arXiv:2301.03318v1 [cs.LG] 9 Jan 2023

原创研究

二元分类的最优输入无关基线：Dutch Draw

Joris Pries[1] | Etienne van de Bijl[1] | Jan Klein[1] |SandjaiBhulai[2] | Rob van der Mei[1,2]

[1]随机学系，荷兰数学与计算机科学研究中心，阿姆斯特丹，北荷兰省，1098 XG，荷兰

[2]数学系，阿姆斯特丹自由大学，阿姆斯特丹，北荷兰省，1081 HV，荷兰

通信作者
Joris Pries，随机学系，荷兰数学与计算机科学研究中心，阿姆斯特丹，北荷兰省，1098 XG，荷兰 电子邮件：joris.pries@cwi.nl

资助信息无额外资助

在将任何二元分类模型投入实际应用之前，必须在适当的测试集上验证其性能。没有基线方法提供参照，就无法确定某个得分是“好”还是“坏”。本文旨在考察所有与特征值无关的基线方法，并确定哪个模型是“最佳”的，以及原因何在。通过确定哪些基线模型是最优的，可以简化评估过程中的一项关键选择决策。我们证明，在样本被随机打乱的前提下，对于所有位置不变的度量（与序列顺序无关），近期提出的 Dutch Draw 基线都是最佳的输入无关分类器（与特征值无关）。这意味着，在这些直观要求下，Dutch Draw 基线是最优基线，因此在实践中应当采用。

关键词

基线，二元分类，基准测试，评估，监督学习

1 | 引言

二元分类模型试图回答以下问题：应将该实例标记为 0 还是 1？这个问题看似简单，但二元分类有许多实际应用，从

预测确诊的新冠肺炎病例（Pirouz 等人，2020年）、检测恶意入侵（Li 等人，2018年），到判断跑步者是否疲劳（Buckley 等人，2017年）。每当针对实际应用开发分类模型时，都必须在测试集上验证其性能。然而，要客观评估所达到的性能，就需要一个基线。没有这一参照标准，只能从结果中得出片面的结论。准确率为 0.9 表示所有预测中有 90% 是正确的。但也有可能模型实际上什么都没有学到，如此高的准确率仅通过预测所有实例为 0 就可以达到。因此，为了客观评估性能，应将其与某种有意义的基准测试方法进行比较，最好是与最先进模型进行比较。

然而，许多最先进的方法都非常依赖具体问题。它们可能迅速变化，而且往往涉及许多经过精细调优的参数。因此，Van de Bijl 等人（2022）主张，在开发过程中有必要额外进行一项检查，即补充一个通用、简单且信息丰富的基线。这个基线应当检验新模型的表现是否确实优于简单模型。如果模型的表现还不如例如加权抛硬币，就应视为一个重大警示信号。二元分类模型可以利用样本的特征值信息，却被一个甚至不考虑这些值的模型击败。那么，这个模型实际上有没有学到有用的东西?

（Van de Bijl 等人，2022）基于 Dutch Draw 分类器提出了一种用于二元分类的理论方法。此类分类器会从所有样本中均匀随机（u.a.r.）抽取一个子集，并将其标记为 1，其余样本则标记为 0。通过优化抽取子集的大小，可获得最优期望性能，即 Dutch Draw 基线。对于最常用的性能度量，文中给出了闭式表达式（Van de Bijl 等人，2022）。

然而，设计基线方法的方式有无穷多种。我们只研究不考虑任何特征信息的预测模型，因为这样得到的基线更通用、更简单。我们称这类模型为输入无关模型。无论输入是什么，这类模型的预测方式都保持不变。任何新开发的模型至少都应优于这类模型的性能，因为输入无关模型无法利用数据中的模式来更准确地预测标签。不过，有时模型可能会碰巧针对标签的某种特定顺序，完美地预测出标签。样本的顺序不应影响模型的“最优性”。因此，我们引入置换最优性的概念。此外，样本的顺序不应改变性能度量的结果（即位置不变）。这并非严格的条件，因为大多数常用度量都具有这一性质。在这些限制条件下，我们证明，对于任何位置不变测度，Dutch Draw 基线在所有输入无关分类器中都是置换最优的。

总而言之，本文:

- 确定通用、信息丰富且简单的基线所需满足的自然要求;
- 证明 Dutch Draw 基线是在这些要求下的最优基线。

这些贡献改进了任何新二元分类方法的评估过程。

本文其余部分安排如下。首先，第2节讨论必要的预备知识和符号。接着，第3节确定通用、简单且信息丰富的基线所需满足的条件。此外，我们还正式定义了这些条件下最优性所包含的含义。第4节给出一种替代定义

荷兰抽签分类器的替代定义，这是主要证明所必需的。第5节证明 Dutch Draw 基线是最优的。最后，第6节总结一般性发现，并探讨未来可能的研究方向。

2 | 预备知识

接下来，我们将介绍一些概念和符号，为主要证明奠定基础。首先，我们将讨论二元分类的二元分类器（第2.1节）和性能度量（第2.2节）。然后，我们将在第2.3节考察排列的性质，这些性质将在主要结果的证明中发挥关键作用。

2.1 | 二元分类器

要为二元分类模型找到合适的基线，我们首先需要讨论二元分类器究竟是什么。为此，设 X 为特征空间（例如，考虑 $\grave{O}^d$）。通常，二元分类器被定义为函数 $h : X \to \{0, 1\}$，它将特征值映射为 0 或 1。然而，这种分类器一次只能对一个样本进行分类。相反，我们关注的是能够同时对多个样本进行分类的分类器：

$$h_M : \mathcal{X}^M \to \{0, 1\}^M,$$

其中 $M \in \hat{I}_{>0}$ 表示被分类的样本数量。这样，分类器就能够精确地预测 M 个样本中有 k 个为正类。请注意，只需对每个样本分别应用分类器，就可以将单样本分类器 h 扩展为同时对 M 个样本进行分类：

$$h_M : (x_1, \ldots, x_M) \mapsto (h(x_1), \ldots, h(x_M)) .$$

令 $H_M = \{ h_M : X^M \to \{0,1\}^M \}$ 表示能够同时对 M 个样本进行分类的所有二元分类器的集合 .

| 二元分类器示例

二元分类器的一个例子是抛硬币：通过抛硬币并判断硬币落在哪一面，对每个样本进行分类。设 $\theta \in [0, 1]$ 为硬币正面朝上的概率，$1 - \theta$ 为反面朝上的概率。假设正面和反面分别分类为 1 和 0，则得到：

$$h_{\text{coin}}^{\text{single}}(\cdot) := \begin{cases} 1 & \text{with probability } \theta, \\ 0 & \text{with probability } 1 - \theta. \end{cases}$$

通过反复抛硬币对 M 个样本进行分类，可以实现：

$$h_{\text{coin}} : (x_1, \ldots, x_M) \mapsto \left(h_{\text{coin}}^{\text{single}}(x_1), \ldots, h_{\text{coin}}^{\text{single}}(x_M) \right) .$$

2.2 | 二元分类的性能度量

要评估二元分类模型的效果，需要选择一种性能度量，用来量化预测标签与真实标签的一致程度。具体而言，每个索引为 i 的样本都有特征值 $\mathbf{x}_i \in \mathrm{X}$ 和相应的标签 $\mathrm{y}_i \in \{0, 1\}$。令 $\mathbf{X} := (\mathbf{x}_1 \dots \mathbf{x}_M) \in \mathrm{X}^M$ 表示 M 个样本的特征值集合。此外，令 $\mathbf{Y} = (\mathrm{y}_1, \dots, \mathrm{y}_M)$ 表示对应的标签。二元分类的性能度量定义为 $\mu : \{0, 1\}^M \times \{0, 1\}^M \to \mathrm{Ò}$，其中 μ 的第一个输入是分类器给出的预测，第二个输入是对应的标签。现在，分类器 h_M 的性能可以写作：$\mu(\mathrm{h}_M(\mathbf{X}), \mathbf{Y})$。

| 性能度量示例

二元分类的一种性能度量示例是准确率（μ_{acc}）。其定义为正确分类的样本总数除以样本总数。对于任意 $\mathrm{h}_M(\mathbf{X}) = (\hat{\mathrm{y}}_1, \dots, \hat{\mathrm{y}}_M) \in \{0, 1\}^M$ 和 $\mathbf{Y} = (\mathrm{y}_1, \dots, \mathrm{y}_M) \in \{0, 1\}^M$，有

$$\mu_{\mathrm{acc}}\left(h_M(\mathbf{X}), \mathbf{Y}\right) = \frac{\sum_{i=1}^{M} 1_{\{\hat{y}_i = y_i\}}}{M}.$$

| 未定义情形

对于 $\mathrm{h}_M(\mathbf{X})$ 和 $\mathbf{Y}$ 的某些特定组合，有些度量未定义。以真正率（Tharwat，2021）为例，它是正确预测的正例数除以实际正例总数。当不存在实际正例时，该度量定义不明确，因为它需要除以零。较不明显的是，阴性预测值（Tharwat，2021）在没有预测负例时未定义，因为其定义为正确预测的负例数除以预测负例总数。令 $x0 := 0$ 对所有 $\mathrm{x} \in \mathrm{Ò}$ 成立，可以解决许多未定义问题。然而，这可能会促使分类器始终预测那些会导致某个度量此前未定义的标签，以使该度量最小化。因此，从现在起，对于每个 $\hat{\mathbf{Y}}, \mathbf{Y} \in \{0,1\}^M$，当 $\mu(\hat{\mathbf{Y}}, \mathbf{Y})$ 未定义时，我们将 μ 重新定义为常数 $\mathrm{C}_{\mathrm{undef}}$。对于每个目标（最大化/最小化），我们分别进行区分。令

$$C_{\mathrm{undef}} := \begin{cases} \max_{\hat{\mathbf{Y}}, \mathbf{Y} \in \{0,1\}^M} \left\{ \mu(\hat{\mathbf{Y}}, \mathbf{Y}) \right\} & \text{if minimizing,} \\ \min_{\hat{\mathbf{Y}}, \mathbf{Y} \in \{0,1\}^M} \left\{ \mu(\hat{\mathbf{Y}}, \mathbf{Y}) \right\} & \text{if maximizing.} \end{cases}$$

因此，分类器预测一个此前未定义的案例总是不利的。通过这样定义 $\mathrm{C}_{\mathrm{undef}}$，我们无需在分析中排除这类分类器。

2.3 | 排列

为了确定哪种二元分类器被视为“最佳”，我们在第3.3.3节中定义了置换最优性，并借助置换来定义“最优性”。本节将考察主要证明中用到的置换的性质（见第5节）。置换是一个从某个集合到其自身的双射函数（Dixon 和 Mortimer，1996年）。这意味着置换并非重新排列后的列表，而是一个用于确定每个元素应被重新排列到何处的函数。

令 S_M den 记一个大小为 M 的集合的所有置换所构成的集合，也称为对称群。更形式化地说 ，

$$S_M := \{\pi : \{1,\dots,M\} \to \{1,\dots,M\} \text{ s.t. } \{\pi(i)\}_{i=1}^{M} = \{1,\dots,M\}\}.$$

| 对称群示例

使用柯西单行记法（柯西，1815年），三个元素的所有可能排列如下：

$$\begin{pmatrix}1 & 2 & 3\end{pmatrix},\quad \begin{pmatrix}1 & 3 & 2\end{pmatrix},\quad \begin{pmatrix}2 & 1 & 3\end{pmatrix},$$

$$\begin{pmatrix}2 & 3 & 1\end{pmatrix},\quad \begin{pmatrix}3 & 1 & 2\end{pmatrix},\quad \begin{pmatrix}3 & 2 & 1\end{pmatrix}.$$

置换 $\begin{pmatrix}2 & 3 & 1\end{pmatrix}$ 将第一个元素移到第二个位置，将第二个元素移到第三个位置，并将第三个元素移到第一个位置。

| 逐样本置换

为了将置换应用于矩阵，我们讨论逐样本置换。对于每个 M ×K 维矩阵 X =($\mathbf{x}_1 \dots \mathbf{x}_M$)，令 X_π 表示在 π 下的逐样本置换。因此，

$$X_\pi := \left(\mathbf{x}_{\pi(1)} \dots \mathbf{x}_{\pi(M)}\right),$$

其中 K $\in \hat{I}_{>0}$ 表示特征的数量。这意味着矩阵 X 按行重新排序。

| 置换的性质

接下来，我们概述 S_M 的一些性质，这些性质将在主要结果的证明中用到。S_M 是一个组，其群运算为函数的复合（记作 ∘），因此必须满足群公理（Dixon 和 Mortimer，1996年；Artin，2011年）。这意味着存在一个单位元 id $\in S_M$，使得对所有 $\pi \in S_M$：

$$\text{id} \circ \pi = \pi = \pi \circ \text{id}.$$

此外，对于每个 $\pi \in S_M$，都存在唯一的逆元 $\pi^{-1} \in S_M$，使得

$$\pi \circ \pi^{-1} = \text{id} = \pi^{-1} \circ \pi.$$

因此，对于每个置换，都存在一个逆置换，用于撤销该置换造成的顺序变化；这一点将在第5节中用到。由于每个逆元都是唯一的，且也属于 S_M，因此

$$\{\pi \in S_M\} = \{\pi^{-1} : \pi \in S_M\}, \tag{1}$$

这意味着 所有置换的集合与这些置换的所有逆置换的集合相同。因此，对

S_M 中的所有置换取期望，与对 S_M 中置换的所有逆置换取期望相同。这一点用于第5节主要结果的证明。

3 | 基本条件

为了证明最优 Dutch Draw 分类器能够产生“最优”基线，我们首先必须定义“最优性”。什么样的基线才被认为是最优的？要确定这一点，必须回答以下两个问题：（**1**）我们比较哪些方法，以及（**2**）我们如何比较它们？为此，我们定义输入无关分类器、位置不变度量和置换最优性的概念。

3.1 | 输入无关分类器

任何二元分类器都可以用作基线。然而，任何良好的标准化基线都应当具备通用、简单且信息丰富的特点（Van de Bijl 等人，2022 年）。因此，它需要适用于任何领域、训练速度快，并且显然仍然可以被超越。为此，我们研究所有不考虑任何特征值的模型，因为它们符合这三项要求。由于不考虑特征值，它们可以应用于任何领域。此外，它们不需要任何训练，因为它们无法学习特征值与相应标签之间的关系。这也使得它们显然仍然可以被超越，因为任何新开发的模型都应利用特征值中的信息来做出更好的预测。

如果对于所有特征空间 X_1^M、X_2^M以及所有特征值$\mathbf{X}_i \in \mathrm{X}_1^M$和 $\mathbf{X}_j \in \mathrm{X}_2^M$，$\mathrm{h}_M(\mathbf{X}_i)$ 与 $\mathrm{h}_M(\mathbf{X}_j)$ 的分布完全相同，则称二元分类器 $\mathrm{h}_M \in \mathrm{H}_M$ 为输入无关分类器。换句话说，

$$h_M(\mathbf{X}_i) \overset{d}{=} h_M(\mathbf{X}_j) \overset{d}{=}: h_M(\cdot),$$

这里用 $\mathrm{h}_M(\cdot)$ 这种记法，是为了直观地表明分类器 h_M 不依赖于输入。根据此定义，输入无关分类器不依赖于特征值，甚至也不依赖于特征域。令 $\mathrm{H}_M^{i.i.} = \{\mathrm{h}_M \in \mathrm{H}_M : \mathrm{h}_M \text{是输入无关的}\}$表示所有输入无关二元分类器的集合。使用相同性能度量进行优化的新开发模型，应始终优于输入无关模型，因为它能从特征值中获取信息。否则，该模型就未能利用这些额外信息来做出更好的预测。

| 输入无关分类器示例

抛硬币（见第2.1节）根据定义是输入无关的。特征值不会影响硬币的概率分布。因此，对于任意 $(\mathrm{x}_1, \ldots, \mathrm{x}_M) \in \mathrm{X}^M$，

$$\left(h_{\text{coin}}^{\text{single}}(x_1), \ldots, h_{\text{coin}}^{\text{single}}(x_M)\right) = \left(h_{\text{coin}}^{\text{single}}(\cdot), \ldots, h_{\text{coin}}^{\text{single}}(\cdot)\right).$$

3.2 | 位置不变测度

为了评估一种方法的性能，需要选择一种度量。合理地说，样本的顺序不应影响该度量的结果。如果一种度量具有这一性质，我们就称其为位置不变。更正式地说，若一个

度量 μ i s 是位置不变的，当且仅当对于每个置换 $\pi \in S_M$ 以及所有 $h_M(\mathbf{X})$、$\mathbf{Y} \in \{0,1\}^M$，都有 ：

$$\mu\left(h_M(\mathbf{X}), \mathbf{Y}\right) = \mu(h_M(\mathbf{X})_\pi, \mathbf{Y}_\pi). \tag{2}$$

这意味着，对耦合的预测标签和真实标签进行任何重新排序，都不会影响性能分数。

这并非严格限制，因为大多数度量都具有这一性质。例如，真正例（TP）、真负例（TN）、假正例（FP）和假负例（FN）的数量都是位置不变的。最常用的度量通常都是这四种度量的函数（Sokolovaand Lapalme，2009），因此它们也具有位置不变性。

| 非位置不变度量的示例

尽管如此，仍然可以定义不具备位置不变性的度量。例如，考虑以下度量

$$\lambda : \{0,1\}^M \times \{0,1\}^M \to \mathbb{R}, (a = (a_1, \ldots, a_M), b) \mapsto a_1,$$

它取决于预测的第一个位置，即

$$\lambda\left(\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}\right) = 0,$$
$$\lambda\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right) = 1.$$

3.3 | 定义最优性

要找到“最优”的基线，首先必须明确“最优性”意味着什么。

3.3.1 | 最优分类器

二元分类器的结果不一定是确定性的。因此，考虑到随机性，如果某个分类器在所有纳入考虑的二元分类器（即 $\tilde{H}_M \subseteq H_M$）中，使期望性能达到最小或最大，我们就认为它是最优的。优化是指最小化还是最大化，取决于问题的目标。因此：

$$h_M^{\min} \in \underset{h_M \in \tilde{\mathcal{H}}_M}{\arg\min} \left\{\mathbb{E}_{h_M(\mathbf{X})}\left[\mu(h_M(\mathbf{X}), \mathbf{Y})\right]\right\}, \tag{3}$$

$$h_M^{\max} \in \underset{h_M \in \tilde{\mathcal{H}}_M}{\arg\max} \left\{\mathbb{E}_{h_M(\mathbf{X})}\left[\mu(h_M(\mathbf{X}), \mathbf{Y})\right]\right\}. \tag{4}$$

例如，当目标是最大化准确率时，$h_M^{\max}$就是$\tilde{H}_M$中所有其他基线里的最优基线。请注意，可能存在多个不同的最优基线。

3.3.2 | 平凡最优解

然而，当我们考虑所有输入无关分类器（$\tilde{\mathrm{H}}_M = \mathrm{H}_M^{i.i.\cdot}$）时，这种“最优性”的定义会导出一个平凡最优解。考虑确定性分类器

$$\tilde{h}_M^{\max}(\cdot) := \hat{\mathbf{Y}}_{\max} \in \underset{\hat{\mathbf{Y}} \in \{0,1\}^M}{\arg\max}\, \mu(\hat{\mathbf{Y}}, \mathbf{Y}),$$

它始终预测一个使度量 μ 达到最大值的向量 $\hat{\mathbf{Y}}_{\max}$。注意，$\tilde{\mathrm{h}}_M^{\max}$显然是输入无关的（见第3节 3.1），因此 $\tilde{\mathrm{h}}_M^{\max} \in \mathrm{H}_M^{i.i.\cdot}$。此外，有

$$\begin{aligned}\max_{h_M \in \tilde{\mathcal{H}}_M} \left\{ \mathbb{E}_{h_M(\mathbf{X})} \left[\mu(h_M(\mathbf{X}), \mathbf{Y}) \right] \right\} &\leq \max_{\hat{\mathbf{Y}} \in \{0,1\}^M} \mu(\hat{\mathbf{Y}}, \mathbf{Y}) \\ &= \mathbb{E}_{\tilde{h}_M^{\max}(\cdot)} \left[\mu(\tilde{h}_M^{\max}(\cdot), \mathbf{Y}) \right].\end{aligned}$$

换言之，$\tilde{\mathrm{h}}_M^{\max}$的期望性能始终高于或等于任何其他分类器。因此，$\tilde{\mathrm{h}}_M^{\max}$被视为最优（见公式 (4)）。对于最小化的情况，同样成立

$$\tilde{h}_M^{\min}(\cdot) := \hat{\mathbf{Y}}_{\min} \in \underset{\hat{\mathbf{Y}} \in \{0,1\}^M}{\arg\min}\, \mu(\hat{\mathbf{Y}}, \mathbf{Y}).$$

本质上，输入无关分类器总能利用真实标签和性能度量作出完美预测。例如，考虑常用的性能度量：准确率。当预测 $\hat{\mathbf{Y}} = \mathbf{Y}$ 时，准确率达到最大值。因此，始终预测 $\mathbf{Y}$ 的分类器 $\tilde{\mathrm{h}}_M^{\max}$ 对于这些给定标签而言是最优的。这表明，应考虑扩展“最优性”的定义。

3.3.3 | 置换最优性

当我们考察所有确定性分类器时，这一最优性质（见方程 (3) 和 (4)）并没有太大启发意义，因为其中总有一个分类器能够做出完美预测。同样，坏掉的时钟一天也会有两次显示正确时间，但我们不应依靠它来确定时间。因此，我们引入一种称为置换最优性的新最优条件。

通常假设测试集是随机打乱的。因此，我们引入置换最优性的概念。我们不再只考虑特征值及其对应标签所给定的特定顺序下的最优性，而是考虑样本的所有置换。如果一个分类器在所考虑的所有二元分类器（$\tilde{\mathrm{H}}_M$）中，能够针对测试集的随机置换使期望性能达到最小值/最大值，则称其为置换最优分类器。因此，

$$h_M^{\min} \in \underset{h_M \in \tilde{\mathcal{H}}_M}{\arg\min} \left\{ \mathbb{E}_{\pi \sim \mathcal{U}(S_M)} \left[\mathbb{E}_{h_M(\mathbf{X}_\pi)} \left[\mu(h_M(\mathbf{X}_\pi), \mathbf{Y}_\pi) \right] \right] \right\}, \tag{5}$$

$$h_M^{\max} \in \underset{h_M \in \tilde{\mathcal{H}}_M}{\arg\max} \left\{ \mathbb{E}_{\pi \sim \mathcal{U}(S_M)} \left[\mathbb{E}_{h_M(\mathbf{X}_\pi)} \left[\mu(h_M(\mathbf{X}_\pi), \mathbf{Y}_\pi) \right] \right] \right\}. \tag{6}$$

4 | 荷兰抽签分类器

Van de Bijl 等人（2022）针对 $\theta \in [0,1]$ 定义了荷兰抽签分类器，如下所示：

$$\sigma_\theta(\mathbf{X}) := (1_E(i))_{i\in\{1,\dots M\}} \text{ with } E \subseteq \{1,\dots M\} \text{ drawn u.a.r. such that } |E| = \lfloor M\cdot\theta\rceil. \tag{7}$$

换言之，该分类器从所有样本中等概率随机抽取一个大小为 $\lfloor M\cdot\theta\rceil$ 的子集 E，并将其标记为 1，其余样本则标记为 0。在本节中，我们将介绍一种用于主要证明的替代定义，并证明所有 Dutch Draw 分类器都是输入无关的。

4.1 | 替代定义

为了简化主要结果的证明，我们不采用公式 (7) 中的定义，而是为 Dutch Draw 分类器引入一种替代定义。给定一个长度为 M 的二进制向量 $(y_1,\dots,y_M) \in \{0,1\}^M$，注意其中 1 的个数可以通过求和 $\sum_{i=1}^{M} y_i$ 计算得出。接下来，我们定义一些二进制向量集合（长度相同），其中的 1 的个数相同。对于所有 $j \in \{0,\dots,M\}$，定义

$$\mathcal{Y}_j := \left\{ \hat{\mathbf{Y}} = (y_1,\dots,y_M) \in \{0,1\}^M \text{ s.t. } \sum_{i=1}^{M} y_i = j \right\}. \tag{8}$$

换言之，$\mathcal{Y}_j$ 包含所有长度为 M 且恰有 j 个 1 和 M − j 个 0 的二进制向量。例如，当 $M = 4$ 时，有

$$\begin{aligned}
\mathcal{Y}_0 &= \{(0,0,0,0)\},\\
\mathcal{Y}_1 &= \{(0,0,0,1),(0,0,1,0),(0,1,0,0),(1,0,0,0)\},\\
\mathcal{Y}_2 &= \{(0,0,1,1),(0,1,0,1),(0,1,1,0),(1,0,0,1),(1,0,1,0),(1,1,0,0)\},\\
\mathcal{Y}_3 &= \{(0,1,1,1),(1,0,1,1),(1,1,0,1),(1,1,1,0)\},\\
\mathcal{Y}_4 &= \{(1,1,1,1)\}.
\end{aligned}$$

荷兰抽签分类器以均匀随机的方式从 M 个样本中选取 E 个，并将这些样本标记为 1，其余样本标记为 0。请注意，这等同于以均匀随机的方式从 $\mathcal{Y}_E$ 中选取一个向量。为简化记号，令 $\mathcal{U}(A)$ 表示有限集合 A 上的均匀分布。因此，当 $X \sim \mathcal{U}(A)$ 时，对于每个 $a \in A$，必有 $\mathbb{P}(X = a) = \frac{1}{|A|}$。现在，荷兰抽签分类器 σ_θ 可以改写为

$$\sigma_\theta(\mathbf{X}) := \hat{\mathbf{Y}} \text{ with } \hat{\mathbf{Y}} \sim \mathcal{U}\left(\mathcal{Y}_{\lfloor M\cdot\theta\rceil}\right). \tag{9}$$

换句话说，Dutch Draw 分类器 σ_θ 从所有具有 $\lfloor M\cdot\theta\rceil$ ones 的向量中，以均匀随机方式选择一个恰好具有 $\lfloor M\cdot\theta\rceil$ ones 的向量作为预测（$\mathbf{Y}_{\lfloor M\cdot\theta\rceil}$）。这种替代定义简化了主要结果的证明。

4.2 | 输入无关性

接下来，我们讨论为什么所有 Dutch Draw 分类器都是输入无关的（见第3节.1）。请注意，Dutch Draw 分类器 σ_θ 与特征值无关，因为它仅取决于 θ 和 M，见公式 (9)。换句话说，任何 Dutch Draw 分类器按定义都是输入无关的。因此，我们可以用 $\sigma_\theta(\cdot)$ 代替 $\sigma_\theta(\mathbf{X})$。总而言之，对于每个 $\theta \in [0,1]$，都有 $\sigma_\theta(\cdot) \in \mathrm{H}_M^{i.i.}$，其中 H 是所有输入无关二元分类器的集合。

4.3 | 最优 Dutch Draw 分类器

最优 Dutch Draw 分类器 $\sigma_{\theta_{\text{opt}}}$ 通过在所有允许的参数值中，对参数 θ 的期望性能进行最小化/最大化来确定 Θ（Van de Bijl 等人，2022）。请注意，某些度量对于特定预测未定义，因此 Θ 并不总是等于 $[0,1]$。例如，精确率这一度量（Tharwat，2021）定义为真正例数除以预测正例总数。因此，如果没有预测正例，该度量便未定义（除以零）。根据第 2.2 节调整每项度量后，所有未定义情形均得到解决，且 $\Theta = [0,1]$ 始终成立。

使用 Dutch Draw 分类器的另一种定义（见公式 (9)），我们得到：

$$\theta^*_{\text{min}} \in \underset{\theta\in[0,1]}{\arg\min} \left\{ \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(y_{\lfloor M\cdot\theta\rceil}\right)} \left[\mu(\hat{\mathbf{Y}}, \mathbf{Y}) \right] \right\}, \tag{10}$$

$$\theta^*_{\text{max}} \in \underset{\theta\in[0,1]}{\arg\max} \left\{ \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(y_{\lfloor M\cdot\theta\rceil}\right)} \left[\mu(\hat{\mathbf{Y}}, \mathbf{Y}) \right] \right\}. \tag{11}$$

根据目标，$\sigma_{\theta^*_{\text{min}}}$ 或 $\sigma_{\theta^*_{\text{max}}}$ 是一个最优 Dutch Draw 分类器。

5 | 定理与证明

在定义了输入无关性（第 3.1 节）、位置不变性（第 3.2 节）、置换最优性（第 3.3.3 节），并为荷兰抽签分类器引入一种替代表述之后，证明以下定理所需的所有要素都已具备。

定理 1（主要结果）对于任意位置不变测度 μ，最优 *Dutch Draw* 分类器 $\sigma_{\theta_{opt}}$ 在所有输入无关分类器 $(\mathcal{H}_M^{i.i.})$ 中是置换最优的。换言之：

$$\sigma_{\theta^*_{min}} \in \underset{h_M\in\mathcal{H}_M^{i.i.}}{\arg\min} \left\{ \mathbb{E}_{\pi\sim\mathcal{U}(S_M)} \left[\mathbb{E}_{h_M(\mathbf{X}_\pi)} \left[\mu(h_M(\mathbf{X}_\pi), \mathbf{Y}_\pi) \right] \right] \right\}, \tag{12}$$

$$\sigma_{\theta^*_{max}} \in \underset{h_M\in\mathcal{H}_M^{i.i.}}{\arg\max} \left\{ \mathbb{E}_{\pi\sim\mathcal{U}(S_M)} \left[\mathbb{E}_{h_M(\mathbf{X}_\pi)} \left[\mu(h_M(\mathbf{X}_\pi), \mathbf{Y}_\pi) \right] \right] \right\}. \tag{13}$$

This 意味着最优 Dutch Draw 分类器是最佳的通用、简单且信息丰富的基线 e.

证明 设 $\mathrm{h}_M \in \mathrm{H}_M^{i.i.}$ 是一个输入无关分类器，μ 是一个位置不变测度。若该分类器在所有输入无关分类器中，对测试集进行随机置换后的期望性能取最小值或最大值，则称其为置换最优（见方程（5）和（6））。

对于任意输入无关分类器 h_M，均有

$$\mathbb{E}_{h_M(\mathbf{X}_\pi)}\left[\mu(h_M(\mathbf{X}_\pi),\mathbf{Y}_\pi)\right] = \mathbb{E}_{h_M(\cdot)}\left[\mu(h_M(\cdot),\mathbf{Y}_\pi)\right]. \tag{14}$$

输入 $\mathbf{X}_\pi$ 与分类无关，因此可以省略。

总共有 2^M 种不同的可能预测，属于 $\{0,1\}^M$。将这些不同的向量记为 $\hat{}\mathbf{Y}_1,\ldots,\hat{}\mathbf{Y}_{2^M}$，使得 $\text{Đ}_{2^M\, i=1}\,\hat{}\mathbf{Y}_\mathbf{i} = \{0,1\}^M$。接下来，公式（14）中的期望可展开为：

$$\mathbb{E}_{h_M(\cdot)}\left[\mu(h_M(\cdot),\mathbf{Y}_\pi)\right] = \sum_{i=1}^{2^M}\mathbb{P}(h_M(\cdot)=\hat{\mathbf{Y}}_\mathbf{i})\cdot\mu(\hat{\mathbf{Y}}_\mathbf{i},\mathbf{Y}_\pi). \tag{15}$$

由于我们需要证明置换最优性，因此必须对所有置换下的公式（15）取期望。利用期望的线性性可得：

$$\begin{aligned}&\mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\sum_{i=1}^{2^M}\mathbb{P}(h_M(\cdot)=\hat{\mathbf{Y}}_\mathbf{i})\cdot\mu(\hat{\mathbf{Y}}_\mathbf{i},\mathbf{Y}_\pi)\right]\\ &\quad=\sum_{i=1}^{2^M}\mathbb{P}(h_M(\cdot)=\hat{\mathbf{Y}}_\mathbf{i})\cdot\mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mu(\hat{\mathbf{Y}}_\mathbf{i},\mathbf{Y}_\pi)\right].\end{aligned} \tag{16}$$

我们现在不再对求和取期望，而是对各项期望求和。

度量 μ 具有位置不变性，因此根据公式（2）可得

$$\mu(\hat{\mathbf{Y}}_\mathbf{i},\mathbf{Y}_\pi) = \mu((\hat{\mathbf{Y}}_\mathbf{i})_{\pi^{-1}},(\mathbf{Y}_\pi)_{\pi^{-1}}) = \mu((\hat{\mathbf{Y}}_\mathbf{i})_{\pi^{-1}},\mathbf{Y}). \tag{17}$$

应用置换不会改变位置不变测度 μ。在这种情况下，我们应用逆置换 π^{-1} 以取回 $\mathbf{Y}$。

根据公式（17），因此也有

$$\mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mu(\hat{\mathbf{Y}}_\mathbf{i},\mathbf{Y}_\pi)\right] = \mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mu((\hat{\mathbf{Y}}_\mathbf{i})_{\pi^{-1}},\mathbf{Y})\right]. \tag{18}$$

公式（1）表明，所有逆置换的集合与所有置换的集合相同。由于这些置换是均匀随机抽取的，对所有逆置换取期望与对所有置换取期望相同。当置换 π 是均匀随机抽取时，对所有 $s \in S_M$，均有 $\text{Đ}(\pi = s) = \text{Đ}(\pi = s^{-1}) = \frac{1}{|S_M|}$。因此，

$$\begin{aligned}\mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mu((\hat{\mathbf{Y}}_\mathbf{i})_{\pi^{-1}},\mathbf{Y})\right] &= \sum_{s\in S_M}\left(\mu((\hat{\mathbf{Y}}_\mathbf{i})_{s^{-1}},\mathbf{Y})\cdot\mathbb{P}(\pi=s)\right)\\ &= \sum_{s\in S_M}\left(\mu((\hat{\mathbf{Y}}_\mathbf{i})_{s^{-1}},\mathbf{Y})\cdot\mathbb{P}(\pi=s^{-1})\right)\\ &= \mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mu((\hat{\mathbf{Y}}_\mathbf{i})_{\pi},\mathbf{Y})\right].\end{aligned} \tag{19}$$

因此， π^{-1} 可以在方程 (18) 中替换为 π。

回顾一下，$\mathrm{Y}j$ 是所有长度为 M 且含有 j 个 1 的二进制向量的集合（见公式（8））。此外，注意，对 $\hat{}\mathbf{Y}_{\mathbf{i}} \in \mathrm{Y}j$ 应用均匀随机选取的置换 $\pi \in \mathrm{S}_M$，与将 $\hat{}\mathbf{Y} \in \mathrm{Y}j$ 作为结果均匀随机选取相同，因为对于每个 $\hat{}\mathbf{Y}_{\star} \in \mathrm{Y}j$，都有

$$\mathbb{P}\left((\hat{\mathbf{Y}}_{\mathbf{i}})_{\pi} = \hat{\mathbf{Y}}_{\star}\right) = \frac{1}{|\mathcal{Y}_j|} \text{ with } \pi \sim \mathcal{U}(S_M),$$

and

$$\mathbb{P}\left(\hat{\mathbf{Y}} = \hat{\mathbf{Y}}_{\star}\right) = \frac{1}{|\mathcal{Y}_j|} \text{ with } \hat{\mathbf{Y}} \sim \mathcal{U}(\mathcal{Y}_j).$$

因此，我们可以将对所有排列的期望 $\mathbb{E}_{\pi\sim\mathcal{U}(S_M)}[\cdot]$ 改写为对一个均匀随机抽取且具有相同数量 1 的向量的期望，方法如下：

$$\mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mu((\hat{\mathbf{Y}}_{\mathbf{i}})_{\pi}, \mathbf{Y})\right] = \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right):\hat{\mathbf{Y}}_{\mathbf{i}}\in\mathcal{Y}_j}\left[\mu(\hat{\mathbf{Y}}, \mathbf{Y})\right]. \tag{20}$$

结合方程 (18)、(19) 和 (20) 与方程 (16)，可得

$$\begin{aligned}&\sum_{i=1}^{2^M} \mathbb{P}(h_M(\cdot) = \hat{\mathbf{Y}}_{\mathbf{i}}) \cdot \mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mu(\hat{\mathbf{Y}}_{\mathbf{i}}, \mathbf{Y}_{\pi})\right] \\ &\quad = \sum_{i=1}^{2^M} \mathbb{P}(h_M(\cdot) = \hat{\mathbf{Y}}_{\mathbf{i}}) \cdot \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right):\hat{\mathbf{Y}}_{\mathbf{i}}\in\mathcal{Y}_j}\left[\mu(\hat{\mathbf{Y}}, \mathbf{Y})\right].\end{aligned}$$

现在我们已从方程中消去了所有排列。请注意，右侧的期望对每个 $\hat{}\mathbf{Y}_{\mathbf{i}} \in \mathrm{Y}j$ 都相同。换言之，对于两个 1 的数量相同的向量，其期望也相同。将 1 的数量相同的向量分组，可得

$$\begin{aligned}&\sum_{i=1}^{2^M} \mathbb{P}(h_M(\cdot) = \hat{\mathbf{Y}}_{\mathbf{i}}) \cdot \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right):\hat{\mathbf{Y}}_{\mathbf{i}}\in\mathcal{Y}_j}\left[\mu(\hat{\mathbf{Y}}, \mathbf{Y})\right] \\ &\quad = \sum_{j=0}^{M} \mathbb{P}(h_M(\cdot) \in \mathcal{Y}_j) \cdot \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\left[\mu(\hat{\mathbf{Y}}, \mathbf{Y})\right].\end{aligned}$$

不再对所有可能的二进制向量 $\hat{}\mathbf{Y}_{\mathbf{i}} \in \{0,1\}^M$ 求和，而是将 1 的数量相同的所有向量归为一组，因为它们的期望相同。分组后，所有这些向量的概率质量也相加。请注意，因此对于分类器而言，唯一相关的是预测 $\mathrm{h}_M(\cdot)$ 属于哪个组 $\mathrm{Y}j$。

对于任意 j $\in \{0, \ldots,$ M } 有 $\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}$h µ($\hat{}$Y, Y)i，其界限可通过对 j 的所有可能取值求最小值/最大值来确定

因此，

$$\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\left[\mu(\hat{\mathbf{Y}},\mathbf{Y})\right] \geq \min_{j'\in\{0,\ldots,M\}} \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_{j'}\right)}\left[\mu(\hat{\mathbf{Y}},\mathbf{Y})\right], \tag{21}$$

$$\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\left[\mu(\hat{\mathbf{Y}},\mathbf{Y})\right] \leq \max_{j'\in\{0,\ldots,M\}} \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_{j'}\right)}\left[\mu(\hat{\mathbf{Y}},\mathbf{Y})\right]. \tag{22}$$

注意，对于每个 j，Í $_{j=0}^{M}$ Đ(h$_M$(·) ∈ Y$_j$) = 1 和 Đ(h$_M$(·) ∈ Y$_j$) ≥ 0 均成立，因此根据方程 (21) 和 (22) 可得

$$\sum_{j=0}^{M} \mathbb{P}(h_M(\cdot)\in\mathcal{Y}_j)\cdot\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\left[\mu(\hat{\mathbf{Y}},\mathbf{Y})\right] \geq \min_{j'\in\{0,\ldots,M\}} \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_{j'}\right)}\left[\mu(\hat{\mathbf{Y}},\mathbf{Y})\right],$$

$$\sum_{j=0}^{M} \mathbb{P}(h_M(\cdot)\in\mathcal{Y}_j)\cdot\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\left[\mu(\hat{\mathbf{Y}},\mathbf{Y})\right] \leq \max_{j'\in\{0,\ldots,M\}} \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_{j'}\right)}\left[\mu(\hat{\mathbf{Y}},\mathbf{Y})\right].$$

因此　我们分别得到了方程 (12) 和 (13) 的下界与上界。具体而言，　，

$$\min_{h_M\in\mathcal{H}_M^{i.i.}} \left\{\mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mathbb{E}_{h_M(\mathbf{X}_\pi)}\left[\mu(h_M(\mathbf{X}_\pi),\mathbf{Y}_\pi)\right]\right]\right\} \geq \min_{j\in\{0,\ldots,M\}} \left\{\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\mu(\hat{\mathbf{Y}},\mathbf{Y})\right\}, \tag{23}$$

$$\max_{h_M\in\mathcal{H}_M^{i.i.}} \left\{\mathbb{E}_{\pi\sim\mathcal{U}(S_M)}\left[\mathbb{E}_{h_M(\mathbf{X}_\pi)}\left[\mu(h_M(\mathbf{X}_\pi),\mathbf{Y}_\pi)\right]\right]\right\} \leq \max_{j\in\{0,\ldots,M\}} \left\{\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\mu(\hat{\mathbf{Y}},\mathbf{Y})\right\}. \tag{24}$$

对于任意分类器h$_M$ ∈ H$_M^{i.i.}$，只有当所有概率质量都分配给 arg min$_{j\in\{0,\ldots,M\}}$ $\left\{\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\mu(\hat{\mathbf{Y}},\mathbf{Y})\right\}$和 arg max$_{j\in\{0,\ldots,M\}}$ $\left\{\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\mu(\hat{\mathbf{Y}},\mathbf{Y})\right\}$时，等号才成立。换言之，只有在以下条件下才能达到最小值：

$$\sum_{j_{\min}\in\arg\min_{j\in\{0,\ldots,M\}}\left\{\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\mu(\hat{\mathbf{Y}},\mathbf{Y})\right\}} \mathbb{P}(h_M(\cdot)\in\mathcal{Y}_{j_{\min}}) = 1, \tag{25}$$

且仅当……时取最大值

$$\sum_{j_{\max}\in\arg\max_{j\in\{0,\ldots,M\}}\left\{\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\mu(\hat{\mathbf{Y}},\mathbf{Y})\right\}} \mathbb{P}(h_M(\cdot)\in\mathcal{Y}_{j_{\max}}) = 1. \tag{26}$$

因此，分类器 h$_M$ ∈ H$_M^{i.i.}$ 只有在所有预测都属于组 Y$_j$，或可能属于多个均使期望最小化/最大化的组时，才能达到最小值/最大值（具体取决于目标）。

记住 Dutch Draw 根据方程 (10) 和 (11) 选择最优分类器，这会导致 to

$$\lfloor M \cdot \theta^*_{\min} \rceil \in \underset{j\in\{0,\ldots,M\}}{\arg\min} \left\{ \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)} \left[\mu(\hat{\mathbf{Y}}, \mathbf{Y}) \right] \right\},$$
$$\lfloor M \cdot \theta^*_{\max} \rceil \in \underset{j\in\{0,\ldots,M\}}{\arg\max} \left\{ \mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)} \left[\mu(\hat{\mathbf{Y}}, \mathbf{Y}) \right] \right\}.$$

将此结果与 Dutch Draw 的另一种定义（方程（9））结合起来，可直接得到

$$\sum_{j_{\min}\in\arg\min_{j\in\{0,\ldots,M\}}\left\{\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\mu(\hat{\mathbf{Y}},\mathbf{Y})\right\}} \mathbb{P}(\sigma_{\theta^*_{\min}}(\cdot) \in \mathcal{Y}_{j_{\min}}) = 1,$$
$$\sum_{j_{\max}\in\arg\max_{j\in\{0,\ldots,M\}}\left\{\mathbb{E}_{\hat{\mathbf{Y}}\sim\mathcal{U}\left(\mathcal{Y}_j\right)}\mu(\hat{\mathbf{Y}},\mathbf{Y})\right\}} \mathbb{P}(\sigma_{\theta^*_{\max}}(\cdot) \in \mathcal{Y}_{j_{\max}}) = 1.$$

结合方程（25）和（26）可知，最优 Dutch Draw 分类器实际上达到了方程（23）和（24）给出的界限。因此，

$$\sigma_{\theta^*_{\min}} \in \underset{h_M\in\mathcal{H}^{i.j.}_M}{\arg\min} \left\{ \mathbb{E}_{\pi\sim\mathcal{U}(S_M)} \left[\mathbb{E}_{h_M(\mathbf{X}_\pi)} \left[\mu(h_M(\mathbf{X}_\pi), \mathbf{Y}_\pi) \right] \right] \right\},$$
$$\sigma_{\theta^*_{\max}} \in \underset{h_M\in\mathcal{H}^{i.j.}_M}{\arg\max} \left\{ \mathbb{E}_{\pi\sim\mathcal{U}(S_M)} \left[\mathbb{E}_{h_M(\mathbf{X}_\pi)} \left[\mu(h_M(\mathbf{X}_\pi), \mathbf{Y}_\pi) \right] \right] \right\}.$$

因此，我们可以得出结论：最优 Dutch Draw 分类器达到了期望性能的最小值/最大值，因此对于所有具有位置不变测度的输入无关分类器而言，它都是置换最优的。

6 | 讨论与结论

基线对于评估预测模型的性能至关重要。然而，设计基线方法的方式有无穷多种。作为开发过程中的一项必要检查，Van de Bijl 等人（2022）主张采用一种通用、简单且信息丰富的补充基线。因此，本文考察了所有不依赖特征值的基线，因为这使它们具有通用性且相对简单。此外，这些基线也能提供信息：如果一个新开发的模型表现不如一个完全不考虑任何特征值的模型，这应当被视为重大警示信号。本文表明，在所有输入无关的二元分类器中，对于任何位置不变测度，Dutch Draw 基线都是置换最优的。我们的研究结果改进了对任何新二元分类方法的评估过程，因为我们证明，Dutch Draw 基线是衡量新开发模型性能分数的理想基准。

接下来，我们讨论两点可能被视为 Dutch Draw 基线的“不公平”优势。首先，本文考察的是能够同时预测 M 个标签的分类器。与逐个对样本进行分类相比，这可能使分类器占据优势，例如，可以恰好将 M 个样本中的 k 个标记为正类。只有在允许分类器跟踪先前的预测，或根据其已进行的

分类次数进行调整时，逐个分类才能做到这一点。即使有这一优势，我们仍认为，所有输入无关模型都显然能够被新开发的模型击败。

其次，对于最常用的度量，无需额外了解正标签的数量 P，即可推导出 Dutch Draw 基线。尽管如此，Van de Bijl 等人（2022）指出，只有在已知 $P \geq M/2$ 是否成立时，才能计算准确率的 Dutch Draw 基线。如果训练集的分布与测试集相同，就可以利用训练集判断 $P \geq M/2$ 是否很可能成立。此外，领域专家可以估计数据集中正样本是否可能多于负样本。例如，在网络安全数据集中，有害样本通常显著少于正常样本（Wheelus et al., 2018）。因此，有许多方法可以估计 $P \geq M/2$ 是否成立。尽管如此，即使 Dutch Draw 基线使用了这些信息（仅用于准确率），我们仍认为任何新开发的模型都应该优于 Dutch Draw 基线，因为它不会利用任何特征值来改进预测。

最后，我们探讨未来的研究机会。本文仅考察了二元分类。一个自然的延伸方向是进一步研究多类别分类（Grandini 等人，2020）。在这种情况下，类似 Dutch Draw 的策略是否仍然最优？能否推导出最优基线的闭式表达式？我们认为，所提出的三项性质（即输入无关、位置不变和置换最优）对于多类别情形仍然适用。这有助于确定何种分类器可被视为最优。Van de Bijl 等人（2022）指出，Dutch Draw 基线可用于缩放现有度量。本文进一步说明了为何应使用 Dutch Draw 基线，而非其他任何输入无关分类器来缩放度量。不过，仍可进一步探讨如何缩放各项度量，以最大限度地提高性能分数的可解释性。

利益冲突声明

作者声明，他们没有需要披露的相关财务或非财务利益。

资金支持

本研究未获得任何经费支持。

数据可用性

本研究未使用任何数据集。

参考资料

Artin, M.（2011年）代数。Prentice Hall，第2版。

van de Bijl, E., Klein, J., Pries, J., Bhulai, S., Hoogendoorn, M. and van der Mei, R.（2022）Dutch Draw：为二元预测模型构建通用基线。网址：https://arxiv.org/abs/2203.13084。

Buckley, C., O'Reilly,M., Whelan, D., Farrell, A. V., Clark, L., Longo, V., Gilchrist, M. 和 Caulfield, B.（2017年）使用单个惯性测量单元对跑步疲劳进行二元分类。载于*2017*年 *IEEE* 第*14*届可穿戴与植入式人体传感器网络国际会议（*BSN*），197–201。

Cauchy, A.-L.（1815年）关于对函数所包含的量以所有可能方式进行置换时，该函数可取值的数量的论文。巴黎综合理工学院学报。

Dixon, J. D. 和 Mortimer, B.（1996年）置换群，第 163 卷。Springer Science & Business Medi a.

Grandini, M.、Bagli, E. 和 Visani, G.（2020）多类别分类指标：综述。

Li, L.、Yu, Y.、Bai, S.、Hou, Y. 和 Chen, X.（2018）一种基于二元分类和 k -nn 的有效两步入侵检测方法。*IEEE Access*，**6**，12060–12073。

Pirouz, B.、Shaffiee Haghshenas, S.、Shaffiee Haghshenas, S. 和 Piro, P.（2020）探究可持续发展进程中的一项严峻挑战：通过人工智能和回归分析进行二元分类，分析新冠肺炎（新型冠状病毒）的确诊病例。《可持续发展》，**12**。网址：https://www.mdpi.com/2071-1050/12/6/2427。

Sokolova, M. 和 Lapalme, G.（2009）分类任务性能度量的系统分析。信息处理与管理，**45**，427–437。网址：https://www.sciencedirect.com/science/article/pii/S0306457309000259。

Tharwat, A.（2021）分类评估方法。应用计算与信息学，**17**，168–192。网址：https://doi.org/10.1016/j.aci.2018.08.003。

Wheelus, C.、Bou-Harb, E. 和 Zhu, X.（2018）应对网络安全数据集中的类别不平衡。载于*2018*年*IEEE*信息复用与集成国际会议（*IRI*），229–232。