方法也采用了保序回归，例如 ENIR [42] ：该方法利用改进的池相邻违背者算法（mPAVA），在线性对数时间[40] 内求解近似保序回归问题的解路径，并结合这些模型所作出的预测。另一种基于保序回归的方法使用排序损失来预测准确概率 [43]。此外，还有一种扩展方法会组合多个二元分类器的输出，以获得校准概率 [44]。

本文证明，对于线性损失，最佳的单峰保序拟合是矩形映射。我们提出一种线性对数时间算法，以跟随领先者的方式找到最优映射。

II. 预备知识

A. 问题定义

我们首先正式定义在存在线性损失时构造最优单峰拟合的问题设定。设有 N 个样本，其对应的索引为 $n \in \{1, 2, \ldots, N\}$。

首先，我们有输入分数 $\{x_n\}_{n=1}^N$，满足

$$x_n \in \overline{\Re}, \qquad \forall n. \tag{1}$$

其次，我们有各自的线性损失 $\{z_n\}_{n=1}^N$，满足

$$z_n \in [Z_l, Z_u], \qquad \forall n, \tag{2}$$

对于某些 Z_l、$Z_u \in \Re$，以及

$$Z_l \leq 0 \leq Z_u. \tag{3}$$

我们的目标是将分数 $\{x_n\}_{n=1}^N$ 最优地映射到 $\{y_n\}_{n=1}^N$，使得

$$C(x_n) = y_n, \qquad \forall n, \tag{4}$$

其中，映射 $C(\cdot)$ 是单峰的，即

1) 我们将这些分数值映射为数值

$$q_n \triangleq C(x_n) \in [Q_0, Q_1]. \tag{5}$$

2) 映射 $C(\cdot)$ 是单峰的，即

$$q_n \geq q_{n'} \text{ if } x_{n^*} \geq x_n \geq x_{n'}, \tag{6}$$

$$q_n \leq q_{n'} \text{ if } x_{n^*} \leq x_n \leq x_{n'}, \tag{7}$$

对于某个 x_{n^*}

对于这一新的设定，我们有如下问题定义。

定义 1. 对于 $\{x_n\}_{n=1}^N$、$\{z_n\}_{n=1}^N$，线性损失最小化表示为

$$\arg\min_{C(\cdot)\in\Omega_u} \sum_{n=1}^{N} z_n C(x_n),$$

其中 Ω_u 是所有映射到区间 $[Q_0, Q_1]$ 的单变量单峰函数构成的类。

我们指出，定义 1 中的问题完全推广了许多分类问题。具体而言，

- 对于二分类，我们有

$$[Q_0, Q_1] = [0, 1]$$

- 如果分类问题不加权，我们有

$$z_n = 1 - 2y_n, \qquad [Z_0, Z_1] = [-1, 1]$$

- 如果存在类别权重，我们有

$$z_n = 1 - (\alpha + 1)y_n, \qquad [Z_0, Z_1] = [-\alpha, 1]$$

- 如果存在样本权重，我们有

$$z_n = \beta_n - \beta_n(\alpha + 1)y_n, \quad [Z_0, Z_1] = [-\alpha\beta, \beta],$$

where $\beta \geq \beta_n, \forall n$.

假设 1。我们有以下两对样本

$$z_1 = Z_u,\ x_1 = -\infty \qquad z_N = Z_u,\ x_N = -\infty,$$

以强制实现单峰性。否则，我们可以任意添加这些虚拟样本，这不会改变原问题的结果。

B. 线性损失下阈值化的最优性

这里，我们说明为什么对于一般的线性损失，分数 x_n 上的最优单峰变换 $C(\cdot)$是一个矩形函数。假设存在一个最优单调变换 $C^*(\cdot)$，它属于 Ω_u，并且使定义 1 最小化；其中，Ω_u是所有映射到区间$[Q_0, Q_1]$的单峰函数所构成的类。

命题 1。如果对所有 n，$z_n \geq 0$。则最优变换对所有 n 都将为 $q_n = Q_0$。如果并非所有 z_n 都是非负数；则存在某个 n_*，使得 $z_{n^*} < 0$，且最优变换将具有 $q_{n^*} = Q_1$，即单峰变换的峰值。

引理 1。如果 $C^*(\cdot)$ 是定义 1 的一个最优变换，其峰值为 n^*，那么

$$C^*(x_n) = q_n^* = \begin{cases} q_{n+1}^*, & z_n < 0, 1 < n < n^* \\ q_{n-1}^*, & z_n > 0, 1 < n < n^* \\ q_{n+1}^*, & z_n > 0, n^* < n < N \\ q_{n-1}^*, & z_n < 0, n^* < n < N \end{cases},$$

对于 $q_1^* = Q_0$、$q_{n^*}^* = Q_1$、$q_N^* = Q_0$，它们是来自假设 1 的虚拟样本及峰值。

证明。证明显然，因为 q_n^* 是以 n^* 为峰值的单峰序列，相应地改变 q_n^* 会降低定义 1。

□

因此，具有相同 q_n^* 的样本会形成若干组。设共有 I 组，其中第 $i \in \{1, \ldots, I\}$ 组包含样本 $n \in \{n_i + 1, \ldots, n_{i+1}\}$（$n_1 = 0$ 和 $n_{I+1} = N$）。令新分组的峰值为 i^*，且 $n^* \in \{n_{i^*} + 1, \ldots, n_{i^*+1}\}$。

引理 2. 若 $C^*(\cdot)$ 是定义 1 中具有峰值样本 n^* 且属于第 i^* 组的最优分类器，则

$$q_i^* = \begin{cases} q_{i+1}^*, & \sum_{n=n_i+1}^{n_{i+1}} z_n < 0, 1 < i < i^* \\ q_{i-1}^*, & \sum_{n=n_i+1}^{n_{i+1}} z_n > 0, 1 < i < i^* \\ q_{i+1}^*, & \sum_{n=n_i+1}^{n_{i+1}} z_n > 0, i^* < i < I \\ q_{i-1}^*, & \sum_{n=n_i+1}^{n_{i+1}} z_n < 0, i^* < i < I \end{cases} \tag{8}$$

对于 $q_1^* = Q_0$、$q_I^* = Q_0$ 和 $q_{i^*}^* = Q_1$。

证明. 证明过程与引理 1 及其证明相同。 □

在确定样本映射会聚为一组之后，我们得到以下结果。

定理 1. 存在一个最优分类器 $C^*(\cdot) \in \Omega_u$（其中 Ω_u 是所有映射到 Q_0、Q_1 的单峰函数所构成的类），它能使定义 1 达到最小值，并满足

$$C^*(\{x_n\}_{n=1}^{\tau_1}) \triangleq q_1^* = Q_0, \tag{9}$$

$$C^*(\{x_n\}_{n=\tau_1+1}^{\tau_2}) \triangleq q_2^* = Q_1, \tag{10}$$

$$C^*(\{x_n\}_{n=\tau_2+1}^{N}) \triangleq q_3^* = Q_0, \tag{11}$$

对于某些 τ_1、$\tau_2 \in \{1, \ldots, N-1\}$。

证明。证明可由引理 2 得出。如果 i^{th} 组的累积线性损失不为零，那么 q_i^*要么是 q_{i-1}^*，要么是 q_{i+1}^*。否则，如果其值为零，则无关紧要。因此，存在三个不同的组，分别对应映射 Q_0、Q_1、Q_0。

□

III. 最优矩形的离线算法

本节中，我们提出能够为定义 1 中的问题找到最优单峰变换（等价地，最优矩形）的算法。

A. 有用定义

定义 2. 设有一个由若干样本$\mathcal{X}=\{x_1,x_2,\ldots,x_N\}$及其对应的线性损失$\mathcal{Z}=\{z_1,z_2,\ldots,z_N\}$组成的有序集合，该集合表示为

$$\mathcal{A}=\{\mathcal{X},\mathcal{Z}\}.$$

给定集合 $\mathcal{A}$，其解概述如下。

定义 3. 对于定义 2 中给定的集合 $\mathcal{A}$，设最优阈值为 x_k、x_m（对应于 $\mathcal{A}$），使得三个相邻集合 $\mathcal{X}_0=\{x_1,\ldots,x_k\}$、$\mathcal{X}_1=\{x_{k+1},\ldots,x_m\}$和 $\mathcal{X}_2=\{x_{m+1},\ldots,x_N\}$的极小化映射分别为$q_1*$、$q_2^*$、$q_3^*$，即分别为 Q_0、Q_1、Q_0。我们将辅助集定义为

$$\mathcal{B}=\{X_1,X_2,L_0,L_1,L_2\},$$

其中，$X_1=x_k, X_2=x_m$是阈值样本，$L_0=\sum_{n=1}^{k}z_n$，$L_1=\sum_{n=k+1}^{m}z_n, L_2=\sum_{n=m+1}^{N}z_n$是对应的累积线性损失。

定义 3 中的集合 $\mathcal{B}$ 完全刻画了该解及其对应的累积损失。给定集合 $\mathcal{B}$，我们的单调变换为 $C(x\le X_1)=Q_0, C(X_1<x\le X_2)=Q_1, C(X_2<x)=Q_0$，由此得到的累积损失为 $L=Q_0L_0+Q_1L_1+Q_0L_2$。

B. 蛮力方法：$O(N^3)$ 中的批量优化

对于定义 2 中给定的 $\mathcal{A}$，要找到定义 3 中的 $\mathcal{B}$，蛮力方法是尝试所有可能的 (k,m) 对作为 $k\in\{1,\ldots,N-1\}$，$m\in\{k+1,\ldots,N\}$，即，

$$\mathcal{B}_{k,m}=\left\{x_k,x_m,\sum_{n=1}^{k}z_n,\sum_{n=k+1}^{m}z_n,\sum_{n=m+1}^{N}z_n\right\},\tag{12}$$

其累积损失为

$$L_{k,m}=Q_0\sum_{n=1}^{k}z_n+Q_1\sum_{n=k+1}^{m}z_n+Q_0\sum_{n=m+1}^{N}z_n\tag{13}$$

然后，我们可以选择损失最小的最优对 (k^*,m^*)，即

$$\mathcal{B}=\mathcal{B}_{k^*,m^*},\tag{14}$$

其中

$$(k^*,m^*)\triangleq\arg\min_{k,m}L_{k,m},\tag{15}$$

其耗时为 $O(N^3)$，因为对于每个 (k,m) 对（共 $O(N^2)$ 个），计算 $L_{k,\ m}$都需要 $O(N)$ 的时间。

C. 迭代方法：在 $O(N^2)$ 中进行批量优化

我们注意到，$L_{k,\ m}$ 可以通过迭代计算得到：

$$L_{k,m+1}=L_{k,m}+z_{m+1}(Q_1-Q_0),\tag{16}$$
$$L_{k,m-1}=L_{k,m}+z_m(Q_0-Q_1),\tag{17}$$
$$L_{k+1,m}=L_{k,m}+z_{k+1}(Q_0-Q_1),\tag{18}$$
$$L_{k-1,m}=L_{k,m}+z_k(Q_1-Q_0).\tag{19}$$

我们可以按如下方式计算每一对：

- Set $k=1$, $m=2$
- Calculate $L_{k,m}$ (in $O(N)$ time)
- WHILE $k+m<2N-1$
  - IF k is odd and $m<N$
    - $m'\leftarrow m+1$
  - ELSE IF k is odd and $m=N$
    - $k'\leftarrow k+1$
  - ELSE IF k is even and $m>k+1$
    - $m'\leftarrow m-1$
  - ELSE
    - $k'\leftarrow k+1$
    - $m'\leftarrow m+1$
  - Calculate $L_{k',m'}$ using $L_{k,m}$ (in $O(1)$ time)
  - $k\leftarrow k'$
  - $m\leftarrow m'$

上述过程通过固定 k 并遍历所有可能的 m，迭代计算所有 (k,m) 对。总之，这种方法在 $O(N^2)$ 中迭代计算所有 $L_{k,\ m}$。最小化也需要 $O(N^2)$。

D. 线性时间方法

接下来，我们提出一种线性时间算法。其步骤如下

- Let $L_{0,1}=z_1, L_{1,1}=0$, $k_1=1$.
- FOR $n\in\{2,\ldots,N\}$
  - $L_{1,n}\leftarrow L_{1,n-1}+z_n$
  - $L_{0,n}\leftarrow L_{0,n-1}$
  - $k_n\leftarrow k_{n-1}$
  - IF $L_{1,n}>0$
    - $L_{0,n}\leftarrow L_{0,n}+L_{1,n}$
    - $k_n\leftarrow n$

上述过程计算前 $0-1$ 个样本的最佳 n 阶跃拟合，适用于 $n\in\{1,\ldots,N\}$；耗时为 $O(N)$。同样，我们可以对最后 n 个样本运行相同的过程，以找到最佳 $1-0$ 阶跃拟合。设相应的损失为$M_{1,\ n}$,$M_{0,\ n}$，最后 n 个样本的阈值为 m_n。然后我们计算以下损失：

$$L_n=Q_0L_{0,n}+Q_1L_{1,n}+Q_1M_{1,N-n}+Q_0M_{0,N-n},\tag{20}$$

耗时为 $O(N)$。设最小化器为

$$n_*=\arg\min_{n}L_n.\tag{21}$$

因此，相应的两个最优阈值由数对 $(k_{n_*},m_{N-n^*}-1)$ 给出。

IV. 在线线性对数时间算法

尽管第三-D节中介绍的方法在批量优化中确定阈值非常高效，但其在线实现效果不佳。观察到一个新样本时，我们必须重新运行批处理算法来寻找最佳阈值，这会导致复杂度为 $O(N^2)$。在此，我们提出一种可在线实现的递归方法。由于每个样本仅排序就需要 $O(\log\ N)$ 时间，因此该方法对每个样本的复杂度为最优的 $O(\log(N))$。

为了求出最优阈值 k^*, m^* （或样本X_1, X_2），我们按如下方式实现递归算法：

1) 在递归的底层，我们有集合$\mathcal{A}^{1,n} = \{\mathcal{X}^{1,n}, \mathcal{Z}^{1,n}\}$ 以及对应的 $\mathcal{B}^{1,n}$，适用于$n \in \{1, \ldots, N\}$，其中 $\mathcal{X}^{1,n} = \{x_n\}$, $\mathcal{Z}^{1,n} = \{z_n\}$。

2) 从层级 $k = 1$ （初始阶段）开始，我们通过合并相邻集合，在每个层级 k 创建集合 $\mathcal{A}^{k+1,n}$ 和 $\mathcal{B}^{k+1,n}$ 。当我们将 $\mathcal{A}^{k,i}$ 与 $\mathcal{A}^{k,i+1}$ （以及它们对应的 $\mathcal{B}$ 集合）合并时，会创建新的集合 $\mathcal{A}^{k+1,j}$和 $\mathcal{B}^{k+1,j}$。如果给定的 $(\mathcal{A}^{k,i}, \mathcal{B}^{k,i})$ 对没有可供合并的相邻集合，那么它们会被移至上一层级，也就是说，$\mathcal{A}^{k+1,j} = \mathcal{A}^{k,i}$ 和 $\mathcal{B}^{k+1,j} = \mathcal{B}^{k,i}$。需要注意的是，$i$ 和 j 分别是在层级 k 和 $k+1$处不同的相对索引。

3) 相应的 $\mathcal{B}$ 集合会随之更新，以反映各自 $\mathcal{A}$ 集合对应的最优映射。除了辅助集 $\mathcal{B}$之外，我们还为每个 $\mathcal{A}$ 跟踪集合 $\mathcal{C}$ 和 $\mathcal{D}$，它们分别反映依据[37]得到的最佳非递减映射和非递增映射。与 $\mathcal{B}$ 集合类似， $\mathcal{C}$ 和 $\mathcal{D}$ 由下式给出

$$\mathcal{C} = \{XC, LC_-, LC_+\}, \quad \mathcal{D} = \{XD, LD_+, LD_-\},$$

其中， XC、 XD 分别为相应的阈值；LC_-, LD_- and LC_+, LD_+ 是 t h累积 lative l损失 o f 分别映射到 Q_0 和 Q_1的样本。

给定一组样本 x_n 及其各自的线性损失 z_n；我们的目标是确定截至目前观测样本的最优阈值。每当有新样本到达时，递归中的集合按如下方式更新：

1) 假设递归已对过去观测到的样本运行，并在每个层级合并集合。新样本到达后，只需更新必要的中间集合。
2) 如果新集合 $\mathcal{A}_m$ 在任意层级 k上不位于一对已合并集合之间，则安排将 $\mathcal{A}_m$ 作为一个整体上移至下一层级 $k+1$。3) 每当任意层级 k上出现一个新集合（中间集合） $\mathcal{A}_m$ ，位于一对已合并的左侧集合 $\mathcal{A}_l$ 和右侧集合 $\mathcal{A}_r$ 之间时，将 $\mathcal{A}_l$ 与 $\mathcal{A}_m$ 合并，而不是与 $\mathcal{A}_r$ 合并（左对齐偏置）；也就是说，将原先 $\mathcal{A}_l$ 与$\mathcal{A}_r$ 的组合更新为 $\mathcal{A}_l$ 与 $\mathcal{A}_m$的新组合。安排将 $\mathcal{A}_r$作为一个整体上移至层级 $k+1$。4) 如果某个待上移集合在某一层级 k有一个相邻且已上移的集合，则在 $k+1$处用它们的组合替换该已上移集合。否则，将待上移集合上移至 $k+1$。
5) 每当任意层级 k 上的中间集合发生更新（新增或更新一个组合）时，更新 $k+1$ 处后续的组合。

接下来，我们证明，对于每个新样本，我们的顺序算法具有对数级时间复杂度。

引理 3。递归的深度为 D，其中 $k \le D$ 对所有 k 均成立；该深度的阶为 $O(\log\ N)$，其中 N 是样本数量。

证明。由于顺序实现的结构，不会有两个相邻集合以自身的形式沿递归向上移动。因此，如果层级 k 上的集合数量为 N_k，那么层级 $k+1$ 上的集合数量至多为 $N_{k+1} \le \frac{2}{3}N_k + \frac{1}{3}$（假设集合按如下情形交替：一个集合向上移动，另外两个集合进行组合），由此得证。 □

引理 4。新样本到达时，我们遍历递归两次，以更新相关的中间集合。

证明。层级 $k+1$ 上的每次更新都与层级 k 上的一次更新相关，包括新的组合。每当有新样本到达时，底层至多发生两次更新；对于每次更新，我们至多分别遍历一次递归，由此可得该引理。 □

引理 5. 假设有两个集合 $\mathcal{A}^0$ 和 $\mathcal{A}^1$ ，它们的 $\mathcal{X}^0$和 $\mathcal{X}^1$ 互斥且相邻，其中 $\mathcal{X}^0$和 $\mathcal{X}^1$ 分别有序，且 $\mathcal{X}^0$的最后一个元素小于或等于 $\mathcal{X}^1$的第一个元素。设 $\mathcal{B}^0 = \{X_1^0, X_2^0, L_0^0, L_1^0, L_2^0\}$和 $\mathcal{B}^1 = \{X_1^1, X_2^1, L_0^1, L_1^1, L_2^1\}$ 分别为 $\mathcal{A}^0$ 和 $\mathcal{A}^1$的辅助集。设 $\mathcal{C}^0 = \{X^0, L_-^0, L_+^0\}$ 和 $\mathcal{D}^1 = \{X^1, L_+^1, L_-^1\}$ 分别为反映 $\mathcal{A}^0$ 的最佳非递减映射和 $\mathcal{A}^1$ 的最佳非递增映射的集合。设 $\mathcal{A} = \{\mathcal{X}^0 \cup \mathcal{X}^1, \mathcal{Z}^0 \cup \mathcal{Z}^1\}$ 为 $\mathcal{A}^0$ 和 $\mathcal{A}^1$的合并。则 $\mathcal{A}$ 的辅助集为:

$$\mathcal{B} = \begin{cases} \{X_1^0, X_2^0, L_0^0, L_1^0, \gamma + L_1^1 + L_2^1\}, & L_1^0 \le \min(L_1^1, \gamma) \\ \{X_1^1, X_2^1, L_0^0 + L_1^0 + \gamma, L_1^1, L_2^1\}, & L_1^1 \le \min(L_1^0, \gamma) \\ \{X^0, X^1, L_-^0, \theta, L_-^1\}, & \gamma \le \min(L_1^0, L_1^1) \end{cases},$$

where $\gamma = L_2^0 + L_0^1$, $\theta = L_+^0 + L_+^1$.

证明。合并两个辅助集 $\mathcal{B}^0$ 和 $\mathcal{B}^1$时，存在三种可能：两个阈值都来自$\mathcal{X}^0$，两个阈值都来自 $\mathcal{X}^1$，或者第一个阈值来自 $\mathcal{X}^0$，第二个阈值来自 $\mathcal{X}^1$。在这些可能性中，损失较小的方案即为最优方案。 □

定理 2. 顺序更新最优阈值的算法对每个样本具有 $O(\log\ N)$ 的平局复杂度。

证明。所述顺序更新方法能够以每个样本 $O(\log\ N)$ 的时间复杂度更新最优阈值。这是因为，根据引理 3 和 4，每个新样本所需的更新总次数为$O(\log\ N)$。此外，根据引理 5，每次更新耗时 $O(1)$。辅助集 $\mathcal{C}$ 和 $\mathcal{D}$ 也会更新，每个样本 [37]所需时间为 $O(\log\ N)$。因此，找到新阈值所需的总时间为 $O(\log\ N)$。 □

这一结论表明，即使是寻找新样本得分输出 x_n 顺序的最优算法，所需时间也至少为 $O(\log\ N)$。因此，由于顺序更新能在 $O(\log\ N)$ 时间内找到新的阈值，所以它是高效的。

V. 结论

总之，我们解决了寻找使线性损失模型最小化的最优单峰变换这一问题。我们证明，最优变换是矩形函数。我们提出了一种线性对数时间顺序递归算法，可根据数据流更新最优矩形拟合。该算法可用于多种场景，例如信号激活检测、目标区域估计和分类。

参考文献

[1] H. V. Poor，《信号检测与估计导论》。新泽西州：Springer，1994。[2] N. Cesa-Bianchi 和 G. Lugosi，《预测、学习与博弈》。剑桥大学出版社，2006。[3] S. Russell 和 P. Norvig，《人工智能：一种现代方法》，第 3 版。Prentice Hall，2010。[4] J. Moody 和 M. Saffell，“通过直接强化学习进行交易”，IEEE 神经网络汇刊，第 12 卷，第 4 期，第 875–889 页，2001 年 7 月。[5] H. R. Berenji 和 P. Khedkar，“通过强化学习和调节模糊逻辑控制器”，IEEE 神经网络汇刊，第 3 卷，第 5 期，第 724–740 页，1992 年 9 月。[6] R. Song、F. L. Lewis 和 Q. Wei，“用于求解非线性连续时间多人非零和博弈的离策略积分强化学习方法”，IEEE 神经网络与学习系统汇刊，第 PP 卷，第 99 期，第 1–10 页，2016。[7] H. S. Chang、J. Hu、M. C. Fu 和 S. I. Marcus，“用于双人零和马尔可夫博弈的自适应对抗多臂老虎机方法”，IEEE 自动控制汇刊，第 55 卷，第 2 期，第 463–468 页，2010 年 2 月。[8] K. Gokcesu 和 H. Gokcesu，“用于单变量函数高效全局优化的低遗憾二元采样方法”，arXiv 预印本 arXiv:2201.07164，2022。[9] M. Zinkevich，“在线凸规划与广义无穷小梯度上升”，载于第 20 届国际机器学习会议（ICML-03）论文集，2003，第 928–936 页。[10] E. Hazan、A. Agarwal 和 S. Kale，“在线凸优化的对数遗憾算法”，《机器学习》，第 69 卷，第 2–3 期，第 169–192 页，2007 年 12 月。[11] K. Gokcesu 和 S. S. Kozat，“利用指数分布族对非平稳源进行在线密度估计”，IEEE 神经网络与学习系统汇刊，第 29 卷，第 9 期，第 4473–4478 页，2018。[12] F. M. J. Willems，“针对二元独立分段同分布源的编码”，IEEE 信息论汇刊，第 42 卷，第 6 期，第 2210–2217 页，1996。[13] K. Gokcesu 和 S. S. Kozat，“非平稳环境中的极小极大最优密度估计在线异常检测”，IEEE 信号处理汇刊，第 66 卷，第 5 期，第 1213–1227 页，2018。[14] G. I. Shamir 和 N. Merhav，“面向分段平稳无记忆源的低复杂度顺序无损编码”，IEEE 信息论汇刊，第 45 卷，第 5 期，第 1498–1519 页，1999 年 7 月。[15] K. Gokcesu 、 M. M. Neyshabouri、H. Gokcesu 和 S. S. Kozat，“基于增量决策树的顺序离群值检测”，IEEE 信号处理汇刊，第 67 卷，第 4 期，第 993–1005 页，2019。[16] H. Ozkan、M. A. Donmez、S. Tunc 和 S. S. Kozat，“在线凸专家混合算法的确定性分析”，IEEE 神经网络与学习系统汇刊，第 26 卷，第 7 期，第 1575–1580 页，2015 年 7 月。[17] A. C. Singer 和 M. Feder，“通过模型阶数加权实现通用线性预测”，IEEE 信号处理汇刊，第 47 卷，第 10 期，第 2685–2699 页，1999 年 10 月。[18] K. Gokcesu 和 H. Gokcesu，“递归专家：动态环境中高效的学习系统最优混合”，arXiv 预印本 arXiv:2009.09249，2020。[19] S. Bubeck 和 N. Cesa-Bianchi，“随机与非随机多臂老虎机问题的遗憾分析”，《机器学习基础与趋势》，第 5 卷，第 1 期，第 1–122 页，2012。[20] M. P. Naeini、G. Cooper 和 M. Hauskrecht，“利用贝叶斯分箱获得校准良好的概率”，载于第二十九届 AAAI 人工智能会议，2015。[21] A. Bella、C. Ferri、J. Hern´andez-Orallo 和 M. J. Ram´ırez-Quintana，“校准对分类器组合的影响”，《应用智能》，第 38 卷，第 4 期，第 566–585 页，2013。

[22] H. Zhang 和 J. Su，“用于排序的朴素贝叶斯分类器”，载于欧洲机器学习会议。Springer，2004，第 501–512 页。[23] L. Jiang、H. Zhang 和 J. Su，“用于排序的 k 近邻朴素贝叶斯学习”，载于国际高级数据挖掘与应用会议。Springer，2005，第 175–185 页。[24] H. B. Hashemi、N. Yazdani、A. Shakery 和 M. P. Naeini，“集成模型在网页排序中的应用”，载于 2010 年第五届国际电信研讨会。IEEE，2010，第 726–731 页。[25] B. Zadrozny 和 C. Elkan，“从决策树和朴素贝叶斯分类器获得校准概率估计”，载于 ICML，第 1 卷。Citeseer，2001，第 609–616 页。[26] J. Platt 等，“支持向量机的概率输出及其与正则化似然方法的比较”，《大间隔分类器进展》，第 10 卷，第 3 期，第 61–74 页，1999。[27] A. Niculescu-Mizil 和 R. Caruana，“利用监督学习预测良好的概率”，载于第 22 届国际机器学习会议论文集，2005，第 625–632 页。[28] P. E. Gill、W. Murray 和 M. H. Wright，《实用优化》。SIAM，2019。[29] B. Zadrozny 和 C. Elkan，“将分类器分数转换为准确的多类别概率估计”，载于第八届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集，系列：KDD ’02。美国纽约州纽约：计算机协会，2002，第 694–699 页。[30] P. N. Bennett，“评估朴素贝叶斯后验估计的校准”，卡内基梅隆大学匹兹堡分校计算机科学学院，技术报告，2000。[31] X. Jiang、M. Osl、J. Kim 和 L. Ohno-Machado，“校准预测模型估计以支持个性化医疗”，《美国医学信息学会杂志》，第 19 卷，274，2012 年第 2 期，第 263 页。[32] K. Gokcesu 和 H. Gokcesu，“用于包络提取、峰值检测和聚类的时间序列非参数极值分析”，arXiv 预印本 arXiv:2109.02082，2021。[33] B. Zadrozny 和 C. Elkan，“当成本和概率均未知时的学习与决策”，载于第七届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集，2001，第 204–213 页。[34] D. Heckerman、D. Geiger 和 D. M. Chickering，“贝叶斯网络学习：知识与统计数据的结合”，《机器学习》，第 20 卷，第 3 期，第 197–243 页，1995。[35] M. P. Naeini、G. F. Cooper 和 M. Hauskrecht，“利用贝叶斯非参数方法进行二元分类器校准”，载于 2015 年 SIAM 国际数据挖掘会议论文集。SIAM，2015，第 208–216 页。[36] T. Robertson、F. T. Wright 和 R. L. Dykstra，《序约束统计推断》。奇切斯特：Wiley，1988，第 229 卷。[37] K. Gokcesu 和 H. Gokcesu，“使分类错误最小化的二元分类器最优高效顺序校准”，arXiv 预印本 arXiv:2108.08780，2021。[38] M. Ayer、H. D. Brunk、G. M. Ewing、W. T. Reid 和 E. Silverman，“基于不完整信息抽样的经验分布函数”，《数理统计年刊》，第 641–647 页，1955。[39] H. Brunk、R. E. Barlow、D. J. Bartholomew 和 J. M. Bremner，“序约束下的统计推断。（保序回归的理论与应用）”，密苏里大学哥伦比亚分校统计系，技术报告，1972。[40] R. J. Tibshirani、H. Hoefling 和 R. Tibshirani，“近似保序回归”，《计量经济学》，第 53 卷，第 1 期，第 54–61 页，2011。[41] K. Gokcesu 和 H. Gokcesu，“严格凸损失下采用保序回归进行校准的高效、随时可用算法”，arXiv 预印本 arXiv:2111.00468，2021。[42] M. P. Naeini 和 G. F. Cooper，“利用近似保序回归模型集成进行二元分类器校准”，载于 2016 年 IEEE 第 16 届国际数据挖掘会议（ICDM）。IEEE，2016，第 360–369 页。[43] A. K. Menon、X. J. Jiang、S. Vembu、C. Elkan 和 L. Ohno-Machado，“利用排序损失预测准确的概率”，载于……国际机器学习会议论文集。国际机器学习会议，第 2012 卷。美国国立卫生研究院公共获取，2012，第 703 页。[44] W. Zhong 和 J. T. Kwok，“多个分类器的准确概率校准”，载于第二十三届国际人工智能联合会议，2013。[45] K. Gokcesu 和 H. Gokcesu，“量化保序 L2 回归的对数线性时间顺序最优校准算法”，arXiv 预印本 arXiv:2206.00744，2022。