
Constrained Binary Decision Making

Daniel Průša Vojtěch Franc
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague
{prusapa1,xfrancv}@fel.cvut.cz

Abstract

Binary statistical decision making involves choosing between two states based on statistical evidence. The optimal decision strategy is typically formulated through a constrained optimization problem, where both the objective and constraints are expressed as integrals involving two Lebesgue measurable functions, one of which represents the strategy being optimized. In this work, we present a comprehensive formulation of the binary decision making problem and provide a detailed characterization of the optimal solution. Our framework encompasses a wide range of well-known and recently proposed decision making problems as specific cases. We demonstrate how our generic approach can be used to derive the optimal decision strategies for these diverse instances. Our results offer a robust mathematical tool that simplifies the process of solving both existing and novel formulations of binary decision making problems which are in the core of many Machine Learning algorithms.

1 Introduction

Binary statistical decision making (BDM) involves selecting between two possible states using statistical evidence. This approach is applied across various fields, from classical statistics to machine learning. A notable example is the Neyman-Pearson problem [18], which establishes the optimal strategy for testing two hypotheses and serves as the foundation for optimal detectors. Beyond classical statistics, BDM problems are used in many machine learning applications, such as detecting out-of-distribution (OOD) samples [11, 14, 6, 5, 15, 10, 2, 21, 20, 22, 23], designing selective classifiers [13, 12, 4, 8, 9], and developing selective classification in the presence of OOD data (SCOD) [24, 16, 7] to name a few.

Instances of BDM determine the best decision-making strategy by solving a constrained optimization problem involving Lebesgue measurable functions in both the objective and constraints. Understanding the structure of this optimal strategy is crucial before deploying efficient methods to find it. The Neyman-Pearson lemma, a well-known example found in many statistics textbooks, states that an optimal detection strategy relies on comparing the likelihood ratio of the underlying distributions with a threshold. Therefore, learning a detector from examples involves effectively approximating the likelihood ratio and then tuning the decision threshold.

However, defining the optimal strategy for various BDM problems is generally challenging, often taking several years to discover. For instance, the Bounded-Improvement and Bounded-Abstention models, serving as formal definitions of optimal selective classifiers, were first introduced in [19]. The optimal strategies for these two BDM problems were initially described in [8], marking a 14-year period of development. Another recent example is the work of [24], who developed a method for learning a selective classifier handling inputs contaminated by out-of-distribution data, so called SCOD problem. Although they formulated an optimal SCOD selection strategy to accept input samples as a solution to a novel BDM problem, they did not analyze the optimal strategy directly.

Instead, they proposed a selector strategy named SIRC, combining two detectors using heuristically chosen non-linear functions. In a recently published paper [16], the authors derived an optimal solution for the SCOD problem, demonstrating that the optimal selective strategy involves a simple linear combination of the two detectors. Empirical evidence confirmed that the optimal linear strategy consistently outperforms the heuristic SIRC, despite its minor change. This example underscores the advantages of understanding the structure of the optimal solution to underlying BDM problems.

In this paper, we present a comprehensive formulation of the BDM problem and thoroughly characterize the optimal strategy. Our framework encompasses various BDM problems as special cases, enabling us to derive optimal decision strategies for these instances. This provides a robust mathematical tool for solving both existing and new BDM problems. The related theorem is highly general, applying to both discrete and continuous instance spaces without requiring the differentiability of decision and loss functions, unlike common proof techniques based on Lagrange duality.

To illustrate the impact of our work, in Section 2, we formulate several practical modifications of existing BDM problems. In Section 4, we demonstrate the derivation of their optimal strategies using our main result, which characterizes the optimal strategies of a generic BDM as presented in Section 3.

2 Examples of existing binary decision making problems

In this section, we present various examples of BDM problems, with a focus on the significant application of designing a selective classifier in the presence of out-of-distribution (SCOD) samples [24, 16]. We note that while selective classification itself is not the contribution of this paper, it serves to illustrate the practical importance and diversity of BDM formulations in addressing a recently emerging problem in the machine learning community.

Consider a given classifier $h: \mathcal{X} \rightarrow \mathcal{Y}$, where \mathcal{X} is the instance space and \mathcal{Y} a finite set of labels. We assume the classifier h was designed to minimize the prediction loss $\ell: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$. Our task is to equip the classifier h with a binary stochastic decision strategy $c: \mathcal{X} \rightarrow [0, 1]$. The strategy c acts as a selector for input samples. The input sample $x \in \mathcal{X}$ is accepted for prediction by h with probability $c(x)$ and reject otherwise. Thus, the pair $(h, c): \mathcal{X} \rightarrow \mathcal{Y} \cup \{\text{reject}\}$ forms a reject option classifier [3]:

$$(h, c)(x) = \begin{cases} h(x) & \text{with probability } c(x), \\ \text{reject} & \text{with probability } 1 - c(x). \end{cases}$$

We assume the following statistical model of the data. Input samples are generated from a mixture of two distributions: the in-distribution (ID) $p_I: \mathcal{X} \rightarrow \mathbb{R}_+$ and the out-of-distribution (OOD) $p_O: \mathcal{X} \rightarrow \mathbb{R}_+$. Thus, the input sample $x \in \mathcal{X}$ is generated from

$$p(x) = p_I(x)(1 - \pi) + p_O(x)\pi, \quad (1)$$

where $\pi \in [0, 1]$ represents the portion of OOD samples in the mixture. Additionally, if the sample $x \in \mathcal{Y}$ is generated from the in-distribution, a latent label $y \in \mathcal{Y}$ is generated according to the distribution $p(y | x)$.

In the outlined statistical model, prediction uncertainty arises from two main sources: i) inherent data uncertainty (aleatoric uncertainty), stemming from the stochastic relationship between the input x and the latent label y , and ii) distribution uncertainty, which arises from input samples potentially coming from either in-distribution or out-of-distribution sources. We will explore six different BDM instances to develop a selector c that admits an input sample x for classification with h only when the prediction uncertainty is below an acceptable threshold. Each instance defines an optimal selector in a unique way. In Section 2.1, we described the Neyman-Pearson formulation of the OOD optimal detector [18], which addresses only distribution uncertainty. Sections 2.2 and 2.3 cover the Bounded-Improvement and Bounded-Abstention models [19], addressing only data uncertainty. In Section 2.4, we describe the SCOD problem [24], which defines the optimal selector addressing both distribution and data uncertainty. Additionally, Sections 2.5 and 2.6 give examples of two novel practical variants of the SCOD problem.

The primary contribution of this paper is a framework for solving a broad range of BDM problems. In Section 3, we detail the main theorem that characterizes the optimal solution for a generic BDM problem. In Section 4, we demonstrate how to apply this theorem to easily find solutions for all the BDM instances described here.

2.1 Neyman-Pearson (NP) detector

Assume our goal is to design a selector strategy $c: \mathcal{X} \rightarrow [0, 1]$ that accepts only in-distribution (ID) samples and rejects out-of-distribution (OOD) samples. In other words, c functions as a binary classifier to distinguish ID and OOD samples. We will characterize the strategy c using two metrics. First, using the True Positive Rate (TPR) representing the probability that ID sample x is correctly accepted by the strategy:

$$\text{tpr}(c) = \int_{\mathcal{X}} p_I(x) c(x) dx . \quad (2)$$

Second, using the False Positive Rate (FPR) representing the probability that OOD sample x is incorrectly accepted by c :

$$\text{fpr}(c) = \int_{\mathcal{X}} p_O(x) c(x) dx . \quad (3)$$

Given maximal acceptable FPR, $\text{fpr}_{\max} > 0$, the NP task is to find a strategy c^* that solves the following BDM problem [18]:

$$\max_{c: \mathcal{X} \rightarrow [0,1]} \text{tpr}(c) \quad \text{s.t.} \quad \text{fpr}(c) \leq \text{fpr}_{\max} . \quad (4)$$

A well-known solution to the NP task 4 is the strategy:

$$c^*(x) = \begin{cases} 1 & \text{if } g(x) < \theta , \\ \tau & \text{if } g(x) = \theta , \\ 0 & \text{if } g(x) > \theta , \end{cases} \quad (5)$$

where $g(x) = \frac{p_O(x)}{p_I(x)}$ is the likelihood ratio, $\theta \in \mathbb{R}$ is a decision threshold, and $\tau \in [0, 1]$ is a probability of acceptance when $g(x) = \theta$.

Knowing the structure of the optimal strategy simplifies its construction significantly. The key to an optimal decision is accurately modeling the likelihood ratio $g(x)$. Once the likelihood ratio is known, finding the optimal strategy involves setting the threshold θ and the randomization parameter τ , typically done by tuning these parameters on a validation sample. Moreover, if $p_I(x)$ and $p_O(x)$ are continuous, the randomization parameter τ can be arbitrary since the event $g(x) = \theta$ has zero probability of occurring.

2.2 Selective Classification: Bounded-abstention model

Let us consider the ideal scenario where the input instances $x \in \mathcal{X}$ are solely generated by the in-distribution $p_I(x)$, meaning the out-of-distribution portion in the model (1) is $\pi = 0$. Our aim is devise a selection strategy $c: \mathcal{X} \rightarrow [0, 1]$ that accepts only those input samples $x \in \mathcal{X}$ where the prediction loss $\ell(y, h(x))$ is likely to be small. We will characterize the strategy c using two metrics. First, we employ the notion of *coverage*:

$$\phi(c) = \int_{\mathcal{X}} p_I(x) c(x) dx \quad (6)$$

which reflects the probability of accepting an input sample. Secondly, we introduce the concept of "selective risk":

$$\text{R}^S(h, c) = \frac{\int_{\mathcal{X}} \sum_{y \in \mathcal{Y}} p_I(x, y) \ell(y, h(x)) c(x) dx}{\phi(c)} . \quad (7)$$

representing the expected prediction loss $\ell: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ on the accepted samples.

Given a minimal acceptable coverage $\omega > 0$, the bounded-abstention model defines the optimal selector strategy as the solution to the following BDM problem [19]:

$$\min_{h \in \mathcal{Y}^{\mathcal{X}}, c \in [0,1]^{\mathcal{X}}} \text{R}^S(h, c) \quad \text{s.t.} \quad \phi(c) \geq \omega . \quad (8)$$

The solution to the bounded-abstention model (8) consists of the Bayes predictor

$$h^*(x) = \operatorname{argmin}_{y'} \sum_y p(y | x) \ell(y, y')$$

and the strategy [8, 9]:

$$c^*(x) = \begin{cases} 1 & \text{if } r(x) < \theta, \\ \tau & \text{if } r(x) = \theta, \\ 0 & \text{if } r(x) > \theta, \end{cases} \quad (9)$$

where

$$r(x) = \sum_{y \in \mathcal{Y}} p_I(y | x) \ell(y, h(x)) \quad (10)$$

is the conditional risk of the classifier h for the input x , $\theta \in \mathbb{R}$ is a decision threshold and $\tau \in [0, 1]$ is a probability of acceptance when $r(x) = \theta$.

Note that the BDM problem (8) was formulated in [19], but the optimal solution (9) was not derived until 14 years later in [8].

The values of θ and τ depend on $p_I(x, y)$ and ω , and in practice, they are tuned on validation data once a good model of the conditional risk $r(x)$ is established. While this process can be challenging and needs to be addressed for each problem separately, it is still significantly easier than tuning an unknown function. For example, discretizing the possible values of θ and τ and conducting an exhaustive search, although not always optimal, has proven effective in practice.

The additivity of the risk R^S allows h to be optimized independently of c , resulting in the Bayes predictor for all the problems discussed in the following sections. Therefore, from this point on, we will focus solely on optimal strategies for c .

2.3 Selective Classification: Bounded-improvement rejection model

A symmetric definition of the optimal selector strategy is provided by the bounded-improvement model. Given a maximum acceptable selective risk $\lambda > 0$, the optimal selector is the solution to the following BDM problem:

$$\max_{h,c} \phi(c) \quad \text{s.t.} \quad R^S(h, c) \leq \lambda. \quad (11)$$

The solution to the bounded-improvement model (11) is similar to the strategy (9), differing only in the specific values of the decision threshold θ and the randomization parameter τ [9].

2.4 Selective Classification in the presence of OOD data (SCOD)

The SCOD problem integrates the objectives of the Neyman-Pearson task and bounded-abstention models. The aim is to design a selector strategy $c \in \mathcal{X} \rightarrow [0, 1]$ that accepts an input sample x if it is likely to be correctly predicted by h and unlikely to be generated from the OOD. Formally, the objective is defined using three metrics: True Positive Rate $\text{tpr}(c)$, equation (2), False Positive Rate $\text{fpr}(c)$, equation (3) and the selective risk $R^S(c)$, equation (7). Given the relative cost $\alpha > 0$ and the minimum acceptable TPR, $\text{tpr}_{\min} > 0$, the goal is to find a strategy c^* that solves the following BDM problem [24]:

$$\min_{h \in \mathcal{Y}^{\mathcal{X}}, c \in [0,1]^{\mathcal{X}}} [(1 - \alpha) R^S(h, c) + \alpha \text{fpr}(c)] \quad \text{s.t.} \quad \text{tpr}(c) \geq \text{tpr}_{\min}. \quad (12)$$

The solution to the SCOD problem (12) is the strategy [16, 7]:

$$c^*(x) = \begin{cases} 1 & \text{if } s(x) > \theta, \\ \tau & \text{if } s(x) = \theta, \\ 0 & \text{if } s(x) < \theta \end{cases}, \quad \text{where } s(x) = r(x) + \frac{\alpha \text{tpr}_{\min}}{1 - \alpha} g(x) \quad (13)$$

for $\alpha \in [0, 1)$ and $s(x) = g(x)$ for $\alpha = 1$. Here, $r(x)$ is the conditional risk (10), $g(x) = \frac{p_O(x)}{p_I(x)}$ is the likelihood ratio, $\theta \in \mathbb{R}$ is a decision threshold, and $\tau \in [0, 1]$ the acceptance probability for the case $s(x) = r(x)$.

The BDM problem (12) was formulated in [24]. The optimal solution for continuous distributions was recently derived in [16], and the solution for the general case was provided in [7].

2.5 SCOD: bounded TPR-FPR

To demonstrate the versatility of our framework, we will present two novel variants of the SCOD problem in this section and the next. While the optimal solutions to these new formulations are unknown, they can be easily derived using the framework proposed in this paper, as we will later show.

The original formulation of the SCOD problem (12) relies on the user-defined relative cost α . However, this assumes that the prediction loss $\ell: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ and the cost for incorrectly accepting the OOD sample share the same units, which is not always practical. In such cases, it can be useful to remove the FPR from the objective and impose a hard constraint on it instead. Specifically, we can alternatively formulate the SCOD problem as follows. Given $\text{tpr}_{\min} > 0$ and $\text{fpr}_{\max} > 0$, the goal is to find a strategy c^* that solves the following BDM problem:

$$\min_{h \in \mathcal{Y}^{\mathcal{X}}, c \in [0,1]^{\mathcal{X}}} R^S(h, c) \quad \text{s.t.} \quad \text{tpr}(c) \geq \text{tpr}_{\min} \quad \text{and} \quad \text{fpr}(c) \leq \text{fpr}_{\max}. \quad (14)$$

In Section 4, we will prove that the optimal solution to the problem (14) is the strategy:

$$c^*(x) = \begin{cases} 1 & \text{if } s(x) < \theta, \\ \tau & \text{if } s(x) = \theta, \\ 0 & \text{if } s(x) > \theta, \end{cases} \quad \text{where } s(x) = r(x) + \beta g(x), \quad (15)$$

with $\theta \in \mathbb{R}$ as the decision threshold, $\tau \in [0, 1]$ as the acceptance probability when $s(x) = \theta$, and $\beta \in \mathbb{R}$ as a constant depending on the problem setup.

2.6 SCOD: bounded Precision-Recall

Assume the portion of OOD samples π is known or can be estimated. In some applications, it is useful to replace FPR with precision, which is the proportion of true ID samples among all accepted inputs. Precision $\text{prec}(c)$ is defined as:

$$\text{prec}(c) = \frac{(1 - \pi) \text{tpr}(c)}{\text{fpr}(c) \pi + \text{tpr}(c) (1 - \pi)}.$$

If we want to avoid defining the relative cost, as in (14), we can formulate the SCOD problem as follows. Given $\text{tpr}_{\min} > 0$ and $\text{prec}_{\min} > 0$, the goal is to find a strategy c^* which solves the following BDM problem:

$$\min_{h \in \mathcal{Y}^{\mathcal{X}}, c \in [0,1]^{\mathcal{X}}} R^S(h, c) \quad \text{s.t.} \quad \text{tpr}(c) \geq \text{tpr}_{\min} \quad \text{and} \quad \text{prec}(c) \geq \text{prec}_{\min} \quad (16)$$

In Section 4, we will prove that the optimal solution to the problem (16) is a strategy similar to (15), differing only in the specific values of the decision threshold τ , the randomization parameter τ and the multiplier β .

2.7 Summary

In the previous sections, we described six BDM problems, each with a unique optimal selector strategy. The first four are established formulations, while the last two are novel SCOD modifications potentially useful for specific applications. As ML applications grow, new formulations are likely to emerge to address specific setups.

Deriving the optimal strategy for a given BDM problem proved to be the essential first step. Although the statistical model is often unknown in practice, understanding the structure of the optimal strategy is crucial for designing efficient algorithms to learn the selector from examples. This knowledge can narrow the search space or simplify the problem using divide-and-conquer approaches.

For instance, consider constructing a selector strategy for the SCOD problem. Whether using the original cost-based formulation (12) or the two novel variants (14) and (16), takes the form (15), the optimal strategy c^* relies on an uncertainty score $s(x) = r(x) + \beta g(x)$. The strategy accepts input x if $s(x)$ is below a threshold θ , and with probability τ if $s(x) = \theta$. This analysis shows that the core problem reduces to approximating the two components: the conditional risk

$$r(x) = \sum_{y \in \mathcal{Y}} p(y | x) \ell(y, h(x))$$

and the OOD/ID likelihood ratio $g(x) = p_O(x)/p_I(x)$. Extensive literature on OOD detection provides methods to approximate g (e.g. [11, 14, 6, 5, 15, 10, 2, 21, 20, 22, 23]), and there are methods to construct good proxies for r (e.g. [13, 12, 4, 8, 9]).

Once r and g are known, constructing the selector strategy involves determining the scalars θ , τ , and β . These parameters can be tuned using held-out data, which is much easier than solving the original problem. Moreover, for continuous distributions, the event $s(x) = \theta$ has zero probability, making τ arbitrary. In the cost-based formulation, the multiplier β has an analytical solution, leaving only θ to be determined.

This example shows that deriving an optimal strategy for BDM problems allows us to use existing methods from OOD detection and selective classification to approximate various SCOD problem formulations. Conversely, skipping the optimal strategy derivation and using heuristic rules, such as the SIRC strategy from the original SCOD paper [24], can lead to sub-optimal performance [16].

3 Main result

In this section, we formulate a generic BDM problem, characterize its optimal solutions, and provide a straightforward instance of the optimal strategy.

For a measurable set \mathcal{X} in a measure space, Lebesgue measurable functions $R, p, q : \mathcal{X} \rightarrow [0, \infty)$ with finite integrals on \mathcal{X} , and fixed non-negative real values δ and σ , we define the following optimization problem:

$$\min_{c: \mathcal{X} \rightarrow [0,1]} \int_{\mathcal{X}} R(x)c(x)dx \quad \text{s.t.} \quad \int_{\mathcal{X}} p(x)c(x)dx \geq \delta \quad \text{and} \quad \int_{\mathcal{X}} q(x)c(x)dx \leq \sigma. \quad (17)$$

Note that we do not pose any additional requirements on the functions, like continuity or differentiability, neither do we differentiate between discrete and continuous sets \mathcal{X} ; they can be arbitrary.

Theorem 1. *For every optimal solution $c^* : \mathcal{X} \rightarrow [0, 1]$ to Problem (17), there exist real numbers λ, μ such that*

$$\int_{\mathcal{X}^<} p(x)c^*(x) dx = \int_{\mathcal{X}^<} p(x) dx, \quad (18)$$

$$\int_{\mathcal{X}^>} p(x)c^*(x) dx = 0, \quad (19)$$

where

$$\mathcal{X}^< = \left\{ x \in \mathcal{X} \mid p(x) > 0 \wedge \frac{R(x)}{p(x)} + \mu \frac{q(x)}{p(x)} < \lambda \right\}, \quad (20)$$

$$\mathcal{X}^> = \left\{ x \in \mathcal{X} \mid p(x) > 0 \wedge \frac{R(x)}{p(x)} + \mu \frac{q(x)}{p(x)} > \lambda \right\}. \quad (21)$$

Informally, equation (18) states that any optimal c^* attains the value 1 on $\mathcal{X}^<$, except on a subset of measure zero. Similarly, by equation (19), c^* attains the value 0 on $\mathcal{X}^>$, again up to a subset of measure zero. The sets $\mathcal{X}^<$ and $\mathcal{X}^>$ contain the points x for which $\frac{R(x)}{p(x)} + \mu \frac{q(x)}{p(x)}$ is below and above the threshold λ , respectively, excluding any insignificant elements where $p(x) = 0$.

Based on these observations, the next lemma suggests a way to construct a single score function that determines an optimal decision strategy.

Lemma 1. *An optimal solution $c^* : \mathcal{X} \rightarrow [0, 1]$ to Problem (17) can be found among the forms*

$$c^*(x) = \begin{cases} 0 & \text{if } s(x) > \lambda, \\ \tau(x) & \text{if } s(x) = \lambda, \\ 1 & \text{if } s(x) < \lambda, \end{cases}$$

where

$$s(x) = \begin{cases} \frac{R(x)}{p(x)} + \mu \frac{q(x)}{p(x)} & \text{if } p(x) > 0, \\ \infty & \text{otherwise,} \end{cases}$$

is a score function, λ, μ are suitable real numbers, and $\tau : \mathcal{X} \rightarrow [0, 1]$ is a function implicitly defined by the problem parameters.

Proof. Let $c : \mathcal{X} \rightarrow [0, 1]$ be an optimal solution to Problem (17). Consider real numbers λ, μ yielded by Theorem 1 and define

$$c^*(x) = \begin{cases} 0 & \text{if } s(x) > \lambda, \\ c(x) & \text{if } s(x) = \lambda, \\ 1 & \text{if } s(x) < \lambda. \end{cases}$$

To show that c^* is optimal, decompose \mathcal{X} into the pairwise disjoint sets $\mathcal{X}^<$ (see (20)), $\mathcal{X}^>$ (see (21)), $\mathcal{X}' = \{x \in \mathcal{X} \mid p(x) = 0\}$ and $\mathcal{X}^= = \{x \in \mathcal{X} \mid s(x) = \lambda\}$. Equalities (18) and (19) ensure that c and c^* are nearly identical on sets $\mathcal{X}^<$ and $\mathcal{X}^>$, respectively, up to subsets of measure zero. By the definition of c^* , they are identical on $\mathcal{X}^=$. Consequently, the criterion and constraints attain the same values for both c and c^* on $\mathcal{X}^< \cup \mathcal{X}^> \cup \mathcal{X}^=$. Finally, the first constraint of (17) is independent of \mathcal{X}' , and $c^*(x) \leq c(x)$ for all $x \in \mathcal{X}'$ ensures

$$\int_{\mathcal{X}'} R(x)c^*(x) \leq \int_{\mathcal{X}'} R(x)c(x), \quad \int_{\mathcal{X}'} q(x)c^*(x) \leq \int_{\mathcal{X}'} q(x)c(x).$$

In conclusion, c^* is an optimal solution. \square

Tool to solve the BDM problems Lemma 1 is a key tool for simplifying the BDM problem (17). It shows that the main challenge is approximating the ratios $\frac{R(x)}{p(x)}$ and $\frac{q(x)}{p(x)}$. Once these ratios are known, determining the score s involves finding the multiplier μ . With the score s , the optimal strategy c^* depends on finding the decision threshold λ and the function $\tau(x)$. However, in cases where \mathcal{X} is a continuous space, the boundary set $\{x \in \mathcal{X} \mid s(x) = \lambda\}$ typically has measure zero, except in some pathological instances. Consequently, $\tau(x)$ can be chosen arbitrarily and does not play a significant role. Thus, in continuous spaces, one only needs to find the scalars μ and λ , which can be tuned using held-out data, making the process much simpler than solving the original problem from scratch.

The problem feasibility In the special case where only one constraint is given, i.e., when either $p(x) = 0$ or $q(x) = 0$, checking feasibility is straightforward. In the general case, to determine the feasibility of Problem (17), we can formulate the task as:

$$\min_{c: \mathcal{X} \rightarrow [0,1]} \int_{\mathcal{X}} q(x)c(x)dx \quad \text{s.t.} \quad \int_{\mathcal{X}} p(x)c(x)dx \geq \delta. \quad (22)$$

We obtain a similar problem but with only one constraint. Problem (17) is feasible if and only if Problem (22) attains a value less than or equal to σ .

Proof technique Problem (17) represents an instance of infinite linear programming, grounded in established theory [17]. Especially, when \mathcal{X} is a finite set, it reduces to a standard linear program, allowing the use of Lagrange duality to derive optimality conditions, as characterized in Theorem 1. However, when \mathcal{X} is arbitrary and the functions R, p, q are essentially unrestricted, this approach becomes substantially more complex. In such cases, a more general duality theorem would be required, and even then, deducing the desired results would involve handling significantly more intricate optimality conditions. Given these challenges, and drawing inspiration from techniques in related works (e.g., [9, 16, 18] and other studies on the Neyman-Pearson problem), we instead pursue a direct proof, provided in Appendix A. By introducing forbidden point configurations, this proof shows that each optimal solution is determined by a line separating the images of the sets $\mathcal{X}^<$ and $\mathcal{X}^>$ in the plane, mapped by the function $\varkappa(x) = \left(\frac{R(x)}{p(x)}, \frac{q(x)}{p(x)} \right)$.

Extensions It remains an open problem whether our results can be generalized to the extended formulation

$$\min_{c: \mathcal{X} \rightarrow [0,1]} \int_{\mathcal{X}} R(x)c(x), dx \quad \text{s.t.} \quad \int_{\mathcal{X}} p(x)c(x), dx \geq \delta \quad \text{and} \quad \int_{\mathcal{X}} q_i(x)c(x), dx \leq \sigma_i, \quad i \in I, \quad (23)$$

where I is a finite index set, the functions q_i are Lebesgue measurable with finite integrals on \mathcal{X} , and σ_i are non-negative bounds. A key question is whether a scoring function with $1 + |I|$ parameters (instead of 2) exists for this case. It is important to note that generalizing the current proof would involve addressing a significantly greater combinatorial complexity in the point configurations.

4 General tasks

In this section, we introduce several special cases of problem (17), whose forms of solutions directly follow from Theorem 1, as we will show. These cases encompass all the examples in Section 2. Thus, all the mentioned BDM problems can be solved in their general form using the single framework proposed in this paper.

To simplify notation, we represent the Lebesgue measurable functions in (17) as $p_i: \mathcal{X} \rightarrow [0, \infty)$, $i \in \{0, 1, 2\}$. Furthermore, for Lebesgue measurable $c: \mathcal{X} \rightarrow [0, 1]$, we define the functionals:

$$F_i(c) = \int_{\mathcal{X}} p_i(x) c(x) dx, \quad i \in \{0, 1, 2\}.$$

The generic BDM problem (17) then requires solving for given non-negative scalars δ and ρ :

$$\min_{c: \mathcal{X} \rightarrow [0,1]} F_0(c) \quad \text{s.t.} \quad F_1(c) \geq \delta \quad \text{and} \quad F_2(c) \leq \sigma. \quad (24)$$

Note that F_0 here does not depend on the classifier h , which is consistent with the assumption that h is the Bayes predictor.

Problem 1: Given $\delta > 0$, the task is to solve:

$$\min_{c: \mathcal{X} \rightarrow [0,1]} F_0(c) \quad \text{s.t.} \quad F_1(c) \geq \delta. \quad (25)$$

A special case is the Neyman-Pearson task (4).

We derive this formulation from (24) by setting $p_2(x) = 0$ for all $x \in \mathcal{X}$. Lemma 1 implies that if the problem is feasible, it has an optimal solution c^* such that

$$c^*(x) = \begin{cases} 0 & \text{if } \frac{p_0(x)}{p_1(x)} > \lambda, \\ \tau(x) & \text{if } \frac{p_0(x)}{p_1(x)} = \lambda, \\ 1 & \text{if } \frac{p_0(x)}{p_1(x)} < \lambda. \end{cases}$$

Note that here, Lemma 1 does not fully replicate the Neyman-Pearson strategy, in which τ is constant.

Problem 2: Given $\delta > 0$, the task is to solve:

$$\min_{c: \mathcal{X} \rightarrow [0,1]} \frac{F_0(c)}{F_1(c)} \quad \text{s.t.} \quad F_1(c) \geq \delta. \quad (26)$$

Special cases are the Bounded-Abstention model (8) and the original SCOD problem (12).

This formulation can equivalently be transformed to (25) by showing that there is an optimal solution c^* such that $F_1(c^*) = \delta$. Indeed, if some optimal $c: \mathcal{X} \rightarrow [0, \infty)$ fulfills $F_1(c) > \delta$, then we can define $c^* = \frac{\delta}{F_1(c)} c$, and it holds that $F_1(c^*) = \delta$ and

$$\frac{F_0(c^*)}{F_1(c^*)} = \frac{\frac{\delta}{F_1(c)} F_0(c)}{\delta} = \frac{F_0(c)}{F_1(c)}.$$

Hence, such a c^* is preserved in the set of optimal solutions of the problem

$$\min_{c: \mathcal{X} \rightarrow [0,1]} \frac{F_0(c)}{\delta} \quad \text{s.t.} \quad F_1(c) \geq \delta.$$

Problem 3: Given $\delta > 0$, the task is to solve:

$$\max_{c: \mathcal{X} \rightarrow [0,1]} F_1(c) \quad \text{s.t.} \quad \frac{F_0(c)}{F_1(c)} \leq \delta. \quad (27)$$

A special case is the Bounded-Improvement model (11).

Let c^* be an optimal solution to (27). Define a problem:

$$\min_{c: \mathcal{X} \rightarrow [0,1]} \frac{F_0(c)}{F_1(c)} \quad \text{s.t.} \quad F_1(c) \geq F_1(c^*).$$

Clearly, c^* is feasible for the new problem, which is of the form (26). Moreover, any optimal solution c to this problem is also optimal to (27) because it attains the maximum $F_1(c^*)$ and satisfies the constraint:

$$\frac{F_0(c)}{F_1(c)} \leq \frac{F_0(c^*)}{F_1(c^*)} \leq \delta.$$

Problem 4: Given $\delta > 0, \sigma > 0$, the task is to solve:

$$\min_{c: \mathcal{X} \rightarrow [0,1]} \frac{F_0(c)}{F_1(c)} \quad \text{s.t.} \quad F_1(c) \geq \delta \quad \text{and} \quad F_2(c) \leq \sigma. \quad (28)$$

A special case is the SCOD problem formulation with constraints on FPR and TPR (14).

In this case, we can transform the problem to (24) in the same way as we did it for the formulation (26).

Problem 5: Given $\delta > 0, \sigma > 0$, the task is to solve:

$$\min_{c: \mathcal{X} \rightarrow [0,1]} \frac{F_0(c)}{F_1(c)} \quad \text{s.t.} \quad F_1(c) \geq \delta \quad \text{and} \quad \frac{F_2(c)}{F_1(c)} \leq \sigma. \quad (29)$$

A special case is the SCOD problem formulation with the constraints on TPR and Precision (16) because the second constraint in (16) can be rewritten as

$$\frac{\text{fpr}(c)}{\text{tpr}(c)} \leq (\pi^{-1} - 1)(\text{prec}_{\min}^{-1} - 1),$$

which matches the form

$$\frac{F_2(c)}{F_1(c)} \leq \sigma.$$

Again, the problem transforms to (24) using the same approach as in the case of the formulation (26).

5 Conclusion

In this paper, we presented a comprehensive framework for solving binary decision-making (BDM) problems, which define optimal decision strategies through constrained optimization involving Lebesgue measurable functions. We characterize all optimal strategies for a generic BDM problem and derive a specific class of optimal strategies based on comparing a single score with a decision threshold. Our framework covers a variety of BDM problems, from well-known instances like the Neyman-Pearson problem to recent developments such as the SCOD problem. We demonstrated the versatility and robustness of our framework by deriving optimal solutions for all the BDM problems discussed in our paper. This work provides a foundational approach that can be adapted to future needs in machine learning and statistical decision making, ensuring more effective and theoretically grounded solutions.

6 Acknowledgement

This research was supported by the CTU institutional support (future fund).

References

- [1] Constantin Carathéodory. Über den variabilitätsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehmen. *Mathematische Annalen*, 64:95–115, 1907.
- [2] Guangyao Chen, Peixi Peng, Xiangqian Wang, and Yonghong Tian. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8065–8081, 2022.

- [3] C. K. Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16(1):41–46, 1970.
- [4] C. Corbiere, N. Thome, A. Bar-Hen, M. Cord, and P. Perez. Addressing failure prediction by learning model confidence. In *Advances in Neural Information Processing Systems*, volume 32, pages 2902–2913, 2019.
- [5] Terrance DeVries and Graham W Taylor. Learning confidence for out-of-distribution detection in neural networks. *arXiv preprint arXiv:1802.04865*, 2018.
- [6] Akshay Raj Dhamija, Manuel Günther, and Terrance Boult. Reducing network agnostophobia. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.
- [7] Vojtech Franc, Jakub Paplham, and Daniel Prusa. SCOD: From heuristics to theory. In *Proceedings of European Conference on Computer Vision*, 2024.
- [8] Vojtech Franc and Daniel Prusa. On discriminative learning of prediction uncertainty. In *Proc. of Machine Learning Research – International Conference on Machine Learning*, June 2019.
- [9] Vojtech Franc, Daniel Prusa, and Vaclav Voracek. Optimal strategies for reject option classifiers. *Journal of Machine Learning Research*, 24(11):1–49, 2023.
- [10] Federica Granese, Marco Romanelli, Daniele Gorla, Catuscia Palamidessi, and Pablo Piantanida. Doctor: A simple method for detecting misclassification errors. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 5669–5681. Curran Associates, Inc., 2021.
- [11] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *Proceedings of International Conference on Learning Representations*, 2017.
- [12] H. Jiang, B. Kim, M. Y. Guan, and M. Gupta. To trust or not to trust a classifier. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 5546–5557, 2018.
- [13] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, pages 6402–6413, 2017.
- [14] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *International Conference on Learning Representations*, 2018.
- [15] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.
- [16] Harikrishna Narasimhan, Aditya Krishna Menon, Wittawat Jitkrittum, and Sanjiv Kumar. Plugin estimator of selective classification with out-of-distribution detection. In *ICLR*, 2024.
- [17] P. Nash and Edward James Anderson. *Linear programming in infinite-dimensional spaces : theory and applications*. Wiley, 1987.
- [18] Jerzy Neyman and Egon Pearson. On the use and interpretation of certain test criteria for purpose of statistical inference. *Biometrika*, pages 175–240, 1928.
- [19] Tadeusz Pietraszek. Optimizing abstaining classifiers using ROC analysis. In *Proceedings of the 22nd International Conference on Machine Learning*, page 665–672, 2005.
- [20] Yue Song, Nicu Sebe, and Wei Wang. Rankfeat: Rank-1 feature removal for out-of-distribution detection. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 17885–17898. Curran Associates, Inc., 2022.
- [21] Yiyun Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 144–157. Curran Associates, Inc., 2021.
- [22] Yiyun Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 20827–20840. PMLR, Jul 2022.

- [23] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. Vim: Out-of-distribution with virtual-logit matching. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4911–4920, 2022.
- [24] Guoxuan Xia and Christos-Savvas Bouganis. Augmenting softmax information for selective classification with out-of-distribution data. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 1995–2012, December 2022.

A Appendix (proof of the main theorem)

For a metric vector space V and its subset $A \subseteq V$, we write $\text{conv}(A)$, $\text{int}(A)$ and $\text{span}(A)$ to denote the convex hull, interior, and span of A , respectively.

Lemma 2. *Let A be a subset of \mathbb{R}^2 with $\dim \text{span}(A) = 2$. Given a point x located in the interior of the convex hull of A , there exists a convex polygon P such that x lies in the interior of P . Moreover, all vertices of the polygon P are elements of the set A .*

Proof. Given that $x \in \text{int}(\text{conv}(A))$, there is a triangle $T \subseteq \text{conv}(A)$ such that $x \in \text{int}(T)$. According to Carathéodory's theorem [1], each vertex of T lies in some simplex whose vertices are in the set A . Let these simplices be denoted as S_1, S_2 , and S_3 . Defining $P = \text{conv}(S_1 \cup S_2 \cup S_3)$ forms a convex polygon P with vertices in A , and x is in the interior of P . \square

Proof of Theorem 1. Let $\bar{\mathcal{X}} = \{x \in \mathcal{X} \mid p(x) > 0\}$. For all $x \in \bar{\mathcal{X}}$, we define $g(x) = \frac{q(x)}{p(x)}$ and $r(x) = \frac{R(x)}{p(x)}$. We will associate elements of $\bar{\mathcal{X}}$ with points in a plane by introducing the mapping $\varkappa : \bar{\mathcal{X}} \rightarrow \mathbb{R}_+^2$ such that $\varkappa(x) = (g(x), r(x))$.

The open ε -ball of $\pi(x)$ in \mathbb{R}^2 is denoted as $B_\varepsilon(x)$.

For a measurable $A \subseteq \mathbb{R}_+^2$, define

$$p(A) = \int_{\varkappa^{-1}(A)} p(x) dx, \quad (30)$$

$$c(A) = \begin{cases} \frac{1}{p(A)} \int_{\varkappa^{-1}(A)} p(x) c(x) dx & \text{if } p(A) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (31)$$

Problem (17) depends only on the subset

$$\mathcal{X}^+ = \{x \in \bar{\mathcal{X}} \mid \forall \varepsilon > 0 : p(B_\varepsilon(x)) > 0\} \subseteq \mathcal{X}.$$

To see this, let us define $Y_1 = \{x \in \mathcal{X} \mid p(x) = 0\}$ and $Y_2 = \{x \in \bar{\mathcal{X}} \mid \exists \varepsilon > 0 : p(B_\varepsilon(x)) = 0\}$. Then we have

$$\int_{\mathcal{X} \setminus \mathcal{X}^+} p(x) dx = \int_{Y_1} p(x) dx + \int_{Y_2} p(x) dx = 0, \quad (32)$$

since, for each open ε -ball $B_\varepsilon(x)$, there exists a circle $Q(x) \subset B_\varepsilon(x)$ that contains x and has rational radius and center coordinates. Thus, Y_2 can be covered by countably many such circles on which the integral of p is zero. As a result, (32) implies that elements of $\mathcal{X} \setminus \mathcal{X}^+$ do not affect the left-hand side of the first constraint and do not negatively impact either the objective function or the second constraint. More precisely, for any feasible function $c : \mathcal{X} \rightarrow [0, 1]$, if

$$c'(x) = \begin{cases} c(x) & \text{if } x \in \mathcal{X}^+, \\ 0 & \text{otherwise,} \end{cases} \quad (33)$$

then it holds $\int_{\mathcal{X}} R(x) c'(x) dx \leq \int_{\mathcal{X}} R(x) c(x) dx$, $\int_{\mathcal{X}} p(x) c'(x) dx = \int_{\mathcal{X}} p(x) c(x) dx$, and $\int_{\mathcal{X}} q(x) c'(x) dx \leq \int_{\mathcal{X}} q(x) c(x) dx$.

For the next considerations, we further distinguish three significant subsets of \mathcal{X}^+ with respect to a fixed optimal solution $c^* : \mathcal{X} \rightarrow [0, 1]$.

$$\begin{aligned} \mathcal{X}_0 &= \{x \in \mathcal{X}^+ \mid \exists \varepsilon > 0 : c^*(B_\varepsilon(x)) = 0\}, \\ \mathcal{X}_1 &= \{x \in \mathcal{X}^+ \mid \exists \varepsilon > 0 : c^*(B_\varepsilon(x)) = 1\}, \\ \mathcal{X}_2 &= \{x \in \mathcal{X}^+ \mid \forall \varepsilon > 0 : 0 < c^*(B_\varepsilon(x)) < 1\}. \end{aligned}$$

Note that

- $\mathcal{X}_0 \cup \mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}^+$,
- $c^*(B_\varepsilon(x)) = 0 \Rightarrow \forall 0 < \varepsilon' \leq \varepsilon : c^*(B_{\varepsilon'}(x)) = 0$, and

- $c^*(B_\varepsilon(x)) = 1 \Rightarrow \forall 0 < \varepsilon' \leq \varepsilon : c^*(B_{\varepsilon'}(x)) = 1$.

To confirm the existence of suitable λ and μ , stated by the theorem for c^* , it is sufficient to demonstrate that the sets

$$A_0 = \{\varkappa(x) \mid x \in \mathcal{X}_0\}, \quad (34)$$

$$A_1 = \{\varkappa(x) \mid x \in \mathcal{X}_1\}, \quad (35)$$

$$A_2 = \{\varkappa(x) \mid x \in \mathcal{X}_2\} \quad (36)$$

are "almost" linearly separable. This necessitates finding a line $L \subset \mathbb{R}^2$ that includes A_2 and linearly separates $A_0 \setminus L$ and $A_1 \setminus L$. The existence of such an L is ensured if

$$\dim \text{span}((\text{conv}(A_0) \cap \text{conv}(A_1)) \cup A_2) < 2. \quad (37)$$

The proof continues by listing some general configurations of points from A_0, A_1, A_2 that contradict the optimality of c^* . These forbidden configurations will be used to confirm the validity of condition (37).

But we firstly establish estimates on $\int_{\varkappa^{-1}(B_\varepsilon(y))} q(x)c^*(x)dx$ and $\int_{\varkappa^{-1}(B_\varepsilon(y))} R(x)c^*(x)dx$, for each $y \in \mathcal{X}$ and $\varepsilon > 0$.

For a given $y \in \mathcal{X}$, let $\varkappa(y) = (a, b) = \left(\frac{q(y)}{p(y)}, \frac{R(y)}{p(y)}\right)$.

Since, for all $x \in \varkappa^{-1}(B_\varepsilon(y))$, it holds

$$a - \frac{\varepsilon}{2} \leq \frac{q(x)}{p(x)} \leq a + \frac{\varepsilon}{2}, \quad (38)$$

$$b - \frac{\varepsilon}{2} \leq \frac{R(x)}{p(x)} \leq b + \frac{\varepsilon}{2}, \quad (39)$$

we easily obtain

$$\left(a - \frac{\varepsilon}{2}\right) p(B_\varepsilon(y))c^*(B_\varepsilon(y)) \leq \int_{\varkappa^{-1}(B_\varepsilon(y))} q(x)c^*(x)dx \leq \left(a + \frac{\varepsilon}{2}\right) p(B_\varepsilon(y))c^*(B_\varepsilon(y)), \quad (40)$$

and

$$\left(b - \frac{\varepsilon}{2}\right) p(B_\varepsilon(y))c^*(B_\varepsilon(y)) \leq \int_{\varkappa^{-1}(B_\varepsilon(y))} R(x)c^*(x)dx \leq \left(b + \frac{\varepsilon}{2}\right) p(B_\varepsilon(y))c^*(B_\varepsilon(y)). \quad (41)$$

Claim 1.1. Let $x_1, x_2 \in \mathcal{X}^+$, $\frac{q(x_1)}{p(x_1)} > \frac{q(x_2)}{p(x_2)}$, and $\frac{R(x_1)}{p(x_1)} > \frac{R(x_2)}{p(x_2)}$. If c^* is optimal, then $x_1 \in \mathcal{X}_0$ or $x_2 \in \mathcal{X}_1$.

Proof of the claim. We proceed by contradiction. Assume that $x_1 \in \mathcal{X}_1 \cup \mathcal{X}_2$ and $x_2 \in \mathcal{X}_0 \cup \mathcal{X}_2$. This implies the existence of an $\varepsilon > 0$ such that $c^*(B_\varepsilon(x_1)) > 0$ and $c^*(B_\varepsilon(x_2)) < 1$. Without loss of generality, choose ε sufficiently small so that

$$\varepsilon \leq \frac{q(x_1)}{p(x_1)} - \frac{q(x_2)}{p(x_2)} \quad (42)$$

and

$$\varepsilon \leq \frac{R(x_1)}{p(x_1)} - \frac{R(x_2)}{p(x_2)}, \quad (43)$$

which further implies $B_\varepsilon(x_1) \cap B_\varepsilon(x_2) = \emptyset$.

For $i \in \{1, 2\}$, denote $B_i = B_\varepsilon(x_i)$, $a_i = \frac{q(x_i)}{p(x_i)}$, and $b_i = \frac{R(x_i)}{p(x_i)}$.

Define a function $c' : \mathcal{X} \rightarrow [0, 1]$ as follows:

$$c'(x) = \begin{cases} c^*(x) - \nu \frac{p(B_1)}{c^*(B_1)} c^*(x) & \text{if } x \in B_1, \\ c^*(x) + \nu \frac{p(B_1)}{p(B_2) - c^*(B_2)} (1 - c^*(x)) & \text{if } x \in B_2, \\ c^*(x) & \text{otherwise.} \end{cases}$$

Here, the parameter ν is defined as:

$$\nu = \min \left\{ c^*(B_1), \frac{p(B_2)}{p(B_1)}(1 - c^*(B_2)) \right\} > 0,$$

which ensures $0 < \nu \frac{p(B_1)}{c^*(B_1)} \leq 1$ as well as $0 < \nu \frac{p(B_1)}{p(B_2) - c^*(B_2)} \leq 1$, hence c' is defined correctly.

By comparing c^* and c' , we will show that c^* is not optimal.

Observe that

$$\begin{aligned} \int_{\mathcal{X}} p(x)c'(x)dx - \int_{\mathcal{X}} p(x)c^*(x)dx &= \\ -\nu \frac{p(B_1)}{c^*(B_1)}c^*(B_1) + \nu \frac{p(B_1)}{p(B_2) - c^*(B_2)}(p(B_2) - c^*(B_2)) &= 0, \end{aligned}$$

Moreover,

$$\begin{aligned} \int_{\mathcal{X}} q(x)c'(x)dx - \int_{\mathcal{X}} q(x)c^*(x)dx &= -\nu \frac{p(B_1)}{c^*(B_1)} \int_{\mathcal{X}^{-1}(B_1)} q(x)c^*(x)dx \\ &+ \nu \frac{p(B_1)}{p(B_2) - c^*(B_2)} \int_{\mathcal{X}^{-1}(B_2)} q(x)(1 - c^*(x))dx \\ &\stackrel{(40)}{\leq} -\nu \frac{p(B_1)}{c^*(B_1)} \left(a_1 - \frac{\varepsilon}{2} \right) c^*(B_1) + \nu \frac{p(B_1)}{p(B_2) - c^*(B_2)} \left(a_2 + \frac{\varepsilon}{2} \right) (p(B_2) - c^*(B_2)) \\ &= \nu p(B_1)(a_2 - a_1 + \varepsilon) \stackrel{(42)}{<} 0. \end{aligned}$$

Analogously, we can derive

$$\int_{\mathcal{X}} R(x)c'(x)dx - \int_{\mathcal{X}} R(x)c^*(x)dx \leq \nu p(B_1)(b_2 - b_1 + \varepsilon) \stackrel{(43)}{<} 0.$$

The relationships above contradict the optimality of c^* . ■

Claim 1.2. Let x_1, x_2, x_3 be distinct elements of \mathcal{X}^+ , and let $\alpha_1, \alpha_2, \beta$ be non-negative real numbers such that $\alpha_1 + \alpha_2 = 1$, $\beta \neq 0$, and $\beta \varkappa(x_3) = \alpha_1 \varkappa(x_1) + \alpha_2 \varkappa(x_2)$. If c^* is optimal, it holds:

- If $\beta < 1$, then either $x_3 \in \mathcal{X}_0$ or $\{x_1, x_2\} \cap \mathcal{X}_1 \neq \emptyset$,
- if $\beta > 1$, then either $x_3 \in \mathcal{X}_1$ or $\{x_1, x_2\} \cap \mathcal{X}_0 \neq \emptyset$.

Proof of the claim. We will give a proof for $\beta < 1$ and note that the steps for $\beta > 1$ are analogous.

For $i \in \{1, 2, 3\}$, denote $\varkappa(x_i) = (a_i, b_i)$. Find an $\varepsilon > 0$ such that

$$\varepsilon < (1 - \beta)a_3, \tag{44}$$

$$\varepsilon < (1 - \beta)b_3, \tag{45}$$

and $B_\varepsilon(x_1), B_\varepsilon(x_2), B_\varepsilon(x_3)$ are pairwise disjoint. By contradiction, assume that $c^*(B_\varepsilon(x_1)) < 1$, $c^*(B_\varepsilon(x_2)) < 1$, and $c^*(B_\varepsilon(x_3)) > 0$.

To simplify the notation, for $i \in \{1, 2, 3\}$, let $B_i = B_\varepsilon(x_i)$, $c_i = c^*(B_i)$, and $p_i = p(B_i)$. Note that, according to the definition of \mathcal{X}^+ , it holds $p_1, p_2, p_3 > 0$.

Define a selective function $c' : \mathcal{X} \rightarrow [0, 1]$ as follows:

$$c'(x) = \begin{cases} c^*(x) + \frac{\nu \alpha_1 p_3}{(1 - c_1) p_1} (1 - c^*(x)) & \text{if } x \in B_1, \\ c^*(x) + \frac{\nu \alpha_2 p_3}{(1 - c_2) p_2} (1 - c^*(x)) & \text{if } x \in B_2, \\ c^*(x) - \frac{\nu}{c_3} c^*(x) & \text{if } x \in B_3, \\ c^*(x) & \text{otherwise.} \end{cases}$$

Taking $\frac{z}{0} = \infty$ for any $z > 0$, the value of ν is given by:

$$\nu = \min \left\{ c_3, \frac{p_1}{\alpha_1 p_3} (1 - c_1), \frac{p_2}{\alpha_2 p_3} (1 - c_2) \right\}.$$

It is easy to see that $\nu > 0$, and $0 < \frac{\nu \alpha_1 p_3}{(1 - c_1) p_1}, \frac{\nu \alpha_2 p_3}{(1 - c_2) p_2}, \frac{\nu}{c_3} \leq 1$.

Now, observe that

$$\begin{aligned} \int_{\mathcal{X}} p(x) c'(x) dx - \int_{\mathcal{X}} p(x) c^*(x) dx &= \int_{\varkappa^{-1}(B_1)} \frac{\nu \alpha_1 p_3}{(1 - c_1) p_1} p(x) (1 - c^*(x)) dx \\ &+ \int_{\varkappa^{-1}(B_2)} \frac{\nu \alpha_2 p_3}{(1 - c_2) p_2} p(x) (1 - c^*(x)) dx - \int_{\varkappa^{-1}(B_3)} \frac{\nu p_3}{c_3} p(x) c^*(x) dx \\ &= \nu \alpha_1 p_3 + \nu \alpha_2 p_3 - \nu p_3 = \nu p_3 (\alpha_1 + \alpha_2 - 1) = 0. \end{aligned}$$

Next, derive

$$\begin{aligned} \int_{\mathcal{X}} q(x) c'(x) dx - \int_{\mathcal{X}} q(x) c^*(x) dx &= \int_{\varkappa^{-1}(B_1)} \frac{\nu \alpha_1 p_3}{(1 - c_1) p_1} q(x) (1 - c^*(x)) dx \\ &+ \int_{\varkappa^{-1}(B_2)} \frac{\nu \alpha_2 p_3}{(1 - c_2) p_2} q(x) (1 - c^*(x)) dx - \int_{\varkappa^{-1}(B_3)} \frac{\nu p_3}{c_3} q(x) c^*(x) dx \\ &\stackrel{(40)}{\leq} \nu \alpha_1 p_3 \left(a_1 + \frac{\varepsilon}{2} \right) + \nu \alpha_2 p_3 \left(a_2 + \frac{\varepsilon}{2} \right) - \nu p_3 \left(a_3 - \frac{\varepsilon}{2} \right) \\ &= \nu p_3 \left(\alpha_1 a_1 + \alpha_2 a_2 - a_3 + \frac{\varepsilon}{2} (\alpha_1 + \alpha_2) + \frac{\varepsilon}{2} \right) \\ &= \nu p_3 (\beta a_3 - a_3 + \varepsilon) \stackrel{(44)}{<} 0 \end{aligned}$$

Quite similarly,

$$\int_{\mathcal{X}} R(x) c'(x) dx - \int_{\mathcal{X}} R(x) c^*(x) dx \leq \nu p_3 (\beta b_3 - b_3 + \varepsilon) \stackrel{(45)}{<} 0.$$

■

Claim 1.3. Let $z_1, \dots, z_n \in \mathcal{X}^+$ and $x \in \mathcal{X}^+$, and let P be a convex polygon with vertices $\varkappa(y_1), \dots, \varkappa(y_n)$, with $\varkappa(x)$ as its inner point. The function c^* is not optimal if one of the following cases holds:

1. All z_i 's are from \mathcal{X}_0 and $x \in \mathcal{X}_1$.
2. All z_i 's are from \mathcal{X}_1 and $x \in \mathcal{X}_0$.

Proof of the claim. For the first case, we identify an edge of the polygon P that is first intersected by the half-line H starting from the origin $(0, 0)$ and passing through the point $\varkappa(x)$ (see Figure 1a). This edge has vertices $\varkappa(y_1)$ and $\varkappa(y_2)$ for some $y_1, y_2 \in \{z_1, \dots, z_n\}$. The points $\varkappa(x)$, $\varkappa(y_1)$, and $\varkappa(y_2)$ are non-collinear and satisfy $\beta \varkappa(x) = \alpha_1 \varkappa(y_1) + \alpha_2 \varkappa(y_2)$ for some non-negative real numbers α_1, α_2 , and β , where $\alpha_1 + \alpha_2 = 1$ and $\beta < 1$. According to Claim 1.2, these points contradict the optimality of c^* .

Similarly, for the second case, we identify an edge of P that is the last one to be intersected by the half-line H (see Figure 1b). Let $\varkappa(y_1)$ and $\varkappa(y_2)$ be vertices of this edge for some $y_1, y_2 \in \{z_1, \dots, z_n\}$. The points $\varkappa(x)$, $\varkappa(y_1)$, and $\varkappa(y_2)$ satisfy $\beta \varkappa(x) = \alpha_1 \varkappa(y_1) + \alpha_2 \varkappa(y_2)$ for some non-negative real numbers α_1, α_2 , and β , where $\alpha_1 + \alpha_2 = 1$ and $\beta > 1$. Again, Claim 1.2 contradicts the optimality of c^* . ■

Claim 1.4. If $x_1, x_2 \in \mathcal{X}_0$ and $x_3, x_4 \in \mathcal{X}_1$, and the line segments with endpoints $\varkappa(x_1), \varkappa(x_2)$, and $\varkappa(x_3), \varkappa(x_4)$ intersect at a point lying in the interior of both segments, then the function c^* is not optimal.

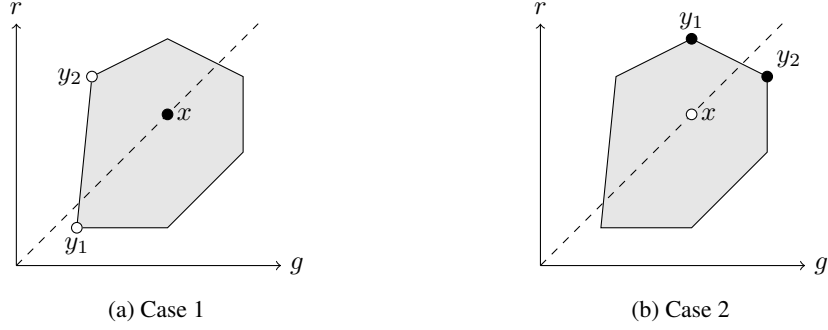


Figure 1: The infeasible configuration described by Claim 1.3. For simplicity, points are denoted by their pre-images from \mathcal{X}^+ .

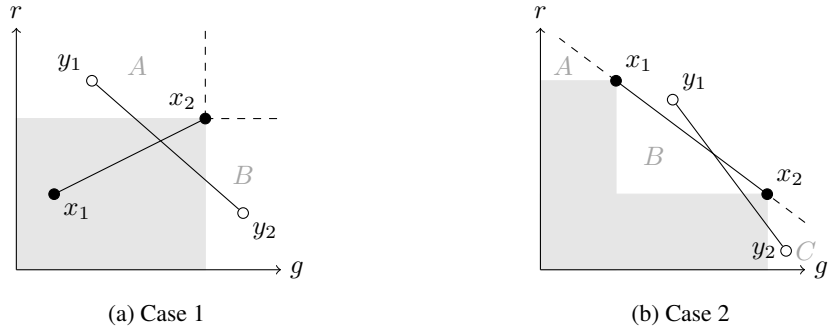


Figure 2: The infeasible configuration described by Claim 1.4.

Proof of the claim.

We consider two possible situations depicted in Figure 2.

First, we examine the scenario where one of the points $\varkappa(x_1)$, $\varkappa(x_2)$ is located lower-left relative to the other one, specifically, we assume $g(x_1) \leq g(x_2)$ and $r(x_1) \leq r(x_2)$, as illustrated in Figure 2a. According to Claim 1.1, the points $\varkappa(y_1)$, $\varkappa(y_2)$ cannot be situated within the shaded rectangular area. Since the line segments intersect, one of the points $\varkappa(y_1)$, $\varkappa(y_2)$ must be positioned upper-left from $\varkappa(x_2)$, while the other lies lower-right from $\varkappa(x_2)$ (these regions are labeled as A and B, respectively, in Figure 1a). However, such an arrangement contradicts Claim 1.2.

Secondly, we consider the configuration depicted in Figure 2b. One of the points $\varkappa(y_1)$, $\varkappa(y_2)$ must lie within the half-plane bounded by the line passing through $\varkappa(x_1)$, $\varkappa(x_2)$, and containing the origin $(0, 0)$. Let's assume this point is y_2 . As per Claim 1.1, $\varkappa(y_2)$ cannot reside in the shaded area. Within this considered half-plane, there are three remaining triangular regions denoted as A, B, and C. Nevertheless, Claim 1.2 prohibits $\varkappa(y_2)$ from lying within region B (by applying the claim to the triple x_1, x_2 , and y_2), region C (consider the triple y_1, y_2 , and x_2), as well as region A (consider the triple y_1, y_2 , and x_1). ■

Claim 1.5. *If $z_1, z_2 \in \mathcal{X}_2$ and $\varkappa(z_1)$, $\varkappa(z_2)$ are distinct points, then the slope of the line L passing through $\varkappa(z_1)$ and $\varkappa(z_2)$ is not positive. Moreover, the open half-plane determined by L and the origin $(0, 0)$ does not contain any point from $A_1 \cup A_2$, and similarly, the opposite open half-plane does not contain any point from $A_0 \cup A_2$.*

Proof of the claim. If the slope of L is positive, as shown in Figure 3a, then the points $\varkappa(z_1)$ and $\varkappa(z_2)$ conform to Claim 1.1, implying that c^* is not optimal.

If the slope of L is not positive, referring to Figure 3b, we deduce that the open half-plane H_0 does not contain any point $\varkappa(x)$ from $A_0 \cup A_2$. By Claim 1.1, such a point P cannot be in the shaded area. It neither can be in areas A, B, or C because in each case, the points x, z_2 , and z_1 , or z_1, z_2 , and x , or z_1, x , and z_2 would conform to Claim 1.2.

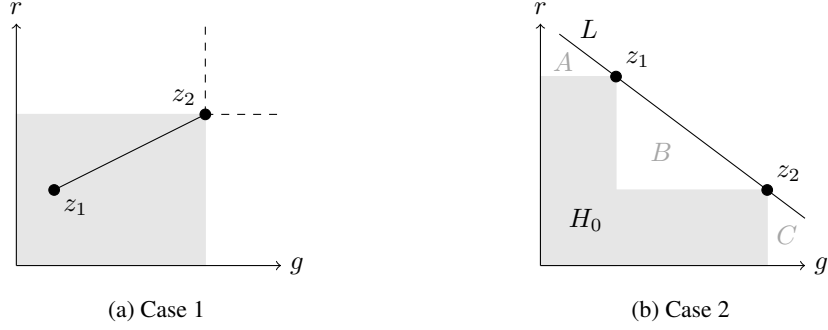


Figure 3: The infeasible configuration described by Claim 1.5.

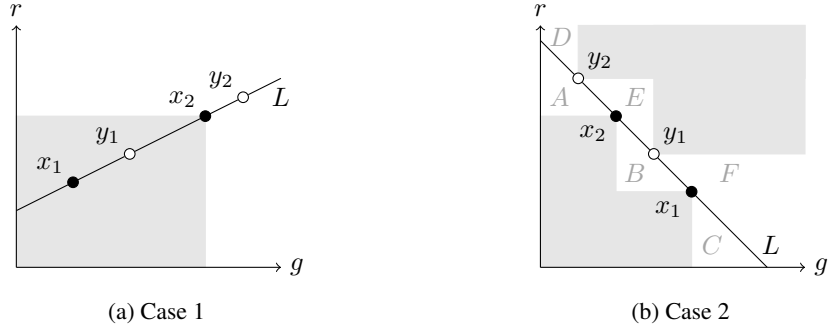


Figure 4: The infeasible configuration described by Claim 1.6.

Similarly, it can be argued that a point from $A_1 \cup A_2$ cannot lie in the opposite open half-plane. ■

Claim 1.6. *Let $x_1, x_2 \in \mathcal{X}_1$, $y_1, y_2 \in \mathcal{X}_0$, $z \in \mathcal{X}_2$. If $\varkappa(x_1)$, $\varkappa(x_2)$, $\varkappa(y_1)$, and $\varkappa(y_2)$ are distinct points lying on a line L in such a way that $\varkappa(y_1)$ is between $\varkappa(x_1)$ and $\varkappa(x_2)$, and $\varkappa(x_2)$ is between $\varkappa(y_1)$ and $\varkappa(y_2)$, then $\varkappa(z)$ lies on L unless the function c^* is not optimal.*

Proof of the claim.

Firstly, observe that the configuration in Figure 4a is not admissible for an optimal c^* , because $\varkappa(y_1)$ is located in the shaded area containing points to the left and down from $\varkappa(x_2)$ (Claim 1.1 applies here).

Hence, it suffices to inspect the case with the opposite slope of L depicted in Figure 4b. Note that, without loss of generality, we assume that the highest point on L is $\varkappa(y_2)$. The case when the highest point is either $\varkappa(x_1)$ or $\varkappa(x_2)$ is handled similarly. If c^* is optimal, then $\varkappa(z)$ cannot lie in the two shaded areas (Claim 1.1). The remaining area is split into 6 different parts by L , denoted from A to F . If $\varkappa(z)$ lies in any of these parts, then we can always select two elements from $\{x_1, x_2, y_1, y_2\}$ that together with z conform to Claim 1.2. These selections are as follows:

$$\begin{array}{lll}
 A: z, y_1, x_2, & B: x_1, x_2, z, & C: z, y_1, x_1, \\
 D: z, x_2, y_2, & E: y_1, y_2, z, & F: z, x_2, y_1.
 \end{array}$$

We are ready to confirm condition (37), by listing all the potential infeasible cases and showing that they contradict the optimality of c^* .

1. $\dim \text{span}(\text{conv}(A_0) \cap \text{conv}(A_1)) = 2$. We distinguish three subcases.

Assume that, for some $x \in \mathcal{X}_0$, $\varkappa(x)$ is in the interior of $\text{conv}(A_0) \cap \text{conv}(A_1)$. By Lemma 2, there is a convex polygon P with vertices in $\varkappa(\mathcal{X}_1)$ such that the interior of P contains x . Apply Claim 1.3.

Next, assume that x belongs to \mathcal{X}_1 instead of \mathcal{X}_0 . Analogously, find a convex polygon P' with vertices in $\mathcal{X}(\mathcal{X}_0)$, containing x in its interior. Apply Claim 1.3.

Finally, if $\text{int}(\text{conv}(A_0) \cap \text{conv}(A_1)) \cap (A_0 \cup A_1) = \emptyset$, take an arbitrary $x \in \text{int}(\text{conv}(A_0) \cap \text{conv}(A_1))$. By Carathéodory's theorem, there exists a simplex S_0 (a line segment or triangle) with vertices in A_0 such that $x \in \text{int}(S_0)$. Similarly, there exists a simplex S_1 (a line segment or triangle) with vertices in A_1 such that $x \in \text{int}(S_1)$. It is not difficult to see that the boundary line segments of S_0 and S_1 intersect, meaning there are two intersecting line segments, one with endpoints in A_0 , and the other with endpoints in A_1 . Hence, Claim 1.4 contradicts the optimality of c^* .

2. $\dim \text{span}(A_2) = 2$. There exist $x_1, x_2, x_3 \in \mathcal{X}_2$ such that $\mathcal{X}(x_1), \mathcal{X}(x_2), \mathcal{X}(x_3)$ are non-collinear. Such a configuration is prohibited by Claim 1.5.
3. $\dim \text{span}(\text{conv}(A_0) \cap \text{conv}(A_1)) = 1$, $\dim \text{span}(A_2) \leq 1$ and $\dim \text{span}((\text{conv}(A_0) \cap \text{conv}(A_1)) \cup A_2) = 2$. There are $x_1, x_2 \in \mathcal{X}_0$, $x_3, x_4 \in \mathcal{X}_1$, $x_5 \in \mathcal{X}_2$ conforming to Claim 1.6.
4. $\dim \text{span}(\text{conv}(A_0) \cap \text{conv}(A_1)) = 0$, $\dim \text{span}(A_2) = 1$, and $\dim \text{span}((\text{conv}(A_0) \cap \text{conv}(A_1)) \cup A_2) = 2$. There are $z_1, z_2 \in \mathcal{X}_2$ such that $\mathcal{X}(z_1)$ and $\mathcal{X}(z_2)$ conform to Claim 1.5, i.e., the line L passing through $\mathcal{X}(z_1)$ and $\mathcal{X}(z_2)$ determines two opposite half-planes, H_0 and H_1 . By Claim 1.5, $A_0 \subseteq H_0$ and $A_1 \subseteq H_1$. This means that $\text{conv}(A_0) \cap \text{conv}(A_1) \subseteq L$, which contradicts the assumption $\dim \text{span}((\text{conv}(A_0) \cap \text{conv}(A_1)) \cup A_2) = 2$.

□

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: The contribution is summarized in the abstract and explained in the introductory section.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: The paper explains that tuning the parameters of the score function that characterizes optimal solutions must be approached individually for each problem. Additionally, it notes that extending the main result to incorporate more constraints remains an open problem.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The full proof of the main theorem is in the appendix, the assumptions are stated in the main paper.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: The paper does not include any experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: The paper does not include any experiments.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: The paper does not include any experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: The paper does not include any experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]

Justification: The paper does not include any experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics [https://neurips.cc/public/EthicsGuidelines?](https://neurips.cc/public/EthicsGuidelines)

Answer: [Yes]

Justification: Our research conducted in the paper conforms with the NeurIPS Code of Ethics which we have reviewed.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: We present a mathematical result which does not have any societal impact.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not utilize any data nor models.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use any assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset’s creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.