UNIVERSITAT DE BARCELONA

FUNDAMENTAL PRINCIPLES OF DATA SCIENCE MASTER'S THESIS

# Distance-based copying of machine learning classifiers

*Author:*
Rubén JIMÉNEZ LUMBRERAS

*Supervisor:*
Dr. Oriol PUJOL VILA

*A thesis submitted in partial fulfillment of the requirements for the degree of MSc in Fundamental Principles of Data Science*

*in the*

Facultat de Matemàtiques i Informàtica

December 21, 2025

UNIVERSITAT DE BARCELONA

# *Abstract*

Facultat de Matemàtiques i Informàtica

MSc

**Distance-based copying of machine learning classifiers**

by Rubén JIMÉNEZ LUMBRERAS

Copying machine learning black box classifiers is a key framework that allows practitioners to upgrade their old models, enriching them with new properties, changing their architectures or adapting them to comply with the current AI legislations. Thanks to the copying techniques and assumptions, these improvements can be done even in settings where retraining the original system from scratch is not possible, due to resource, protocol or availability constraints. In this work, we propose the use of signed distances to the decision boundary as a replacement of the black box hard labels used to build the copies, and introduce two different algorithms to compute these distances. In addition, we observe that distance-based copying could behave as a model-agnostic regularization technique and develop a flexible framework to reduce the generalization error of the copies. Then, we validate these proposals through a series of experiments on synthetic datasets and real problems. Results show that distance-based copying is successful across multiple relevant settings and evaluation metrics. Furthermore, results also validate the quality of the predicted distances and their potential as uncertainty measures.

# *Acknowledgements*

I would like to deeply thank Dr. Oriol Pujol Vila for all the support, time, dedication and help he has given me throughout this project, which has made it possible.

I would also like to thank my master's classmates and professors for their help and the good moments I had the pleasure of sharing with them during this past year.

Finally, I also want to thank my family for all their support, help, and advice. Thanks to them, I have been able to get to where I am and become who I am.

# Contents

# Chapter 1

# Introduction

Nowadays, society is going through a period of very fast development of machine learning models, where the state of the art is evolving at an increasing rate (Lu, 2019; Azad and Banu, 2024). Parallel to this expansion, new legislation about transparency (Lund et al., 2025), privacy (Ye et al., 2024; Korobenko, Nikiforova, and Sharma, 2024; Parinandi et al., 2024) or explainability for AI systems (Panigutti et al., 2023; Nannini, Balayn, and Smith, 2023; Gade et al., 2019) is also being implemented around the world, aiming to answer the growing societal concerns.

As a consequence, it is vital for practitioners to be able to react to these changes and keep themselves up to date, upgrading their old models to more suitable architectures or adapting them to comply with new legislation (Wong, Chong, and Aspegren, 2023). Nevertheless, these are challenging problems that do not have a simple and immediate solution, because in many occasions the data that was used to train the system is no longer available or for example because it may be necessary to add or remove features from their previous models.

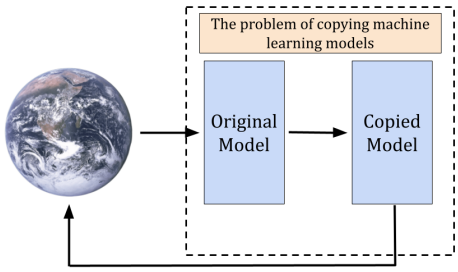



FIGURE 1.1: Graphical representation of the copying problem.

All of these issues can, in fact, be regarded as specific instances of the broader problem of copying[1] classifiers (see Fig. 1.1). In this setting, given a black box classifier that cannot be inspected and only outputs hard labels, the goal is to replicate its decision boundary. That is, the aim is to build a new classifier that has the same decision boundary as the original one. This framework has been successfully applied to address the aforementioned challenges in numerous cases (Wood-Doughty, Cachola, and Dredze, 2022; Unceta, Nin, and Pujol, 2020c; Unceta, Nin, and Pujol, 2021; Goldsteen et al., 2022) although, in practice, practitioners often care not only about fidelity (how closely the copy reproduces the black box) but also about its accuracy.

Still, the fact that these black boxes only provide class labels, without any additional information, remains as a key limitation. This lack of context is even more

[1]Throughout this work, the term copying is used to denote model-agnostic and data-agnostic distillation, as introduced by (Unceta, Nin, and Pujol, 2020a).

noticeable for example when the decision boundary that should be copied has complex yet smooth shapes, that the copy cannot exploit because it lacks the additional guidance needed to fully leverage its inductive biases and recover the global structure. Moreover, another challenging scenario where this absence of information is especially damaging are high dimensional settings, where the combination of the unavoidable data scarcity and the curse of dimensionality (Altman and Krzywinski, 2018; Peng, Gui, and Wu, 2025; Worel, 2023) significantly limits the ability to find reliable copies. To solve these problems, past efforts have mainly been focused on how to choose the synthetic dataset that is used to build the copy (Unceta et al., 2020; Heo et al., 2019), trying for example to explore better the boundary. Until now, the hard labels themselves have received little attention.

In this thesis, we propose a way to enrich these labels, substituting them by signed distances to the decision boundary of the classifier. The objective of this change is to mitigate the downsides of working with a black box but without actually violating that assumption. In addition and more generally, having access to predicted distances instead of classes enriches these copies through differential replication[2], providing them with an uncertainty measure that can be used to assess the confidence in their predictions. Not only that, but thanks to the mathematical regularity of distances, working with them makes the resulting copies smooth, something that can produce accurate and naturally regularized copies.

We summarize the main contributions of this thesis as:

- We propose the use of signed distances to the decision boundary as a replacement for the hard black box labels, aiming to facilitate the copying process and to enrich models with a useful uncertainty measure.
- We introduce and compare two different algorithms to compute these signed distances and discuss the trade-off between their quality and quantity, in a model-agnostic and data-agnostic setting.
- We explore the regularization effect exhibited by distance copies, analysing its causes and implications. We use these ideas to introduce a new model-agnostic and implicit regularization approach based on distance copying.
- We validate these distance copying proposals through experimentation in two-dimensional synthetic problems, designed to show the strengths and weaknesses of these copies, as well as in real UCI datasets.

The rest of this report is organized as follows. In Ch. 2 we present a literature survey of related work. From there, we introduce the proposal in Ch. 3, explaining it and exploring two main approaches to computing the distances. Once this new copying framework has been set, in Ch. 4 we validate it on several two-dimensional synthetic datasets, that make possible to perform a visual inspection of the copies, as well as on various UCI problems. After that, in Ch. 5, we analyse the obtained results in detail, examining the behaviour of these copies and discussing their strengths and limitations. Finally, the project ends with a summary of the main conclusions.

Finally, regarding the use of LLM, this thesis employed ChatGPT and Microsoft Copilot to improve language clarity, rephrase certain sentences in an academic tone and generate initial code scripts. All data analysis, interpretation, and argumentation were conducted by the author.

[2]Differential replication refers to the process by which copies not only reproduce the black box's decision boundary, but also incorporate additional characteristics and properties that extend or modify the model's functionality in ways suited to the current environment. (Unceta, Nin, and Pujol, 2020b)

# Chapter 2

# Related Work

The copying problem for black box models has already been studied in many past works, although, as already explained, the emphasis has mainly been placed on developing its theory (Unceta, Nin, and Pujol, 2020a) and analysing the copy building (Statuto et al., 2023; Unceta, Nin, and Pujol, 2019) and data sampling processes (Unceta et al., 2020; Heo et al., 2019). Nevertheless, this is not the first time that the idea of working with one model to train another appears in the literature, something that has been studied under many perspectives and names, from teacher-student problems (Xu et al., 2022) and adversarial training (Papernot et al., 2017; Lowd and Meek, 2005) to model compression (Bucila, Caruana, and Niculescu-Mizil, 2006), model extraction (Bastani, Kim, and Bastani, 2019) or model distillation (Lian, Huang, and Wang, 2023; Tan et al., 2018; Hinton, Vinyals, and Dean, 2015) among others (Zeng and Martinez, 2000). In this context, it is also common to assume that the original model is not a black box, and there are many articles exploring how to use this extra knowledge (Jin, Wang, and Lin, 2023; Guo et al., 2024).

In addition, outside the scope of the copying problem itself, this proposal has also led to a new regularization technique through distance copying, that is, a way to control the complexity of the black box and improve its generalization ability on the original problem. In this sense, the topic of regularization is vast (Moradi, Berangi, and Minaei, 2020), encompassing many methodologies. Among them, the most common approach is to introduce an additive penalty term into the loss function (Kolluri et al., 2020; Wu and Xu, 2020), such as in $L1$ or $L2$ regularization, which explicitly constrains parameter magnitudes. Nevertheless, the regularization through distillation explored in this work constitutes an implicit form of regularization, achieved by modifying the dataset and targets used to build the new model. These methods have also been examined previously in the literature (Mosca and Magoulas, 2017; Mobahi, Farajtabar, and Bartlett, 2020; Borup and Andersen, 2021; Pareek, Du, and Oh, 2024), but past works rarely assume model-agnostic, black box constraints or provide a parametric way to control the extent of the regularization.

Apart from that, the idea of exploiting the distance to a classifier's decision boundary is not new. For example, prior works on this topic have enforced large distances during training to improve model robustness (Elsayed et al., 2018; Ding et al., 2020), estimated boundary proximity for sample selection (Ducoffe and Precioso, 2018), and analysed margin distributions to study model generalization gaps (Jiang et al., 2019). Furthermore, the use of distances as uncertainty measures has also been explored (Hashimoto, Kamigaito, and Watanabe, 2025; Liu et al., 2020), although some of these techniques can be susceptible to the curse of dimensionality (Pestov, 2013; Aggarwal, Hinneburg, and Keim, 2001), where pairwise distances tend to become increasingly similar and thereby weaken the reliability of such metrics. Nevertheless, none of the above approaches use distances as the direct supervised signal, leveraging them for classification, regularization, and uncertainty estimation.

# Chapter 3

# Distance-based copying

## 3.1 Proposal

To introduce the proposal, let $\mathcal{X}$ and $\mathcal{T} = \{-1, 1\}$ be the input and label spaces of the problem respectively, we can consider $f_{\mathcal{O}} : \mathcal{X} \longrightarrow \mathcal{T}$ the black box classifier to be copied, that only provides hard labels as predictions. Then, aiming to build this copy, we could sample a sequence $\mathscr{Z} = \{z_i\}_{i=1}^n$ on $\mathcal{X}$, following an appropriate method or distribution. From here, we would use the original model $f_{\mathcal{O}}$ to label the data points, building the synthetic dataset $\mathscr{Z}' = \{(z_i, f_{\mathcal{O}}(z_i))\}_{i=1}^n$ that then could be used to find[1] the copy $f_{\mathcal{C}} : \mathcal{X} \longrightarrow \mathcal{T}$.

In this scenario, we propose to train the copy $f_{\mathcal{C}}$ using signed distances to the decision boundary of the black box, instead of working with the hard labels $f_{\mathcal{O}}(z_i)$. Specifically, for any point $z_i$, the aim is to compute an approximation of its distance to the boundary and then multiply it by $f_{\mathcal{O}}(z_i)$ to find its signed distance $\ell_i$.

$$\ell_i = f_{\mathcal{O}}(z_i) \cdot \xi(z_i), \quad \xi(z_i) = \inf_{\substack{x \in \mathcal{X} \\ f_{\mathcal{O}}(x) \neq f_{\mathcal{O}}(z_i)}} d(z_i, x) \tag{3.1}$$

Then, using these targets, that encode both the class[2] and the distance in a single number, one can build a new synthetic dataset $\mathscr{Z}^* = \{(z_i, \ell_i)\}_{i=1}^n$ that would be used to train the copy instead of $\mathscr{Z}'$.

## 3.2 Algorithms

Once the idea has been introduced, we discuss how these distances to the decision boundary $\ell_i$ for the points in $\mathscr{Z}$ can be computed, proposing and comparing two distance, sampling and model-agnostic approaches.

On the one hand, a possibility is to put the focus on the quality of the computed distances, ensuring that they are very similar to the actual ones. To achieve this goal, we can apply Algorithm 1 (see Appendix A for the pseudocode), that iterates through each point $z_i$ in the synthetic dataset $\mathscr{Z}$ and computes the desired distance for each of them. To construct these approximations, we start with $z_i$ stored in a

[1] This approach to solving the copying problem and finding the corresponding copy follows the single-pass framework, which simplifies the task by framing it as a standard training procedure. Nevertheless, in general, the actual copying problem can be addressed through a dual optimization of both the copy's parameters and the synthetic samples, as implemented, for example, in (Unceta, Nin, and Pujol, 2019). While this alternative method remains fully compatible with the proposal of the project, we do not explore this direction for the sake of simplicity. From this point onward, we assume that copies are generated as the result of a training process on the synthetic dataset.

[2] This proposal seems to restrict ourselves to two class classification problems, but, in fact, one could extend these ideas to multiple classes, transforming the multi-class scenario into a series of two class problems with usual techniques. So, this assumption is not an actual limitation.

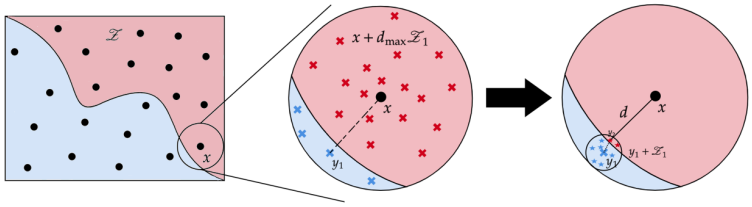


FIGURE 3.1: Graphical representation of Algorithm 1

variable $c$ and repeat the following procedure $it_{\max}$ times. First, we generate a cloud of points $c + d_{\max}\mathscr{B}$ of size $m$ centred at $c$, where $\mathscr{B}$ denotes a set sampled[3] once beforehand in the unit ball. Next, we identify the point in this cloud that minimizes $d(x, c),\ x \in c + d_{\max}\mathscr{B}$ such that $f_{\mathcal{O}}(x) \neq f_{\mathcal{O}}(c)$ and we store it in $c$. Finally, we decrease the cloud radius and repeat the process. After completing all iterations, we compute the distance between the final point $c$ and the original $z_i$. This process is shown in Fig. 3.1.

Nevertheless, even though Alg. 1 can produce high quality distances, its computational complexity is considerable, because iterating through all the dataset $\mathscr{Z}$ and finding the distances one by one requires many calls to the black box $f_{\mathcal{O}}$. Specifically, assuming that the maximum number of iterations $it_{\max}$ is always reached and that the cost of computing the distance $d$ is proportional to the dimension dim, this algorithm requires $n(it_{\max} \cdot m + 1)$ black box evaluations ($it_{\max} \cdot m + 1$ per labelled point) and around $O(it_{\max} \cdot n \cdot m \cdot \dim)$ additional operations.

To solve this problem, on the other hand, an alternative approach can be to put the focus on the quantity of points that we are able to label, trying to maximize it at the cost of reducing the precision on the distances. As an example of this idea, we propose Algorithm 2 (see Appendix A for the pseudocode), where we sample the synthetic dataset $\mathscr{Z}$ in small clusters that can be labelled at the same time. In detail, we start by sampling a first dataset $\mathscr{C}$ of size $n_c$, that determines the positions of the small clusters. Then, for each of these positions $c$, we center a small $\mathscr{B}_{\text{in}}$ and a big $\mathscr{B}_{\text{out}}$ cloud of points around it, labelling both of them with the black box $f_{\mathcal{O}}$. Finally, we use the outer cloud to compute the distances to the boundary of the points in the inner one, taking the closest point with a different label (see Fig. 3.2).

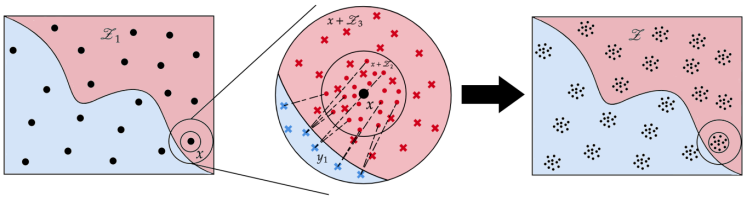


FIGURE 3.2: Graphical representation of Algorithm 2

Thanks to the use of these clusters, and denoting by $n_{\text{in}}$ and $n_{\text{out}}$ the number of points in the clouds $\mathscr{B}_{\text{in}}$ and $\mathscr{B}_{\text{out}}$ respectively, this algorithm requires $n_c(n_{\text{in}} + n_{\text{out}})$ black box evaluations ($1 + (n_{\text{out}}/n_{\text{in}})$ per labelled point) and around $O(n_c \cdot n_{\text{in}} \cdot n_{\text{out}} \cdot \dim)$ additional operations if we consider that the cost of computing the distance $d$

[3]The way in which the points of the set $\mathscr{B}$ have been sampled by the algorithm may seem unusual, because their distribution in the unit ball is neither uniform nor another distribution that evenly spaces the points in some sense. The reason behind this is that, with this choice, more points are generated around the centre of the ball rather than on its outskirts, something that reduces the error for small distances at the cost of increasing it for bigger ones. That controls the relative error of the computations and lowers the number of issues around the decision boundary, where misclassification is more likely.

scales linearly with the dimension dim. Since the roles of $m$ and $n_{\text{out}}$ are comparable between Alg. 1 and 2, this represents a significant reduction in computational cost: the number of black box calls decreases by a factor of $it_{\max} \cdot n_{\text{in}}$ compared to Alg. 1.

## 3.3 Regularization effect

In this subsection, we discuss the regularization effect that distance-based copying has and introduce a proposal to leverage it. However, before proceeding, it is important to clarify the meaning of regularization in this context, since it is a notion that goes beyond the scope of the copying problem itself. We define this effect as occurring when the copy $f_{\mathcal{C}}$ has a lower complexity and a simpler decision boundary than the black box, something that could lead to $f_{\mathcal{C}}$ having a lower generalization error on the original problem. In particular, here we treat the copy as a new independent model, whose performance on the initial problem is of direct interest.

As already mentioned in Ch. 2, this use of self-distillation as an implicit regularization tool is not new (Mosca and Magoulas, 2017; Mobahi, Farajtabar, and Bartlett, 2020; Borup and Andersen, 2021; Pareek, Du, and Oh, 2024), and many experiments have shown that the copy often outperforms the original model on the problem. In the black box setting, this effect can still appear because, first, the teacher's hard labels can be cleaner and more consistent than the original dataset. Second, although the copy aims to imitate the black box, it rarely replicates all its irregularities, since it only observes the box's outputs on a large but finite set. Consequently, the copy learns a simplified approximation of the original boundary, capturing its structure but removing the small variations present in a negligible fraction of samples.

In this situation, the use of distances can strengthen even further this regularization, since signed distances vary smoothly and that imposes a continuity constraint on the target function. This property facilitates learning and reduces the impact of small irregularities, that are down-weighted in the loss and can be ignored by the copy. Moreover, computed distances usually come with numerical errors, something that reduces the amount of detail that can be copied. This effect, negative from the copying point of view, can improve the generalization, as these details are often the ones produced by overfitting. Finally, having to predict signed distances instead of the class is a more demanding task that requires richer outputs and thus a larger portion of the capacity of the model is used, reducing its ability to overfit.

To leverage these effects, we propose a model-agnostic regularization framework based on distance copying, introducing a parameter $\alpha$ to control the strength of the regularization. Specifically, denoting the original model by $f_{\mathcal{O}}$, the approach applies a copying or self-distillation procedure by training a new model $f_{\mathcal{C}}$ with the loss:

$$\frac{1}{N}\sum_{i=1}^{N} L(f_{\mathcal{C}}(x_i), f_{\mathcal{O}}(x_i)|\ell_i|^{\alpha}) \tag{3.2}$$

where $L$ denotes a regression loss such as MAE or MSE. As a consequence, the choice of $\alpha$ enables the method to shift from standard regression-based replication ($\alpha = 0$) to distance-based copying ($\alpha = 1$), directly impacting the smoothness of the learned function as formalized in the theorem below.

**Theorem 1** (Regularity of $\alpha$ signed distances)**.** *Let $f : \mathcal{X} \longrightarrow \{-1, 1\}$ be a function and let $\alpha > 0$, we consider $d$ a distance in $\mathcal{X}$ bounded by $D > 0$ and define $l_\alpha(x) = f(x)d(x, A_x)^\alpha$ for all $x \in \mathcal{X}$, where $A_x = \{y \in \mathcal{X} \mid f(y) \neq f(x)\}$. Then, we can conclude that for all $x, y \in \mathcal{X}$:*

- *If* $\alpha \leq 1$, *we have that* $|l_\alpha(x) - l_\alpha(y)| \leq 2d(x,y)^\alpha$.
- *If* $\alpha \geq 1$, *we have that* $|l_\alpha(x) - l_\alpha(y)| \leq 2\alpha D^{\alpha-1} d(x,y)$.

*Proof.* To show this result, given any $x, y \in \mathcal{X}$, we can start by distinguishing two different cases. On the one hand, if $f(x) \neq f(y)$, then we have by definition that $d(x, A_x) \leq d(x,y)$ and $d(y, A_y) \leq d(x,y)$, something that implies that $|l_\alpha(x) - l_\alpha(y)| = d(x, A_x)^\alpha + d(y, A_y)^\alpha \leq 2d(x,y)^\alpha$. In particular, if $\alpha \geq 1$, we can also deduce that $|l_\alpha(x) - l_\alpha(y)| \leq 2d(x,y)^\alpha \leq 2\alpha D^{\alpha-1} d(x,y)$.

From here, assuming now that $f(x) = f(y)$, we can observe that $A_x = A_y$. As a consequence, for all $z \in A_x$, we have that $d(x, A_x) \leq d(x,z) \leq d(x,y) + d(y,z)$, so we can take the infimum over $z$ to conclude that $d(x, A_x) \leq d(x,y) + d(y, A_y)$. Then, the same argument exchanging the roles of $x$ and $y$ shows that $|d(x, A_x) - d(y, A_y)| \leq d(x,y)$.

Finally, we distinguish the two additional cases depending on the value of $\alpha$:

**Case 1:** On the one hand, assuming that $\alpha \leq 1$, we can apply the fact that the expression $\alpha t^{\alpha-1}$ is decreasing on $t$ (since $\alpha \leq 1$) to deduce that:

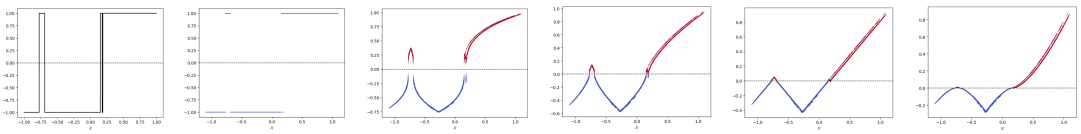
$$\begin{gathered} |l_\alpha(x) - l_\alpha(y)| = |d(x, A_x)^\alpha - d(y, A_y)^\alpha| = \\ \int_{d(y,A_y)}^{d(x,A_x)} \alpha t^{\alpha-1} dt \leq \int_{d(y,A_y)}^{d(x,A_x)} \alpha (t - d(y, A_y))^{\alpha-1} dt = \\ ((t - d(y, A_y))^\alpha |_{d(y,A_y)}^{d(x,A_x)} = |d(x, A_x) - d(y, A_y)|^\alpha \leq d(x,y)^\alpha \end{gathered} \tag{3.3}$$

where we have assumed without loss of generality that $d(y, A_y) \leq d(x, A_x)$. That completes the proof of this case.

**Case 2:** On the other hand, if $\alpha \geq 1$, we can conclude that:

$$\begin{gathered} |l_\alpha(x) - l_\alpha(y)| = |\int_{d(y,A_y)}^{d(x,A_x)} \alpha t^{\alpha-1} dt| \leq \\ |\int_{d(y,A_y)}^{d(x,A_x)} \alpha D^{\alpha-1} dt| \leq \alpha D^{\alpha-1} |d(x, A_x) - d(y, A_y)| \leq 2\alpha D^{\alpha-1} d(x,y) \end{gathered} \tag{3.4}$$

something that finishes the proof. □



(A) Black box (B) $\alpha = 0$ (C) $\alpha = 0.33$ (D) $\alpha = 0.67$ (E) $\alpha = 1$ (F) $\alpha = 1.5$

FIGURE 3.3: Comparison between a black box and several distance-based labellings of a synthetic dataset $\mathscr{X}$.

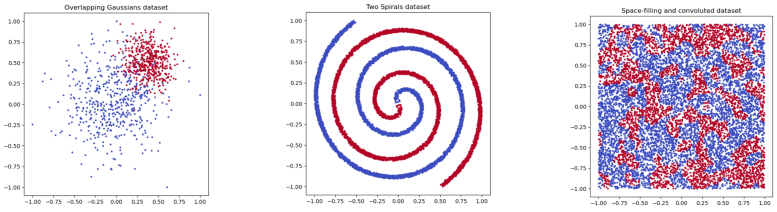
This impact of the parameter $\alpha$ and the aforementioned factors explaining why distance-copying can strengthen regularization are illustrated in Fig. 3.3, where one can observe that larger values of $\alpha$ result in a smoother target. As a result, small irregularities in the black box have a minimal effect, and the target remains close to 0 when they occur, making it easier for the copy to disregard them during training. Additionally, due to numerical errors, some of the computed distances may fail to capture these irregularities, something this is negative from the copying perspective, but that helps to forget these damaging details. These errors appear in the plot as the double line on the right side, formed by points whose distance estimation either included or ignored the anomaly.

# Chapter 4

# Empirical validation

## 4.1 Setup and experiments

**Datasets** In this section, several datasets are introduced to validate the explained proposal in a series of experiments, using both synthetic data in dimension two as well as real datasets extracted from the UCI machine learning repository (Kelly, Longjohn, and Nottingham, n.d.).




(A) Dataset 1 (B) Dataset 2 (C) Dataset 3

FIGURE 4.1: Synthetic datasets considered in dimension 2.

On the one hand, regarding the former, we have considered the three synthetic datasets shown in Fig. 4.1, aimed at visualizing the behaviours exhibited by the copies in different scenarios. Among them, Dataset 1 is generated through two colliding Gaussian distributions and thus its classes are not easily separable, something that may lead the black boxes to overfit and to exhibit irregular decisions boundaries. Moreover, we have also considered Dataset 2, that describes the two spirals shape. Thanks to this dataset, we will be able to asses the ability of the distance-based copies to replicate complex but smooth decision boundaries. Similarly, Dataset 3 is a convoluted dataset aimed at studying how these copies perform in difficult scenarios where the decision boundaries may still be relatively smooth.

On the other hand, to test this approach on real high dimensional problems, we have also considered three datasets extracted from the UCI machine learning repository. In detail, as Dataset 4, we have chosen the Breast Cancer Wisconsin (Diagnostic) dataset, which is often used as a benchmark for machine learning systems. Moreover, we have also worked with the Rice (Cammeo and Osmancik) and the Connectionist bench (mines vs rocks) datasets as Dataset 5 and 6 respectively. Together, these datasets exhibit a variety of dimensionalities aimed at analysing how distance-based copies are affected by them.

In each case, the datasets have been divided with a train/test proportion of 80/20, using the former part to train the black boxes while keeping the latter to test the accuracies of the copies (see Table 4.1).

**Models and model parameters** We have considered different types of models implemented with Scikit-learn (Pedregosa et al., 2011) and Keras (Chollet et al., 2015).

TABLE 4.1: Information of the datasets used in this project.

| Dataset | $\lvert\mathscr{D}_{\text{tr}}\rvert$ | $\lvert\mathscr{D}_{\text{te}}\rvert$ | Dim. |
|---|---|---|---|
| Dataset 1: Colliding Gaussians | 800 | 200 | 2 |
| Dataset 2: Two spirals | 8000 | 2000 | 2 |
| Dataset 3: Space-filling curves | 8000 | 2000 | 2 |
| Dataset 4: Breast Cancer | 455 | 114 | 30 |
| Dataset 5: Rice | 3048 | 762 | 7 |
| Dataset 6: Connectionist bench (M vs R) | 166 | 42 | 60 |

Specifically, as black boxes, we have worked with random forests, neural networks and boosting machines, which produce a diverse set of decision boundaries.

**Random Forest (RF):** initialized with 100 trees of maximum depth 10 and a minimum of 5 samples per leaf.

**Gradient Boosting Machines (GB):** trained with a 0.1 learning rate and using trees with a maximum of 31 leaves and a minimum of 20 samples in each of them.

**Multilayer Perceptron (NN):** following a 128-64-32-16-1 architecture and trained with a learning rate of 0.01 and batch size of 32 during 50 epochs.

In addition, to perform the copies, we have mainly chosen neural networks, because they excel at regressing smooth functions, such as distances. To train them, we have used a 0.1 learning rate, a batch size of 32 and $\text{int}(100 \cdot 20^{1-\log_{1000}(|\mathscr{Z}|)})$ epochs[1].

**Small (SNN):** 32-16-1 architecture.

**Medium (MNN):** 128-64-32-16-1 architecture.

**Large (LNN):** 512-256-256-128-64-32-16-1 architecture.

Nevertheless, for comparison purposes, we have also considered boosting machine regressor copies, with the same parameters as the GB black boxes.

**Performance metrics** As metrics, we have used the empirical fidelity error $R_{\text{em}}^{\mathcal{S}}$ on a uniform dataset $\mathcal{S} = \{x_j\}_{j=1}^{n}$ and the accuracy $\mathcal{A}_{\mathcal{C}}$. The former can be defined as:

$$R_{\text{em}}^{\mathcal{S}} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}[f_{\mathcal{O}}(x_j) \neq f_{\mathcal{C}}(x_j)] \tag{4.1}$$

and it is a measure of how well the copy replicates the original black box. In contrast, assuming that a real test dataset $\mathscr{D}_{\text{te}}$ is available, the latter can be used to asses the performance of the copy as a regular machine learning model.

In addition, to test the quality of the distances predicted by these copies, we have used the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE).

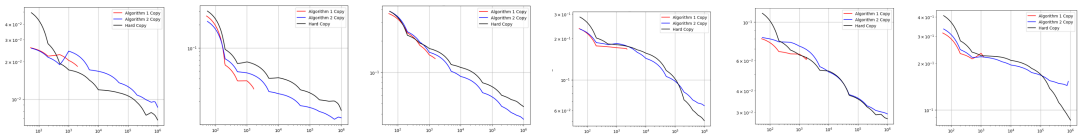
**Experimental settings** We have decided to design **three experiments** aimed at analysing the fidelity and performance of these copies, the regularization effect induced by distance-based copying, and the quality of the predicted distances.

**Experiment 1:** As the main experiment, setting a limit of 1,000,000 synthetic points and 240 seconds, we have sampled and labelled with distances synthetic datasets $\mathscr{Z}$, taking snapshots of the computations at predefined dataset sizes. That

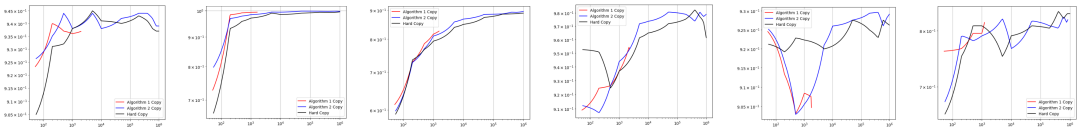
[1]The number of epochs has been adjusted according to the size of the synthetic dataset, with larger datasets leading to fewer epochs. For example, the formula assigns 100 epochs to synthetic datasets of 1000 samples, while for datasets of 1,000,000 points the number of epochs is reduced to 5.

TABLE 4.2: Comparison of algorithms across multiple datasets. Results show the number of wins out of 9 black box - NN copy combinations (draws counted as a win for the algorithms involved).

| Copy | Metric | Dat. 1 | Dat. 2 | Dat. 3 | Dat. 4 | Dat. 5 | Dat. 6 | Row Total |
|---|---|---|---|---|---|---|---|---|
| Alg. 1 cp. | $R^{\mathcal{S}}_{\text{em}}$ | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Alg. 2 cp. | $R^{\mathcal{S}}_{\text{em}}$ | 2 | **9** | **8** | 0 | 0 | 0 | 19 |
| Hard cp. | $R^{\mathcal{S}}_{\text{em}}$ | **7** | 0 | 1 | **9** | **9** | **9** | **35** |
| Alg. 1 cp. | $\mathcal{A}_{\mathcal{C}}$ | 2 | 2 | 0 | 1 | 2 | 2 | 9 |
| Alg. 2 cp. | $\mathcal{A}_{\mathcal{C}}$ | **6** | **9** | **9** | **6** | **7** | 3 | **40** |
| Hard cp. | $\mathcal{A}_{\mathcal{C}}$ | 2 | 3 | 0 | 5 | 6 | **5** | 21 |



(A) Dat. 1-$R^{\mathcal{S}}_{\text{em}}$ (B) Dat. 2-$R^{\mathcal{S}}_{\text{em}}$ (C) Dat. 3-$R^{\mathcal{S}}_{\text{em}}$ (D) Dat. 4-$R^{\mathcal{S}}_{\text{em}}$ (E) Dat. 5-$R^{\mathcal{S}}_{\text{em}}$ (F) Dat. 6-$R^{\mathcal{S}}_{\text{em}}$



(G) Dat. 1-$\mathcal{A}_{\mathcal{C}}$ (H) Dat. 2-$\mathcal{A}_{\mathcal{C}}$ (I) Dat. 3-$\mathcal{A}_{\mathcal{C}}$ (J) Dat. 4-$\mathcal{A}_{\mathcal{C}}$ (K) Dat. 5-$\mathcal{A}_{\mathcal{C}}$ (L) Dat. 6-$\mathcal{A}_{\mathcal{C}}$

FIGURE 4.2: Evolution of the metrics as a function of the number of training points. Plots made with a GB black box and a LNN copy.

has produced several size/metric plots, that compare how the performance of the algorithms and their corresponding copies evolves over time.

**Experiment 2:** In order to show the regularization effect linked to distance copying, we have also trained these copies for multiple values of the parameter $\alpha$, following the framework introduced in Ch. 3, and compared their results to the ones of the original black box and the hard copy. All copies have been trained on the same 1,000,000 synthetics points labelled with Alg. 2.

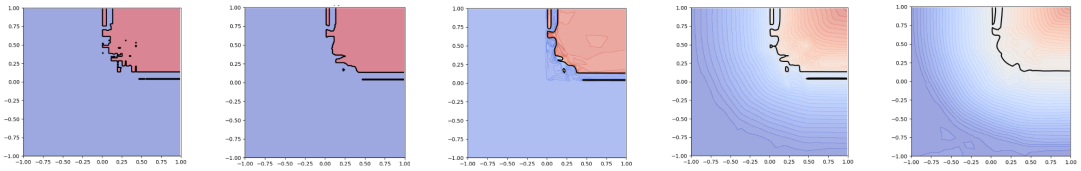
**Experiment 3:** To asses the uncertainty measure that distance-based copies have, in this experiment we have analysed the quality of their predicted distances. To this end, we have sampled a uniform dataset $\mathcal{S}^{'}$ and subset of the test dataset $\mathcal{D}^{'}_{\text{te}}$. Then, for each point in these datasets, we have computed their distance to the decision boundary using Alg. 1, treating these values as the ground-truth and comparing them to the corresponding predictions produced by the copies of Experiment 1.

All results produced by the above experiments have been averaged over 5 runs with different seeds. In addition, these experiments have employed translated Sobol sequences (Sobol, 1967) to sample the corresponding synthetic datasets.

## 4.2 Results

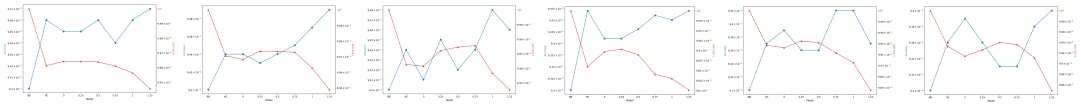
### 4.2.1 Experiment 1. Global comparison

Looking at Table 4.2 and Fig. 4.2 (extended versions available in Appendix B and C), we can start by recalling that Dataset 1 had overlapping classes and thus the black boxes trained on it tend to exhibit complex and overfitted decision boundaries. As a



(A) GB black box (B) Hard copy (C) $\alpha = 0$ (D) $\alpha = 0.5$ (E) $\alpha = 1$

FIGURE 4.3: Regularization produced by distance copying on Dataset 1. MNN copies trained on the same 1,000,000 synthetic points.

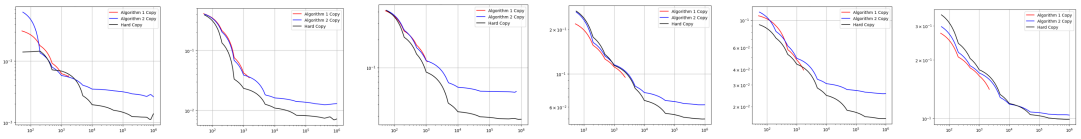


(A) RF/SNN (B) RF/MNN (C) RF/LNN (D) GB/SNN (E) GB/MNN (F) GB/LNN

FIGURE 4.4: Plots showing the evolution of the two metrics as a function of the regularization parameter $\alpha$ in Dataset 1.

consequence, here distance-based copies tend to have lower fidelities than the corresponding hard copy, due to their higher regularization. Nevertheless, thanks to that regularization, the accuracy of these copies is better than the one of the hard copy. Not only that, but analysing the plots, we can also conclude that these copies can outperform the hard ones when the size of the synthetic dataset is small. In addition, when we consider datasets with smoother decisions boundaries, like Dataset 2 and 3, distance-based copies usually outperform the hard one.

Regarding the results on the UCI datasets, generally, the empirical fidelity error of the hard copy is smaller than the ones of distance-based copies by a narrow margin. Nevertheless, the proposed approach tends to be on par or outperform (especially with the LNN copy, as shown in Fig. 4.2) the hard copies when restricted to small-sized synthetic datasets, a useful property in high dimensional but memory constrained settings. Apart from that, if we move the focus from fidelity to accuracy, here distance-based copies have a small advantage over the hard ones.



(A) Dat. 1-$R^{\mathcal{S}}_{\text{em}}$ (B) Dat. 2-$R^{\mathcal{S}}_{\text{em}}$ (C) Dat. 3-$R^{\mathcal{S}}_{\text{em}}$ (D) Dat. 4-$R^{\mathcal{S}}_{\text{em}}$ (E) Dat. 5-$R^{\mathcal{S}}_{\text{em}}$ (F) Dat. 6-$R^{\mathcal{S}}_{\text{em}}$

FIGURE 4.5: Evolution of the metric as a function of the number of points used to train the copy (GB) from the black box (RF).

Finally, as a comment, above we have put the focus on the results achieved by multilayer perceptrons copies, since this copying framework seems to be especially well-suited to inherently smooth copies (such as neural networks) that are trained on a continuous input space. However, in other contexts, for example when we consider gradient boosting copies, their performance is generally worse than the one of the corresponding hard copy, at least in regards to fidelity (see Fig. 4.5).

### 4.2.2 Experiment 2. Regularization effect

Focusing on the second experiment, whose results can be seen in Fig. 4.3 and Fig. 4.4 (extended versions available in Appendix C), it seems that when we increase the

value of the parameter $\alpha$, the corresponding model becomes smoother and thus it generalizes better. In particular, this regularization still grows in strength for values of $\alpha > 0$, something that shows that this effect does not only come from the traditional self-distillation and thus the value of the proposed approach is justified.

From a qualitative point of view, this regularization can be seen by looking at the decision boundaries of these copies, that get smoother as we increase the value of the parameter $\alpha$, disregarding the undesired details. As a consequence, thanks to this effect, the generalization ability of the copies tends to improve, together with their accuracies on the original dataset. Nevertheless, this loss of details can come at the cost of also increasing the empirical fidelity error of the copy, since they are no longer able to copy the target decision boundary with a high precision. This phenomenon is especially noticeable in datasets with irregular decision boundaries and overlapping classes, where these details (that are mainly produced by overfitting) are present. This explains why we have centred Fig. 4.3 and Fig. 4.4 around Dataset 1, that was chosen to highlight this type of behaviours.

TABLE 4.3: Number of wins, draws and losses of the distance-based copies with $\alpha > 0$, compared against the hard and $\alpha = 0$ copies (out of 12 black box - copy combinations).

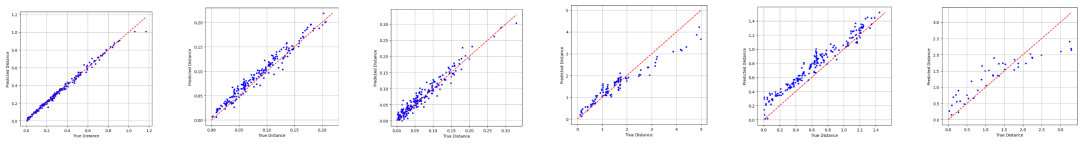
| Dataset | Win/Draw/Loss $R^{\mathcal{S}}_{\text{em}}$ | Win/Draw/Loss $\mathcal{A}_{\mathcal{C}}$ |
|---|---|---|
| Dataset 1 | 8/2/2 | 8/4/0 |
| Dataset 2 | 9/0/3 | 8/3/1 |
| Dataset 3 | 9/0/3 | 9/1/2 |
| Dataset 4 | 6/0/6 | 9/2/1 |
| Dataset 5 | 4/0/8 | 8/2/2 |
| Dataset 6 | 10/0/2 | 10/1/1 |

In contrast, looking at the global results, summarized in Table 4.3, we can observe that in the two dimensional datasets the distance-based copies, given by values of $\alpha$ greater than 0, have outperformed the hard ones both in fidelity and accuracy. This outcome is consistent with the results of the previous experiment, where we saw that the copy for $\alpha = 1$ had an edge in these datasets compared to the hard one. In addition, the above is also not contradictory with the previous observations for Dataset 1, since even though the fidelity of the distance-based copies has a tendency to decrease when we increase the value of $\alpha$, this decrease is not strict and there can be smaller values of $\alpha > 0$ that still offer higher fidelities than the hard copies.

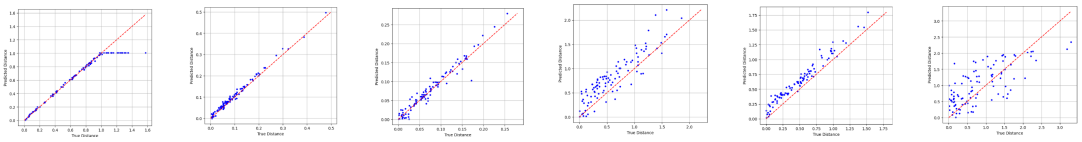
On the other hand, for the high-dimensional datasets, unlike in the previous experiment, we have trained all copies on synthetic datasets built with Alg. 2, consisting of families of small clusters. As a consequence, since these datasets are not uniform, training reliable copies becomes more difficult, especially in these high-dimensional settings. This fact, together with the observations from Experiment 1, is consistent with the obtained results: the distance-based copies outperform the hard ones in terms of accuracy, thanks to their regularization effect, but they exhibit a higher tendency to lose in terms of fidelity as the dimension decreases. In particular, in Dataset 6, the one with the highest dimensionality and where the negative effects of the clusters are more noticeable, distance-based copies prevail in regards fidelity, something that shows their edge in sample-constrained scenarios.

TABLE 4.4: Error of the predicted distances across different algorithms, metrics and datasets. Results averaged over every black box - copy - test dataset combination.

| Copy | Metric | Dat. 1 | Dat. 2 | Dat. 3 | Dat. 4 | Dat. 5 | Dat. 6 | Row Avg. |
|---|---|---|---|---|---|---|---|---|
| Alg. 1 cp. | MAE | 0.022 | 0.013 | 0.020 | 0.283 | 0.083 | 0.499 | 0.153 |
| Alg. 2 cp. | MAE | 0.040 | 0.008 | 0.012 | 0.296 | 0.152 | 0.443 | 0.158 |
| Alg. 1 cp. | RMSE | 0.030 | 0.017 | 0.026 | 0.391 | 0.107 | 0.669 | 0.207 |
| Alg. 2 cp. | RMSE | 0.072 | 0.010 | 0.016 | 0.413 | 0.176 | 0.628 | 0.219 |



(A) Dat. 1-$\mathscr{D}'_{\text{te}}$ (B) Dat. 2-$\mathscr{D}'_{\text{te}}$ (C) Dat. 3-$\mathscr{D}'_{\text{te}}$ (D) Dat. 4-$\mathscr{D}'_{\text{te}}$ (E) Dat. 5-$\mathscr{D}'_{\text{te}}$ (F) Dat. 6-$\mathscr{D}'_{\text{te}}$



(G) Dat. 1-$\mathcal{S}'$ (H) Dat. 2-$\mathcal{S}'$ (I) Dat. 3-$\mathcal{S}'$ (J) Dat. 4-$\mathcal{S}'$ (K) Dat. 5-$\mathcal{S}'$ (L) Dat. 6-$\mathcal{S}'$

FIGURE 4.6: Scatter plots of distance predictions against ground truth distances. Results are shown for a RF black box and a MNN copy performed using Alg. 2.

### 4.2.3 Experiment 3. Quality of the distances

Finally, looking at the results of this last experiment shown in Table 4.4 and Fig. 4.6 (extended versions available in Appendix B and C), we can begin by noting that the distances predicted by the copies closely approximate the ground truth in the two-dimensional datasets, a observation that we can extend to Dataset 5, which has the lowest dimensionality among the UCI problems. Nevertheless, the quality of these approximations decreases in higher-dimensional settings, where the error metrics increase and the points in the scatter plots become more dispersed.

In addition to dimensionality, performance also depends on the type of model used as the black box. Specifically, when a neural network serves as the black box, results are generally better: the points in the scatter plots are less dispersed and align more closely with the target line, something especially noticeable in these challenging high-dimensional settings. However, with Alg. 2, this improvement does not always translate into better metrics. This happens because, in this scenario, the models can more accurately predict distances, so they more often learn and respect the maximum distance threshold present in the synthetic training data, a behaviour that can paradoxically lead to lower scores.

Nevertheless, this distortion that the maximum distance threshold imposed in Alg. 2 produces on the metrics, that can be felt across most model and dataset combinations, does not compromise the value of the predicted distances as an uncertainty measure, as they still indicate whether small perturbations could change the label of a given sample. This explains why, even though the metrics may seem unfavourable, the plots show that Alg. 2 tends to outperform Alg. 1, probably due to the higher number of labelled points.

# Chapter 5

# Discussion

The experiments that have been performed in the previous section collectively highlight the strengths and limitations of distance-based copies.

On the one hand, on problems with smoother decision boundaries, the results show that the natural smoothness of distances and the greater context awareness they give to the copy can be leveraged to get an edge when working with this type of datasets, both in terms of accuracy as well as empirical fidelity error on the uniform dataset $\mathcal{S}$. However, in high dimensional settings, we have observed that distance-based copies have a lower fidelity than hard ones. A possible reason for that is that the errors in the computed distances may increase, something that combined with a higher complexity of the target decision boundary can make the problem more challenging for these copies. Moreover, the proposed algorithms label with signed distances few points (Alg. 1) or they label them in small clusters (Alg. 2), which means that distance-based copies are essentially working with less training points and being at a disadvantage. To mitigate these problems, we have explored the possibility of performing two-stage distance copies, but for reasons of scope and length, the detailed development of this idea is presented in Appendix D.

In this context, an additional cause that could explain the aforementioned disparity between the fidelities of the proposed copies and the traditional ones is the curse of dimensionality, that usually has a negative effect in distance-based machine learning models. This can be the case because, as already mentioned, in high dimensional settings the intuition behind distances may become weaker and given any pair of points, the distance between them tends become increasingly similar. Nevertheless, here that means that $\ell_i$ is close to $cf_{\mathcal{O}}(z_i)$, where $c > 0$ is the typical distance at which any pair of points tends to be, and thus the copy is trained on $\{(z_i, cf_{\mathcal{O}}(z_i))\}$, which is a change of labels from the original $\mathcal{T} = \{-1, 1\}$ to $\mathcal{T}' = \{-c, c\}$. This observation implies that, even though in these high-dimensional scenarios the predicted distances can be less informative and unable to capture meaningful geometries, something that can explain in part the observed decrease in performance, there are reasons to conclude that distance-based copies can still work and keep their ability to replicate the desired decision boundary.

Thanks to the results provided by the experiments, we can also compare the two algorithms we proposed to compute the required distances. In coherence with their theoretical cost, we have observed that Alg. 1 is much slower than Alg. 2, even though the quality of the distances produced by the former may be higher, something that limits the number of synthetic samples that we can generate to train the copy in a given timeframe.

Nevertheless, it is worth mentioning that Alg. 2 also presents several disadvantages, as this gain in speed sacrifices some freedom in the choice of the synthetic dataset distribution, since we are constrained to work with small clusters, and exactitude in the distance computations. However, in practice, as we can control how

the cluster locations $\mathscr{Z}_1$ are sampled and ensure that each cluster contains only a few points, the dataset can still remain dispersed. Moreover, thanks to distance regularity and the increased number of samples, it seems that the aforementioned numerical errors may balance out among neighbouring points without causing a significant drop in performance. In other words, that means that, in general, it is preferable to label these bigger synthetic datasets and thus Alg. 2 seems to offer the best approach to distance computation.

Finally, regarding the observed decrease in performance when we work with non-smooth regressor copies, this situation may happen since, when we use this framework, we shift from copying a sharp classification boundary to approximating a regular distance-measuring function. As a result, if the regressor is not able to learn such smooth functions, maybe because it is a random forest or a gradient boosting regressor whose inductive biases lead to stepwise approximations, it cannot reliably replicate the target distance function. Consequently, the predicted distances will not be accurate, and performance might deteriorate. Nevertheless, this copying approach could still have value in these scenarios, as a regularization technique or as a way to equip the copy with the uncertainty measure distances provide.

A part from that, in the context of the implicit regularization effect distance-based copies exhibit, the results have shown that the parameter $\alpha$ offers the possibility to control the trade-off between fidelity and accuracy, allowing the practitioner to prioritize the fidelity of the copy (small $\alpha$), its accuracy on the original problem (large $\alpha$) or to find a balance between them.

Nevertheless, it is worth mentioning that taking values of $\alpha$ larger than 1 can increase even more the regularization, but may also lead to numerical instabilities. This happens because distances around the boundary are, as expected, small and thus when we raise them to a large value of $\alpha$, they can get dangerously smaller. Consequently, this creates a region surrounding the decision boundary where all the labels $f_{\mathcal{O}}(x_i)|\ell_i|^{\alpha}$ become essentially 0, something that results in a loss of contrast that makes the copy unable to learn, with the corresponding decrease in accuracy and fidelity.

Finally, regarding the quality of the distances predicted by distance-based copies, we have observed that it decreases with the dimensionality, a tendency may be explained by the fact that, as the dimensionality increases, computing reliable approximations of the distances to the decision becomes increasingly difficult, due to the inherent sparsity and complexity of these spaces. In addition, the accuracy of these predictions also improves when copying multilayer perceptrons black boxes, since they tend to produce simpler, smoother decision boundaries, which facilitate more reliable distance approximations.

Despite these differences, the results lead to the conclusion that distance-based copies are generally capable of accurately predicting the distances to the decision boundary for the target points, even though they have not been explicitly trained on them. Consequently, the uncertainty measure they provide is meaningful and can be used to assess the confidence of their predictions, something that, for instance, could enable the construction of effective rejection classifiers.

As a summary, taken together, these findings confirm that distance-based copies offer a flexible and effective approach, particularly when regularization and uncertainty estimation are desirable.

# Chapter 6

# Conclusions

In this project, we have introduced and validated a novel framework for the copying problem, that focuses on enriching the labels used to train the copies. To do it, we have augmented them with information about the corresponding distances to the decision boundary, aiming to address the limitations of hard labels by providing additional context. This enhancement enables a more effective training process, while still working under the black box assumptions.

Moreover, by incorporating distance awareness into the replication process, we have explored how these copies are enriched with a robust uncertainty measure via differential replication. This measure serves as a valuable tool for assessing prediction confidence and can be used to analyse the behaviour of the copy when subjected to input perturbations, something shown in the experiments.

Furthermore, thanks to this additional information encoded in the labels, we have observed that distance-based copies can outperform traditional methods in memory-constrained scenarios, where only small synthetic datasets are available to train the copies. This advantage could be relevant in high dimensional settings, where data scarcity is a common issue.

Finally, as a side effect of the natural smoothness of distances and the numerical errors that arise in their computations, we have also seen that distance copying applies a significant regularization to the copies, improving their generalization ability and accuracy on the original problem. In addition, we have analysed how this effect could be framed as an implicit and model-agnostic regularization technique, whose strength could be controlled by a parameter $\alpha$.

Thanks to this phenomenon, these copies are encouraged to exhibit smooth decision boundaries, something that can be leveraged to increase their performance when the aim is to replicate this type of boundaries. In support of this, the experiments have shown that distance-based copies enjoy an advantage when copying complex but regular datasets, that are usually challenging for the traditional approach.

However, a downside of this regularization is that these copies can struggle to accurately reproduce irregular and overfitted decision boundaries. These boundaries often contain small, undesired details that this method smooths out, resulting in lower fidelities than the ones of classical copies. Additionally, the distance-based strategy can also suffer a small drop in fidelity in certain high dimensional settings, where the numerical errors on distances, their loss of interpretability and the computational cost limit the quality of these models.

# Appendix A

# Algorithms

In this appendix, we present the pseudocode of the algorithms described in Ch. 3.

**Algorithm 1** Individual distance computation

**Require:** The black box model $f_{\mathcal{O}}$, the region of interest $R$, a distance $d$ and the parameters $n, m, d_{\text{max}}, d_{\text{min}}, it_{\text{max}}$.

1: Sample $n$ points $\mathscr{Z}$ in the region $R$ and label them with $f_{\mathcal{O}}$.
2: Sample $m$ points $\mathscr{B}$ in the unit ball with uniform directions and uniform radius.
3: **for** $z$ belonging to $\mathscr{Z}$ **do**
4: Store $z$ in a new variable $c$.
5: Center $\mathscr{B}$ in $c$ and rescale it by $d_{\text{max}}$, labelling the points in the resulting set with $f_{\mathcal{O}}$.
6: Find the closest point of $c + d_{\text{max}}\mathscr{B}$ to $c$ with a different label and store it in $c$. If there are none, store $d_{\text{max}} f_{\mathcal{O}}(z)$ and continue the loop stated in 3.
7: **for** $i$ belonging to $\{2, \ldots, it_{\text{max}}\}$ **do**
8: Repeat steps 5 and 6 using $d_{\text{min}}$ instead of $d_{\text{max}}$.
9: **end for**
10: Store the signed distance $d(c, z) f_{\mathcal{O}}(z)$ between the original $z$ and the current point in $c$.
11: **end for**
12: **return** The set $\mathscr{Z} = \{z_i\}_i$ and the signed distances $\{\ell_i\}_i$.

**Algorithm 2** Grouped distance computation

**Require:** The black box model $f_{\mathcal{O}}$, the region of interest $R$, a distance $d$ and the parameters $n_c, n_{\text{in}}, n_{\text{out}}, d_{\text{in}}, d_{\text{out}}$.

1: Sample $n_c$ points $\mathscr{C}$ in the region of interest $R$.
2: Sample $n_{\text{in}}$ and $n_{\text{out}}$ points in the unit ball with uniform directions and uniform radius. Rescale them by $d_{\text{in}}$ and $d_{\text{out}}$ to obtain the sets $\mathscr{B}_{\text{in}}$ and $\mathscr{B}_{\text{out}}$ respectively.
3: **for** $c$ belonging to $\mathscr{C}$ **do**
4: Center $\mathscr{B}_{\text{in}}$ and $\mathscr{B}_{\text{out}}$ in $c$ and label them with $f_{\mathcal{O}}$.
5: **for** $p$ belonging to $c + \mathscr{B}_{\text{in}}$ **do**
6: Find $y$ the closest point of $c + \mathscr{B}_{\text{out}}$ to $p$ with a different label and store $d(p, y) f_{\mathcal{O}}(p)$. If there are none, store $d_{\text{out}} f_{\mathcal{O}}(p)$.
7: **end for**
8: **end for**
9: **return** The set $\mathscr{Z} = \bigcup_{c \in \mathscr{C}} (c + \mathscr{B}_{\text{in}}) = \{z_i\}_i$ and their corresponding signed distances to the boundary $\{\ell_i\}_i$.

# Appendix B

# Supplemental material. Tables

In this appendix, we present the complete version of the tables that appeared in the paper, showing the results for every combination of Black box, copy and dataset we analysed with the format mean ± std.

## Table 4.2

| Copy | $f_O/f_C$ | Dataset 1 | | | Dataset 2 | | | Dataset 3 | | | $\mathcal{A}_C/R^{\mathcal{S}}_{\text{em}}$ Dat. 1-3 | $\mathcal{A}_C/R^{\mathcal{S}}_{\text{em}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{A}_O$ | $\mathcal{A}_C$ | $R^{\mathcal{S}}_{\text{em}}$ | $\mathcal{A}_O$ | $\mathcal{A}_C$ | $R^{\mathcal{S}}_{\text{em}}$ | $\mathcal{A}_O$ | $\mathcal{A}_C$ | $R^{\mathcal{S}}_{\text{em}}$ | | |
| Algo. 1 copy | RF/SNN | 0.94 | 0.945±0.000 | 0.0053±0.0007 | 0.99 | 0.883±0.047 | 0.1162±0.0360 | 0.87 | 0.720±0.020 | 0.2038±0.0160 | 3/3 | 2.5/3 |
| Algo. 2 copy | RF/SNN | 0.94 | **0.948±0.002** | 0.0047±0.0007 | 0.99 | **0.991±0.004** | **0.0248±0.0012** | 0.87 | **0.802±0.003** | 0.1147±0.0072 | **1**/1.67 | **1.33**/1.83 |
| Hard copy | RF/SNN | 0.94 | 0.947±0.002 | **0.0037±0.0007** | 0.99 | 0.986±0.005 | 0.0273±0.0059 | 0.87 | 0.798±0.010 | **0.1142±0.0028** | 2/**1.33** | 1.83/**1.17** |
| Algo. 1 copy | GB/SNN | 0.93 | 0.941±0.009 | 0.0200±0.0032 | 1.00 | 0.949±0.022 | 0.0682±0.0167 | 0.90 | 0.713±0.016 | 0.2517±0.0108 | 3/3 | 2.67/3 |
| Algo. 2 copy | GB/SNN | 0.93 | **0.947±0.005** | 0.0178±0.0029 | 1.00 | **0.995±0.002** | **0.0215±0.0027** | 0.90 | **0.798±0.009** | **0.1569±0.0086** | **1**/**1.33** | **1.17**/1.67 |
| Hard copy | GB/SNN | 0.93 | 0.944±0.002 | **0.0112±0.0022** | 1.00 | 0.991±0.002 | 0.0233±0.0017 | 0.90 | 0.788±0.011 | 0.1680±0.0102 | 2/1.67 | 1.5/**1.33** |
| Algo. 1 copy | NN/SNN | 0.94 | **0.940±0.005** | **0.0014±0.0007** | 1.00 | 0.996±0.003 | 0.0398±0.0097 | 0.83 | 0.745±0.010 | 0.1442±0.0117 | 2/2.33 | 2/2.67 |
| Algo. 2 copy | NN/SNN | 0.94 | **0.940±0.003** | 0.0015±0.0003 | 1.00 | **1.000±0.000** | **0.0173±0.0030** | 0.83 | **0.807±0.007** | **0.0576±0.0065** | **1**/**1.33** | **1.33**/1.67 |
| Hard copy | NN/SNN | 0.94 | 0.939±0.002 | **0.0014±0.0008** | 1.00 | **1.000±0.001** | 0.0190±0.0037 | 0.83 | 0.786±0.010 | 0.0850±0.0119 | 1.67/1.67 | **1.33**/**1.33** |
| Algo. 1 copy | RF/MNN | 0.94 | **0.946±0.002** | 0.0062±0.0003 | 0.99 | 0.983±0.007 | 0.0355±0.0080 | 0.87 | 0.783±0.009 | 0.1352±0.0108 | 2.33/3 | 2.33/3 |
| Algo. 2 copy | RF/MNN | 0.94 | 0.945±0.003 | 0.0037±0.0004 | 0.99 | **0.993±0.002** | **0.0132±0.0013** | 0.87 | **0.862±0.003** | **0.0444±0.0031** | **1.33**/**1.33** | **1.17**/1.67 |
| Hard copy | RF/MNN | 0.94 | 0.943±0.002 | **0.0030±0.0010** | 0.99 | 0.989±0.001 | 0.0173±0.0017 | 0.87 | 0.856±0.007 | 0.0475±0.0043 | 2.33/1.67 | 2.67/**1.33** |
| Algo. 1 copy | GB/MNN | 0.93 | 0.942±0.007 | 0.0216±0.0047 | 1.00 | 0.995±0.003 | 0.0277±0.0044 | 0.90 | 0.798±0.009 | 0.1615±0.0026 | 2.67/3 | 2.83/3 |
| Algo. 2 copy | GB/MNN | 0.93 | **0.946±0.004** | 0.0094±0.0007 | 1.00 | **0.998±0.001** | **0.0125±0.0015** | 0.90 | **0.890±0.004** | **0.0496±0.0040** | **1**/**1.33** | **1**/1.67 |
| Hard copy | GB/MNN | 0.93 | 0.937±0.002 | **0.0057±0.0008** | 1.00 | 0.996±0.001 | 0.0155±0.0032 | 0.90 | 0.880±0.009 | 0.0657±0.0065 | 2.33/1.67 | 2.17/**1.33** |
| Algo. 1 copy | NN/MNN | 0.94 | 0.939±0.004 | 0.0018±0.0009 | 1.00 | **1.000±0.000** | 0.0193±0.0057 | 0.82 | 0.803±0.010 | 0.0559±0.0070 | 2.33/3 | 2/3 |
| Algo. 2 copy | NN/MNN | 0.94 | 0.941±0.004 | **0.0011±0.0002** | 1.00 | **1.000±0.000** | **0.0087±0.0023** | 0.82 | **0.819±0.008** | **0.0185±0.0020** | **1.33**/**1** | 1.83/**1.5** |
| Hard copy | NN/MNN | 0.94 | **0.942±0.002** | 0.0013±0.0005 | 1.00 | **1.000±0.000** | 0.0157±0.0032 | 0.82 | 0.817±0.009 | 0.0264±0.0041 | **1.33**/2 | **1.33**/**1.5** |
| Algo. 1 copy | RF/LNN | 0.94 | 0.945±0.003 | 0.0062±0.0009 | 0.99 | 0.973±0.008 | 0.0386±0.0058 | 0.87 | 0.802±0.013 | 0.1174±0.0111 | 2.67/3 | 2.33/3 |
| Algo. 2 copy | RF/LNN | 0.94 | **0.948±0.004** | 0.0036±0.0007 | 0.99 | **0.988±0.008** | **0.0120±0.0014** | 0.87 | **0.865±0.002** | **0.0348±0.0036** | **1**/**1.33** | **1.33**/1.67 |
| Hard copy | RF/LNN | 0.94 | 0.944±0.002 | **0.0035±0.0006** | 0.99 | 0.982±0.010 | 0.0214±0.0042 | 0.87 | 0.864±0.003 | 0.0381±0.0028 | 2.33/1.67 | 2/**1.33** |
| Algo. 1 copy | GB/LNN | 0.93 | 0.937±0.007 | 0.0181±0.0012 | 1.00 | 0.994±0.002 | 0.0287±0.0044 | 0.90 | 0.828±0.007 | 0.1327±0.0040 | 2.67/3 | 2.67/3 |
| Algo. 2 copy | GB/LNN | 0.93 | **0.939±0.005** | 0.0086±0.0012 | 1.00 | **0.997±0.001** | **0.0123±0.0020** | 0.90 | **0.895±0.002** | **0.0355±0.0024** | **1**/**1.33** | **1.33**/1.67 |
| Hard copy | GB/LNN | 0.93 | 0.937±0.002 | **0.0068±0.0006** | 1.00 | 0.995±0.001 | 0.0154±0.0035 | 0.90 | 0.889±0.004 | 0.0470±0.0022 | 2/1.67 | 1.83/**1.33** |
| Algo. 1 copy | NN/LNN | 0.94 | 0.939±0.006 | 0.0033±0.0040 | 1.00 | **1.000±0.000** | 0.0202±0.0036 | 0.83 | 0.814±0.010 | 0.0555±0.0130 | 2/3 | 2.17/3 |
| Algo. 2 copy | NN/LNN | 0.94 | 0.937±0.005 | **0.0015±0.0008** | 1.00 | **1.000±0.000** | **0.0088±0.0017** | 0.83 | **0.823±0.011** | **0.0159±0.0024** | 1.67/**1** | 1.67/**1.5** |
| Hard copy | NN/LNN | 0.94 | **0.941±0.002** | 0.0019±0.0013 | 1.00 | **1.000±0.000** | 0.0143±0.0010 | 0.83 | 0.822±0.011 | 0.0238±0.0035 | **1.33**/2 | **1.33**/**1.5** |
| Algo. 1 copy | RF/GB | 0.94 | 0.944±0.006 | 0.0055±0.0005 | 0.99 | 0.981±0.007 | 0.0337±0.0037 | 0.87 | 0.801±0.004 | 0.1127±0.0040 | 2.33/3 | 1.83/3 |
| Algo. 2 copy | RF/GB | 0.94 | **0.945±0.003** | 0.0026±0.0003 | 0.99 | **0.989±0.004** | 0.0130±0.0012 | 0.87 | 0.844±0.003 | 0.0540±0.0014 | **1.33**/2 | **1.5**/2 |
| Hard copy | RF/GB | 0.94 | 0.941±0.002 | **0.0014±0.0002** | 0.99 | **0.989±0.004** | **0.0072±0.0011** | 0.87 | **0.856±0.004** | **0.0277±0.0005** | 1.67/**1** | 1.83/**1** |
| Algo. 1 copy | GB/GB | 0.93 | **0.940±0.003** | 0.0139±0.0004 | 1.00 | 0.993±0.001 | 0.0325±0.0020 | 0.90 | 0.829±0.004 | 0.1165±0.0047 | 2.33/3 | **1.67**/3 |
| Algo. 2 copy | GB/GB | 0.93 | 0.939±0.002 | 0.0069±0.0005 | 1.00 | **0.998±0.001** | 0.0123±0.0005 | 0.90 | 0.861±0.002 | 0.0687±0.0027 | **1.67**/2 | 1.83/2 |
| Hard copy | GB/GB | 0.93 | 0.933±0.002 | **0.0031±0.0003** | 1.00 | 0.997±0.001 | **0.0060±0.0005** | 0.90 | **0.882±0.003** | **0.0325±0.0020** | 2/**1** | **1.67**/**1** |
| Algo. 1 copy | NN/GB | 0.94 | 0.937±0.004 | 0.0027±0.0005 | 1.00 | **1.000±0.000** | 0.0340±0.0021 | 0.83 | 0.811±0.011 | 0.0649±0.0046 | 2/3 | 2.33/3 |
| Algo. 2 copy | NN/GB | 0.94 | 0.937±0.008 | 0.0017±0.0001 | 1.00 | **1.000±0.000** | 0.0219±0.0010 | 0.83 | 0.812±0.012 | 0.0516±0.0068 | 1.67/2 | **1.5**/1.83 |
| Hard copy | NN/GB | 0.94 | **0.939±0.002** | **0.0008±0.0001** | 1.00 | **1.000±0.000** | **0.0125±0.0006** | 0.83 | **0.826±0.011** | **0.0315±0.0035** | **1**/**1** | **1.5**/**1.17** |

| Copy | $f_{\mathcal{O}}/f_{\mathcal{C}}$ | Dataset 4 | | | Dataset 5 | | | Dataset 6 | | | $\mathcal{A}_{\mathcal{C}}/R^{\mathcal{S}}_{\text{em}}$ Dat. 4-6 | $\mathcal{A}_{\mathcal{C}}/R^{\mathcal{S}}_{\text{em}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{A}_{\mathcal{O}}$ | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ | $\mathcal{A}_{\mathcal{O}}$ | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ | $\mathcal{A}_{\mathcal{O}}$ | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ | | |
| Algo. 1 copy | RF/SNN | 0.97 | 0.949±0.030 | 0.1678±0.0082 | 0.92 | 0.920±0.009 | 0.0523±0.0052 | 0.83 | **0.829±0.053** | 0.2218±0.0093 | 2/3 | 2.5/3 |
| Algo. 2 copy | RF/SNN | 0.97 | 0.965±0.006 | 0.0762±0.0017 | 0.92 | **0.929±0.004** | 0.0306±0.0033 | 0.83 | 0.814±0.032 | 0.1244±0.0066 | **1.67**/2 | **1.33**/1.83 |
| Hard copy | RF/SNN | 0.97 | **0.967±0.007** | **0.0613±0.0029** | 0.92 | **0.929±0.007** | **0.0272±0.0032** | 0.83 | 0.810±0.040 | **0.1154±0.0101** | **1.67**/**1** | 1.83/**1.17** |
| Algo. 1 copy | GB/SNN | 0.98 | 0.960±0.023 | 0.1606±0.0043 | 0.92 | 0.905±0.010 | 0.0541±0.0016 | 0.87 | 0.781±0.059 | 0.2484±0.0101 | 2.33/3 | 2.67/3 |
| Algo. 2 copy | GB/SNN | 0.98 | 0.977±0.017 | 0.0769±0.0043 | 0.92 | **0.927±0.004** | 0.0349±0.0034 | 0.87 | **0.838±0.044** | 0.1264±0.0098 | 1.33/2 | **1.17**/1.67 |
| Hard copy | GB/SNN | 0.98 | **0.979±0.012** | **0.0674±0.0048** | 0.92 | **0.927±0.004** | **0.0341±0.0021** | 0.87 | **0.838±0.041** | **0.1108±0.0107** | **1**/**1** | 1.5/**1.33** |
| Algo. 1 copy | NN/SNN | 0.98 | 0.979±0.015 | 0.0298±0.0024 | 0.92 | **0.925±0.004** | 0.0135±0.0010 | 0.89 | 0.838±0.032 | 0.0892±0.0041 | 2/3 | 2/2.67 |
| Algo. 2 copy | NN/SNN | 0.98 | **0.981±0.010** | 0.0190±0.0012 | 0.92 | 0.924±0.005 | 0.0107±0.0013 | 0.89 | 0.848±0.039 | 0.0441±0.0029 | 1.67/2 | **1.33**/1.67 |
| Hard copy | NN/SNN | 0.98 | **0.981±0.010** | **0.0176±0.0015** | 0.92 | **0.925±0.007** | **0.0083±0.0011** | 0.89 | **0.862±0.038** | **0.0417±0.0031** | **1**/**1** | **1.33**/**1.33** |
| Algo. 1 copy | RF/MNN | 0.97 | 0.965±0.019 | 0.1771±0.0104 | 0.93 | 0.926±0.006 | 0.0501±0.0048 | 0.81 | 0.786±0.050 | 0.2105±0.0164 | 2.33/3 | 2.33/3 |
| Algo. 2 copy | RF/MNN | 0.97 | **0.968±0.007** | 0.0613±0.0023 | 0.93 | **0.928±0.004** | 0.0230±0.0014 | 0.81 | **0.810±0.040** | 0.1440±0.0060 | **1**/2 | **1.17**/1.67 |
| Hard copy | RF/MNN | 0.97 | **0.968±0.009** | **0.0493±0.0016** | 0.93 | **0.928±0.006** | **0.0210±0.0015** | 0.81 | 0.800±0.072 | **0.1021±0.0043** | 1.33/**1** | 2.67/**1.33** |
| Algo. 1 copy | GB/MNN | 0.98 | 0.958±0.039 | 0.1798±0.0092 | 0.92 | 0.919±0.006 | 0.0582±0.0025 | 0.87 | 0.786±0.026 | 0.2396±0.0124 | 3/3 | 2.83/3 |
| Algo. 2 copy | GB/MNN | 0.98 | **0.981±0.012** | 0.0627±0.0062 | 0.92 | **0.927±0.003** | 0.0302±0.0028 | 0.87 | **0.833±0.026** | 0.1504±0.0144 | **1**/2 | **1**/1.67 |
| Hard copy | GB/MNN | 0.98 | 0.979±0.012 | **0.0508±0.0039** | 0.92 | 0.926±0.004 | **0.0285±0.0016** | 0.87 | 0.824±0.024 | **0.0874±0.0097** | 2/**1** | 2.17/**1.33** |
| Algo. 1 copy | NN/MNN | 0.98 | **0.984±0.010** | 0.0351±0.0090 | 0.92 | **0.925±0.004** | 0.0129±0.0009 | 0.88 | 0.829±0.028 | 0.0791±0.0061 | 1.67/3 | 2/3 |
| Algo. 2 copy | NN/MNN | 0.98 | 0.979±0.007 | 0.0160±0.0017 | 0.92 | 0.924±0.004 | 0.0078±0.0006 | 0.88 | 0.862±0.032 | 0.0384±0.0016 | 2.33/2 | 1.83/**1.5** |
| Hard copy | NN/MNN | 0.98 | 0.983±0.010 | **0.0138±0.0016** | 0.92 | **0.925±0.004** | **0.0065±0.0010** | 0.88 | **0.886±0.038** | **0.0313±0.0012** | 1.33/1 | **1.33**/**1.5** |
| Algo. 1 copy | RF/LNN | 0.97 | 0.953±0.026 | 0.1821±0.0107 | 0.93 | 0.921±0.005 | 0.0569±0.0128 | 0.80 | **0.810±0.072** | 0.2157±0.0163 | 2/3 | 2.33/3 |
| Algo. 2 copy | RF/LNN | 0.97 | **0.970±0.009** | 0.0630±0.0018 | 0.93 | **0.928±0.004** | 0.0233±0.0032 | 0.80 | 0.800±0.039 | 0.1324±0.0074 | **1.67**/2 | **1.33**/1.67 |
| Hard copy | RF/LNN | 0.97 | 0.960±0.012 | **0.0496±0.0018** | 0.93 | **0.928±0.006** | **0.0228±0.0032** | 0.80 | 0.805±0.044 | **0.0971±0.0042** | **1.67**/**1** | 2/**1.33** |
| Algo. 1 copy | GB/LNN | 0.98 | 0.956±0.029 | 0.1683±0.0126 | 0.92 | 0.907±0.027 | 0.0599±0.0115 | 0.87 | 0.819±0.044 | 0.2163±0.0097 | 2.67/3 | 2.67/3 |
| Algo. 2 copy | GB/LNN | 0.98 | **0.979±0.009** | 0.0648±0.0050 | 0.92 | **0.929±0.004** | 0.0294±0.0020 | 0.87 | 0.814±0.038 | 0.1415±0.0096 | **1.67**/2 | **1.33**/1.67 |
| Hard copy | GB/LNN | 0.98 | 0.961±0.015 | **0.0504±0.0037** | 0.92 | 0.926±0.004 | **0.0275±0.0012** | 0.87 | **0.833±0.040** | **0.0858±0.0067** | **1.67**/**1** | **1.83**/1.33 |
| Algo. 1 copy | NN/LNN | 0.97 | 0.972±0.007 | 0.0342±0.0067 | 0.92 | 0.925±0.003 | 0.0137±0.0027 | 0.85 | 0.838±0.051 | 0.0822±0.0111 | 2.33/3 | 2.17/3 |
| Algo. 2 copy | NN/LNN | 0.97 | **0.975±0.007** | 0.0156±0.0016 | 0.92 | **0.927±0.006** | 0.0077±0.0009 | 0.85 | 0.833±0.050 | 0.0371±0.0019 | 1.67/2 | 1.67/**1.5** |
| Hard copy | NN/LNN | 0.97 | **0.975±0.007** | **0.0140±0.0017** | 0.92 | 0.926±0.005 | **0.0070±0.0018** | 0.85 | **0.843±0.039** | **0.0296±0.0011** | **1.33**/**1** | 1.33/**1.5** |
| Algo. 1 copy | RF/GB | 0.97 | **0.970±0.009** | 0.0933±0.0028 | 0.93 | 0.927±0.004 | 0.0388±0.0018 | 0.82 | **0.805±0.071** | 0.1402±0.0079 | **1.33**/3 | 1.83/3 |
| Algo. 2 copy | RF/GB | 0.97 | 0.967±0.007 | 0.0621±0.0017 | 0.93 | **0.928±0.005** | 0.0255±0.0022 | 0.82 | 0.800±0.070 | 0.1040±0.0043 | 1.67/2 | **1.5**/2 |
| Hard copy | RF/GB | 0.97 | 0.967±0.007 | **0.0498±0.0024** | 0.93 | **0.928±0.003** | **0.0160±0.0011** | 0.82 | 0.791±0.076 | **0.0988±0.0047** | 2/**1** | 1.83/**1** |
| Algo. 1 copy | GB/GB | 0.98 | **0.975±0.015** | 0.0938±0.0023 | 0.92 | **0.927±0.005** | 0.0376±0.0025 | 0.87 | **0.848±0.019** | 0.1354±0.0061 | **1**/3 | **1.67**/3 |
| Algo. 2 copy | GB/GB | 0.98 | 0.970±0.009 | 0.0594±0.0027 | 0.92 | 0.926±0.005 | 0.0278±0.0023 | 0.87 | **0.848±0.054** | 0.0838±0.0014 | 2/2 | 1.83/2 |
| Hard copy | GB/GB | 0.98 | 0.972±0.009 | **0.0428±0.0042** | 0.92 | **0.927±0.005** | **0.0206±0.0024** | 0.87 | **0.848±0.044** | **0.0691±0.0045** | 1.33/**1** | **1.67**/**1** |
| Algo. 1 copy | NN/GB | 0.98 | 0.972±0.015 | 0.0887±0.0079 | 0.92 | 0.924±0.006 | 0.0298±0.0038 | 0.89 | 0.786±0.067 | 0.1543±0.0016 | 2.67/3 | 2.33/3 |
| Algo. 2 copy | NN/GB | 0.98 | **0.984±0.010** | 0.0752±0.0062 | 0.92 | **0.925±0.006** | 0.0238±0.0033 | 0.89 | 0.848±0.036 | **0.1311±0.0009** | **1.33**/1.67 | **1.5**/1.83 |
| Hard copy | NN/GB | 0.98 | 0.977±0.016 | **0.0744±0.0064** | 0.92 | 0.923±0.005 | **0.0206±0.0032** | 0.89 | **0.852±0.028** | 0.1339±0.0008 | 2/**1.33** | **1.5**/**1.17** |

# Table 4.4

## Metric: MAE

| Copy | $f_{\mathcal{O}}/f_{\mathcal{C}}$ | Metric | Dataset 1 | | Dataset 2 | | Dataset 3 | | Avg. Dat. 1-3 | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | | |
| Alg. 1 copy | RF/SNN | MAE | 0.018±0.003 | 0.019±0.002 | 0.027±0.004 | 0.024±0.002 | 0.028±0.001 | 0.031±0.003 | 0.024 | 0.163 |
| Alg. 2 copy | RF/SNN | MAE | 0.016±0.001 | 0.059±0.002 | 0.009±0.001 | 0.009±0.001 | 0.019±0.002 | 0.019±0.001 | 0.022 | 0.152 |
| Alg. 1 copy | GB/SNN | MAE | 0.043±0.003 | 0.040±0.004 | 0.019±0.003 | 0.019±0.003 | 0.029±0.001 | 0.030±0.001 | 0.030 | 0.159 |
| Alg. 2 copy | GB/SNN | MAE | 0.036±0.002 | 0.062±0.005 | 0.010±0.001 | 0.009±0.001 | 0.019±0.001 | 0.019±0.001 | 0.026 | 0.134 |
| Alg. 1 copy | NN/SNN | MAE | 0.006±0.001 | 0.007±0.001 | 0.016±0.005 | 0.012±0.002 | 0.024±0.002 | 0.027±0.003 | 0.015 | 0.118 |
| Alg. 2 copy | NN/SNN | MAE | 0.011±0.000 | 0.053±0.003 | 0.009±0.001 | 0.009±0.001 | 0.014±0.001 | 0.015±0.002 | 0.018 | 0.164 |
| Alg. 1 copy | RF/MNN | MAE | 0.017±0.001 | 0.019±0.002 | 0.011±0.002 | 0.012±0.003 | 0.020±0.001 | 0.020±0.000 | 0.016 | 0.170 |
| Alg. 2 copy | RF/MNN | MAE | 0.016±0.002 | 0.058±0.003 | 0.008±0.001 | 0.008±0.001 | 0.010±0.001 | 0.011±0.002 | 0.018 | 0.152 |
| Alg. 1 copy | GB/MNN | MAE | 0.039±0.003 | 0.037±0.004 | 0.010±0.001 | 0.010±0.002 | 0.020±0.001 | 0.020±0.002 | 0.023 | 0.168 |
| Alg. 2 copy | GB/MNN | MAE | 0.034±0.001 | 0.062±0.003 | 0.007±0.000 | 0.008±0.000 | 0.009±0.001 | 0.008±0.000 | 0.021 | 0.147 |
| Alg. 1 copy | NN/MNN | MAE | 0.006±0.002 | 0.008±0.003 | 0.008±0.002 | 0.008±0.002 | 0.013±0.001 | 0.012±0.001 | 0.009 | 0.122 |
| Alg. 2 copy | NN/MNN | MAE | 0.013±0.001 | 0.054±0.002 | 0.007±0.001 | 0.008±0.001 | 0.008±0.001 | 0.008±0.000 | 0.016 | 0.163 |
| Alg. 1 copy | RF/LNN | MAE | 0.017±0.002 | 0.019±0.005 | 0.012±0.001 | 0.011±0.001 | 0.020±0.003 | 0.020±0.002 | 0.016 | 0.162 |
| Alg. 2 copy | RF/LNN | MAE | 0.017±0.002 | 0.060±0.002 | 0.007±0.001 | 0.007±0.001 | 0.010±0.001 | 0.010±0.001 | 0.018 | 0.157 |
| Alg. 1 copy | GB/LNN | MAE | 0.037±0.003 | 0.043±0.010 | 0.010±0.001 | 0.010±0.001 | 0.018±0.001 | 0.018±0.001 | 0.023 | 0.155 |
| Alg. 2 copy | GB/LNN | MAE | 0.035±0.002 | 0.061±0.004 | 0.008±0.001 | 0.009±0.001 | 0.008±0.001 | 0.007±0.000 | 0.021 | 0.149 |
| Alg. 1 copy | NN/LNN | MAE | 0.008±0.005 | 0.009±0.003 | 0.007±0.001 | 0.007±0.001 | 0.014±0.003 | 0.013±0.003 | 0.010 | 0.132 |
| Alg. 2 copy | NN/LNN | MAE | 0.013±0.002 | 0.054±0.003 | 0.007±0.001 | 0.007±0.001 | 0.008±0.001 | 0.007±0.001 | 0.016 | 0.166 |
| Alg. 1 copy | RF/GB | MAE | 0.019±0.002 | 0.019±0.002 | 0.011±0.000 | 0.013±0.002 | 0.021±0.001 | 0.020±0.002 | 0.017 | 0.151 |
| Alg. 2 copy | RF/GB | MAE | 0.016±0.001 | 0.058±0.003 | 0.008±0.000 | 0.008±0.000 | 0.013±0.000 | 0.014±0.001 | 0.020 | 0.159 |
| Alg. 1 copy | GB/GB | MAE | 0.040±0.001 | 0.037±0.004 | 0.012±0.001 | 0.012±0.001 | 0.018±0.001 | 0.018±0.001 | 0.023 | 0.131 |
| Alg. 2 copy | GB/GB | MAE | 0.035±0.002 | 0.065±0.003 | 0.009±0.000 | 0.008±0.000 | 0.013±0.000 | 0.014±0.001 | 0.024 | 0.155 |
| Alg. 1 copy | NN/GB | MAE | 0.010±0.005 | 0.009±0.001 | 0.017±0.001 | 0.012±0.001 | 0.015±0.001 | 0.014±0.001 | 0.013 | 0.211 |
| Alg. 2 copy | NN/GB | MAE | 0.012±0.001 | 0.053±0.003 | 0.014±0.001 | 0.010±0.000 | 0.013±0.001 | 0.013±0.001 | 0.019 | 0.208 |

| Copy | $f_{\mathcal{O}}/f_{\mathcal{C}}$ | Metric | Dataset 4 | | Dataset 5 | | Dataset 6 | | Avg. Dat. 4-6 | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | | |
| Alg. 1 copy | RF/SNN | MAE | 0.342±0.026 | 0.295±0.023 | 0.083±0.007 | 0.087±0.006 | 0.493±0.089 | 0.514±0.038 | 0.302 | 0.163 |
| Alg. 2 copy | RF/SNN | MAE | 0.295±0.049 | 0.240±0.005 | 0.131±0.020 | 0.146±0.027 | 0.529±0.091 | 0.344±0.014 | 0.281 | 0.151 |
| Alg. 1 copy | GB/SNN | MAE | 0.246±0.030 | 0.299±0.018 | 0.121±0.013 | 0.088±0.026 | 0.461±0.090 | 0.510±0.046 | 0.288 | 0.159 |
| Alg. 2 copy | GB/SNN | MAE | 0.239±0.024 | 0.231±0.018 | 0.180±0.022 | 0.139±0.012 | 0.333±0.054 | 0.326±0.034 | 0.241 | 0.134 |
| Alg. 1 copy | NN/SNN | MAE | 0.185±0.042 | 0.177±0.035 | 0.057±0.008 | 0.053±0.003 | 0.486±0.057 | 0.370±0.047 | 0.221 | 0.118 |
| Alg. 2 copy | NN/SNN | MAE | 0.290±0.056 | 0.333±0.074 | 0.151±0.015 | 0.154±0.013 | 0.597±0.065 | 0.328±0.025 | 0.309 | 0.164 |
| Alg. 1 copy | RF/MNN | MAE | 0.410±0.033 | 0.309±0.019 | 0.082±0.017 | 0.085±0.016 | 0.591±0.087 | 0.463±0.035 | 0.323 | 0.170 |
| Alg. 2 copy | RF/MNN | MAE | 0.303±0.030 | 0.238±0.009 | 0.141±0.011 | 0.150±0.009 | 0.461±0.082 | 0.416±0.028 | 0.285 | 0.152 |
| Alg. 1 copy | GB/MNN | MAE | 0.310±0.067 | 0.327±0.027 | 0.106±0.009 | 0.097±0.009 | 0.559±0.072 | 0.474±0.048 | 0.312 | 0.168 |
| Alg. 2 copy | GB/MNN | MAE | 0.261±0.040 | 0.261±0.009 | 0.204±0.018 | 0.166±0.018 | 0.371±0.041 | 0.372±0.040 | 0.272 | 0.147 |
| Alg. 1 copy | NN/MNN | MAE | 0.171±0.019 | 0.218±0.077 | 0.060±0.006 | 0.053±0.005 | 0.542±0.087 | 0.360±0.055 | 0.234 | 0.122 |
| Alg. 2 copy | NN/MNN | MAE | 0.302±0.038 | 0.316±0.036 | 0.167±0.017 | 0.152±0.017 | 0.621±0.046 | 0.303±0.018 | 0.310 | 0.163 |
| Alg. 1 copy | RF/LNN | MAE | 0.385±0.075 | 0.310±0.028 | 0.090±0.030 | 0.091±0.015 | 0.523±0.072 | 0.445±0.037 | 0.307 | 0.162 |
| Alg. 2 copy | RF/LNN | MAE | 0.392±0.041 | 0.240±0.017 | 0.128±0.006 | 0.129±0.014 | 0.504±0.053 | 0.386±0.033 | 0.296 | 0.157 |
| Alg. 1 copy | GB/LNN | MAE | 0.258±0.027 | 0.311±0.012 | 0.121±0.020 | 0.107±0.014 | 0.486±0.071 | 0.430±0.019 | 0.286 | 0.155 |
| Alg. 2 copy | GB/LNN | MAE | 0.270±0.026 | 0.265±0.026 | 0.204±0.019 | 0.149±0.011 | 0.418±0.066 | 0.348±0.027 | 0.276 | 0.149 |
| Alg. 1 copy | NN/LNN | MAE | 0.241±0.053 | 0.225±0.038 | 0.057±0.007 | 0.052±0.005 | 0.583±0.079 | 0.368±0.046 | 0.254 | 0.132 |
| Alg. 2 copy | NN/LNN | MAE | 0.311±0.036 | 0.317±0.023 | 0.156±0.011 | 0.137±0.009 | 0.622±0.070 | 0.356±0.060 | 0.316 | 0.166 |
| Alg. 1 copy | RF/GB | MAE | 0.339±0.043 | 0.224±0.015 | 0.086±0.006 | 0.079±0.006 | 0.561±0.078 | 0.418±0.038 | 0.285 | 0.151 |
| Alg. 2 copy | RF/GB | MAE | 0.490±0.046 | 0.215±0.013 | 0.119±0.012 | 0.131±0.013 | 0.503±0.063 | 0.330±0.015 | 0.298 | 0.159 |
| Alg. 1 copy | GB/GB | MAE | 0.253±0.024 | 0.221±0.026 | 0.111±0.010 | 0.077±0.008 | 0.395±0.068 | 0.370±0.041 | 0.238 | 0.131 |
| Alg. 2 copy | GB/GB | MAE | 0.408±0.059 | 0.232±0.011 | 0.205±0.024 | 0.148±0.017 | 0.442±0.031 | 0.276±0.016 | 0.285 | 0.155 |
| Alg. 1 copy | NN/GB | MAE | 0.359±0.077 | 0.382±0.040 | 0.075±0.010 | 0.071±0.005 | 0.943±0.125 | 0.626±0.042 | 0.409 | 0.211 |
| Alg. 2 copy | NN/GB | MAE | 0.350±0.042 | 0.317±0.023 | 0.150±0.025 | 0.119±0.011 | 0.856±0.069 | 0.587±0.041 | 0.396 | 0.208 |

## Metric: RMSE

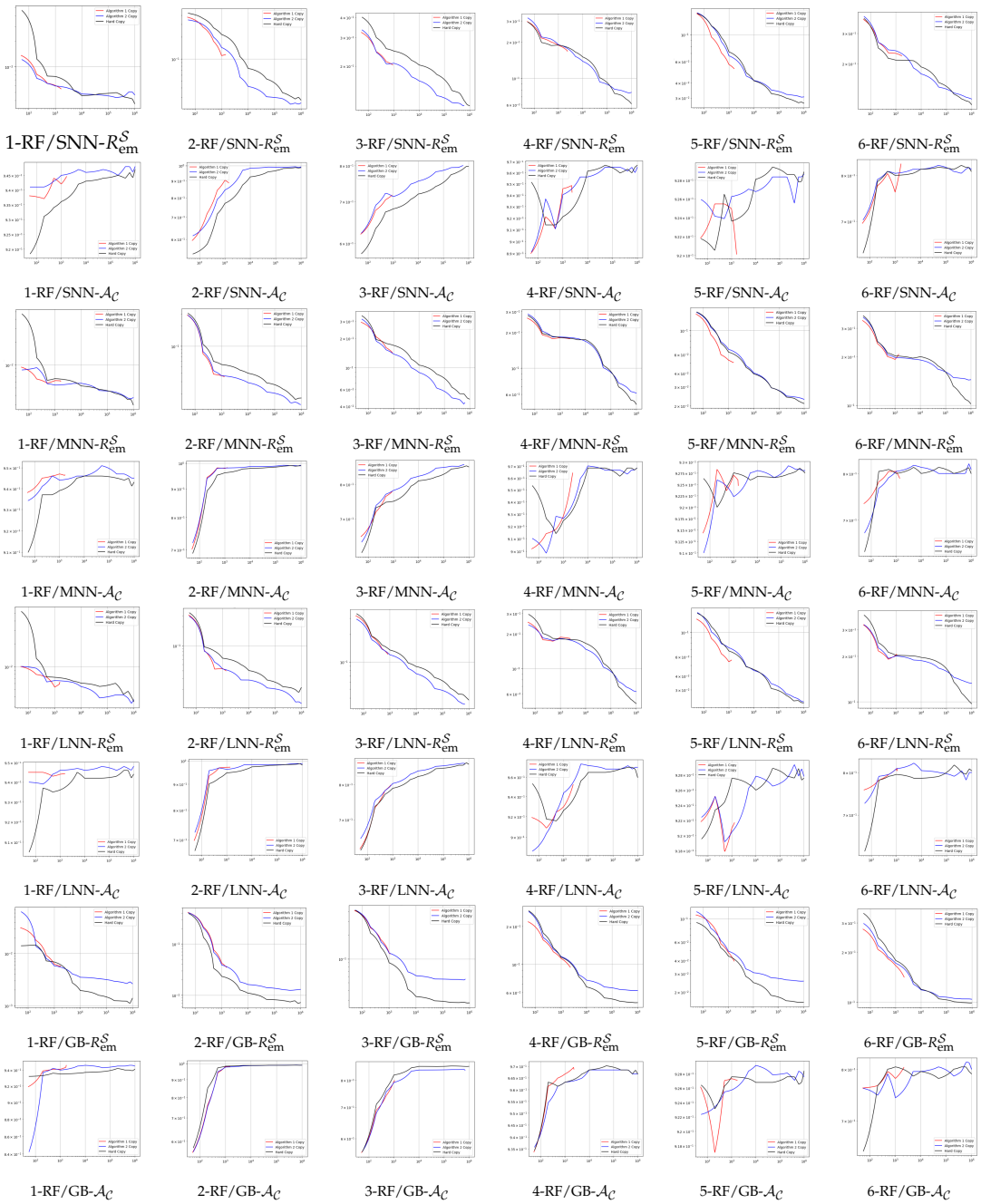
| Copy | $f_{\mathcal{O}}/f_{\mathcal{C}}$ | Metric | Dataset 1 | | Dataset 2 | | Dataset 3 | | Avg. Dat. 1-3 | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | | |
| Alg. 1 copy | RF/SNN | RMSE | 0.024±0.003 | 0.026±0.004 | 0.035±0.005 | 0.033±0.003 | 0.037±0.001 | 0.039±0.004 | 0.032 | 0.211 |
| Alg. 2 copy | RF/SNN | RMSE | 0.024±0.002 | 0.118±0.001 | 0.011±0.001 | 0.012±0.001 | 0.024±0.002 | 0.025±0.001 | 0.036 | 0.211 |
| Alg. 1 copy | GB/SNN | RMSE | 0.058±0.004 | 0.055±0.004 | 0.025±0.004 | 0.026±0.003 | 0.037±0.001 | 0.038±0.001 | 0.040 | 0.205 |
| Alg. 2 copy | GB/SNN | RMSE | 0.050±0.003 | 0.110±0.007 | 0.012±0.001 | 0.012±0.001 | 0.024±0.001 | 0.024±0.002 | 0.039 | 0.174 |
| Alg. 1 copy | NN/SNN | RMSE | 0.009±0.001 | 0.009±0.001 | 0.021±0.005 | 0.017±0.003 | 0.032±0.002 | 0.034±0.003 | 0.020 | 0.165 |
| Alg. 2 copy | NN/SNN | RMSE | 0.017±0.001 | 0.111±0.005 | 0.011±0.001 | 0.011±0.001 | 0.018±0.001 | 0.019±0.002 | 0.031 | 0.249 |
| Alg. 1 copy | RF/MNN | RMSE | 0.023±0.002 | 0.025±0.002 | 0.015±0.003 | 0.017±0.005 | 0.027±0.001 | 0.026±0.001 | 0.022 | 0.218 |
| Alg. 2 copy | RF/MNN | RMSE | 0.024±0.002 | 0.118±0.003 | 0.011±0.001 | 0.011±0.002 | 0.013±0.001 | 0.015±0.004 | 0.032 | 0.212 |
| Alg. 1 copy | GB/MNN | RMSE | 0.054±0.003 | 0.053±0.005 | 0.013±0.001 | 0.013±0.002 | 0.026±0.002 | 0.025±0.002 | 0.031 | 0.215 |
| Alg. 2 copy | GB/MNN | RMSE | 0.049±0.003 | 0.111±0.005 | 0.009±0.000 | 0.010±0.000 | 0.012±0.001 | 0.011±0.001 | 0.034 | 0.192 |
| Alg. 1 copy | NN/MNN | RMSE | 0.008±0.002 | 0.012±0.003 | 0.010±0.002 | 0.010±0.002 | 0.017±0.002 | 0.017±0.002 | 0.012 | 0.180 |
| Alg. 2 copy | NN/MNN | RMSE | 0.018±0.002 | 0.112±0.003 | 0.008±0.001 | 0.009±0.001 | 0.010±0.001 | 0.010±0.000 | 0.028 | 0.246 |
| Alg. 1 copy | RF/LNN | RMSE | 0.023±0.003 | 0.025±0.005 | 0.014±0.001 | 0.015±0.001 | 0.026±0.004 | 0.026±0.004 | 0.022 | 0.210 |
| Alg. 2 copy | RF/LNN | RMSE | 0.025±0.003 | 0.121±0.004 | 0.008±0.001 | 0.009±0.001 | 0.013±0.001 | 0.014±0.002 | 0.032 | 0.227 |
| Alg. 1 copy | GB/LNN | RMSE | 0.054±0.005 | 0.059±0.010 | 0.013±0.001 | 0.013±0.001 | 0.024±0.001 | 0.023±0.002 | 0.031 | 0.200 |
| Alg. 2 copy | GB/LNN | RMSE | 0.051±0.003 | 0.108±0.005 | 0.009±0.001 | 0.011±0.001 | 0.010±0.001 | 0.009±0.001 | 0.033 | 0.196 |
| Alg. 1 copy | NN/LNN | RMSE | 0.012±0.005 | 0.013±0.004 | 0.009±0.001 | 0.009±0.001 | 0.019±0.003 | 0.018±0.005 | 0.013 | 0.190 |
| Alg. 2 copy | NN/LNN | RMSE | 0.018±0.001 | 0.111±0.003 | 0.008±0.000 | 0.008±0.001 | 0.010±0.001 | 0.009±0.001 | 0.027 | 0.256 |
| Alg. 1 copy | RF/GB | RMSE | 0.027±0.003 | 0.027±0.003 | 0.016±0.001 | 0.018±0.003 | 0.028±0.001 | 0.027±0.003 | 0.024 | 0.219 |
| Alg. 2 copy | RF/GB | RMSE | 0.025±0.001 | 0.115±0.004 | 0.010±0.001 | 0.010±0.000 | 0.016±0.000 | 0.018±0.002 | 0.032 | 0.206 |
| Alg. 1 copy | GB/GB | RMSE | 0.059±0.004 | 0.051±0.007 | 0.016±0.001 | 0.017±0.001 | 0.023±0.001 | 0.023±0.001 | 0.032 | 0.182 |
| Alg. 2 copy | GB/GB | RMSE | 0.051±0.003 | 0.115±0.005 | 0.011±0.000 | 0.010±0.001 | 0.017±0.000 | 0.018±0.000 | 0.037 | 0.193 |
| Alg. 1 copy | NN/GB | RMSE | 0.013±0.003 | 0.012±0.001 | 0.021±0.001 | 0.016±0.001 | 0.020±0.001 | 0.018±0.001 | 0.017 | 0.291 |
| Alg. 2 copy | NN/GB | RMSE | 0.018±0.001 | 0.110±0.005 | 0.017±0.001 | 0.013±0.001 | 0.017±0.001 | 0.016±0.001 | 0.032 | 0.273 |

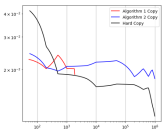
| Copy | $f_{\mathcal{O}}/f_{\mathcal{C}}$ | Metric | Dataset 4 | | Dataset 5 | | Dataset 6 | | Avg. Dat. 4-6 | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | $\mathscr{D}'_{\text{te}}$ real data | $\mathcal{S}'$ uniform data | | |
| Alg. 1 copy | RF/SNN | RMSE | 0.420±0.034 | 0.379±0.023 | 0.101±0.007 | 0.112±0.008 | 0.669±0.134 | 0.662±0.064 | 0.390 | 0.211 |
| Alg. 2 copy | RF/SNN | RMSE | 0.424±0.081 | 0.302±0.013 | 0.150±0.018 | 0.172±0.026 | 0.837±0.187 | 0.430±0.021 | 0.386 | 0.211 |
| Alg. 1 copy | GB/SNN | RMSE | 0.310±0.030 | 0.383±0.030 | 0.151±0.016 | 0.113±0.003 | 0.622±0.117 | 0.643±0.050 | 0.370 | 0.205 |
| Alg. 2 copy | GB/SNN | RMSE | 0.296±0.023 | 0.290±0.023 | 0.224±0.026 | 0.165±0.013 | 0.458±0.090 | 0.415±0.045 | 0.308 | 0.174 |
| Alg. 1 copy | NN/SNN | RMSE | 0.268±0.095 | 0.280±0.120 | 0.076±0.012 | 0.069±0.007 | 0.621±0.079 | 0.543±0.108 | 0.310 | 0.165 |
| Alg. 2 copy | NN/SNN | RMSE | 0.425±0.169 | 0.571±0.228 | 0.164±0.015 | 0.169±0.015 | 0.925±0.130 | 0.545±0.107 | 0.466 | 0.249 |
| Alg. 1 copy | RF/MNN | RMSE | 0.484±0.044 | 0.394±0.026 | 0.107±0.023 | 0.109±0.017 | 0.766±0.128 | 0.615±0.039 | 0.413 | 0.218 |
| Alg. 2 copy | RF/MNN | RMSE | 0.496±0.090 | 0.289±0.012 | 0.159±0.011 | 0.167±0.010 | 0.712±0.145 | 0.525±0.030 | 0.391 | 0.212 |
| Alg. 1 copy | GB/MNN | RMSE | 0.388±0.071 | 0.424±0.034 | 0.139±0.010 | 0.122±0.007 | 0.700±0.076 | 0.619±0.077 | 0.399 | 0.215 |
| Alg. 2 copy | GB/MNN | RMSE | 0.350±0.068 | 0.316±0.014 | 0.246±0.020 | 0.192±0.019 | 0.511±0.072 | 0.478±0.043 | 0.349 | 0.192 |
| Alg. 1 copy | NN/MNN | RMSE | 0.238±0.028 | 0.367±0.193 | 0.077±0.007 | 0.072±0.009 | 0.738±0.169 | 0.594±0.221 | 0.348 | 0.180 |
| Alg. 2 copy | NN/MNN | RMSE | 0.443±0.079 | 0.508±0.168 | 0.182±0.017 | 0.166±0.017 | 1.016±0.091 | 0.469±0.042 | 0.464 | 0.246 |
| Alg. 1 copy | RF/LNN | RMSE | 0.486±0.095 | 0.388±0.036 | 0.116±0.038 | 0.118±0.019 | 0.691±0.102 | 0.584±0.035 | 0.397 | 0.210 |
| Alg. 2 copy | RF/LNN | RMSE | 0.653±0.112 | 0.292±0.022 | 0.147±0.005 | 0.148±0.014 | 0.802±0.135 | 0.486±0.043 | 0.421 | 0.227 |
| Alg. 1 copy | GB/LNN | RMSE | 0.341±0.039 | 0.396±0.016 | 0.160±0.023 | 0.143±0.016 | 0.606±0.074 | 0.561±0.029 | 0.368 | 0.200 |
| Alg. 2 copy | GB/LNN | RMSE | 0.372±0.047 | 0.328±0.027 | 0.243±0.022 | 0.171±0.010 | 0.583±0.119 | 0.454±0.047 | 0.358 | 0.196 |
| Alg. 1 copy | NN/LNN | RMSE | 0.323±0.082 | 0.378±0.129 | 0.073±0.010 | 0.071±0.009 | 0.816±0.129 | 0.536±0.092 | 0.366 | 0.190 |
| Alg. 2 copy | NN/LNN | RMSE | 0.558±0.109 | 0.468±0.058 | 0.169±0.010 | 0.150±0.010 | 1.031±0.202 | 0.525±0.121 | 0.484 | 0.256 |
| Alg. 1 copy | RF/GB | RMSE | 0.619±0.067 | 0.296±0.020 | 0.110±0.006 | 0.105±0.012 | 0.765±0.074 | 0.580±0.045 | 0.413 | 0.219 |
| Alg. 2 copy | RF/GB | RMSE | 0.610±0.047 | 0.269±0.017 | 0.139±0.012 | 0.154±0.015 | 0.661±0.120 | 0.443±0.009 | 0.379 | 0.206 |
| Alg. 1 copy | GB/GB | RMSE | 0.382±0.039 | 0.297±0.034 | 0.142±0.010 | 0.104±0.009 | 0.555±0.091 | 0.514±0.055 | 0.332 | 0.182 |
| Alg. 2 copy | GB/GB | RMSE | 0.473±0.062 | 0.289±0.017 | 0.250±0.022 | 0.178±0.016 | 0.536±0.022 | 0.366±0.028 | 0.349 | 0.193 |
| Alg. 1 copy | NN/GB | RMSE | 0.656±0.234 | 0.487±0.056 | 0.097±0.013 | 0.090±0.005 | 1.260±0.252 | 0.799±0.045 | 0.565 | 0.291 |
| Alg. 2 copy | NN/GB | RMSE | 0.453±0.041 | 0.444±0.038 | 0.171±0.028 | 0.140±0.012 | 1.122±0.137 | 0.745±0.052 | 0.513 | 0.273 |

# Appendix C

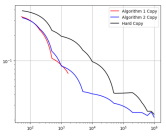
# Supplemental material. Figures

## Figures 4.2 and 4.5 (Dataset - Black box / Copy - Metric)

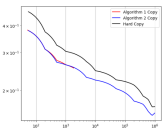




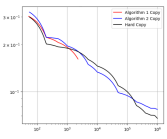
1-GB/SNN-$R^{\mathcal{S}}_{\text{em}}$



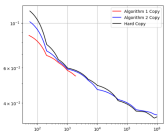
2-GB/SNN-$R^{\mathcal{S}}_{\text{em}}$



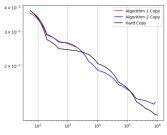
3-GB/SNN-$R^{\mathcal{S}}_{\text{em}}$



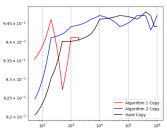
4-GB/SNN-$R^{\mathcal{S}}_{\text{em}}$



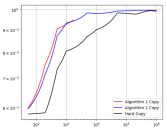
5-GB/SNN-$R^{\mathcal{S}}_{\text{em}}$



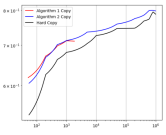
6-GB/SNN-$R^{\mathcal{S}}_{\text{em}}$



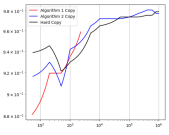
1-GB/SNN-$\mathcal{A}_{\mathcal{C}}$



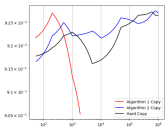
2-GB/SNN-$\mathcal{A}_{\mathcal{C}}$



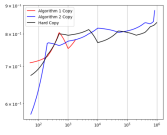
3-GB/SNN-$\mathcal{A}_{\mathcal{C}}$



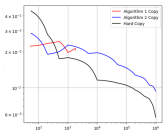
4-GB/SNN-$\mathcal{A}_{\mathcal{C}}$



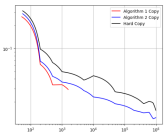
5-GB/SNN-$\mathcal{A}_{\mathcal{C}}$



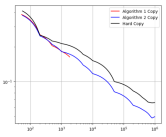
6-GB/SNN-$\mathcal{A}_{\mathcal{C}}$



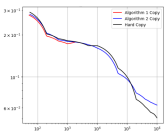
1-GB/MNN-$R^{\mathcal{S}}_{\text{em}}$



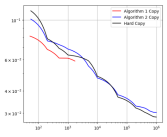
2-GB/MNN-$R^{\mathcal{S}}_{\text{em}}$



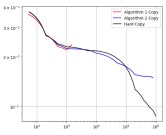
3-GB/MNN-$R^{\mathcal{S}}_{\text{em}}$



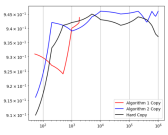
4-GB/MNN-$R^{\mathcal{S}}_{\text{em}}$



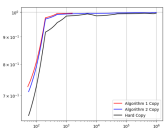
5-GB/MNN-$R^{\mathcal{S}}_{\text{em}}$



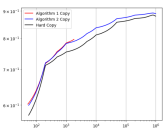
6-GB/MNN-$R^{\mathcal{S}}_{\text{em}}$



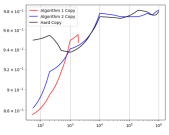
1-GB/MNN-$\mathcal{A}_{\mathcal{C}}$



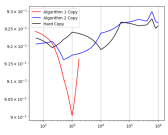
2-GB/MNN-$\mathcal{A}_{\mathcal{C}}$



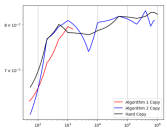
3-GB/MNN-$\mathcal{A}_{\mathcal{C}}$



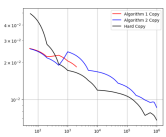
4-GB/MNN-$\mathcal{A}_{\mathcal{C}}$



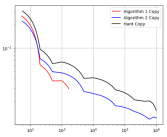
5-GB/MNN-$\mathcal{A}_{\mathcal{C}}$



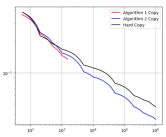
6-GB/MNN-$\mathcal{A}_{\mathcal{C}}$



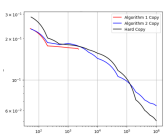
1-GB/LNN-$R^{\mathcal{S}}_{\text{em}}$



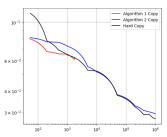
2-GB/LNN-$R^{\mathcal{S}}_{\text{em}}$



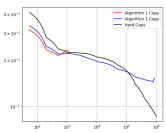
3-GB/LNN-$R^{\mathcal{S}}_{\text{em}}$



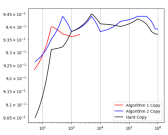
4-GB/LNN-$R^{\mathcal{S}}_{\text{em}}$



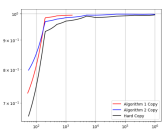
5-GB/LNN-$R^{\mathcal{S}}_{\text{em}}$



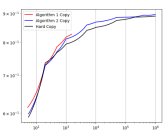
6-GB/LNN-$R^{\mathcal{S}}_{\text{em}}$



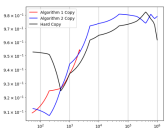
1-GB/LNN-$\mathcal{A}_{\mathcal{C}}$



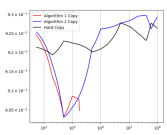
2-GB/LNN-$\mathcal{A}_{\mathcal{C}}$



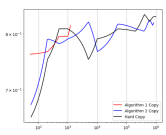
3-GB/LNN-$\mathcal{A}_{\mathcal{C}}$



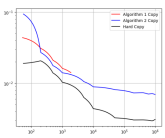
4-GB/LNN-$\mathcal{A}_{\mathcal{C}}$



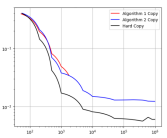
5-GB/LNN-$\mathcal{A}_{\mathcal{C}}$



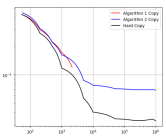
6-GB/LNN-$\mathcal{A}_{\mathcal{C}}$



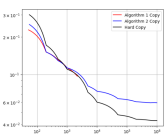
1-GB/GB-$R^{\mathcal{S}}_{\text{em}}$



2-GB/GB-$R^{\mathcal{S}}_{\text{em}}$



3-GB/GB-$R^{\mathcal{S}}_{\text{em}}$



4-GB/GB-$R^{\mathcal{S}}_{\text{em}}$



5-GB/GB-$R^{\mathcal{S}}_{\text{em}}$



6-GB/GB-$R^{\mathcal{S}}_{\text{em}}$



1-GB/GB-$\mathcal{A}_{\mathcal{C}}$



2-GB/GB-$\mathcal{A}_{\mathcal{C}}$



3-GB/GB-$\mathcal{A}_{\mathcal{C}}$



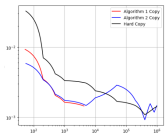
4-GB/GB-$\mathcal{A}_{\mathcal{C}}$



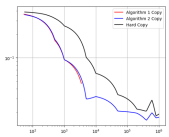
5-GB/GB-$\mathcal{A}_{\mathcal{C}}$



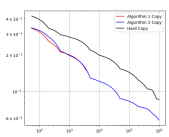
6-GB/GB-$\mathcal{A}_{\mathcal{C}}$



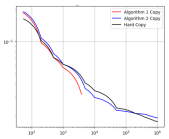
1-NN/SNN-$R^{\mathcal{S}}_{\text{em}}$



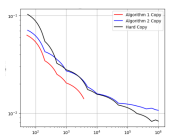
2-NN/SNN-$R^{\mathcal{S}}_{\text{em}}$



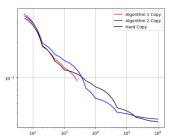
3-NN/SNN-$R^{\mathcal{S}}_{\text{em}}$



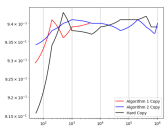
4-NN/SNN-$R^{\mathcal{S}}_{\text{em}}$



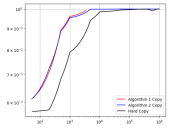
5-NN/SNN-$R^{\mathcal{S}}_{\text{em}}$



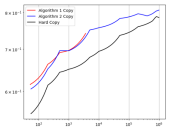
6-NN/SNN-$R^{\mathcal{S}}_{\text{em}}$



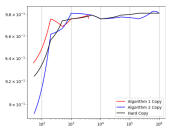
1-NN/SNN-$\mathcal{A}_{\mathcal{C}}$



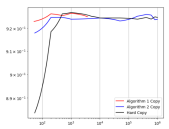
2-NN/SNN-$\mathcal{A}_{\mathcal{C}}$



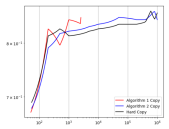
3-NN/SNN-$\mathcal{A}_{\mathcal{C}}$



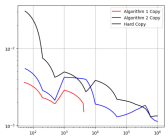
4-NN/SNN-$\mathcal{A}_{\mathcal{C}}$



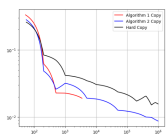
5-NN/SNN-$\mathcal{A}_{\mathcal{C}}$



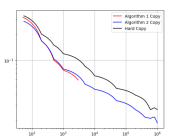
6-NN/SNN-$\mathcal{A}_{\mathcal{C}}$



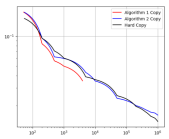
1-NN/MNN-$R^{\mathcal{S}}_{\text{em}}$



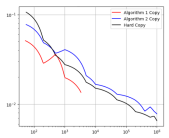
2-NN/MNN-$R^{\mathcal{S}}_{\text{em}}$



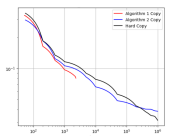
3-NN/MNN-$R^{\mathcal{S}}_{\text{em}}$



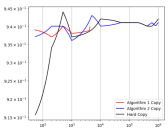
4-NN/MNN-$R^{\mathcal{S}}_{\text{em}}$



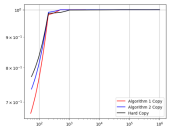
5-NN/MNN-$R^{\mathcal{S}}_{\text{em}}$



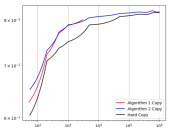
6-NN/MNN-$R^{\mathcal{S}}_{\text{em}}$



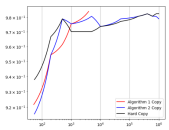
1-NN/MNN-$\mathcal{A}_{\mathcal{C}}$



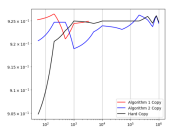
2-NN/MNN-$\mathcal{A}_{\mathcal{C}}$



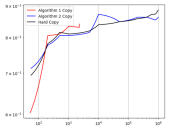
3-NN/MNN-$\mathcal{A}_{\mathcal{C}}$



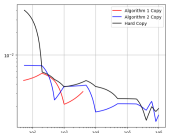
4-NN/MNN-$\mathcal{A}_{\mathcal{C}}$



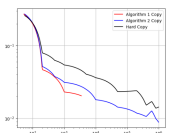
5-NN/MNN-$\mathcal{A}_{\mathcal{C}}$



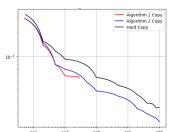
6-NN/MNN-$\mathcal{A}_{\mathcal{C}}$



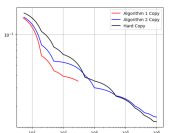
1-NN/LNN-$R^{\mathcal{S}}_{\text{em}}$



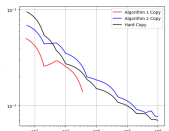
2-NN/LNN-$R^{\mathcal{S}}_{\text{em}}$



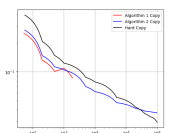
3-NN/LNN-$R^{\mathcal{S}}_{\text{em}}$



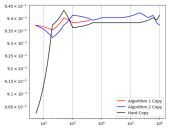
4-NN/LNN-$R^{\mathcal{S}}_{\text{em}}$



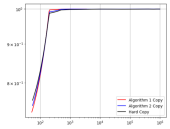
5-NN/LNN-$R^{\mathcal{S}}_{\text{em}}$



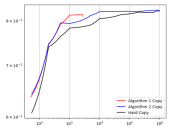
6-NN/LNN-$R^{\mathcal{S}}_{\text{em}}$



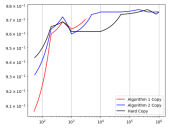
1-NN/LNN-$\mathcal{A}_{\mathcal{C}}$



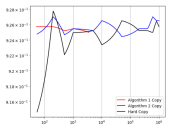
2-NN/LNN-$\mathcal{A}_{\mathcal{C}}$



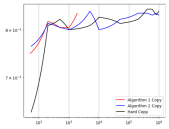
3-NN/LNN-$\mathcal{A}_{\mathcal{C}}$



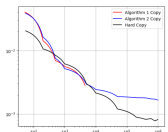
4-NN/LNN-$\mathcal{A}_{\mathcal{C}}$



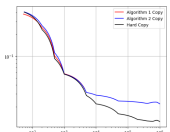
5-NN/LNN-$\mathcal{A}_{\mathcal{C}}$



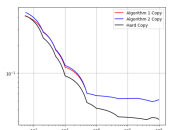
6-NN/LNN-$\mathcal{A}_{\mathcal{C}}$



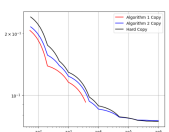
1-NN/GB-$R^{\mathcal{S}}_{\text{em}}$



2-NN/GB-$R^{\mathcal{S}}_{\text{em}}$



3-NN/GB-$R^{\mathcal{S}}_{\text{em}}$



4-NN/GB-$R^{\mathcal{S}}_{\text{em}}$



5-NN/GB-$R^{\mathcal{S}}_{\text{em}}$



6-NN/GB-$R^{\mathcal{S}}_{\text{em}}$



1-NN/GB-$\mathcal{A}_{\mathcal{C}}$



2-NN/GB-$\mathcal{A}_{\mathcal{C}}$



3-NN/GB-$\mathcal{A}_{\mathcal{C}}$



4-NN/GB-$\mathcal{A}_{\mathcal{C}}$

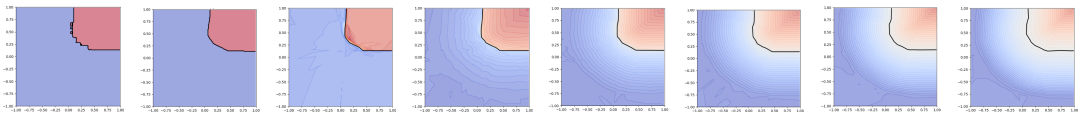


5-NN/GB-$\mathcal{A}_{\mathcal{C}}$

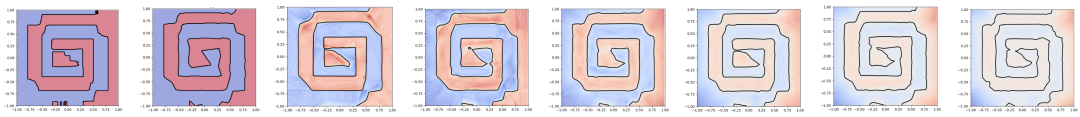


6-NN/GB-$\mathcal{A}_{\mathcal{C}}$

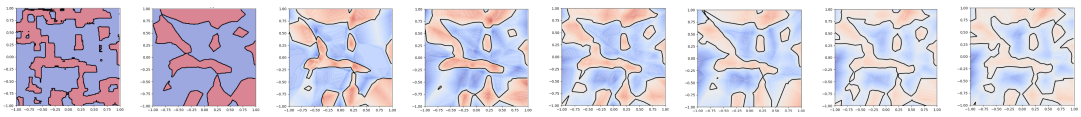
# Figure 4.3 (Dataset - Black box / Copy - Regularization)



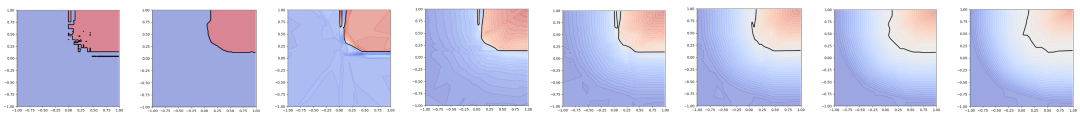
1-RF/SNN-Bb 1-RF/SNN-Hc 1-RF/SNN-0 1-RF/SNN-0.25 1-RF/SNN-0.5 1-RF/SNN-0.75 1-RF/SNN-1 1-RF/SNN-1.25



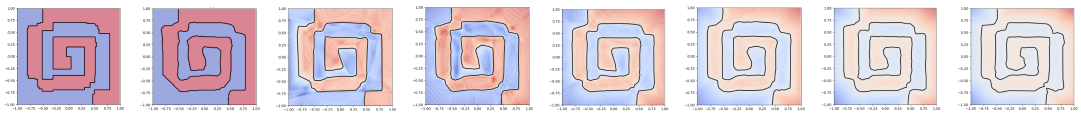
2-RF/SNN-Bb 2-RF/SNN-Hc 2-RF/SNN-0 2-RF/SNN-0.25 2-RF/SNN-0.5 2-RF/SNN-0.75 2-RF/SNN-1 2-RF/SNN-1.25



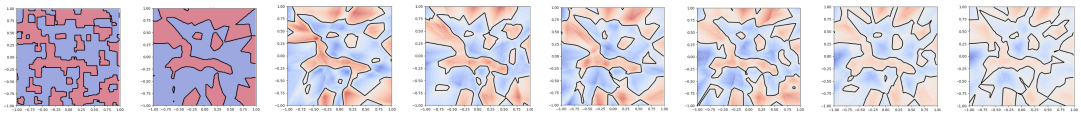
3-RF/SNN-Bb 3-RF/SNN-Hc 3-RF/SNN-0 3-RF/SNN-0.25 3-RF/SNN-0.5 3-RF/SNN-0.75 3-RF/SNN-1 3-RF/SNN-1.25



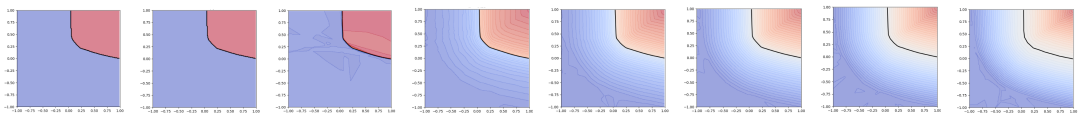
1-GB/SNN-Bb 1-GB/SNN-Hc 1-GB/SNN-0 1-GB/SNN-0.25 1-GB/SNN-0.5 1-GB/SNN-0.75 1-GB/SNN-1 1-GB/SNN-1.25



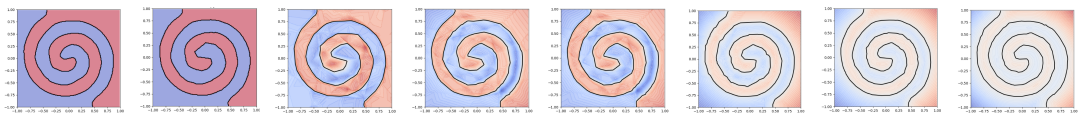
2-GB/SNN-Bb 2-GB/SNN-Hc 2-GB/SNN-0 2-GB/SNN-0.25 2-GB/SNN-0.5 2-GB/SNN-0.75 2-GB/SNN-1 2-GB/SNN-1.25



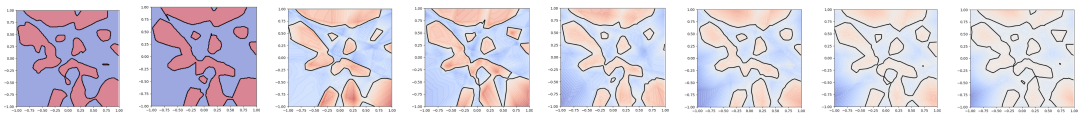
3-GB/SNN-Bb 3-GB/SNN-Hc 3-GB/SNN-0 3-GB/SNN-0.25 3-GB/SNN-0.5 3-GB/SNN-0.75 3-GB/SNN-1 3-GB/SNN-1.25



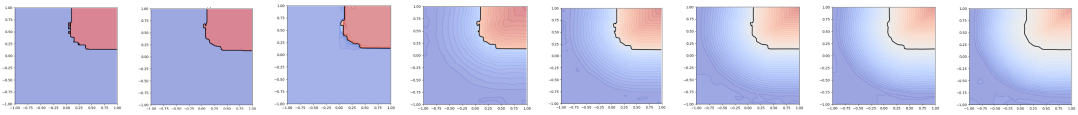
1-NN/SNN-Bb 1-NN/SNN-Hc 1-NN/SNN-0 1-NN/SNN-0.25 1-NN/SNN-0.5 1-NN/SNN-0.75 1-NN/SNN-1 1-NN/SNN-1.25



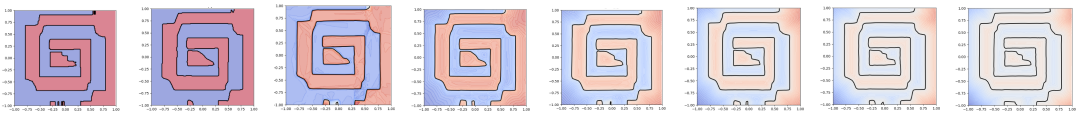
2-NN/SNN-Bb 2-NN/SNN-Hc 2-NN/SNN-0 2-NN/SNN-0.25 2-NN/SNN-0.5 2-NN/SNN-0.75 2-NN/SNN-1 2-NN/SNN-1.25



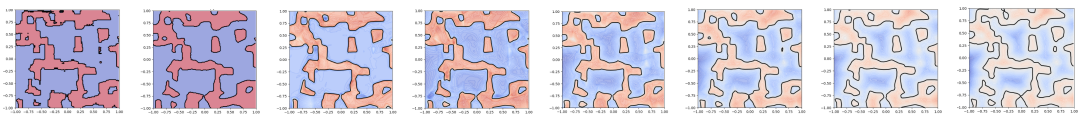
3-NN/SNN-Bb 3-NN/SNN-Hc 3-NN/SNN-0 3-NN/SNN-0.25 3-NN/SNN-0.5 3-NN/SNN-0.75 3-NN/SNN-1 3-NN/SNN-1.25



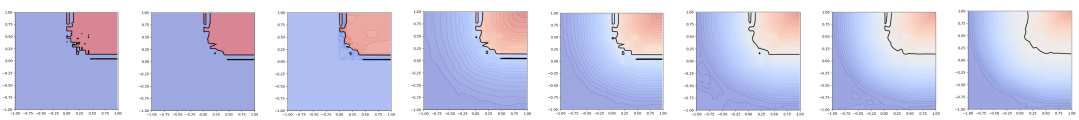
1-RF/MNN-Bb 1-RF/MNN-Hc 1-RF/MNN-0 1-RF/MNN-0.25 1-RF/MNN-0.5 1-RF/MNN-0.75 1-RF/MNN-1 1-RF/MNN-1.25



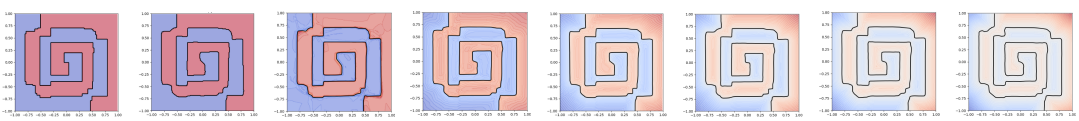
2-RF/MNN-Bb 2-RF/MNN-Hc 2-RF/MNN-0 2-RF/MNN-0.25 2-RF/MNN-0.5 2-RF/MNN-0.75 2-RF/MNN-1 2-RF/MNN-1.25



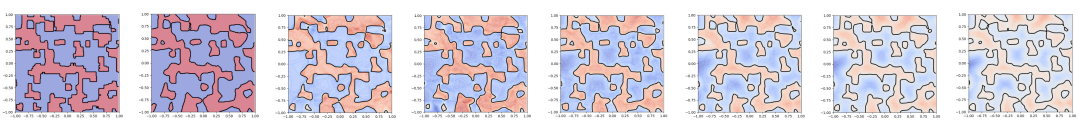
3-RF/MNN-Bb 3-RF/MNN-Hc 3-RF/MNN-0 3-RF/MNN-0.25 3-RF/MNN-0.5 3-RF/MNN-0.75 3-RF/MNN-1 3-RF/MNN-1.25



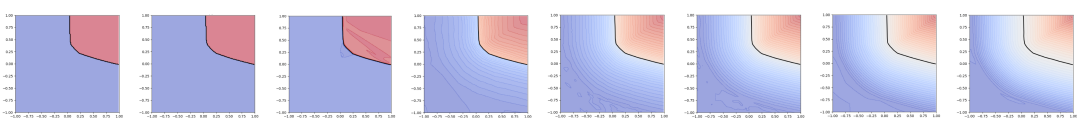
1-GB/MNN-Bb 1-GB/MNN-Hc 1-GB/MNN-0 1-GB/MNN-0.25 1-GB/MNN-0.5 1-GB/MNN-0.75 1-GB/MNN-1 1-GB/MNN-1.25



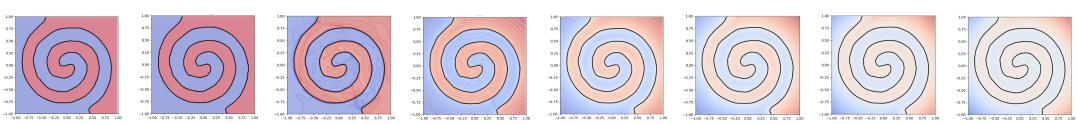
2-GB/MNN-Bb 2-GB/MNN-Hc 2-GB/MNN-0 2-GB/MNN-0.25 2-GB/MNN-0.5 2-GB/MNN-0.75 2-GB/MNN-1 2-GB/MNN-1.25



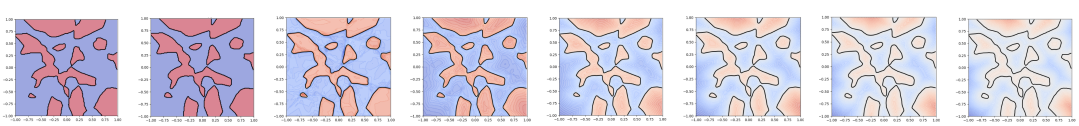
3-GB/MNN-Bb 3-GB/MNN-Hc 3-GB/MNN-0 3-GB/MNN-0.25 3-GB/MNN-0.5 3-GB/MNN-0.75 3-GB/MNN-1 3-GB/MNN-1.25



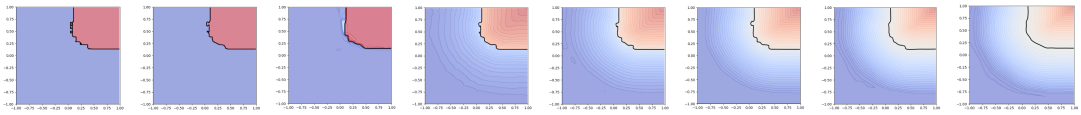
1-NN/MNN-Bb 1-NN/MNN-Hc 1-NN/MNN-0 1-NN/MNN-0.25 1-NN/MNN-0.5 1-NN/MNN-0.75 1-NN/MNN-1 1-NN/MNN-1.25



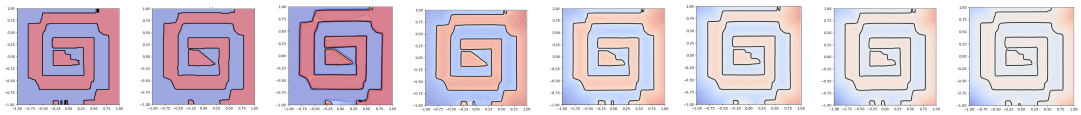
2-NN/MNN-Bb 2-NN/MNN-Hc 2-NN/MNN-0 2-NN/MNN-0.25 2-NN/MNN-0.5 2-NN/MNN-0.75 2-NN/MNN-1 2-NN/MNN-1.25



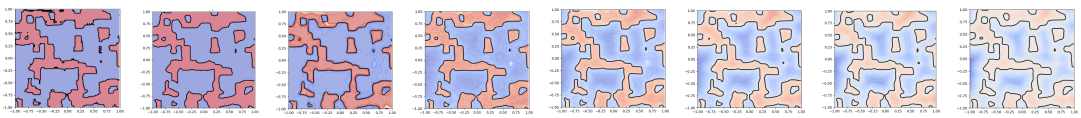
3-NN/MNN-Bb 3-NN/MNN-Hc 3-NN/MNN-0 3-NN/MNN-0.25 3-NN/MNN-0.5 3-NN/MNN-0.75 3-NN/MNN-1 3-NN/MNN-1.25



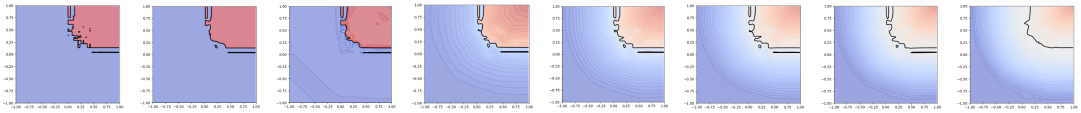
1-RF/LNN-Bb 1-RF/LNN-Hc 1-RF/LNN-0 1-RF/LNN-0.25 1-RF/LNN-0.5 1-RF/LNN-0.75 1-RF/LNN-1 1-RF/LNN-1.25



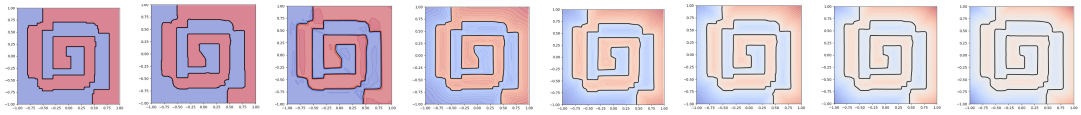
2-RF/LNN-Bb 2-RF/LNN-Hc 2-RF/LNN-0 2-RF/LNN-0.25 2-RF/LNN-0.5 2-RF/LNN-0.75 2-RF/LNN-1 2-RF/LNN-1.25



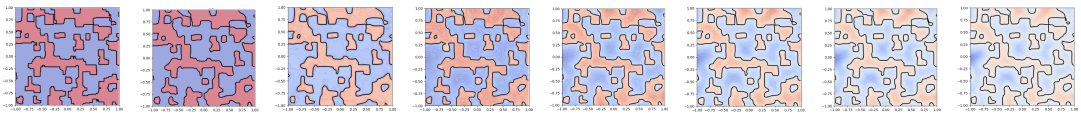
3-RF/LNN-Bb 3-RF/LNN-Hc 3-RF/LNN-0 3-RF/LNN-0.25 3-RF/LNN-0.5 3-RF/LNN-0.75 3-RF/LNN-1 3-RF/LNN-1.25



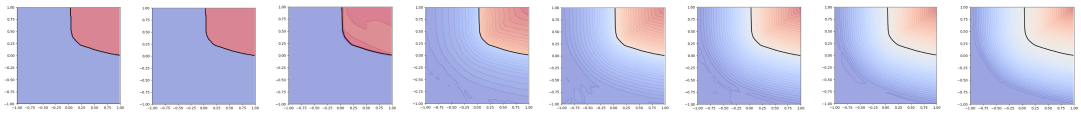
1-GB/LNN-Bb 1-GB/LNN-Hc 1-GB/LNN-0 1-GB/LNN-0.25 1-GB/LNN-0.5 1-GB/LNN-0.75 1-GB/LNN-1 1-GB/LNN-1.25



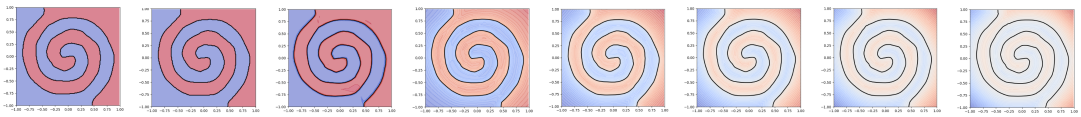
2-GB/LNN-Bb 2-GB/LNN-Hc 2-GB/LNN-0 2-GB/LNN-0.25 2-GB/LNN-0.5 2-GB/LNN-0.75 2-GB/LNN-1 2-GB/LNN-1.25



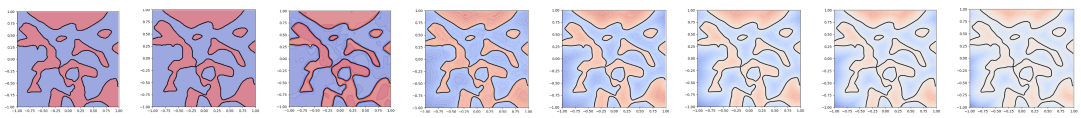
3-GB/LNN-Bb 3-GB/LNN-Hc 3-GB/LNN-0 3-GB/LNN-0.25 3-GB/LNN-0.5 3-GB/LNN-0.75 3-GB/LNN-1 3-GB/LNN-1.25



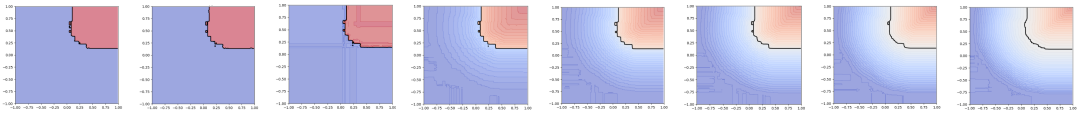
1-NN/LNN-Bb 1-NN/LNN-Hc 1-NN/LNN-0 1-NN/LNN-0.25 1-NN/LNN-0.5 1-NN/LNN-0.75 1-NN/LNN-1 1-NN/LNN-1.25



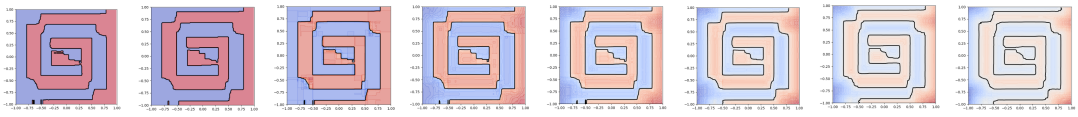
2-NN/LNN-Bb 2-NN/LNN-Hc 2-NN/LNN-0 2-NN/LNN-0.25 2-NN/LNN-0.5 2-NN/LNN-0.75 2-NN/LNN-1 2-NN/LNN-1.25



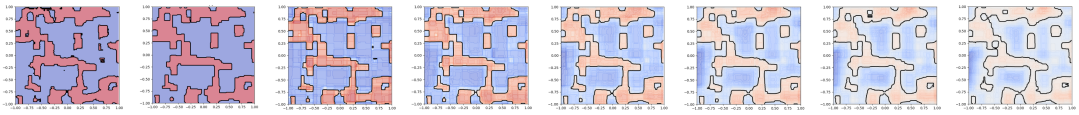
3-NN/LNN-Bb 3-NN/LNN-Hc 3-NN/LNN-0 3-NN/LNN-0.25 3-NN/LNN-0.5 3-NN/LNN-0.75 3-NN/LNN-1 3-NN/LNN-1.25



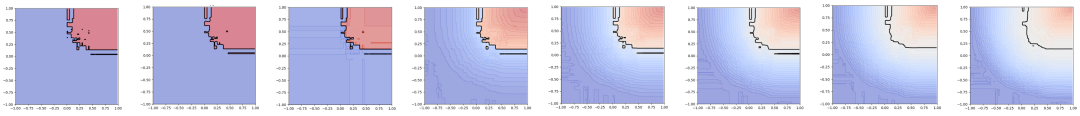
1-RF/GB-Bb 1-RF/GB-Hc 1-RF/GB-0 1-RF/GB-0.25 1-RF/GB-0.5 1-RF/GB-0.75 1-RF/GB-1 1-RF/GB-1.25



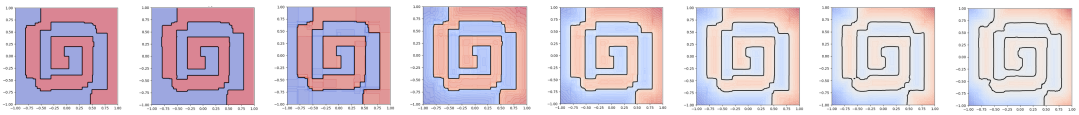
2-RF/GB-Bb 2-RF/GB-Hc 2-RF/GB-0 2-RF/GB-0.25 2-RF/GB-0.5 2-RF/GB-0.75 2-RF/GB-1 2-RF/GB-1.25



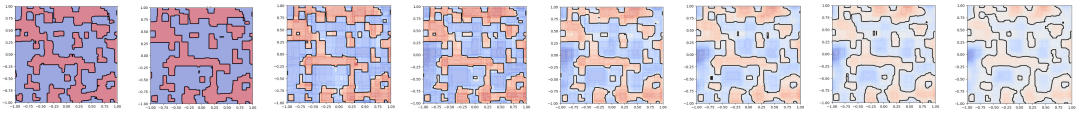
3-RF/GB-Bb 3-RF/GB-Hc 3-RF/GB-0 3-RF/GB-0.25 3-RF/GB-0.5 3-RF/GB-0.75 3-RF/GB-1 3-RF/GB-1.25



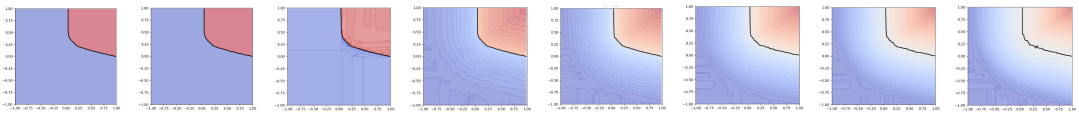
1-GB/GB-Bb 1-GB/GB-Hc 1-GB/GB-0 1-GB/GB-0.25 1-GB/GB-0.5 1-GB/GB-0.75 1-GB/GB-1 1-GB/GB-1.25



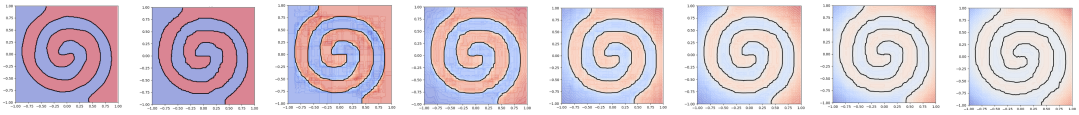
2-GB/GB-Bb 2-GB/GB-Hc 2-GB/GB-0 2-GB/GB-0.25 2-GB/GB-0.5 2-GB/GB-0.75 2-GB/GB-1 2-GB/GB-1.25



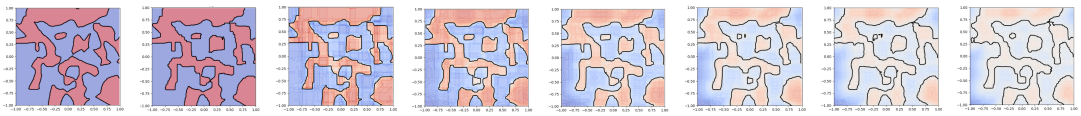
3-GB/GB-Bb 3-GB/GB-Hc 3-GB/GB-0 3-GB/GB-0.25 3-GB/GB-0.5 3-GB/GB-0.75 3-GB/GB-1 3-GB/GB-1.25



1-NN/GB-Bb 1-NN/GB-Hc 1-NN/GB-0 1-NN/GB-0.25 1-NN/GB-0.5 1-NN/GB-0.75 1-NN/GB-1 1-NN/GB-1.25

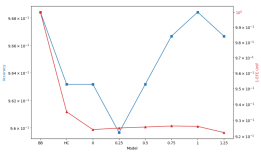


2-NN/GB-Bb 2-NN/GB-Hc 2-NN/GB-0 2-NN/GB-0.25 2-NN/GB-0.5 2-NN/GB-0.75 2-NN/GB-1 2-NN/GB-1.25

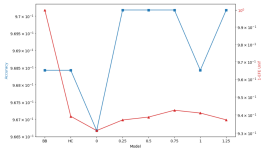


3-NN/GB-Bb 3-NN/GB-Hc 3-NN/GB-0 3-NN/GB-0.25 3-NN/GB-0.5 3-NN/GB-0.75 3-NN/GB-1 3-NN/GB-1.25

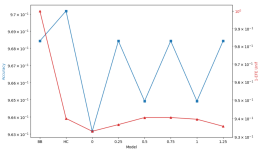
# Figure 4.4 (Dataset - Black box / Copy)



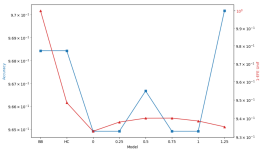
4-RF/SNN



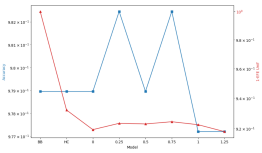
4-RF/MNN



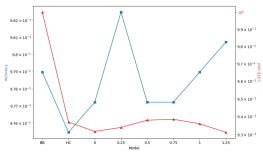
4-RF/LNN



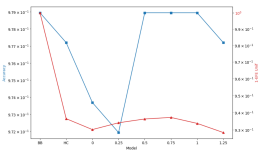
4-RF/GB



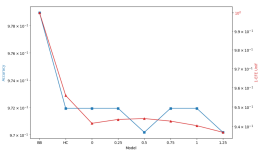
4-GB/SNN



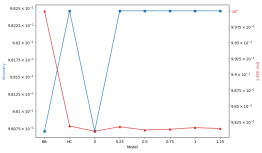
4-GB/MNN



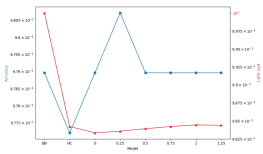
4-GB/LNN



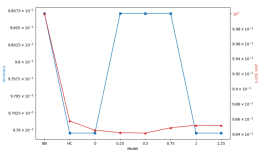
4-GB/GB



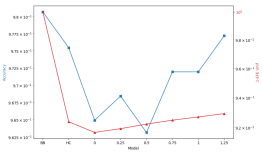
4-NN/SNN



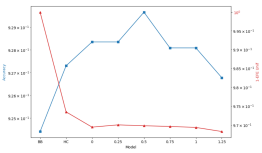
4-NN/MNN



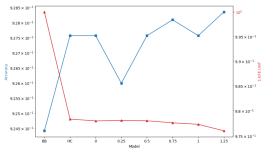
4-NN/LNN



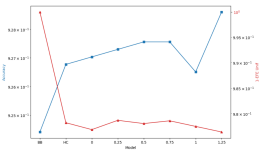
4-NN/GB



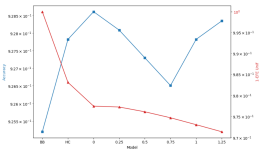
5-RF/SNN



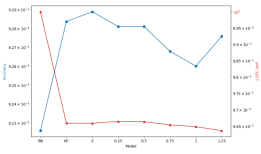
5-RF/MNN



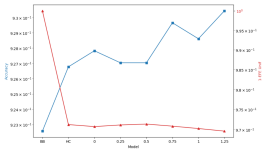
5-RF/LNN



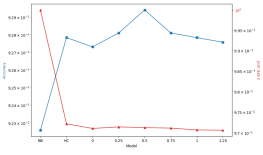
5-RF/GB



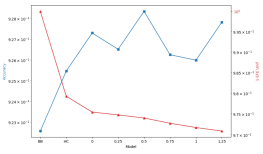
5-GB/SNN



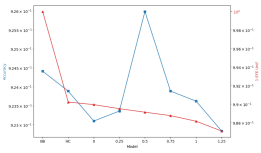
5-GB/MNN



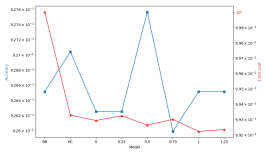
5-GB/LNN



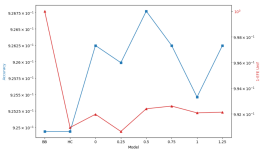
5-GB/GB



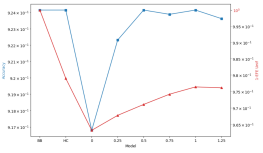
5-NN/SNN



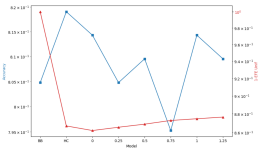
5-NN/MNN



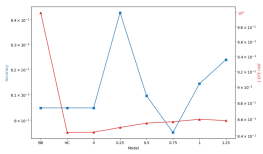
5-NN/LNN



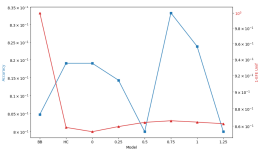
5-NN/GB



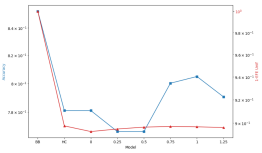
6-RF/SNN



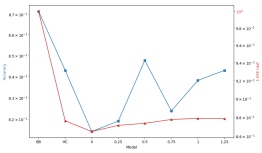
6-RF/MNN



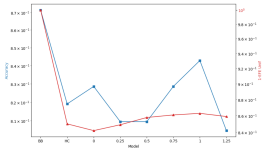
6-RF/LNN



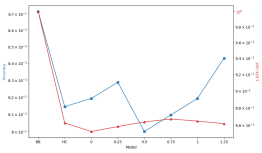
6-RF/GB



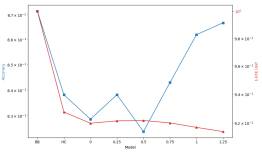
6-GB/SNN



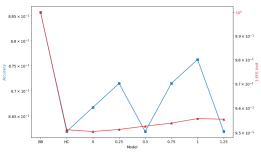
6-GB/MNN



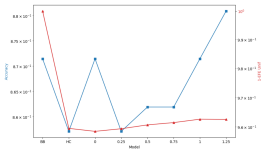
6-GB/LNN



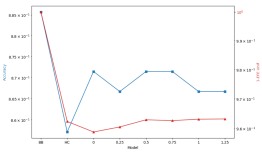
6-GB/GB



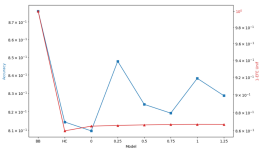
6-NN/SNN



6-NN/MNN

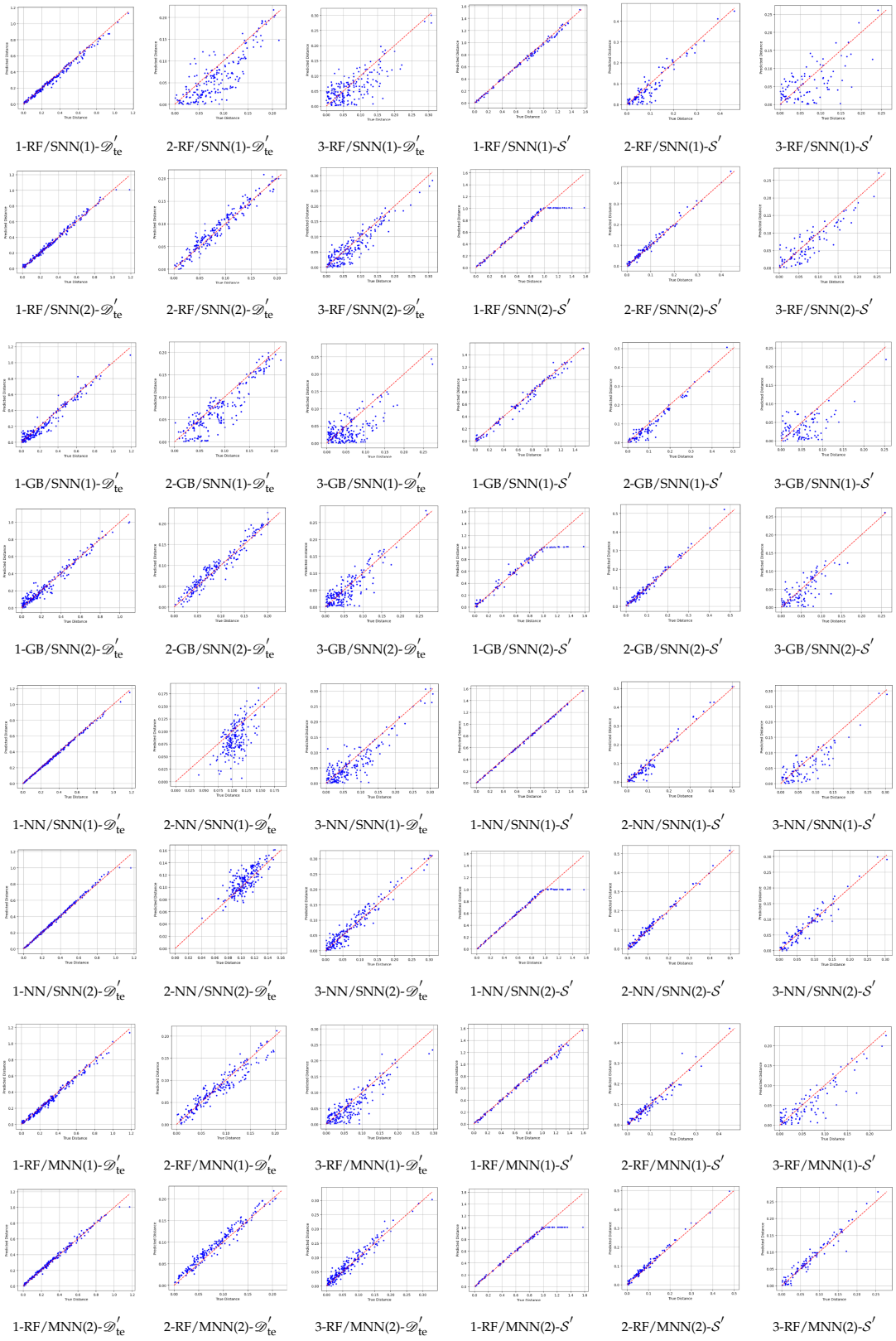


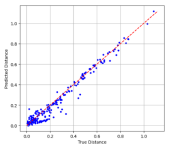
6-NN/LNN



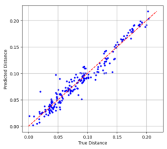
6-NN/GB

# Figure 4.6 (Dataset - Black box/Copy (Algorithm) - Samples)

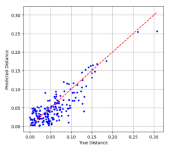




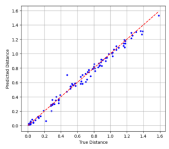
1-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



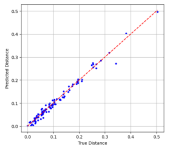
2-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



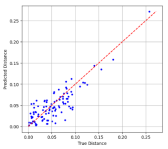
3-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



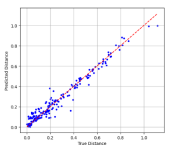
1-GB/MNN(1)-$\mathcal{S}'$



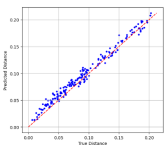
2-GB/MNN(1)-$\mathcal{S}'$



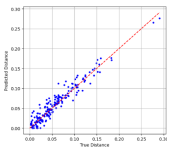
3-GB/MNN(1)-$\mathcal{S}'$



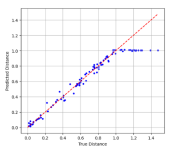
1-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



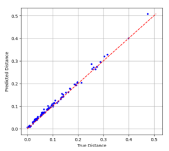
2-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



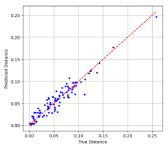
3-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



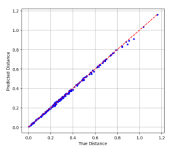
1-GB/MNN(2)-$\mathcal{S}'$



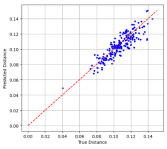
2-GB/MNN(2)-$\mathcal{S}'$



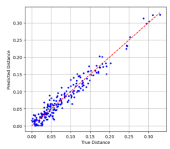
3-GB/MNN(2)-$\mathcal{S}'$



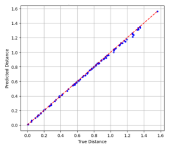
1-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



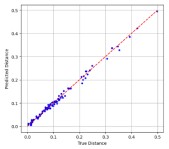
2-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



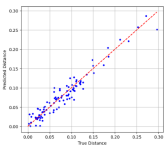
3-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



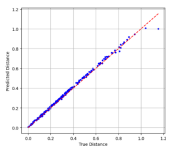
1-NN/MNN(1)-$\mathcal{S}'$



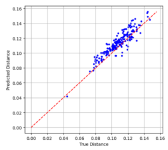
2-NN/MNN(1)-$\mathcal{S}'$



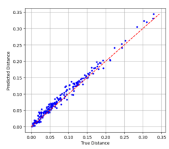
3-NN/MNN(1)-$\mathcal{S}'$



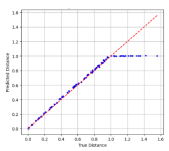
1-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



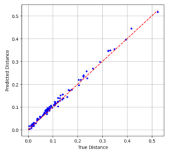
2-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



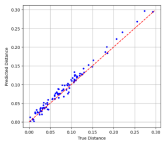
3-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



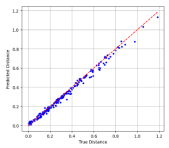
1-NN/MNN(2)-$\mathcal{S}'$



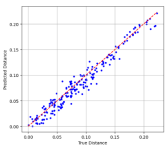
2-NN/MNN(2)-$\mathcal{S}'$



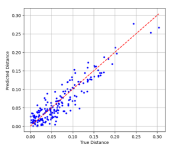
3-NN/MNN(2)-$\mathcal{S}'$



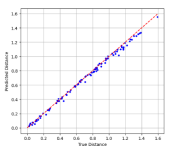
1-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



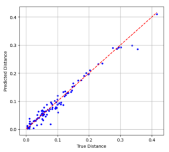
2-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



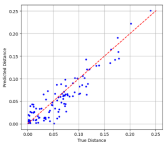
3-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



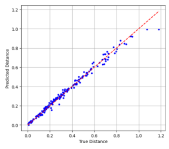
1-RF/LNN(1)-$\mathcal{S}'$



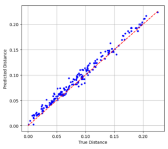
2-RF/LNN(1)-$\mathcal{S}'$



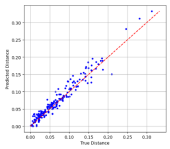
3-RF/LNN(1)-$\mathcal{S}'$



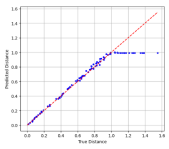
1-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



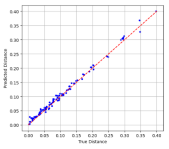
2-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



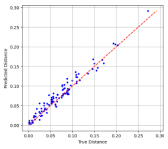
3-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



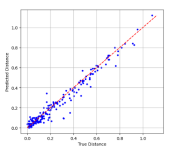
1-RF/LNN(2)-$\mathcal{S}'$



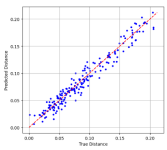
2-RF/LNN(2)-$\mathcal{S}'$



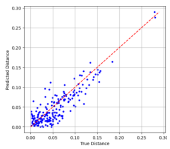
3-RF/LNN(2)-$\mathcal{S}'$



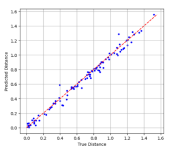
1-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



2-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



3-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



1-GB/LNN(1)-$\mathcal{S}'$



2-GB/LNN(1)-$\mathcal{S}'$



3-GB/LNN(1)-$\mathcal{S}'$



1-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



2-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



3-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



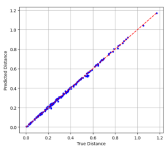
1-GB/LNN(2)-$\mathcal{S}'$



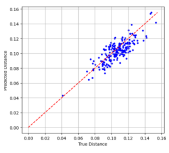
2-GB/LNN(2)-$\mathcal{S}'$



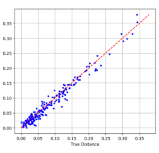
3-GB/LNN(2)-$\mathcal{S}'$



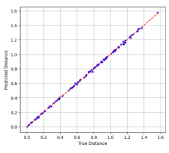
1-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



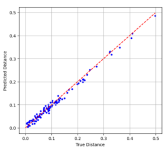
2-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



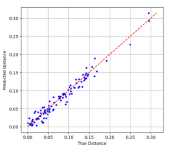
3-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



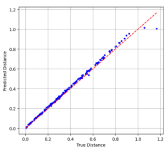
1-NN/LNN(1)-$\mathcal{S}'$



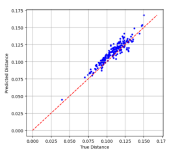
2-NN/LNN(1)-$\mathcal{S}'$



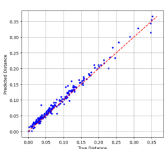
3-NN/LNN(1)-$\mathcal{S}'$



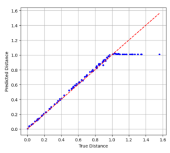
1-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



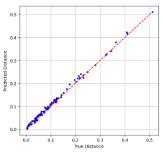
2-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



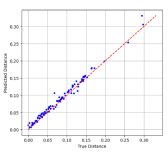
3-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



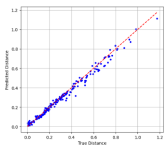
1-NN/LNN(2)-$\mathcal{S}'$



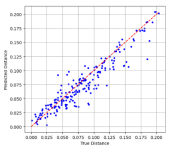
2-NN/LNN(2)-$\mathcal{S}'$



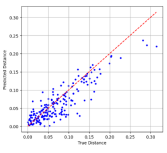
3-NN/LNN(2)-$\mathcal{S}'$



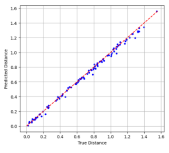
1-RF/GB(1)-$\mathscr{D}'_{\text{te}}$



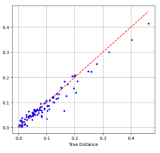
2-RF/GB(1)-$\mathscr{D}'_{\text{te}}$



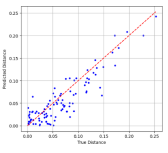
3-RF/GB(1)-$\mathscr{D}'_{\text{te}}$



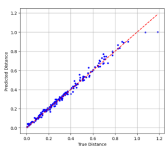
1-RF/GB(1)-$\mathcal{S}'$



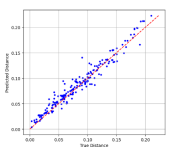
2-RF/GB(1)-$\mathcal{S}'$



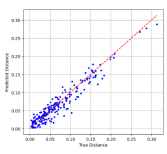
3-RF/GB(1)-$\mathcal{S}'$



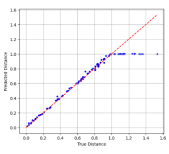
1-RF/GB(2)-$\mathscr{D}'_{\text{te}}$



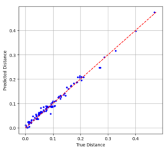
2-RF/GB(2)-$\mathscr{D}'_{\text{te}}$



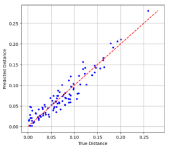
3-RF/GB(2)-$\mathscr{D}'_{\text{te}}$



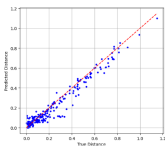
1-RF/GB(2)-$\mathcal{S}'$



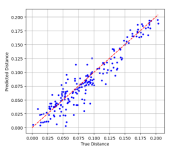
2-RF/GB(2)-$\mathcal{S}'$



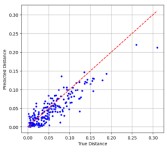
3-RF/GB(2)-$\mathcal{S}'$



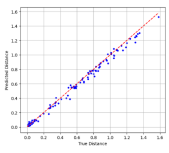
1-GB/GB(1)-$\mathscr{D}'_{\text{te}}$



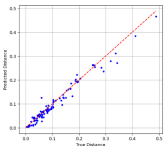
2-GB/GB(1)-$\mathscr{D}'_{\text{te}}$



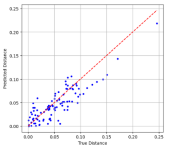
3-GB/GB(1)-$\mathscr{D}'_{\text{te}}$



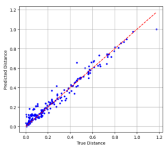
1-GB/GB(1)-$\mathcal{S}'$



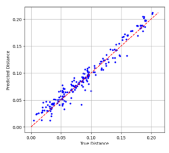
2-GB/GB(1)-$\mathcal{S}'$



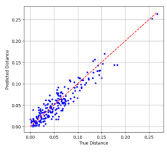
3-GB/GB(1)-$\mathcal{S}'$



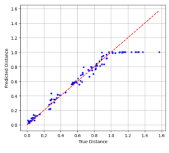
1-GB/GB(2)-$\mathscr{D}'_{\text{te}}$



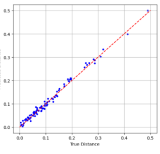
2-GB/GB(2)-$\mathscr{D}'_{\text{te}}$



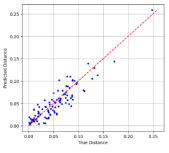
3-GB/GB(2)-$\mathscr{D}'_{\text{te}}$



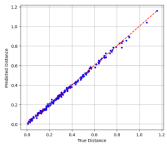
1-GB/GB(2)-$\mathcal{S}'$



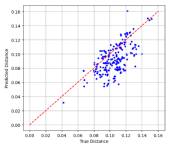
2-GB/GB(2)-$\mathcal{S}'$



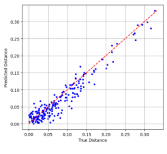
3-GB/GB(2)-$\mathcal{S}'$



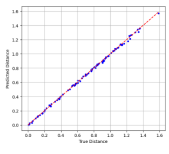
1-NN/GB(1)-$\mathscr{D}'_{\text{te}}$



2-NN/GB(1)-$\mathscr{D}'_{\text{te}}$



3-NN/GB(1)-$\mathscr{D}'_{\text{te}}$



1-NN/GB(1)-$\mathcal{S}'$



2-NN/GB(1)-$\mathcal{S}'$



3-NN/GB(1)-$\mathcal{S}'$



1-NN/GB(2)-$\mathscr{D}'_{\text{te}}$



2-NN/GB(2)-$\mathscr{D}'_{\text{te}}$



3-NN/GB(2)-$\mathscr{D}'_{\text{te}}$



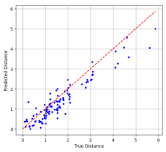
1-NN/GB(2)-$\mathcal{S}'$



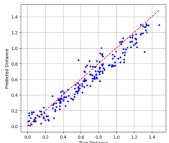
2-NN/GB(2)-$\mathcal{S}'$



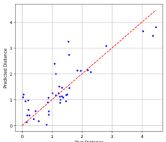
3-NN/GB(2)-$\mathcal{S}'$



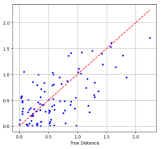
4-RF/SNN(1)-$\mathscr{D}'_{\text{te}}$



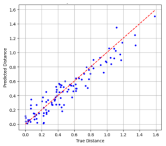
5-RF/SNN(1)-$\mathscr{D}'_{\text{te}}$



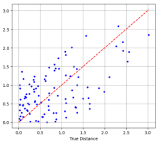
6-RF/SNN(1)-$\mathscr{D}'_{\text{te}}$



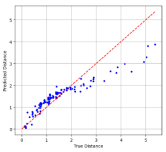
4-RF/SNN(1)-$\mathcal{S}'$



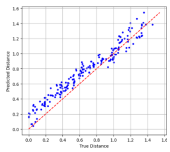
5-RF/SNN(1)-$\mathcal{S}'$



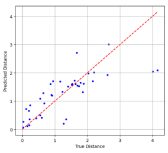
6-RF/SNN(1)-$\mathcal{S}'$



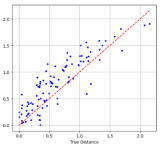
4-RF/SNN(2)-$\mathscr{D}'_{\text{te}}$



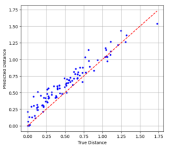
5-RF/SNN(2)-$\mathscr{D}'_{\text{te}}$



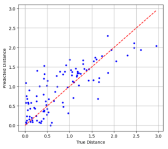
6-RF/SNN(2)-$\mathscr{D}'_{\text{te}}$



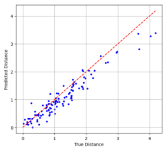
4-RF/SNN(2)-$\mathcal{S}'$



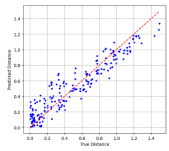
5-RF/SNN(2)-$\mathcal{S}'$



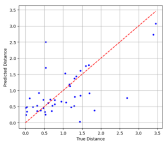
6-RF/SNN(2)-$\mathcal{S}'$



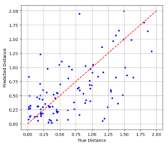
4-GB/SNN(1)-$\mathscr{D}'_{\text{te}}$



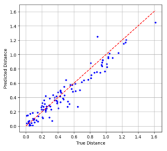
5-GB/SNN(1)-$\mathscr{D}'_{\text{te}}$



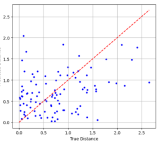
6-GB/SNN(1)-$\mathscr{D}'_{\text{te}}$



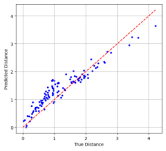
4-GB/SNN(1)-$\mathcal{S}'$



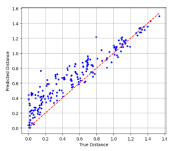
5-GB/SNN(1)-$\mathcal{S}'$



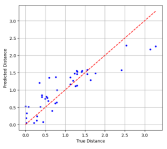
6-GB/SNN(1)-$\mathcal{S}'$



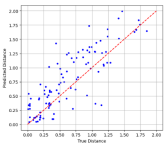
4-GB/SNN(2)-$\mathscr{D}'_{\text{te}}$



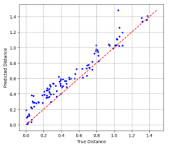
5-GB/SNN(2)-$\mathscr{D}'_{\text{te}}$



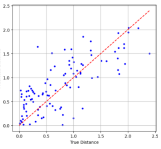
6-GB/SNN(2)-$\mathscr{D}'_{\text{te}}$



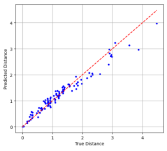
4-GB/SNN(2)-$\mathcal{S}'$



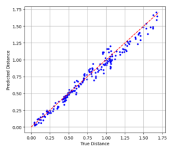
5-GB/SNN(2)-$\mathcal{S}'$



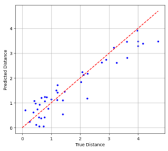
6-GB/SNN(2)-$\mathcal{S}'$



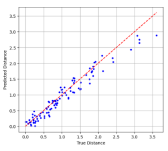
4-NN/SNN(1)-$\mathscr{D}'_{\text{te}}$



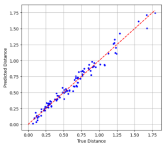
5-NN/SNN(1)-$\mathscr{D}'_{\text{te}}$



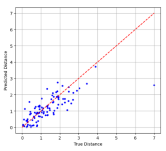
6-NN/SNN(1)-$\mathscr{D}'_{\text{te}}$



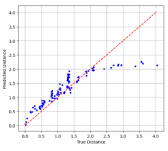
4-NN/SNN(1)-$\mathcal{S}'$



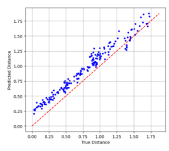
5-NN/SNN(1)-$\mathcal{S}'$



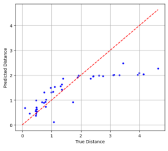
6-NN/SNN(1)-$\mathcal{S}'$



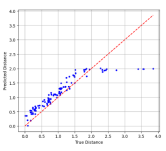
4-NN/SNN(2)-$\mathscr{D}'_{\text{te}}$



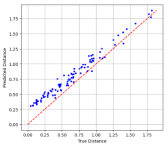
5-NN/SNN(2)-$\mathscr{D}'_{\text{te}}$



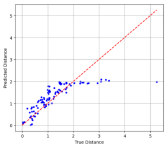
6-NN/SNN(2)-$\mathscr{D}'_{\text{te}}$



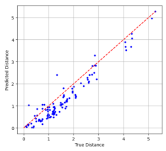
4-NN/SNN(2)-$\mathcal{S}'$



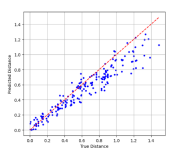
5-NN/SNN(2)-$\mathcal{S}'$



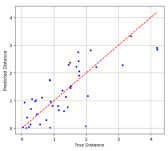
6-NN/SNN(2)-$\mathcal{S}'$



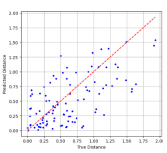
4-RF/MNN(1)-$\mathscr{D}'_{\text{te}}$



5-RF/MNN(1)-$\mathscr{D}'_{\text{te}}$



6-RF/MNN(1)-$\mathscr{D}'_{\text{te}}$



4-RF/MNN(1)-$\mathcal{S}'$



5-RF/MNN(1)-$\mathcal{S}'$



6-RF/MNN(1)-$\mathcal{S}'$



4-RF/MNN(2)-$\mathscr{D}'_{\text{te}}$



5-RF/MNN(2)-$\mathscr{D}'_{\text{te}}$



6-RF/MNN(2)-$\mathscr{D}'_{\text{te}}$



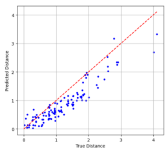
4-RF/MNN(2)-$\mathcal{S}'$



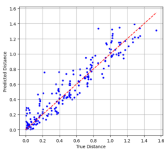
5-RF/MNN(2)-$\mathcal{S}'$



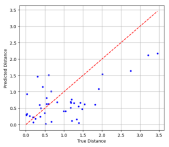
6-RF/MNN(2)-$\mathcal{S}'$



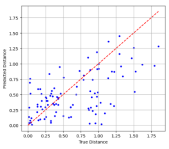
4-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



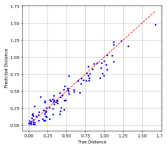
5-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



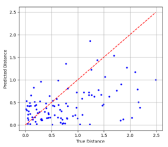
6-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



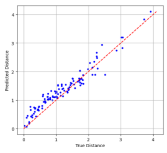
4-GB/MNN(1)-$\mathcal{S}'$



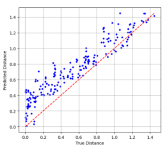
5-GB/MNN(1)-$\mathcal{S}'$



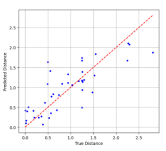
6-GB/MNN(1)-$\mathcal{S}'$



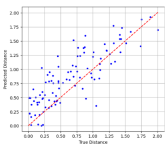
4-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



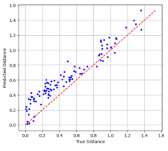
5-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



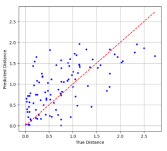
6-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



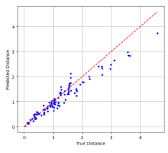
4-GB/MNN(2)-$\mathcal{S}'$



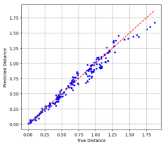
5-GB/MNN(2)-$\mathcal{S}'$



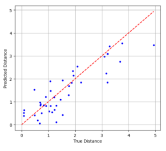
6-GB/MNN(2)-$\mathcal{S}'$



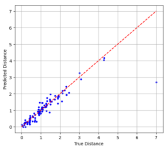
4-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



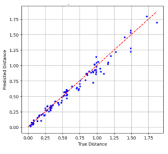
5-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



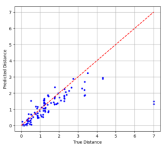
6-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



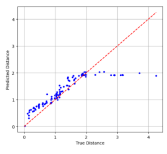
4-NN/MNN(1)-$\mathcal{S}'$



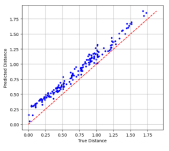
5-NN/MNN(1)-$\mathcal{S}'$



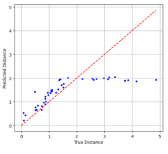
6-NN/MNN(1)-$\mathcal{S}'$



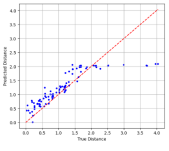
4-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



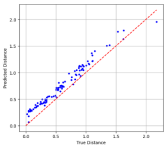
5-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



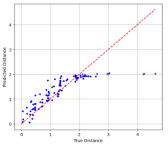
6-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



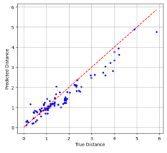
4-NN/MNN(2)-$\mathcal{S}'$



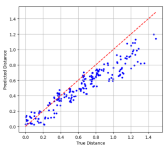
5-NN/MNN(2)-$\mathcal{S}'$



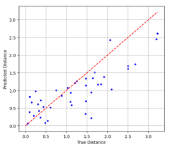
6-NN/MNN(2)-$\mathcal{S}'$



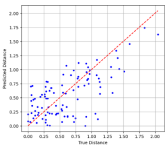
4-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



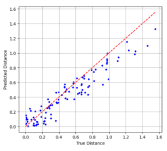
5-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



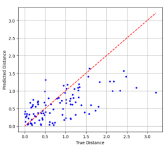
6-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



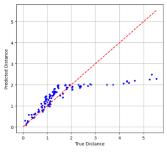
4-RF/LNN(1)-$\mathcal{S}'$



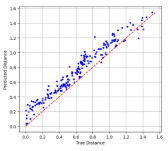
5-RF/LNN(1)-$\mathcal{S}'$



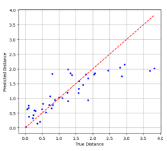
6-RF/LNN(1)-$\mathcal{S}'$



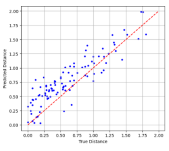
4-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



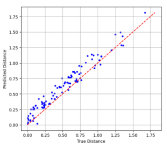
5-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



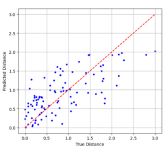
6-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



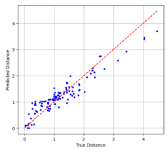
4-RF/LNN(2)-$\mathcal{S}'$



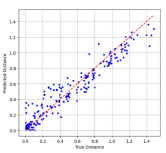
5-RF/LNN(2)-$\mathcal{S}'$



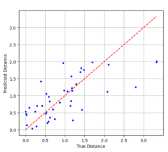
6-RF/LNN(2)-$\mathcal{S}'$



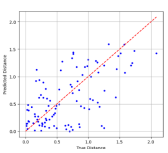
4-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



5-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



6-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



4-GB/LNN(1)-$\mathcal{S}'$



5-GB/LNN(1)-$\mathcal{S}'$



6-GB/LNN(1)-$\mathcal{S}'$



4-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



5-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



6-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



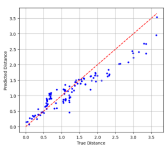
4-GB/LNN(2)-$\mathcal{S}'$



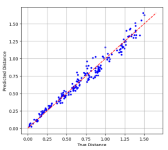
5-GB/LNN(2)-$\mathcal{S}'$



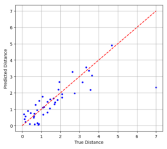
6-GB/LNN(2)-$\mathcal{S}'$



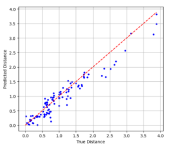
4-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



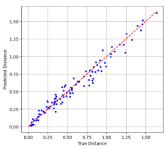
5-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



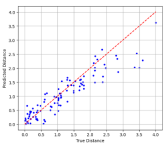
6-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



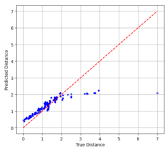
4-NN/LNN(1)-$\mathcal{S}'$



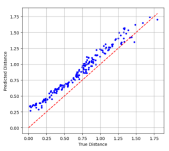
5-NN/LNN(1)-$\mathcal{S}'$



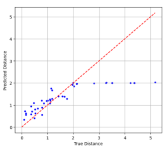
6-NN/LNN(1)-$\mathcal{S}'$



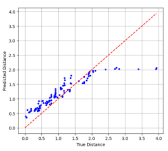
4-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



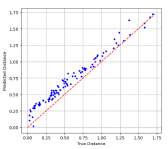
5-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



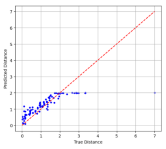
6-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



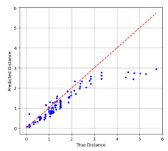
4-NN/LNN(2)-$\mathcal{S}'$



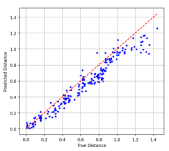
5-NN/LNN(2)-$\mathcal{S}'$



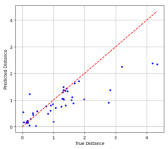
6-NN/LNN(2)-$\mathcal{S}'$



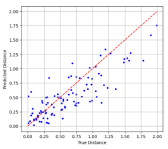
4-RF/GB(1)-$\mathscr{D}'_{\text{te}}$



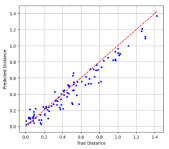
5-RF/GB(1)-$\mathscr{D}'_{\text{te}}$



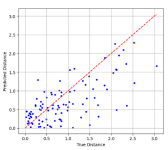
6-RF/GB(1)-$\mathscr{D}'_{\text{te}}$



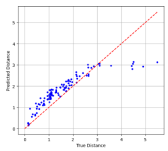
4-RF/GB(1)-$\mathcal{S}'$



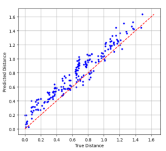
5-RF/GB(1)-$\mathcal{S}'$



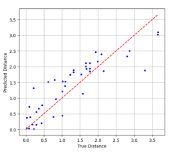
6-RF/GB(1)-$\mathcal{S}'$



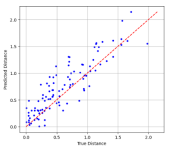
4-RF/GB(2)-$\mathscr{D}'_{\text{te}}$



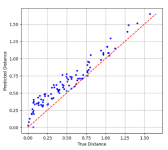
5-RF/GB(2)-$\mathscr{D}'_{\text{te}}$



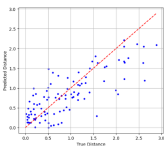
6-RF/GB(2)-$\mathscr{D}'_{\text{te}}$



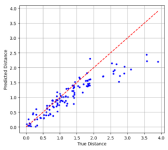
4-RF/GB(2)-$\mathcal{S}'$



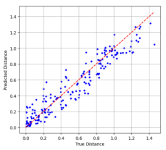
5-RF/GB(2)-$\mathcal{S}'$



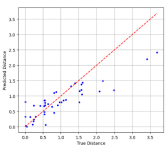
6-RF/GB(2)-$\mathcal{S}'$



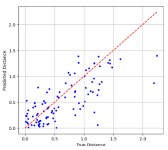
4-GB/GB(1)-$\mathscr{D}'_{\text{te}}$



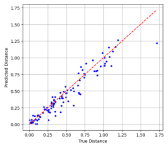
5-GB/GB(1)-$\mathscr{D}'_{\text{te}}$



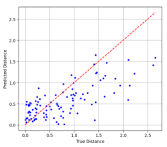
6-GB/GB(1)-$\mathscr{D}'_{\text{te}}$



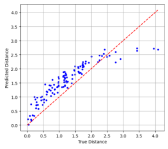
4-GB/GB(1)-$\mathcal{S}'$



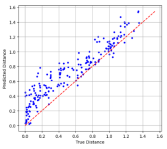
5-GB/GB(1)-$\mathcal{S}'$



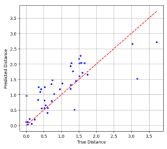
6-GB/GB(1)-$\mathcal{S}'$



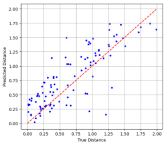
4-GB/GB(2)-$\mathscr{D}'_{\text{te}}$



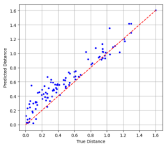
5-GB/GB(2)-$\mathscr{D}'_{\text{te}}$



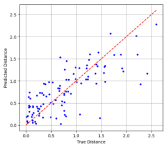
6-GB/GB(2)-$\mathscr{D}'_{\text{te}}$



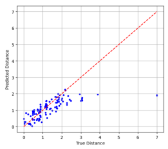
4-GB/GB(2)-$\mathcal{S}'$



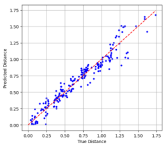
5-GB/GB(2)-$\mathcal{S}'$



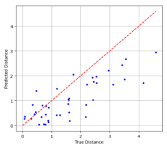
6-GB/GB(2)-$\mathcal{S}'$



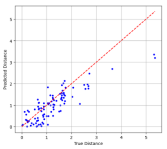
4-NN/GB(1)-$\mathscr{D}'_{\text{te}}$



5-NN/GB(1)-$\mathscr{D}'_{\text{te}}$



6-NN/GB(1)-$\mathscr{D}'_{\text{te}}$



4-NN/GB(1)-$\mathcal{S}'$



5-NN/GB(1)-$\mathcal{S}'$



6-NN/GB(1)-$\mathcal{S}'$



4-NN/GB(2)-$\mathscr{D}'_{\text{te}}$



5-NN/GB(2)-$\mathscr{D}'_{\text{te}}$



6-NN/GB(2)-$\mathscr{D}'_{\text{te}}$



4-NN/GB(2)-$\mathcal{S}'$



5-NN/GB(2)-$\mathcal{S}'$



6-NN/GB(2)-$\mathcal{S}'$

# Appendix D

# Two-stage distance-copying extension

In this final annex, we briefly discuss an extension to the distance-based copying framework introduced in the above paper, with the aim of improving the achieved results.

Specifically, even though this work has highlighted several contexts and situations where it could be valuable to use distance-based copies instead of the traditional ones, the results have also shown scenarios where these copies are slightly inferior to the classical ones, something that mainly occurs in the high dimensional UCI datasets. As already mentioned, one of the reasons that explains why this may happen is that these copies are trained in (essentially) smaller synthetic datasets than the hard ones, due to the cost of computing the required distances. So, to mitigate this problem, in this extension we propose a two-stage approach to approximate these distances and to build copies with them.

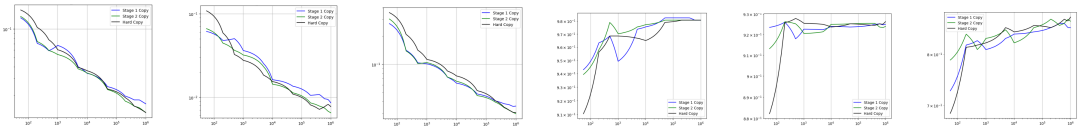
So far, first we sampled the synthetic dataset $\mathscr{X}$ and then computed the corresponding signed distances to the boundary (applying Alg. 1 or Alg. 2) that were used to train the copy. The key observation here is that this copy can be seen as a tool to approximate these distances for previously unlabelled points, by taking the absolute value of its predictions. Consequently, in this situation, we can sample a new synthetic dataset (with the desired size and distribution) and then label it using this copy. Finally, to build the second-stage copy, we proceed to train it on this new and more suitable dataset, in such a way that it benefits from the properties and behaviours of distance copying while not suffering from their associated data scarcity. As expected, this two-stage process is more costly than the previous framework, but it may also increase the performance of distance-based copies in these adverse high dimensional datasets, bridging the gap between them and the traditional copies.

From here, to validate this new approach, we repeat the previous **Experiment 1** from Ch. 4 in the paper, aiming to compare the accuracies and fidelities of the different stages to the ones of the hard copy, something that will show the gains brought by this composition of models. In this case, to make the experiment more informative, here we only use Alg. 2 to label the first synthetic dataset, because it is the one that achieved better results in the past. In addition, from now on will only work with the UCI datasets, that are the ones where this new extension to distance copying could be valuable.

Looking at the obtained results, that are summarized in Table D.1 and Fig. D.1 (extended versions available at the end of the appendix), we can start by observing that this combined approach improves the fidelity of distance-based copies, since the stage-two copy generally performs better than the first one. This performance boost is especially noticeable in some instances of Dataset 6, the one with the highest dimensionality, something that suggests that this framework is valuable in high

TABLE D.1: Performance of the stages across datasets 4-6. Computations limited to 1,000,000 points and 600 seconds.

| Copy | $f_{\mathcal{O}}/f_{\mathcal{C}}$ | Dataset 4 | | Dataset 5 | | Dataset 6 | |
|---|---|---|---|---|---|---|---|
| | | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ |
| St. 1 cp | GB/MNN | 0.98 | 0.060 | 0.93 | 0.030 | 0.82 | 0.121 |
| St. 2 cp | GB/MNN | 0.98 | 0.054 | 0.93 | 0.029 | 0.82 | 0.102 |
| Hard cp | GB/MNN | 0.98 | 0.052 | 0.93 | 0.028 | 0.84 | 0.090 |
| St. 1 cp | NN/SNN | 0.98 | 0.020 | 0.93 | 0.012 | 0.86 | 0.045 |
| St. 2 cp | NN/SNN | 0.98 | 0.019 | 0.93 | 0.011 | 0.86 | 0.044 |
| Hard cp | NN/SNN | 0.98 | 0.018 | 0.92 | 0.009 | 0.84 | 0.043 |
| St. 1 cp | RF/LNN | 0.97 | 0.056 | 0.93 | 0.023 | 0.80 | 0.122 |
| St. 2 cp | RF/LNN | 0.97 | 0.051 | 0.93 | 0.021 | 0.82 | 0.103 |
| Hard cp | RF/LNN | 0.97 | 0.049 | 0.93 | 0.021 | 0.80 | 0.095 |

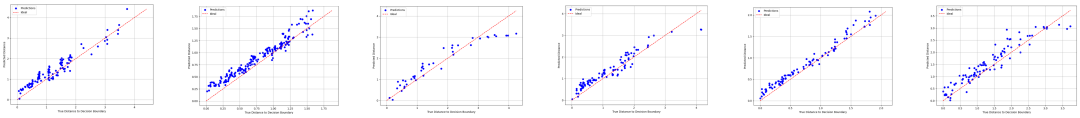


Dataset 4 - $R^{\mathcal{S}}_{\text{em}}$ | Dataset 5 - $R^{\mathcal{S}}_{\text{em}}$ | Dataset 6 - $R^{\mathcal{S}}_{\text{em}}$ | Dataset 4 - $\mathcal{A}_{\mathcal{C}}$ | Dataset 5 - $\mathcal{A}_{\mathcal{C}}$ | Dataset 6 - $\mathcal{A}_{\mathcal{C}}$

FIGURE D.1: Evolution of metrics as functions of the number of points used to train the copy (LNN) from the black box (NN).

dimensional settings, where the size of the synthetic dataset $\mathscr{Z}$ becomes more important due to data scarcity. Moreover, even thought the above improvements may still seem limited, usually they are significant enough to make the new copy be much closer or even on par with the corresponding hard copy in these adverse scenarios, while preserving all the advantages of working with distances.

Finally, since one of these key advantages that may lead us to consider the use of distance-based copies is their uncertainty measures, we can also analyse the reliability of the distances predicted by these two-stage copies, with the aim of studying if this extended framework also has an impact on the quality of these confidence measures. To this end, we have repeated the **Experiment 3** developed in Ch. 4, keeping the same experimental setting and metrics as in the paper, but comparing now the stage-one and stage-two distance-based copies built in the previous experiment of this annex.



Dataset 4- $\mathscr{D}'_{\text{te}}$ | Dataset 5 - $\mathscr{D}'_{\text{te}}$ | Dataset 6 - $\mathscr{D}'_{\text{te}}$ | Dataset 4- $\mathcal{S}'$ | Dataset 5 - $\mathcal{S}'$ | Dataset 6 - $\mathcal{S}'$

FIGURE D.2: Comparison of the distance predictions to the actual distances, for the NN black box and St. 2 MNN copy.

Looking at the obtained results appearing in Table D.2 and Fig. D.2 (extended versions available at the end of the appendix), we can conclude that, in general, the stage-two copies achieve better results than their stage-one counterparts, exhibiting

TABLE D.2: Quality of the predicted distances. Metrics corresponding to the NN black box and MNN copy.

| Copy | Dataset N. | $\mathscr{D}'_{\text{te}}$ real data | | $\mathcal{S}'$ uniform data | |
|---|---|---|---|---|---|
| | | MAE | RMSE | MAE | RMSE |
| Stage 1 copy | 4 | 0.222 | 0.257 | 0.269 | 0.368 |
| Stage 2 copy | 4 | 0.226 | 0.262 | 0.257 | 0.356 |
| Stage 1 copy | 5 | 0.168 | 0.185 | 0.144 | 0.160 |
| Stage 2 copy | 5 | 0.164 | 0.181 | 0.142 | 0.159 |
| Stage 1 copy | 6 | 0.360 | 0.572 | 0.288 | 0.363 |
| Stage 2 copy | 6 | 0.355 | 0.533 | 0.263 | 0.333 |

lower error metrics and less dispersed scatter plots. As a consequence, this shows that this extended framework can be used to improve the quality of these measures of uncertainty, at the cost of a higher computational time.

*Table D.1*

| Copy | $f_{\mathcal{O}}/f_{\mathcal{C}}$ | Dataset 4 | | | Dataset 5 | | | Dataset 6 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{A}_{\mathcal{O}}$ | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ | $\mathcal{A}_{\mathcal{O}}$ | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ | $\mathcal{A}_{\mathcal{O}}$ | $\mathcal{A}_{\mathcal{C}}$ | $R^{\mathcal{S}}_{\text{em}}$ |
| St. 1 cp | RF/SNN | 0.97 | 0.970±0.004 | 0.0754±0.0023 | 0.93 | 0.927±0.006 | 0.0316±0.0034 | 0.81 | 0.814±0.061 | 0.1202±0.0065 |
| St. 2 cp | RF/SNN | 0.97 | 0.961±0.004 | 0.0737±0.0017 | 0.93 | 0.927±0.006 | 0.0315±0.0039 | 0.81 | 0.795±0.060 | 0.1182±0.0063 |
| Hard cp | RF/SNN | 0.97 | 0.961±0.007 | 0.0624±0.0045 | 0.93 | 0.928±0.004 | 0.0269±0.0035 | 0.81 | 0.814±0.061 | 0.1113±0.0088 |
| St. 1 cp | GB/SNN | 0.98 | 0.979±0.012 | 0.0781±0.0047 | 0.92 | 0.927±0.005 | 0.0358±0.0032 | 0.87 | 0.824±0.024 | 0.1208±0.0092 |
| St. 2 cp | GB/SNN | 0.98 | 0.981±0.012 | 0.0764±0.0043 | 0.92 | 0.927±0.003 | 0.0346±0.0026 | 0.87 | 0.829±0.028 | 0.1181±0.0080 |
| Hard cp | GB/SNN | 0.98 | 0.977±0.013 | 0.0653±0.0042 | 0.92 | 0.926±0.007 | 0.0338±0.0019 | 0.87 | 0.829±0.035 | 0.1109±0.0094 |
| St. 1 cp | NN/SNN | 0.98 | 0.979±0.007 | 0.0196±0.0010 | 0.93 | 0.926±0.006 | 0.0117±0.0019 | 0.89 | 0.862±0.035 | 0.0446±0.0016 |
| St. 2 cp | NN/SNN | 0.98 | 0.979±0.007 | 0.0191±0.0010 | 0.93 | 0.926±0.006 | 0.0105±0.0013 | 0.89 | 0.857±0.021 | 0.0442±0.0021 |
| Hard cp | NN/SNN | 0.98 | 0.983±0.008 | 0.0177±0.0017 | 0.93 | 0.924±0.005 | 0.0094±0.0025 | 0.89 | 0.843±0.032 | 0.0430±0.0019 |
| St. 1 cp | RF/MNN | 0.97 | 0.967±0.009 | 0.0559±0.0004 | 0.93 | 0.926±0.005 | 0.0235±0.0022 | 0.82 | 0.810±0.062 | 0.1203±0.0071 |
| St. 2 cp | RF/MNN | 0.97 | 0.967±0.009 | 0.0505±0.0013 | 0.93 | 0.927±0.006 | 0.0223±0.0025 | 0.82 | 0.814±0.063 | 0.1045±0.0056 |
| Hard cp | RF/MNN | 0.97 | 0.967±0.007 | 0.0490±0.0018 | 0.93 | 0.928±0.007 | 0.0214±0.0024 | 0.82 | 0.814±0.063 | 0.0985±0.0050 |
| St. 1 cp | GB/MNN | 0.98 | 0.981±0.009 | 0.0598±0.0054 | 0.92 | 0.926±0.051 | 0.0303±0.0023 | 0.87 | 0.824±0.041 | 0.1212±0.0162 |
| St. 2 cp | GB/MNN | 0.98 | 0.983±0.011 | 0.0538±0.0037 | 0.92 | 0.926±0.052 | 0.0287±0.0022 | 0.87 | 0.824±0.032 | 0.1021±0.0116 |
| Hard cp | GB/MNN | 0.98 | 0.979±0.009 | 0.0521±0.0032 | 0.92 | 0.927±0.054 | 0.0281±0.0013 | 0.87 | 0.838±0.028 | 0.0901±0.0104 |
| St. 1 cp | NN/MNN | 0.98 | 0.981±0.007 | 0.0168±0.0012 | 0.92 | 0.928±0.005 | 0.0082±0.0009 | 0.89 | 0.876±0.041 | 0.0372±0.0010 |
| St. 2 cp | NN/MNN | 0.98 | 0.979±0.007 | 0.0148±0.0010 | 0.92 | 0.927±0.005 | 0.0067±0.0003 | 0.89 | 0.881±0.034 | 0.0325±0.0011 |
| Hard cp | NN/MNN | 0.98 | 0.979±0.007 | 0.0141±0.0009 | 0.92 | 0.926±0.005 | 0.0063±0.0005 | 0.89 | 0.881±0.045 | 0.0316±0.0012 |
| St. 1 cp | RF/LNN | 0.97 | 0.972±0.010 | 0.0559±0.0008 | 0.93 | 0.927±0.005 | 0.0227±0.0025 | 0.81 | 0.800±0.061 | 0.1215±0.0024 |
| St. 2 cp | RF/LNN | 0.97 | 0.970±0.009 | 0.0506±0.0011 | 0.93 | 0.928±0.005 | 0.0210±0.0023 | 0.81 | 0.819±0.054 | 0.1025±0.0011 |
| Hard cp | RF/LNN | 0.97 | 0.967±0.004 | 0.0490±0.0018 | 0.93 | 0.929±0.005 | 0.0209±0.0020 | 0.81 | 0.805±0.059 | 0.0951±0.0027 |
| St. 1 cp | GB/LNN | 0.98 | 0.979±0.012 | 0.0607±0.0063 | 0.92 | 0.929±0.008 | 0.0299±0.0023 | 0.87 | 0.819±0.068 | 0.1302±0.0145 |
| St. 2 cp | GB/LNN | 0.98 | 0.979±0.012 | 0.0540±0.0042 | 0.92 | 0.927±0.006 | 0.0278±0.0019 | 0.87 | 0.829±0.028 | 0.1049±0.0123 |
| Hard cp | GB/LNN | 0.98 | 0.977±0.009 | 0.0512±0.0039 | 0.92 | 0.927±0.004 | 0.0275±0.0019 | 0.87 | 0.833±0.040 | 0.0866±0.0073 |
| St. 1 cp | NN/LNN | 0.98 | 0.981±0.007 | 0.0162±0.0020 | 0.92 | 0.925±0.007 | 0.0086±0.0011 | 0.88 | 0.857±0.052 | 0.0357±0.0031 |
| St. 2 cp | NN/LNN | 0.98 | 0.981±0.007 | 0.0131±0.0014 | 0.92 | 0.924±0.006 | 0.0065±0.0005 | 0.88 | 0.871±0.049 | 0.0296±0.0027 |
| Hard cp | NN/LNN | 0.98 | 0.981±0.007 | 0.0131±0.0019 | 0.92 | 0.926±0.006 | 0.0077±0.0009 | 0.88 | 0.881±0.050 | 0.0304±0.0031 |

*Table D.2*

| Copy | $f_{\mathcal{O}}/f_{\mathcal{C}}$ | Dataset 4 | | | | Dataset 5 | | | | Dataset 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathscr{D}'_{\text{te}}$ real data | | $\mathcal{S}'$ uniform data | | $\mathscr{D}'_{\text{te}}$ real data | | $\mathcal{S}'$ uniform data | | $\mathscr{D}'_{\text{te}}$ real data | | $\mathcal{S}'$ uniform data | |
| | | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| St. 1 copy | RF/SNN | 0.22±0.02 | 0.29±0.03 | 0.30±0.03 | 0.37±0.02 | 0.14±0.02 | 0.16±0.02 | 0.17±0.02 | 0.20±0.02 | 0.41±0.03 | 0.55±0.04 | 0.36±0.02 | 0.46±0.02 |
| St. 2 copy | RF/SNN | 0.26±0.02 | 0.33±0.03 | 0.28±0.02 | 0.35±0.02 | 0.14±0.03 | 0.16±0.03 | 0.17±0.02 | 0.20±0.02 | 0.41±0.02 | 0.53±0.04 | 0.33±0.02 | 0.43±0.02 |
| St. 1 copy | GB/SNN | 0.26±0.04 | 0.31±0.04 | 0.28±0.03 | 0.35±0.03 | 0.20±0.01 | 0.25±0.01 | 0.16±0.01 | 0.19±0.01 | 0.33±0.02 | 0.43±0.03 | 0.35±0.01 | 0.43±0.02 |
| St. 2 copy | GB/SNN | 0.29±0.06 | 0.33±0.07 | 0.26±0.03 | 0.33±0.03 | 0.19±0.01 | 0.24±0.01 | 0.15±0.01 | 0.18±0.01 | 0.32±0.03 | 0.42±0.04 | 0.33±0.01 | 0.42±0.02 |
| St. 1 copy | NN/SNN | 0.26±0.03 | 0.33±0.07 | 0.29±0.05 | 0.39±0.10 | 0.15±0.02 | 0.17±0.02 | 0.15±0.02 | 0.16±0.02 | 0.36±0.04 | 0.52±0.07 | 0.34±0.03 | 0.48±0.11 |
| St. 2 copy | NN/SNN | 0.26±0.05 | 0.32±0.09 | 0.29±0.06 | 0.38±0.11 | 0.15±0.02 | 0.17±0.02 | 0.15±0.02 | 0.17±0.02 | 0.37±0.05 | 0.52±0.08 | 0.32±0.03 | 0.47±0.12 |
| St. 1 copy | RF/MNN | 0.22±0.03 | 0.29±0.05 | 0.31±0.04 | 0.37±0.04 | 0.15±0.03 | 0.18±0.03 | 0.16±0.03 | 0.18±0.03 | 0.35±0.02 | 0.49±0.04 | 0.39±0.02 | 0.49±0.02 |
| St. 2 copy | RF/MNN | 0.23±0.03 | 0.29±0.06 | 0.29±0.03 | 0.35±0.03 | 0.14±0.03 | 0.17±0.02 | 0.15±0.03 | 0.18±0.03 | 0.32±0.03 | 0.44±0.05 | 0.32±0.02 | 0.42±0.03 |
| St. 1 copy | GB/MNN | 0.23±0.06 | 0.28±0.06 | 0.32±0.04 | 0.38±0.04 | 0.22±0.00 | 0.26±0.00 | 0.15±0.01 | 0.18±0.01 | 0.33±0.04 | 0.44±0.06 | 0.37±0.02 | 0.46±0.03 |
| St. 2 copy | GB/MNN | 0.23±0.05 | 0.29±0.06 | 0.30±0.04 | 0.36±0.03 | 0.21±0.01 | 0.25±0.01 | 0.15±0.01 | 0.17±0.01 | 0.29±0.02 | 0.39±0.04 | 0.32±0.02 | 0.41±0.03 |
| St. 1 copy | NN/MNN | 0.22±0.02 | 0.26±0.02 | 0.27±0.02 | 0.37±0.06 | 0.17±0.02 | 0.18±0.02 | 0.14±0.03 | 0.16±0.03 | 0.36±0.03 | 0.57±0.11 | 0.29±0.02 | 0.36±0.02 |
| St. 2 copy | NN/MNN | 0.23±0.02 | 0.26±0.02 | 0.25±0.02 | 0.36±0.06 | 0.16±0.02 | 0.18±0.02 | 0.14±0.03 | 0.16±0.03 | 0.35±0.03 | 0.55±0.10 | 0.26±0.01 | 0.33±0.01 |
| St. 1 copy | RF/LNN | 0.27±0.05 | 0.34±0.05 | 0.30±0.04 | 0.36±0.04 | 0.15±0.03 | 0.18±0.03 | 0.16±0.02 | 0.18±0.02 | 0.44±0.04 | 0.65±0.07 | 0.40±0.03 | 0.52±0.05 |
| St. 2 copy | RF/LNN | 0.24±0.03 | 0.31±0.03 | 0.29±0.04 | 0.34±0.04 | 0.15±0.03 | 0.17±0.03 | 0.15±0.01 | 0.18±0.02 | 0.41±0.04 | 0.59±0.07 | 0.33±0.03 | 0.44±0.04 |
| St. 1 copy | GB/LNN | 0.24±0.03 | 0.30±0.03 | 0.31±0.02 | 0.37±0.02 | 0.21±0.02 | 0.26±0.02 | 0.15±0.01 | 0.18±0.01 | 0.33±0.03 | 0.43±0.02 | 0.36±0.06 | 0.46±0.01 |
| St. 2 copy | GB/LNN | 0.23±0.03 | 0.29±0.05 | 0.28±0.01 | 0.35±0.01 | 0.20±0.02 | 0.25±0.02 | 0.14±0.01 | 0.17±0.01 | 0.30±0.02 | 0.41±0.02 | 0.32±0.02 | 0.40±0.02 |
| St. 1 copy | NN/LNN | 0.25±0.03 | 0.30±0.04 | 0.28±0.02 | 0.38±0.06 | 0.17±0.03 | 0.19±0.03 | 0.15±0.02 | 0.17±0.02 | 0.34±0.03 | 0.55±0.09 | 0.32±0.02 | 0.43±0.07 |
| St. 2 copy | NN/LNN | 0.26±0.03 | 0.31±0.04 | 0.27±0.01 | 0.37±0.06 | 0.17±0.03 | 0.19±0.03 | 0.15±0.02 | 0.17±0.03 | 0.33±0.03 | 0.54±0.08 | 0.31±0.03 | 0.41±0.07 |

*Figure D.1 (Dataset - Black box / Copy - Metric)*

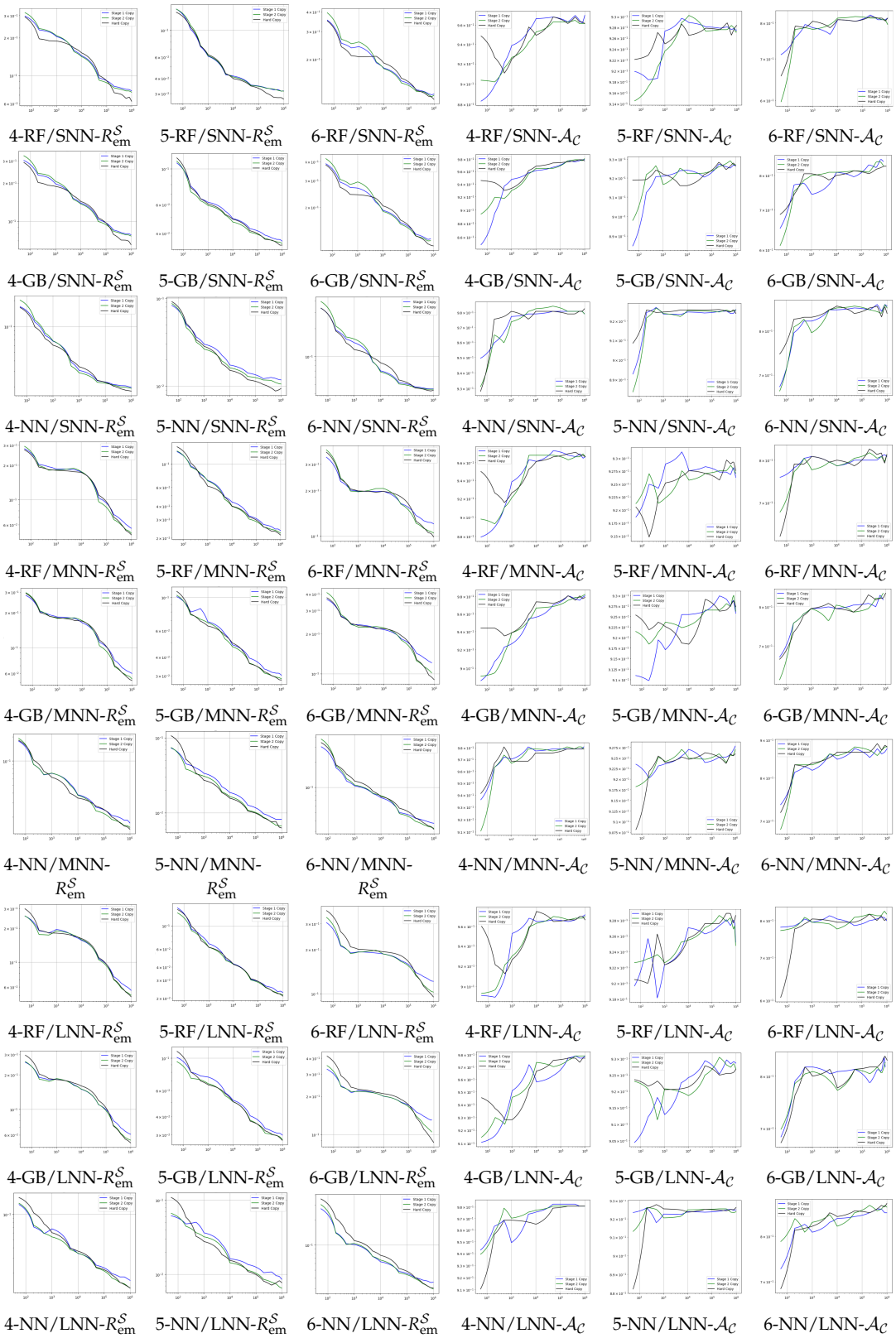
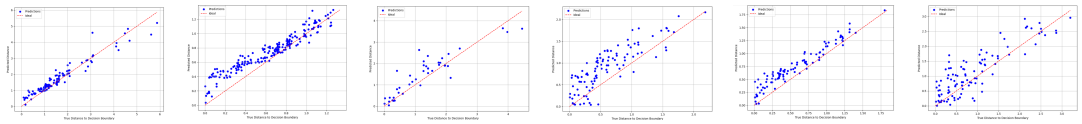
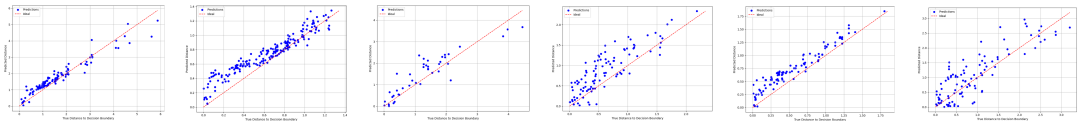


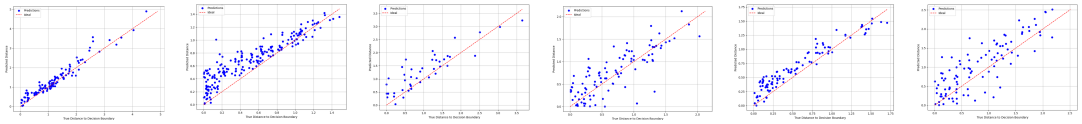
*Figure D.2 (Dataset - Black box/Copy (Stage) - Samples)*



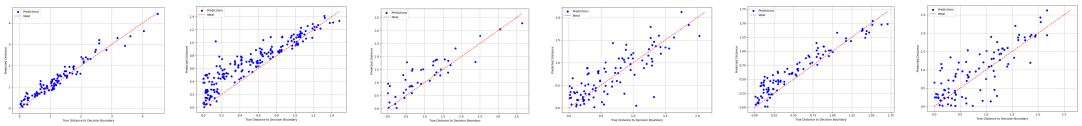
4-RF/SNN(1)-$\mathscr{D}'_{\text{te}}$ 5-RF/SNN(1)-$\mathscr{D}'_{\text{te}}$ 6-RF/SNN(1)-$\mathscr{D}'_{\text{te}}$ 4-RF/SNN(1)-$\mathcal{S}'$ 5-RF/SNN(1)-$\mathcal{S}'$ 6-RF/SNN(1)-$\mathcal{S}'$



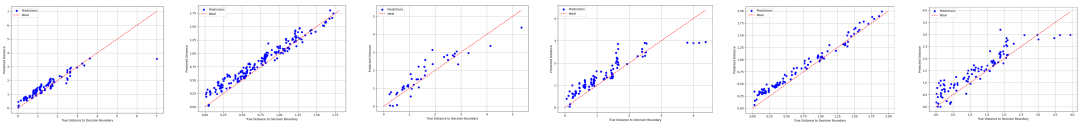
4-RF/SNN(2)-$\mathscr{D}'_{\text{te}}$ 5-RF/SNN(2)-$\mathscr{D}'_{\text{te}}$ 6-RF/SNN(2)-$\mathscr{D}'_{\text{te}}$ 4-RF/SNN(2)-$\mathcal{S}'$ 5-RF/SNN(2)-$\mathcal{S}'$ 6-RF/SNN(2)-$\mathcal{S}'$



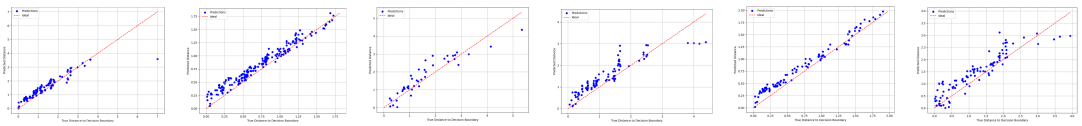
4-GB/SNN(1)-$\mathscr{D}'_{\text{te}}$ 5-GB/SNN(1)-$\mathscr{D}'_{\text{te}}$ 6-GB/SNN(1)-$\mathscr{D}'_{\text{te}}$ 4-GB/SNN(1)-$\mathcal{S}'$ 5-GB/SNN(1)-$\mathcal{S}'$ 6-GB/SNN(1)-$\mathcal{S}'$



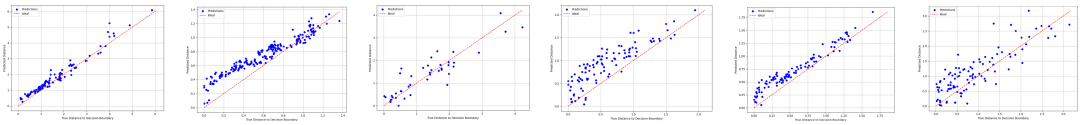
4-GB/SNN(2)-$\mathscr{D}'_{\text{te}}$ 5-GB/SNN(2)-$\mathscr{D}'_{\text{te}}$ 6-GB/SNN(2)-$\mathscr{D}'_{\text{te}}$ 4-GB/SNN(2)-$\mathcal{S}'$ 5-GB/SNN(2)-$\mathcal{S}'$ 6-GB/SNN(2)-$\mathcal{S}'$



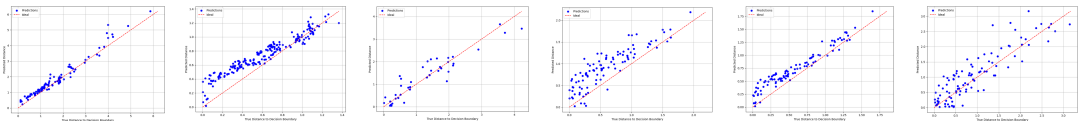
4-NN/SNN(1)-$\mathscr{D}'_{\text{te}}$ 5-NN/SNN(1)-$\mathscr{D}'_{\text{te}}$ 6-NN/SNN(1)-$\mathscr{D}'_{\text{te}}$ 4-NN/SNN(1)-$\mathcal{S}'$ 5-NN/SNN(1)-$\mathcal{S}'$ 6-NN/SNN(1)-$\mathcal{S}'$



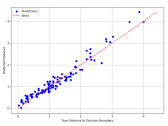
4-NN/SNN(2)-$\mathscr{D}'_{\text{te}}$ 5-NN/SNN(2)-$\mathscr{D}'_{\text{te}}$ 6-NN/SNN(2)-$\mathscr{D}'_{\text{te}}$ 4-NN/SNN(2)-$\mathcal{S}'$ 5-NN/SNN(2)-$\mathcal{S}'$ 6-NN/SNN(2)-$\mathcal{S}'$



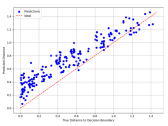
4-RF/MNN(1)-$\mathscr{D}'_{\text{te}}$ 5-RF/MNN(1)-$\mathscr{D}'_{\text{te}}$ 6-RF/MNN(1)-$\mathscr{D}'_{\text{te}}$ 4-RF/MNN(1)-$\mathcal{S}'$ 5-RF/MNN(1)-$\mathcal{S}'$ 6-RF/MNN(1)-$\mathcal{S}'$



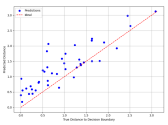
4-RF/MNN(2)-$\mathscr{D}'_{\text{te}}$ 5-RF/MNN(2)-$\mathscr{D}'_{\text{te}}$ 6-RF/MNN(2)-$\mathscr{D}'_{\text{te}}$ 4-RF/MNN(2)-$\mathcal{S}'$ 5-RF/MNN(2)-$\mathcal{S}'$ 6-RF/MNN(2)-$\mathcal{S}'$



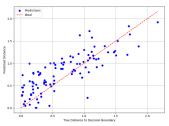
4-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



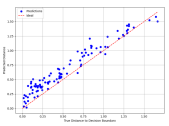
5-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



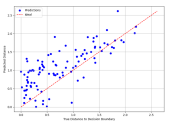
6-GB/MNN(1)-$\mathscr{D}'_{\text{te}}$



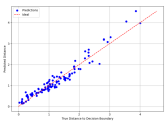
4-GB/MNN(1)-$\mathscr{S}'$



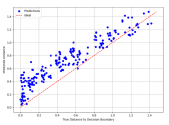
5-GB/MNN(1)-$\mathscr{S}'$



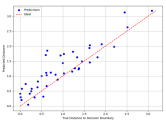
6-GB/MNN(1)-$\mathscr{S}'$



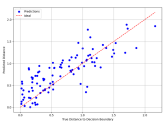
4-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



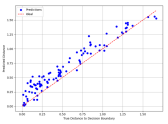
5-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



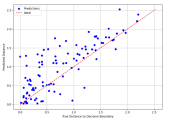
6-GB/MNN(2)-$\mathscr{D}'_{\text{te}}$



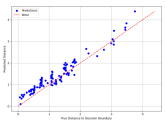
4-GB/MNN(2)-$\mathscr{S}'$



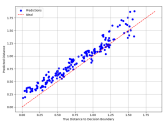
5-GB/MNN(2)-$\mathscr{S}'$



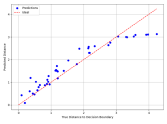
6-GB/MNN(2)-$\mathscr{S}'$



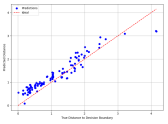
4-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



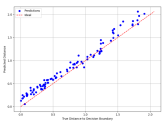
5-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



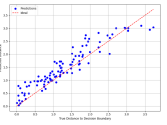
6-NN/MNN(1)-$\mathscr{D}'_{\text{te}}$



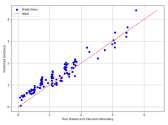
4-NN/MNN(1)-$\mathscr{S}'$



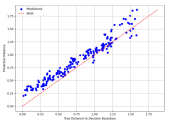
5-NN/MNN(1)-$\mathscr{S}'$



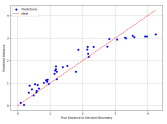
6-NN/MNN(1)-$\mathscr{S}'$



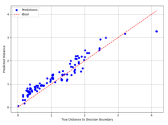
4-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



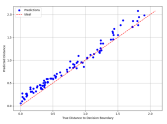
5-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



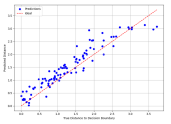
6-NN/MNN(2)-$\mathscr{D}'_{\text{te}}$



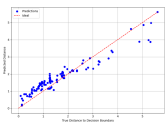
4-NN/MNN(2)-$\mathscr{S}'$



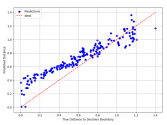
5-NN/MNN(2)-$\mathscr{S}'$



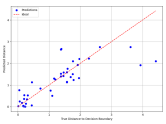
6-NN/MNN(2)-$\mathscr{S}'$



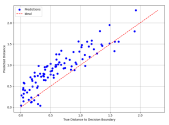
4-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



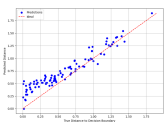
5-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



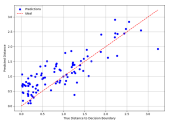
6-RF/LNN(1)-$\mathscr{D}'_{\text{te}}$



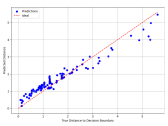
4-RF/LNN(1)-$\mathscr{S}'$



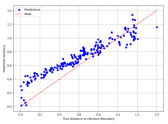
5-RF/LNN(1)-$\mathscr{S}'$



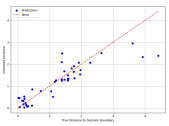
6-RF/LNN(1)-$\mathscr{S}'$



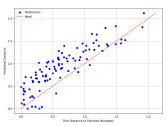
4-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



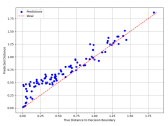
5-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



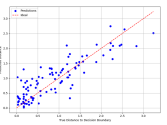
6-RF/LNN(2)-$\mathscr{D}'_{\text{te}}$



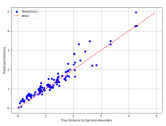
4-RF/LNN(2)-$\mathscr{S}'$



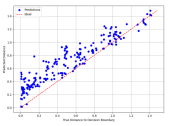
5-RF/LNN(2)-$\mathscr{S}'$



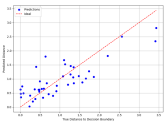
6-RF/LNN(2)-$\mathscr{S}'$



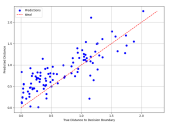
4-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



5-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



6-GB/LNN(1)-$\mathscr{D}'_{\text{te}}$



4-GB/LNN(1)-$\mathscr{S}'$



5-GB/LNN(1)-$\mathscr{S}'$



6-GB/LNN(1)-$\mathscr{S}'$



4-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



5-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



6-GB/LNN(2)-$\mathscr{D}'_{\text{te}}$



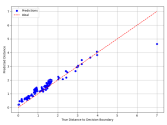
4-GB/LNN(2)-$\mathscr{S}'$



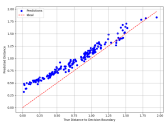
5-GB/LNN(2)-$\mathscr{S}'$



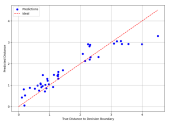
6-GB/LNN(2)-$\mathscr{S}'$



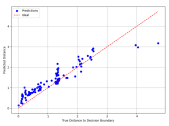
4-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



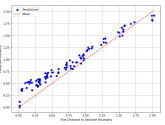
5-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



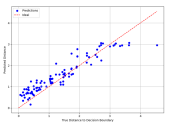
6-NN/LNN(1)-$\mathscr{D}'_{\text{te}}$



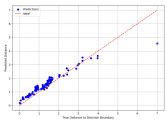
4-NN/LNN(1)-$\mathcal{S}'$



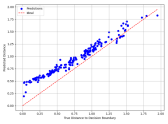
5-NN/LNN(1)-$\mathcal{S}'$



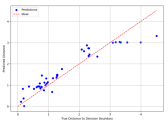
6-NN/LNN(1)-$\mathcal{S}'$



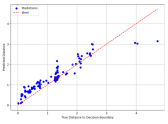
4-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



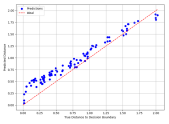
5-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



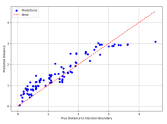
6-NN/LNN(2)-$\mathscr{D}'_{\text{te}}$



4-NN/LNN(2)-$\mathcal{S}'$



5-NN/LNN(2)-$\mathcal{S}'$



6-NN/LNN(2)-$\mathcal{S}'$

# Appendix E

# Source Code Repository

The code used in this project is available at the following GitHub repository:
https://mat.ub.edu/

# Bibliography


Aggarwal, C.C., A. Hinneburg, and D.A. Keim (Oct. 2001). "On the Surprising Behavior of Distance Metrics in High Dimensional Space". In: *International Conference on Database Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 420–434. DOI: 10.1007/3-540-44503-X_27.

Altman, N. and M. Krzywinski (May 2018). "The curse(s) of dimensionality". In: *Nature Methods* 15.6, pp. 399–400. DOI: 10.1038/s41592-018-0019-x.

Azad, A. and A. Banu (Nov. 2024). *Publication Trends in Artificial Intelligence Conferences: The Rise of Super Prolific Authors*. DOI: 10.48550/arXiv.2412.07793. arXiv: 2412.07793 [cs.DL].

Bastani, O., C. Kim, and H. Bastani (Jan. 2019). *Interpreting blackbox models via model extraction*. DOI: 10.48550/arXiv.1705.08504. arXiv: 1705.08504v6 [cs.LG].

Borup, K. and L.N. Andersen (Dec. 2021). "Even your Teacher Needs Guidance: Ground-Truth Targets Dampen Regularization Imposed by Self-Distillation". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., pp. 5316–5327. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/2adcefe38fbcd3dcd45908fbab1bf628-Paper.pdf.

Bucila, C., R. Caruana, and A. Niculescu-Mizil (Aug. 2006). "Model compression". In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, pp. 535–541. DOI: 10.1145/1150402.1150464.

Chollet, F. et al. (2015). *Keras*. https://keras.io.

Ding, G.W. et al. (Apr. 2020). "MMA Training: Direct Input Space Margin Maximization through Adversarial Training". In: *Proceedings of the 8th International Conference on Learning Representations*. Appleton, WI, USA: ICLR, pp. 1709–1736. URL: https://openreview.net/forum?id=HkeryxBtPB.

Ducoffe, M. and F. Precioso (Feb. 2018). *Adversarial Active Learning for Deep Networks: a Margin Based Approach*. DOI: 10.48550/arXiv.1802.09841. arXiv: 1802.09841 [cs.LG].

Elsayed, G.F. et al. (Dec. 2018). "Large Margin Deep Networks for Classification". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2018/file/42998cf32d552343bc8e460416382dca-Paper.pdf.

Gade, K. et al. (July 2019). "Explainable AI in Industry". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, pp. 3203–3204. DOI: 10.1145/3292500.3332281.

Goldsteen, A. et al. (Sept. 2022). "Data minimization for GDPR compliance in machine learning models". In: *AI and Ethics* 2.3, pp. 477–91. DOI: 10.1007/s43681-021-00095-8.

Guo, Z. et al. (Dec. 2024). "Leveraging logit uncertainty for better knowledge distillation". In: *Scientific Reports* 14.1, pp. 31249–31260. DOI: 10.1038/s41598-024-82647-6.

Hashimoto, W., H. Kamigaito, and T. Watanabe (Apr. 2025). "Efficient Nearest Neighbor based Uncertainty Estimation for Natural Language Processing

Tasks". In: *Findings of the Association for Computational Linguistics*. Albuquerque, New Mexico: ACL, pp. 4350–4366. DOI: 10.18653/v1/2025.findings-naacl.246.

Heo, B. et al. (July 2019). "Knowledge Distillation with Adversarial Samples Supporting Decision Boundary". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.1, pp. 3771–3778. DOI: 10.1609/aaai.v33i01.33013771.

Hinton, G., O. Vinyals, and J. Dean (Mar. 2015). *Distilling the knowledge in a neural network*. DOI: /10.48550/arXiv.1503.02531. arXiv: 1503.02531 [stat.ML].

Jiang, Y. et al. (May 2019). "Predicting the Generalization Gap in Deep Networks with Margin Distributions". In: *Proceedings of the 7th International Conference on Learning Representations*. Appleton, WI, USA: ICLR, pp. 2294–2313. URL: https://openreview.net/forum?id=HJlQfnCqKX.

Jin, Y., J. Wang, and D. Lin (June 2023). "Multi-Level Logit Distillation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE, pp. 24276–24285. DOI: 10.1109/CVPR52729.2023.02325.

Kelly, M., R. Longjohn, and K. Nottingham (n.d.). *The UCI Machine Learning Repository*. https://archive.ics.uci.edu.

Kolluri, J. et al. (July 2020). "Reducing Overfitting Problem in Machine Learning Using Novel L1/4 Regularization Method". In: *2020 4th International Conference on Trends in Electronics and Informatics*. Piscataway, NJ, USA: IEEE, pp. 934–938. DOI: 10.1109/ICOEI48184.2020.9142992.

Korobenko, D., A. Nikiforova, and R. Sharma (June 2024). "Towards a Privacy and Security-Aware Framework for Ethical AI: Guiding the Development and Assessment of AI Systems". In: *Proceedings of the 25th Annual International Conference on Digital Government Research*. New York, NY, USA: ACM, pp. 740–753. DOI: 10.1145/3657054.3657141.

Lian, X., Z. Huang, and C. Wang (Aug. 2023). "AKD: Using Adversarial Knowledge Distillation to Achieve Black-box Attacks". In: *Proceedings of the International Joint Conference on Neural Networks*. Piscataway, NJ, USA: IEEE, pp. 1–7. DOI: 10.1109/IJCNN54540.2023.10191087.

Liu, J. et al. (Dec. 2020). "Simple and Principled Uncertainty Estimation with Deterministic Deep Learning via Distance Awareness". In: *Advances in Neural Information Processing Systems*. Vol. 33. Red Hook, NY, USA: Curran Associates, Inc., pp. 7498–7512. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/543e83748234f7cbab21aa0ade66565f-Paper.pdf.

Lowd, D. and C. Meek (Aug. 2005). "Adversarial learning". In: *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*. New York, NY, USA: ACM, pp. 641–647. DOI: 10.1145/1081870.1081950.

Lu, Y. (Feb. 2019). "Artificial intelligence: a survey on evolution, models, applications and future trends". In: *Journal of Management Analytics* 6.1, pp. 1–29. DOI: 10.1080/23270012.2019.1570365.

Lund, B. et al. (Jan. 2025). "Standards, frameworks, and legislation for artificial intelligence (AI) transparency". In: *AI and Ethics* 5.4, pp. 3639–3655. DOI: 10.1007/s43681-025-00661-4.

Mobahi, H., M. Farajtabar, and P. Bartlett (Dec. 2020). "Self-Distillation Amplifies Regularization in Hilbert Space". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., pp. 3351–3361. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/2288f691b58edecadcc9a8691762b4fd-Paper.pdf.

Moradi, R., R. Berangi, and B. Minaei (Aug. 2020). "A survey of regularization strategies for deep models". In: *Artificial Intelligence Review* 53.6, pp. 3947–3986. DOI: 10.1007/s10462-019-09784-7.

Mosca, A. and G.D. Magoulas (Oct. 2017). "Distillation of Deep Learning Ensembles as a Regularisation Method". In: *Advances in Hybridization of Intelligent Methods: Models, Systems and Applications*. Cham: Springer International Publishing, pp. 97–118. DOI: 10.1007/978-3-319-66790-4_6.

Nannini, L., A. Balayn, and A.L. Smith (June 2023). "Explainability in AI Policies: A Critical Review of Communications, Reports, Regulations, and Standards in the EU, US, and UK". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: ACM, pp. 1198–1212. DOI: 10.1145/3593013.3594074.

Panigutti, C. et al. (June 2023). "The role of explainable AI in the context of the AI Act". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: ACM, pp. 1139–1150. DOI: 10.1145/3593013.3594069.

Papernot, N. et al. (Apr. 2017). "Practical Black-Box Attacks against Machine Learning". In: *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*. New York, NY, USA: ACM, pp. 506–519. DOI: 10.1145/3052973.3053009.

Pareek, D., S.S. Du, and S. Oh (Dec. 2024). "Understanding the Gains from Repeated Self-Distillation". In: *Advances in Neural Information Processing Systems*. Vol. 37. Curran Associates, Inc., pp. 7759–7796. DOI: 10.52202/079017-0249.

Parinandi, S. et al. (June 2024). "Investigating the politics and content of US State artificial intelligence legislation". In: *Business and Politics* 26.2, pp. 240–262. DOI: 10.1017/bap.2023.40.

Pedregosa, F. et al. (Oct. 2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830. URL: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

Peng, D., Z. Gui, and H. Wu (Mar. 2025). *Interpreting the Curse of Dimensionality from Distance Concentration and Manifold Effect*. DOI: 10.48550/arXiv.2401.00422. arXiv: 2401.00422v3 [cs.LG].

Pestov, V. (May 2013). "Is the k-NN classifier in high dimensions affected by the curse of dimensionality?" In: *Computers and Mathematics with Applications* 65.10, pp. 1427–1437. DOI: https://doi.org/10.1016/j.camwa.2012.09.011.

Sobol, I.M. (Jan. 1967). "On the distribution of points in a cube and the approximate evaluation of integrals". In: *USSR Computational Mathematics and Mathematical Physics* 7.4, pp. 86–112. DOI: 10.1016/0041-5553(67)90144-9.

Statuto, N. et al. (Dec. 2023). "A Scalable and Efficient Iterative Method for Copying Machine Learning Classifiers". In: *Journal of Machine Learning Research* 24.390, pp. 1–34. URL: http://jmlr.org/papers/v24/23-0135.html.

Tan, S. et al. (Dec. 2018). "Distill-and-Compare: Auditing Black-Box Models Using Transparent Model Distillation". In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. New York, NY, USA: ACM, pp. 303–310. DOI: 10.1145/3278721.3278725.

Unceta, I., J. Nin, and O. Pujol (Sept. 2019). "From batch to online learning using copies". In: *Proceedings of the 22nd International Conference of the Catalan Association for Artificial Intelligence*. Netherlands: IOS Press, pp. 125–134. DOI: 10.3233/FAIA190115.

— (Aug. 2020a). "Copying Machine Learning Classifiers". In: *IEEE Access* 8, pp. 160268–160284. DOI: 10.1109/ACCESS.2020.3020638.

Unceta, I., J. Nin, and O. Pujol (July 2020b). *Differential replication in machine learning*. DOI: 10.48550/arXiv.2007.07981. arXiv: 2007.07981 [cs.LG].

— (Nov. 2020c). "Risk mitigation in algorithmic accountability: The role of machine learning copies". In: *PLOS ONE* 15.11, pp. 1–26. DOI: 10.1371/journal.pone.0241286.

— (Mar. 2021). "Differential Replication for Credit Scoring in Regulated Environments". In: *Entropy* 23.4. DOI: 10.3390/e23040407.

Unceta, I. et al. (Aug. 2020). "Sampling Unknown Decision Functions to Build Classifier Copies". In: *Modeling Decisions for Artificial Intelligence*. Cham: Springer International Publishing, pp. 192–204. DOI: 10.1007/978-3-030-57524-3_16.

Wong, R.Y., A. Chong, and R.C. Aspegren (Apr. 2023). "Privacy Legislation as Business Risks: How GDPR and CCPA are Represented in Technology Companies' Investment Risk Disclosures". In: *Proceedings of the ACM on Human-Computer Interaction* 7.CSCW1, pp. 1–26. DOI: 10.1145/3579515.

Wood-Doughty, Z., I. Cachola, and M. Dredze (May 2022). "Model Distillation for Faithful Explanations of Medical Code Predictions". In: *Proceedings of the 21st Workshop on Biomedical Language Processing*. Dublin, Ireland: ACL, pp. 412–425. DOI: 10.18653/v1/2022.bionlp-1.41.

Worel, C.F. (May 2023). *The Curse of Dimensionality: When Too Many Features Break Your Machine Learning Model*. DOI: 10.2139/ssrn.5206704. ssrn: 5206704.

Wu, D. and J. Xu (Dec. 2020). "On the Optimal Weighted $\ell_2$ Regularization in Overparameterized Linear Regression". In: *Advances in Neural Information Processing Systems*. Vol. 33. Red Hook, NY, USA: Curran Associates, Inc., pp. 10112–10123. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/72e6d3238361fe70f22fb0ac624a7072-Paper.pdf.

Xu, Q. et al. (Oct. 2022). "Student Surpasses Teacher: Imitation Attack for Black-Box NLP APIs". In: *Proceedings of the 29th International Conference on Computational Linguistics*. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, pp. 2849–2860. URL: https://aclanthology.org/2022.coling-1.251/.

Ye, X. et al. (Nov. 2024). "Privacy and personal data risk governance for generative artificial intelligence: A Chinese perspective". In: *Telecommunications Policy* 48.10, pp. 102851–102866. DOI: 10.1016/j.telpol.2024.102851.

Zeng, X. and T.R. Martinez (Dec. 2000). "Using a Neural Network to Approximate an Ensemble of Classifiers". In: *Neural Processing Letters* 12.3, pp. 225–237. DOI: 10.1023/A:1026530200837.