1-NN learning on Hanan sets*

Przemysław Śliwiński† Paweł Wachel† Jerzy W. Rozenblit‡

Abstract

The output of the nearest neighbor (1-NN) classification rule, $g_{\mathcal{S},q}(\mathbf{x})$, depends on a given learning set $\mathcal{S}_N$ and on a distance function $\rho_q(\mathbf{x},\mathbf{X})$. We show that transforming $\mathcal{S}_N$ into a set $\mathcal{A}_N$ whose patterns have a Hanan grid-like structure, results in the equivalence $g_{\mathcal{A},q}(\mathbf{x}) = g_{\mathcal{A},p}(\mathbf{x})$ that holds for any NN classifier with distance functions $\|\mathbf{x}-\mathbf{X}\|_q$ and with any $q \in (0,\infty)$. Thanks to the equivalence, $\mathcal{A}_N$ can be used to learn $g_{\mathcal{A},q}(\mathbf{x})$ to mimic a behavior of the classifier $g_{\mathcal{S},p}(\mathbf{x})$ based on the original set $\mathcal{S}_N$ even when q is unknown (and varying). Possible application of the proposed framework (inspired also by a time-varying stimuli perception phenomenon) in autism spectrum disorder (ASD) therapeutic tools design is discussed.

Index Terms

Nearest neighbor algorithm, Hanan set, classifier equivalence, autistic perception model

I. Introduction and problem statement

Usually, in machine learning, we are free to select a classifier or design a new one. Here, we have a set of training data, $\mathcal{S}_N = \{(\mathbf{X}_n, Y_n)\}, n = 1,\ldots,N$ (where $\mathbf{X}_n$'s are D dimensional patterns and Y_n's are their class indices) that is shared by a pair of nearest neighbor (1-NN) classifiers, $g_{\mathcal{S},p}(\mathbf{x})$ and $g_{\mathcal{S},q}(\mathbf{x})$ with the distance functions

$$\rho_p(\mathbf{x},\mathbf{X}) = \|\mathbf{x}-\mathbf{X}\|_p \text{ and } \rho_q(\mathbf{x},\mathbf{X}) = \|\mathbf{x}-\mathbf{X}\|_q, p,q \in (0,\infty), \tag{1}$$

respectively. The first classifier, $g_{\mathcal{S},p}(\mathbf{x})$, is known and its outputs are available, while the outputs of the other, $g_{\mathcal{S},q}(\mathbf{x})$, are inaccessible and the distance function parameter q is unknown and may vary in time.

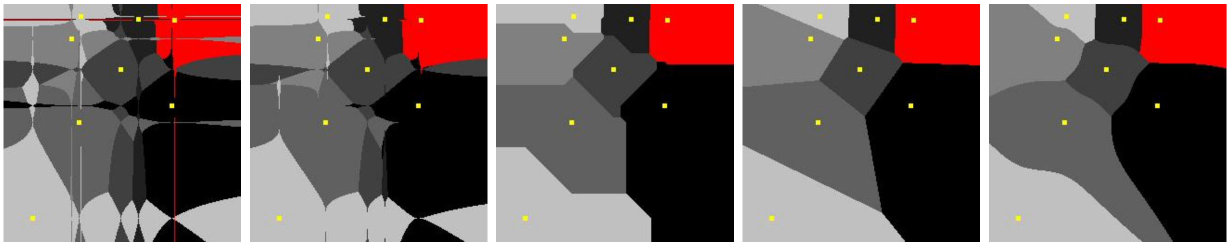
In general, for a given set of training patterns, $\mathcal{S}_N$, such classifiers do not agree, that is, $g_{\mathcal{S},p}(\mathbf{x}) = g_{\mathcal{S},q}(\mathbf{x})$ does not hold for arbitrary $\mathbf{x}$ if $p \neq q$ (*cf.* the shapes of the 2D Voronoi cells in Fig. 1). Hence, without knowing q we can only assume that their decisions match for the training patterns, *i.e.* for $\mathbf{x} \in \mathcal{S}_N$. We will show however that by transforming the original set $\mathcal{S}_N$ into its Hanan grid[1] counterpart $\mathcal{A}_N \supseteq \mathcal{S}_N$, we will obtain the equivalence $g_{\mathcal{A},q}(\mathbf{x}) = g_{\mathcal{A},p}(\mathbf{x})$ for any $\mathbf{x}$. Next, we use this observation to construct algorithms solving the following problem and its variants:

*A preliminary version of the paper (without a formal proof of the main algorithm properties and without a reference to the actually observed phenomenon) was presented at the ICCS'21 conference; [19].

†Przemysław Śliwiński and Paweł Wachel are with Wrocław University of Science and Technology, Wrocław, Poland (e-mails: przemyslaw.sliwinski@pwr.edu.pl, pawel.wachel@pwr.edu.pl).

‡Jerzy W. Rozenblit is with University of Arizona, Tucson, USA, (e-mail: jerzyr@arizona.edu).

[1]If $\mathcal{S}_N$ is a set of points on a plane (2D patterns), then the corresponding Hanan grid, $\mathcal{A}_N$, consists of all patterns of $\mathcal{S}_N$ and all points located at intersections of vertical and horizontal lines that pass through these original patterns; *cf.* [10], [20], [8].



(a) $q = \frac{1}{4}$ (b) $q = \frac{1}{2}$ (c) $q = 1$ (d) $q = 2$ (e) $q = 4$

Fig. 1: Voronoi diagrams for the same set $\mathcal{S}_N$ of $N = 8$ patterns and distance functions $\rho_{\mathcal{S},q}(\mathbf{x}, \mathbf{X})$ with various q

Problem 1. *Given a set of training patterns, $\mathcal{S}_N$, find a new training set $\mathcal{A}_N$, for which the unknown classifier, $g_{\mathcal{A},q}(\mathbf{x})$, agrees with the known one, $g_{\mathcal{A},p}(\mathbf{x})$, for any $\mathbf{x}$.*

Problem 2. *Update $\mathcal{A}_N$ to $\mathcal{A}_{N+1}$ recursively when a new pattern $(\mathbf{X}_{N+1}, Y_{N+1})$ is added to $\mathcal{S}_N$, so that the equivalence $g_{\mathcal{A},q}(\mathbf{x}) = g_{\mathcal{A},p}(\mathbf{x})$ is preserved.*

Problem 3. *Refine $\mathcal{A}_N$ to increase a set of patterns $\mathbf{x}$ for which $g_{\mathcal{A},q}(\mathbf{x}) = g_{\mathcal{S},p}(\mathbf{x})$ holds.*

The algorithm that solves Problem 1 will allow for exact prediction of the unknown classifier decisions even if its parameter q varies. The algorithm addressing Problem 2 will maintain the grid-like structure of the training set $\mathcal{A}_N$, and thus the ability to predict the output of the unknown classifier, each time $\mathcal{S}_N$ is updated. Finally, since the sets $\mathcal{A}_N$ and $\mathcal{S}_N$ are not the same (unless $\mathcal{S}_N$ is already of a grid-like structure), the algorithm for Problem 3 is helpful to make the behavior of the $\mathcal{A}_N$ -based unknown classifier, $g_{\mathcal{A},q}(\mathbf{x})$, more and more similar to the known one, $g_{\mathcal{S},p}(\mathbf{x})$ for the original set $\mathcal{S}_N$.

The problems were inspired by the observed time- and context-dependent perception phenomenon where:

> "[...] tonic levels of DA [dopamine] control the precision of stimulus encoding, which is weighed against contextual information when making decisions. When DA levels are high, the animal relies more heavily on the (highly precise) stimulus encoding, whereas when DA levels are low, the context affects decisions more strongly." [15].

Moreover, the way the training set $\mathcal{A}_N$ is constructed could serve as design guidelines for therapies aimed at non-verbal individuals with autism spectrum disorder (ASD).

A. 1-NN classifier

For convenience, we shortly recall the nearest neighbor algorithm and its basic asymptotic properties; see [7, Ch. 5.1], [3, Ch. 18] and the works cited there. Let $\mathbf{x}$ be a new pattern and let

$$\mathcal{S}_{p,N}(\mathbf{x}) = \left\{ \left(\mathbf{X}_{(1)}, Y_{(1)}\right), \ldots, \left(\mathbf{X}_{(N)}, Y_{(N)}\right) \right\}$$

be a sequence of the training pairs from $\mathcal{S}_N$ sorted w.r.t. their increasing distances to $\mathbf{x}$, *i.e.*, such that $\rho_p\left(\mathbf{x}, \mathbf{X}_{(k)}\right) \leq \rho_p\left(\mathbf{x}, \mathbf{X}_{(l)}\right)$ for $k < l$. The 1-NN rule assigns $\mathbf{x}$ to the class indicated by the first pattern in the ordered sequence:

$$g_{\mathcal{S},p}(\mathbf{x}) = Y_{(1)}.$$

In spite of its simplicity, the algorithm has relatively good asymptotic properties. In particular, for an arbitrary distribution of $\mathbf{X}$, the following upper bound holds for the expected error probability of the binary 1-NN classifier

$$L_{NN} = \lim_{N\to\infty} P\left\{g_{\mathcal{S},p}(\mathbf{X}) \neq Y\right\} \leq 2L^*,$$

where $L^* = E\left\{2\eta(\mathbf{X})(1-\eta(\mathbf{X}))\right\}$ and $\eta(\mathbf{x}) = P(Y=1|\mathbf{X}=\mathbf{x})$ are *the Bayes error* and *a posteriori probability* of error, respectively. In other words, the error of the 1-NN classifier is asymptotically at most twice as large as of the optimal classifier.[2] The 1-NN algorithm is also universal: its asymptotic performance does not depend on the choice of the distance functions $\rho_p(\mathbf{x}, \mathbf{X})$ if it is derived from an l_p-norm in R^D, $p > 1, D < \infty$, [7, Pr. 5.1].

Remark 1. *Note that we focus a non-asymptotic behavior of the 1-NN classifiers and, since p in (1) can be any positive number, our distance functions are not necessary norm-induced.*

II. Algorithms

In what follows, we assume that $D = 2$, however, the presented algorithms[3] 1-3 and the main result in Theorem 1 can be extended to any dimension $D \geq 2$.

Algorithm 1 Transformation of $\mathcal{S}_N$ into a grid learning set $\mathcal{A}_N$

function GENERATE-GRID-SET($\mathcal{S}_N, p$)
    $\mathcal{A} \leftarrow \emptyset$
    $\{X_{1,n}\}, \{X_{2,n}\} \leftarrow$ sets of coordinates of patterns $\mathbf{X}_n$ from $\mathcal{S}_N$
    $\mathcal{H}_N \leftarrow \{X_{1,n}\} \times \{X_{2,n}\}$ ▷ Cartesian product
    for all $\mathbf{H}_n$ in $\mathcal{H}_N$ **do**
        $\mathcal{A} \leftarrow \mathcal{A} \cup \{(\mathbf{H}_n, g_{\mathcal{S},p}(\mathbf{H}_n))\}$ ▷ Pattern classification
    end for
    $\mathcal{A}_N \leftarrow \mathcal{A}$
    return $\mathcal{A}_N$
end function

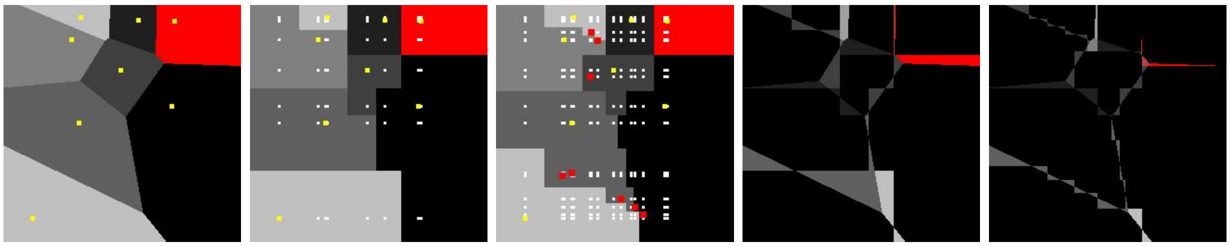
Theorem 1. *If $\mathcal{A}_N$ is a set constructed from $\mathcal{S}_N$ by Algorithm 1, then $g_{\mathcal{A},p}(\mathbf{x}) = g_{\mathcal{A},q}(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^D$ and $p, q \in (0, \infty)$.*

Proof. See Appendix. □

[2] Thus, if $L^* \ll 1$, then the nearest neighbor algorithms can still be considered as reliable; [7, Ch. 2.1 and 5.2].

[3] See https://github.com/Bahrd/ALLY for their Python implementations

The theorem says that for the resulting set $\mathcal{A}_N$, any classifier $g_{\mathcal{A},p}(\mathbf{x})$ with $p \in (0,\infty)$, will classify any new pattern $\mathbf{x}$ in the same way (one can also say that the unknown classifier generalizes in the same way the known one does) and the parameter q can be unknown and can vary in time.



(a) (b) (c) (d) (e)

Fig. 2: The original diagram of $g_{\mathcal{S},2}(\mathbf{x})$ with $N = 8$ original patterns (a). Two distance function-independent Hanan grid-like approximations (white points obtained by Algorithm 1 and red by Algorithm 3 for $L = 8$ additional patterns (b)-(c)); The differences between decisions $g_{\mathcal{S},2}(\mathbf{x})$ and $g_{\mathcal{A},2}(\mathbf{x})$ for $\mathcal{A}_N$ and $\mathcal{A}_{N+L}$, respectively (d)-(e)

Any modification of the original set $\mathcal{S}_N$ requires the set $\mathcal{A}_N$ to be updated as well. In particular (Problem 2), if a new pattern is added to $\mathcal{S}_N$, *i.e.*, $\mathcal{S}_{N+1} = \mathcal{S}_N \cup \{(\mathbf{X}_{N+1}, Y_{N+1})\}$, then the following modification of the Algorithm 1 can be used:

Algorithm 2 Updating the on-grid learning set from $\mathcal{A}_N$ to $\mathcal{A}_{N+1}$

$\mathcal{S}_{N+1} = \mathcal{S}_N \cup \{(\mathbf{X}_{N+1}, Y_{N+1})\}$
function UPDATE-GRID-SET($\mathcal{S}_{N+1}, \mathcal{A}_N, \mathcal{H}_N, p$)
  $\mathcal{A} \leftarrow \mathcal{A}_N$
  $\{H_{1,n}\}, \{H_{2,n}\} \leftarrow$ sets of coordinates of patterns $\mathbf{X}_n \in \mathcal{H}_N$
  $X_1, X_2 \leftarrow$ coordinates of the new pattern $(\mathbf{X}_{N+1}, Y_{N+1})$
  $\mathcal{H}_{N+1} \leftarrow \{H_{1,n}\} \cup \{X_1\} \times \{H_{2,n}\} \cup \{X_2\}$ ▷ Cartesian product
  for all $\mathbf{A}_n \in \mathcal{H}_{N+1} \setminus \mathcal{H}_N$ **do**
    $\mathcal{A} \leftarrow \mathcal{A} \cup \{(\mathbf{A}_n, g_{\mathcal{S},p}(\mathbf{A}_n))\}$ ▷ Pattern classification
  end for
  $\mathcal{A}_{N+1} \leftarrow \mathcal{A}_N \cup \mathcal{A}$
  return $\mathcal{A}_{N+1}$
end function

The Voronoi cells generated by the grid set $\mathcal{A}_N$ are approximations of those produced by the original set $\mathcal{S}_N$ (Problem 3) and the decisions made by $g_{\mathcal{A},q}(\mathbf{x})$ and $g_{\mathcal{S},p}(\mathbf{x})$ (which are based on $\mathcal{A}_N$ and $\mathcal{S}_N$, respectively), may differ for some patterns $\mathbf{x}$; *cf.* Fig. 2. The set of such $\mathbf{x}$'s can be reduced with the help of Algorithm 3.

Remark 2. *If, in general, $\mathcal{S}_N$ is an N-element set of D-dimensional patterns, then the number of training patterns in $\mathcal{A}_N$ will, at worst, be N^D (this upper limit will occur when patterns in $\mathcal{S}_N$ have no common features;* cf.

Algorithm 3 Create $\mathcal{A}_{N+L}$ to make $g_{\mathcal{A},q}(\mathbf{x})$ better approximate $g_{\mathcal{S},p}(\mathbf{x})$

$\mathcal{L}_L \leftarrow$ set of L new (random) patterns $\mathbf{x}$ such that $g_{\mathcal{A},q}(\mathbf{x}) \neq g_{\mathcal{S},p}(\mathbf{x})$,
function Refine-Grid-Set($\mathcal{S}_N, \mathcal{A}_N, \mathcal{H}_N, \mathcal{L}_L, p$)
  $\mathcal{A} \leftarrow \mathcal{A}_N$
  $\{L_{1,l}\}, \{L_{2,l}\} \leftarrow$ sets of coordinates of patterns $\mathbf{X}_l \in \mathcal{L}_L$
  $\{H_{1,n}\}, \{H_{2,n}\} \leftarrow$ sets of coordinates of patterns $\mathbf{X}_n \in \mathcal{H}_N$
  $\mathcal{H}_{N+L} \leftarrow \{H_{1,n}\} \cup \{L_{1,l}\} \times \{H_{2,n}\} \cup \{L_{2,l}\}$ ▷ Cartesian product
  for all $\mathbf{A}_n \in \mathcal{H}_{N+L} \setminus \mathcal{H}_N$ **do**
    $\mathcal{A} \leftarrow \mathcal{A} \cup \{(\mathbf{A}_n, g_{\mathcal{S},p}(\mathbf{A}_n))\}$ ▷ Pattern classification
  end for
  $\mathcal{A}_{N+L} \leftarrow \mathcal{A}_N \cup \mathcal{A}$
  return $\mathcal{A}_{N+L}$
end function

also Fig. 3). However, if $\mathcal{S}_N$ has already a Hanan grid structure (or can be designed as such), then $\mathcal{A}_N = \mathcal{S}_N\,(= \mathcal{A}_{N+L})\,, L = 1, 2, \ldots$, and no new patterns need to be added.

III. Final comments and possible applications

While there is still no definition of autism (see *e.g.* [16], [4]), atypical learning and perception abnormalities are its key characteristics; [11], [17], [21]. Below we discuss the presented results in the context of autism therapy and discuss their applications as the new tools that could be used to learn and interact with the persons with autism. We focus especially on low-functioning non-verbal individuals; *cf.* [18].

A. Guidelines for autism therapies

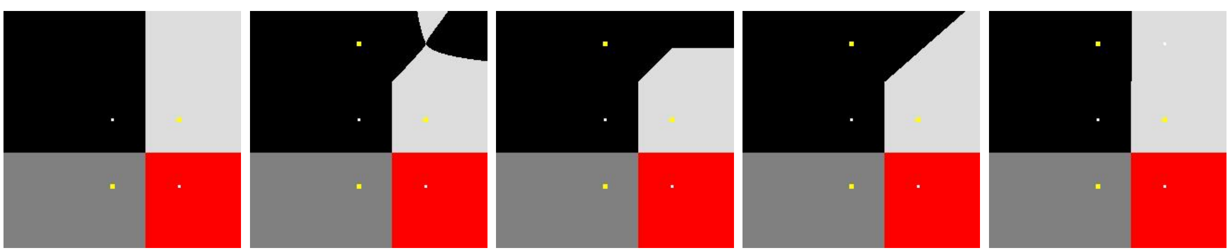
The following assumptions are used to locate the presented algorithms in the autism-related framework:

1) Perception of an individual with autism is represented by the 1-NN classifier with $q \ll 1$ modeling their lack of generalization ability; [6], [11]. Inaccessibility of both the true value of q and the classifier outputs corresponds with the inability to communicate verbally.
2) Conversely, the 1-NN classifier with p known and fixed is a model of a teacher/therapist, whose decisions can be queried and/or easily predicted. The ability to generalize or abstract[4] is represented by the classifier with the distance function L_p, with $p \gg 1$ (since the larger p the less important small differences in patterns).
3) The set $\mathcal{S}_N$ represents the knowledge (*e.g.* static items like objects, scenes and their attributes, and/or dynamic ones, like sequences and scenarios of events or activities) that we want to transfer.
4) Because $\mathcal{S}_N \subseteq \mathcal{A}_N$, all the original ('real') patterns from $\mathcal{S}_N$ are present in $\mathcal{A}_N$. The auxiliary patterns in $\mathcal{A}_N$ can be interpreted as a 'proper context' that helps making the same decisions by the 1-NN classifiers with different q.

[4]"*The idea is that* [...] *abstract space ignores irrelevant details* [*e.g.* negligible pattern features]" [9]

The following guidelines can be derived from these assumptions:

1) In case of low-functioning non-verbal individuals with autism only ‘naturally abstract’ classes like *known or unknown* and *pleasant or unpleasant* are applicable *a priori*.
2) The role of the teacher/therapist remains multifold and encompasses:
   - ‘Supervised’ learning of the real patterns from $\mathcal{S}_N$ and of the context ones from $\mathcal{A}_N$ before allowing an individual to classify a new pattern $\mathbf{x}$ without assistance (Algorithms 1 -3).
   - Controlling the timing (and circumstances) of new items introduction (to avoid, in particular, exposition to new patterns when the individual’s behavior suggests fatigue or anxiety, or when a medical treatment is changed).
   - Selecting features of patterns in the refinement Algorithm 3 and controlling their presentation order; *cf.* [11].
3) The therapeutic tools should be based on realistic, interactive, computer generated 3D objects, scenes and scenarios rather than on simplified (abstract) communication-based ones (like *e.g.* PECS; [5]). Moreover, as the number of context patterns in $\mathcal{A}_N$ can grow fast (see Remark 2), creating them virtually seems to be the only manageable way.
4) Implementation the proposed framework in a form of distributed, available at-home and non-invasive software could help addressing one of the most challenging (especially in the context of autism) problem in reproducible research (see *e.g.* [1], [2]), that is, making a cohort sufficiently large, the experiment conditions consistent and the environmental contribution minimized (so their outcomes can be assessed statistically).



(a) $\mathcal{A}_N, N = 2$, any q (b) $q = \frac{1}{2}$ (c) $q = 1$ (d) $q = 2$ (e) $\mathcal{A}_N, N = 3$, any q

Fig. 3: Adding to the existing grid set (a) a pattern with a single pattern feature (but *without* the context patterns) invalidates its lattice structure (and the subsequent equivalence of the classifiers with different q (b-d)). A single context pattern rectifies the problem in this case (e).

B. Other applications

The proposed framework can also be used in the (non-invasive) experiments to verify or reject the specific hypotheses, for instance:

1) Theorem 1 implies that, in order to maintain the same classification results, all context patterns have to be learned beforehand – see Algorithms 1 and 2. It could suggest that an ’intuitive’ approach that is based on

'a-single-new-change-at-time' principle (where a new pattern with only one feature not shared with others is added) can (if it invalidates the grid structure; see Fig. 3) fail to preserve the classifiers equivalence.

2) Headphones and sunglasses are often used to prevent sensory overload [17], [21]. One could want to verify if such overload is context-dependent and, for instance, whether it occurs when the individuals are exposed to loud and bright scenes and scenarios that are already known to them (and accepted as pleasant).
3) If the already known scene or scenario is associated with unpleasant sensations, then – taking advantage of the properties of distance functions l_q for small q – the new patterns that masks the unpleasant one can be created (together with a pleasant experience) by introducing small changes to the pattern features (that is, without serious rearrangement of the existing environment).
4) There are experimental results suggesting that, in case of high-functioning and fluently speaking individuals, learning by repetition is not effective; [6], [11]. It could therefore be interesting to verify if, instead of repeating the original patterns from $\mathcal{S}_N$, presenting their slightly modified versions from $\mathcal{A}_N$ is a more effective approach.

The framework could further be considered in a more general context as, for instance, recommendations for human-machine interface designs that will potentially be more robust against inconsistent decision making driven by the aforementioned DA levels and/or related to mood or fatigue; *cf. e.g.* [12], [13], [14].

Appendix

Proof of Theorem 1

Proof. Let $\mathbf{x}$ and $\mathbf{X}$ be a pair of patterns. Observe that the distance functions $\rho_p(\mathbf{x}, \mathbf{X}) = \|\mathbf{x} - \mathbf{X}\|_p$ will produce the same value (equal distance between them) for all $p \in (0, \infty)$ if, and only if $\mathbf{x}$ and $\mathbf{X}$ differ at a single feature. In the context of the 1-NN algorithm, it implies that:

- the nearest patterns form a grid, and
- the decision boundaries are straight lines.

In the considered 2D case the patterns can conveniently be illustrated as points on a plane. Let $\mathbf{X} = (x_1, x_2)$ and $\mathbf{Y} = (y_1, y_2)$ be a pair of points. A point $\mathbf{m} = (m_1, m_2)$ is located on a decision boundary if it is equidistant to $\mathbf{X}$ and $\mathbf{Y}$, that is, if $\|\mathbf{X} - \mathbf{m}\|_p = \|\mathbf{Y} - \mathbf{m}\|_p$. Assume $\mathbf{X}$ and $\mathbf{Y}$ differ at the first coordinate. Then $m_2 = x_2 = y_2$ and

$$
\begin{aligned}
\left(|x_1 - m_1|^p + |x_2 - m_2|^p\right)^{\frac{1}{p}} &= \left(|y_1 - m_1|^p + |y_2 - m_2|^p\right)^{\frac{1}{p}} \\
|x_1 - m_1|^p &= |y_1 - m_1|^p + \underbrace{|y_2 - m_2|^p}_{=0} - \underbrace{|x_2 - m_2|^p}_{=0}
\end{aligned}
$$

yields $|x_1 - m_1| = |y_1 - m_1|$. The point $\mathbf{m}$ has thus the following coordinates

$$m_1 = \tfrac{x_1 + y_1}{2}, \; m_2 = x_2 \, (= y_2) \, . \tag{2}$$

To verify that the decision boundary is a straight line, perpendicular to the one intersecting $\mathbf{X}$ and $\mathbf{Y}$ it suffices to take an arbitrary point, $\mathbf{m}' = \left(\frac{x_1+y_1}{2}, m_2'\right)$, on that line and compare the distances $\rho_p\left(\mathbf{m}', \mathbf{X}\right)$ and $\rho_p\left(\mathbf{m}', \mathbf{Y}\right)$:

$$\left|x_1 - \frac{x_1+y_1}{2}\right|^p + \left|x_2 - m_2'\right|^p = \left|y_1 - \frac{x_1+y_1}{2}\right|^p + \left|y_2 - m_2'\right|^p$$

$$\left|x_2 - m_2'\right|^p = \left|y_2 - m_2'\right|^p + \underbrace{\left|\frac{y_1-x_1}{2}\right|^p - \left|\frac{x_1-y_1}{2}\right|^p}_{=0}$$

and because $x_2 = y_2$, then $\left|x_2 - m_2'\right|^p = \left|y_2 - m_2'\right|^p$ holds for any m_2'. From this and from the fact that (2) does not depend on p, we conclude that the decision boundary between any points $\mathbf{X}$ and $\mathbf{Y}$ which differ only at a single coordinate, is the same for all 1-NN algorithms with $\rho_p\left(\mathbf{x}, \mathbf{X}\right)$, any $p \in (0, \infty)$. $\square$

References

[1] "Supporting computational reproducibility through code review," *Nature Human Behaviour*, vol. 5, pp. 965–966, 2021, https://doi.org/10.1038/s41562-021-01190-w.

[2] O. B. Amaral and K. Neves, "Reproducibility: expect less of the scientific paper," 2021.

[3] G. Biau and L. Devroye, *Lectures on the Nearest Neighbor Method*. Springer, 2015, vol. 246.

[4] R. Chapman and W. Veit, "The essence of autism: fact or artefact?" *Molecular Psychiatry*, vol. 26, no. 5, pp. 1440–1441, 2021.

[5] M. H. Charlop-Christy, M. Carpenter, L. Le, L. A. LeBlanc, and K. Kellet, "Using the picture exchange communication system (PECS) with children with autism: Assessment of PECS acquisition, speech, social-communicative behavior, and problem behavior," *Journal of applied behavior analysis*, vol. 35, no. 3, pp. 213–231, 2002, https://onlinelibrary.wiley.com/doi/abs/10.1901/jaba.2002.35-213.

[6] A. B. de Marchena, I.-M. Eigsti, and B. E. Yerys, "Brief report: Generalization weaknesses in verbally fluent children and adolescents with autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 45, no. 10, pp. 3370–3376, 2015.

[7] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. New York: Springer Verlag, 1996.

[8] D.-Z. Du and F. K. Hwang, "Reducing the steiner problem in a normed space," *SIAM Journal on Computing*, vol. 21, no. 6, pp. 1001–1007, 1992.

[9] F. Giunchiglia and T. Walsh, "A theory of abstraction," *Artificial Intelligence*, vol. 57, no. 2-3, pp. 323–389, 1992.

[10] M. Hanan, "On Steiner's problem with rectilinear distance," *SIAM Journal on Applied Mathematics*, vol. 14, no. 2, pp. 255–265, 1966.

[11] H. Harris, D. Israeli, N. Minshew, Y. Bonneh, D. J. Heeger, M. Behrmann, and D. Sagi, "Perceptual learning in autism: over-specificity and possible remedies," *Nature Neuroscience*, vol. 18, no. 11, pp. 1574–1576, 2015.

[12] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1052–1068, 2004.

[13] W.-C. Lin and C.-C. Lee, "Computational analyses of thin-sliced behavior segments in session-level affect perception," *IEEE Transactions on Affective Computing*, vol. 11, no. 4, pp. 560–573, 2018.

[14] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, E. Kruijff, and J. Schöning, "Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 12, pp. 3389–3401, 2020.

[15] J. G. Mikhael and S. J. Gershman, "Impulsivity and risk-seeking as Bayesian inference under dopaminergic control," *Neuropsychopharmacology*, vol. 47, no. 2, p. 465–476, 2022, https://doi.org/10.1038/s41386-021-01125-z.

[16] L. Mottron and D. Bzdok, "Autism spectrum heterogeneity: fact or artifact?" *Molecular Psychiatry*, vol. 25, no. 12, pp. 3178–3185, 2020.

[17] M. O'Neill and R. S. Jones, "Sensory-perceptual abnormalities in autism: a case for more research?" *Journal of Autism and Developmental Disorders*, vol. 27, no. 3, pp. 283–293, 1997.

[18] S. Piana, C. Malagoli, M. C. Usai, and A. Camurri, "Effects of computerized emotional training on children with high functioning autism," *IEEE Transactions on Affective Computing*, 2019.

[19] P. Śliwiński, "On autistic behavior model," in *2019 Spring Simulation Conference (SpringSim)*. IEEE, 2019, pp. 1–8.

[20] T. L. Snyder, "On the exact location of Steiner points in general dimension," *SIAM Journal on Computing*, vol. 21, no. 1, pp. 163–180, 1992.

[21] L. D. Wiggins, D. L. Robins, R. Bakeman, and L. B. Adamson, "Brief report: sensory abnormalities as distinguishing symptoms of autism spectrum disorders in young children," *Journal of Autism and Developmental Disorders*, vol. 39, no. 7, pp. 1087–1091, 2009.