Mimicking learning for 1-NN classifiers

Przemysław Śliwiński[1 [0000−0003−3839−1580]], Paweł Wachel[1 [0000−0002−7353−2310]], and Jerzy W. Rozenblit[2 [0000−0002−7348−4128]]

[1] Wrocław University of Science and Technology, 50-370 Wrocław, Poland
{przemyslaw.sliwinski,pawel.wachel}@pwr.edu.pl
[2] University of Arizona, Tucson, AZ 85721, USA, jerzyr@arizona.edu

Abstract. We consider the problem of mimicking the behavior of the nearest neighbor algorithm with an unknown distance measure. Our goal is, in particular, to design and update a learning set so that two NN algorithms with various distance functions ρ_p and ρ_q, $0 < p, q < \infty$, classify in the same way, and to approximate the behavior of one classifier by the other. The autism disorder-related motivation of the problem is presented.

1 Introduction and problem statement

We investigate a problem of mimicking the behavior of a nearest neighbor binary classifier with a distance measure function $\rho_p(\mathbf{x}, \mathbf{X}) = \|\mathbf{x} - \mathbf{X}\|_p$, $p \in (0, \infty)$. The training set consists of pairs $\mathcal{S} = \{(\mathbf{X}_n, Y_n)\}$, where $\mathbf{X}_n \in R^2$ is a two-dimensional vector (a pattern), and $Y_n \in \{0, 1\}$ denotes a class the vector belongs to.

Unlike most of the classification problems, where the goal is to construct an effective classifier (and investigate its properties), we assume that there exists a fixed classifier, $g_q(\cdot)$, implementing the nearest neighbor rule based on a distance measure ρ_q, with unknown q. A fundamental assumption here is that neither its classification rule nor the distance function can be replaced.

Within such a framework, the following problems are examined:

1. Given a training set $\mathcal{S} = \{(\mathbf{X}_n, Y_n)\}$, $n = 1, \ldots, N$, design a new (augmented) training set $\mathcal{A} \supset \mathcal{S}$ that allows the unknown classifier $g_q(\cdot)$ to approximate (mimic) the behavior of a known classifier $g_p(\cdot)$.
2. For the known classifier $q_p(\cdot)$ and a new data pair $\{(\mathbf{X}_{N+1}, Y_{N+1})\}$, modify the training set $\mathcal{A}$ to make the unknown one, $g_q(\cdot)$, work in the same way.

Our problem formulation is based on the autistic perception model (presented in more detail in Section 4).

2 1-NN classifiers and on-grid learning sets

Here we shortly recall a rule implemented in the nearest neighbor classifier and its basic asymptotic properties; see [2, Ch. 5.1] and the works cited therein.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_7

2.1 Nearest neighbor classifier

Let $\mathbf{x}$ be a new pattern and let

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

be a sequence in which the training pairs from $\mathcal{S}$ are sorted w.r.t. increasing distances $\rho_p(\mathbf{x}, \mathbf{X}_n)$. The NN rule assigns $\mathbf{x}$ to the class indicated by the first pattern in the ordered sequence:

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

In spite of its simplicity, the algorithm has relatively good asymptotic properties. In particular, the following upper bound holds for the expected error probability

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

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 for an arbitrary distribution of $\mathbf{X}$, respectively. In other words, the error of the NN classifier is asymptotically at most twice as large as of the optimal classifier.[3] The NN algorithm is universal in the sense that its asymptotic performance does not depend on the choice of the distance measure ρ if it is an arbitrary norm in R^d, $d < \infty$, [2, Pr. 5.1].

The formal results presented above are thus encouraging, however, we are interested in a non-asymptotic behavior of the NN classifier and in distance measures that are not derived from norms, that is, we admit ρ_p, ρ_q with $p, q < 1$.

Fig. 1: Voronoi diagrams for various distance measures $\rho_p, p = 1/4, 1, 2$ and 4, from left to right. $N = 8$ (for improved visibility the Voronoi cells have different colors, in spite of the fact that we examine a dichotomy problem, *i.e.* the binary classifiers only)

[3] Hence, if L^{*} is small, the performance of the NN algorithm can be acceptable, [2, Ch. 2.1 and 5.2].

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_7

2.2 Voronoi cells

For two-dimensional patterns, a working principle of the nearest neighbor classifier can conveniently be illustrated in the form of the Voronoi diagrams where each training pattern $\{\mathbf{X}_n\}$ determines a set (called a Voronoi cell) of its closest neighbors (with respect to the selected distance measure).[4]

2.3 On-grid training set

The various shapes of the Voronoi cells – as clearly seen in Fig. 1 – imply that for the same training set S, the classifiers with various distance functions L_q will make different decisions. In particular, if the parameter q in ρ_q is unknown, these decisions can be difficult to predict, especially for $q < 1$.

Observe, however, that if the training patterns $\{\mathbf{X}_n\}$, $n = 1, \ldots, N \times N$, form a grid,[5] that is, they are all located at the crossing of lines parallel to OX and OY axes, and determined by their coordinates, then, for any $p \in (0, \infty)$, the shapes of the resulting Voronoi diagrams will be the same.

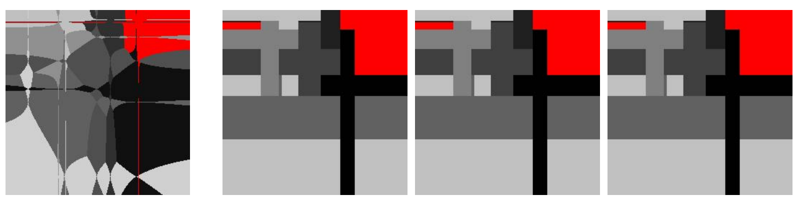

Fig. 2: The original Voronoi diagram for $\rho_p, p = 1/4$, $N = 8$ (left) and its (identical) distance measure-independent approximations, generated for $q = 1, 2$, and 4

We will only present a sketch of the proof of this property here. Consider a case $p = 2$ and take a pair of points. Then, using the *ruler and compass*, construct the line which is equidistant to them. Observe that:

- The line splits the plane and creates two Voronoi cells,
- The construction is not based on the radius of the drawn circles but on their mirror symmetry with respect to the constructed line. Since all circles in ρ_p possess this symmetry property, the splitting line will have the same location for any p.

Repeating the construction for all pairs of adjacent points (corresponding with patterns $\{\mathbf{X}_n\}$) will yield the Voronoi diagram the same for all ρ_p, $p \in (0, \infty)$; see Figs. 2 and 3.

[4] See `https://github.com/Bahrd/Voronoi` for the Python scripts.
[5] Note that, due to randomness of the patterns $\{X_n\}$, the grid points are not equidistant.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_7

Remark 1. Observe that, by virtue of the construction, the same behavior of the NN classifiers for various values p holds only if points are located on the grid. It seems to be a serious restriction, however in what follows we assume that forming a grid is a necessary condition for two classifiers with different p and q to behave in the same way.

3 Proposed solutions

The solution to the first and second problems can now be immediately derived as follows:

1. Given a set $\mathcal{S}$ of N learning pairs, add $N(N-1)$ new pairs that create an $N \times N$ on-grid set $\mathcal{A}$ with new points classified according to the known classifier $g_p(\cdot)$. If the approximation appears too crude (*i.e.* the shapes of the new Voronoi cells are not sufficently similar to the original ones), then one can add new L points and create a denser grid $(L+N) \times (L+N)$, *cf.* Figs. 3. The obvious expense is the quadratic growth of the number of training points.
2. In order to add a new pattern $(\mathbf{X}_{N\times N+1}, Y_{N\times N+1})$ to the existing on-grid set $\mathcal{A}$, the set of accompanying patterns has to be added as well in order to maintain a grid structure of the training set. That is, the new set of patterns $\{(\mathbf{X}_{N\times N+n}, Y_{N\times N+n})\}$, $n = 1, \ldots, N+1$, need to be added with each new pattern classified by the known classifier $g_p(\cdot)$.

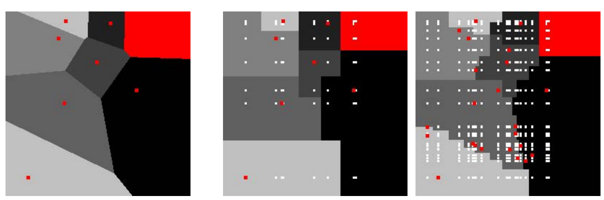

Fig. 3: The diagram of the $q_2(\cdot)$ classifier with $N = 8$ training pairs and its distance measure-independent approximations: 'crude' for $N = 8 \times 8$ and 'fine' for $N = 16 \times 16$ grid points (red dots – the initial set $\mathbf{S}$, white – augmented on-grid sets $\mathbf{A}$)

4 Autistic learning context and final remarks

The specific assumptions made in this note are derived from observations of autistic persons and from hypotheses based on the phenomena published in the literature; see *e.g.*: [1], [7], [4], [9], [8]. For instance:

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_7

- The attention to details and perception of minute changes, that is characteristic for autistic persons, can correspond to the distance measures ρ_q with $q < 1$. Note that q can be unknown, different for each person, and (because, for instance, of varying fatigue level) can also vary in time, *cf.* [8].
- An increased attention can result in sensory overload and chronic fatigue-like state, decreasing learning abilities and increasing impatience can be represent by the *greedy* 1-NN algorithm (rather than by the k-NN one[6]).
- The binary classification can be applied to generic, yet still useful classes, like known vs. unknown/unpleasant scenes/situations/objects.
- On-grid patterns points can be seen, in general, as a serious restriction, however, in a controlled environment, the new samples can be, in principle, provided and augmented by a therapist.

We believe that our approach, starting from observations, to hypotheses, to models and algorithms (together with their formal and simulation-based verification), resulting eventually in treatment and therapy proposals, is compatible with the *computational psychiatry* methodology; [7] *cf.* [3], [5], [6]. Early examination of formal properties of the proposed models can also limit a number of actual experiments the autistic persons are involved in. The primary reasons behind this motivation is that trials and tests are:

- Expensive, as gathering numerous and representative autistic cohort without disturbing their routine (and subsequently, interfere with the outcome of the experiment) is laborious.
- Time-consuming, because getting the autistic person used to the new environment requires sometimes a months-long preparation period.
- Difficult to assess, since autistic persons may not be able to communicate the results, and therefore some indirect and noninvasive measurement methods have to be applied.

Acknowledgments

Autors want to thank the Reviewers for their comments and suggestions.

References

1. Association, A.P., et al.: Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub (2013)
2. Devroye, L., Györfi, L., Lugosi, G.: A Probabilistic Theory of Pattern Recognition. Springer Verlag, New York (1996)
3. Galitsky, B.: Computational Autism. Springer (2016)

[6] The k-NN algorithm can however be used to model a known classifier (a 'teacher').

[7] This relatively new scientific discipline aims at developing and examining theoretical and computational models that could serve as a basis for new therapies and/or medicines.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_7

4. Maurer, D., Werker, J.F.: Perceptual narrowing during infancy: A comparison of language and faces. Developmental Psychobiology **56**(2), 154–178 (2014)
5. Palmer, C.J., Lawson, R.P., Hohwy, J.: Bayesian approaches to autism: Towards volatility, action, and behavior. Psychological Bulletin **143**(5), 521 (2017)
6. Śliwiński, P.: On autistic behavior model. In: 2019 Spring Simulation Conference (SpringSim). pp. 1–8. IEEE (2019)
7. Wiggins, L.D., Robins, D.L., Bakeman, R., Adamson, L.B.: Breif report: sensory abnormalities as distinguishing symptoms of autism spectrum disorders in young children. Journal of Autism and Developmental Disorders **39**(7), 1087–1091 (2009)
8. Wilson, R.C., Shenhav, A., Straccia, M., Cohen, J.D.: The eighty five percent rule for optimal learning. Nature communications **10**(1), 1–9 (2019)
9. Yang, J., Kanazawa, S., Yamaguchi, M.K., Motoyoshi, I.: Pre-constancy vision in infants. Current Biology **25**(24), 3209–3212 (2015)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_7