PowerHC: non linear normalization of distances for advanced nearest neighbor classification

Manuele Bicego[1] and Mauricio Orozco-Alzate[2]

[1]University of Verona, Italy

[2]Universidad Nacional de Colombia - Sede Manizales, Colombia

Explainable/interpretable Pattern Recognition and advanced nearest neighbor rules

- Explainable/interpretable Pattern Recognition means giving importance not just to the accurate decision (the "what") but also to its reason (the "why") [3].
- The nearest neighbor rule (1-NN) and its variants are very interpretable since the nearest training objects, which assign the label, can be visualized.
- Two advanced 1-NN variants: the Hypersphere Classifier (HC) [2] and the Adaptive Nearest Neighbor rule (ANN) [5]. They scale the distances using radii.

HC and ANN

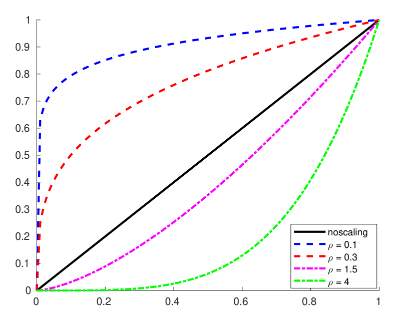
For $\mathbf{x}_i$, let r_i be the distance (radius of its influence hypersphere) to the nearest training object belonging to a different class, then:

$$d_{HC}(\mathbf{x}, \mathbf{x}_i) = d(\mathbf{x}, \mathbf{x}_i) - r_i$$

$$d_{ANN}(\mathbf{x}, \mathbf{x}_i) = \frac{d(\mathbf{x}, \mathbf{x}_i)}{r_i}$$

- HC and ANN are logarithmically related [4].
- It seems promising to investigate the effect of other non linear scalings, such as the power transformation [1]: $d(\mathbf{x}_i, \mathbf{x}_j)^\rho, \quad \rho > 0$.

Power transformation



The PowerHC rule

- Let p_i the power-radius of $\mathbf{x}_i$; i.e. the radius of its hypersphere computed using the power of the distances, then:

$$d_{PowerHC}(\mathbf{x}, \mathbf{x}_i) = d(\mathbf{x}, \mathbf{x}_i)^\rho - p_i.$$

- Given this corrected distance, as in the case of ANN and HC, the classification is then performed using the NN rule (or the KNN rule).
- We tested PowerHC under several conditions of dimensionality, cardinality and number of classes, as well as on a challenging classification of volcano-seismic signals.

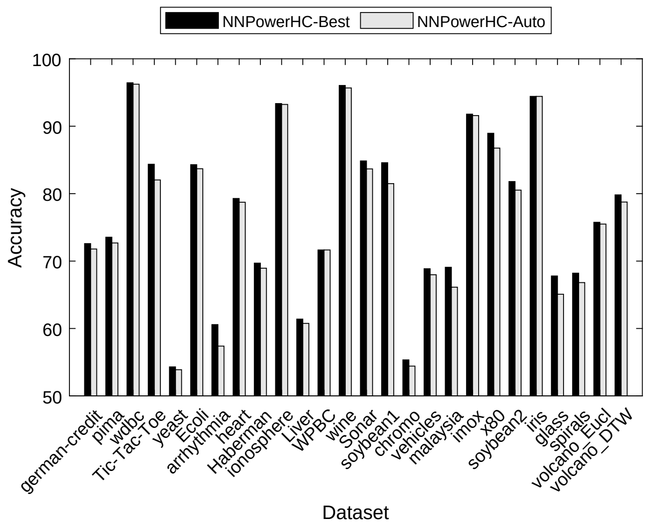
Results: 50 rep., 50-50 train-test with best ρ

- The best ρ is the one which minimizes the LOO 1-NN error on the training set.

Method / Dataset	Accuracies ($\rho \in \{0.2, 0.4, \dots, 9.8, 10\}$) A NN	B NN-ANN	C NN-HC	D NN-PowerHC	t-tests B vs D	C vs D
★ `german-credit`	68.72±0.29	71.32±0.29	71.58±0.29	72.59±0.28 ($\rho = 6.2$)	Reject $\nearrow$	Reject $\nearrow$
★ `pima`	69.70±0.33	72.43±0.32	72.70±0.32	73.54±0.32 ($\rho = 6.0$)	Reject $\nearrow$	Reject $\nearrow$
★ `tic-tac-toe`	79.52±0.26	80.86±0.25	83.04±0.24	84.36±0.23 ($\rho = 5.8$)	Reject $\nearrow$	Reject $\nearrow$
★ `yeast`	51.02±0.26	53.53±0.26	53.78±0.26	54.31±0.26 ($\rho = 3.4$)	Reject $\nearrow$	Reject $\nearrow$
★ `arrhythmia`	57.88±0.48	55.07±0.49	58.02±0.48	60.58±0.48 ($\rho = 2.0$)	Reject $\nearrow$	Reject $\nearrow$
★ `heart`	76.55±0.49	78.23±0.48	78.50±0.48	79.28±0.47 ($\rho = 3.2$)	Reject $\nearrow$	Reject $\nearrow$
★ `haberman`	66.32±0.54	68.70±0.53	68.70±0.53	69.70±0.53 ($\rho = 9.6$)	Reject $\nearrow$	Reject $\nearrow$
■ `wdbc`	95.06±0.18	96.16±0.16	96.36±0.16	96.44±0.16 ($\rho = 1.6$)	Reject $\nearrow$	Accept
■ `ecoli`	81.79±0.42	83.52±0.40	84.14±0.40	84.30±0.40 ($\rho = 1.8$)	Reject $\nearrow$	Accept
■ `volcano_DTW`	72.55±0.27	78.41±0.25	79.75±0.25	79.82±0.25 ($\rho = 1.6$)	Reject $\nearrow$	Accept
■ `glass`	68.50±0.64	66.91±0.64	67.66±0.64	67.79±0.64 ($\rho = 1.2$)	Reject $\nearrow$	Accept
▲ `sonar`	83.44±0.52	84.82±0.50	84.49±0.50	84.85±0.50 ($\rho = 0.2$)	Accept	Reject $\nearrow$
▲ `iris`	93.33±0.41	94.40±0.38	93.89±0.39	94.43±0.37 ($\rho = 0.2$)	Accept	Reject $\nearrow$
▲ `liver`	61.45±0.52	61.40±0.52	61.06±0.52	61.40±0.52 ($\rho = 0.6$)	Accept	Reject $\nearrow$
▲ `vehicles`	69.11±0.32	68.79±0.32	68.74±0.32	68.86±0.32 ($\rho = 0.8$)	Accept	Reject $\nearrow$
▲ `malaysia`	70.64±0.53	69.05±0.54	68.72±0.54	69.08±0.54 ($\rho = 0.2$)	Accept	Reject $\nearrow$
◇ `ionosphere`	85.21±0.38	93.36±0.27	93.19±0.27	93.36±0.27 ($\rho = 0.2$)	Accept	Accept
◇ `wpbc`	65.59±0.68	71.46±0.65	71.18±0.65	71.65±0.65 ($\rho = 3.4$)	Accept	Accept
◇ `wine`	95.00±0.33	95.93±0.30	96.00±0.29	96.05±0.29 ($\rho = 0.6$)	Accept	Accept
◇ `chromo`	55.34±0.29	55.24±0.29	55.28±0.29	55.35±0.29 ($\rho = 0.8$)	Accept	Accept
◇ `volcano_Eucl`	73.91±0.27	75.73±0.26	75.69±0.26	75.75±0.26 ($\rho = 0.4$)	Accept	Accept
◇ `soybean1`	85.16±0.44	84.24±0.45	84.41±0.44	84.59±0.44 ($\rho = 1.6$)	Accept	Accept
◇ `imox`	92.94±0.37	91.52±0.40	91.73±0.40	91.79±0.40 ($\rho = 0.8$)	Accept	Accept
◇ `x80`	94.38±0.68	88.29±0.95	88.29±0.95	88.95±0.92 ($\rho = 3.6$)	Accept	Accept
◇ `soybean2`	82.03±0.66	81.62±0.66	81.62±0.66	81.79±0.66 ($\rho = 1.2$)	Accept	Accept
△ `spirals`	74.25±0.63	68.56±0.67	67.44±0.67	68.21±0.67 ($\rho = 0.2$)	$\nwarrow$Reject	Reject $\nearrow$

Results: Best ρ vs. automatic tuning

- The automatic tuning approach is based on cross-validation.



- Automatic tuning is only slightly worse than using the best ρ.

Results were also computed performed for KNN and its advanced variants (KNN-ANN, KNN-HC and KNN-PowerHC), reporting the best values among the results for $K \in \{1, 3, ..., 27, 29\}$; see the conference papers for details.

Conclusions

- We investigated the suitability of non linear scaling of distances to improve standard as well as advanced Nearest Neighbor approaches.
- In particular we studied PowerHC, a method which normalizes distances with a power transformation prior to applying the HC classifier.
- The experimental evaluation confirms the suitability of the proposed approach.
- Remarkably, we have shown that on a real world challenging application related to the classification of volcano-seismic signals, advanced nearest neighbor techniques —and especially the PowerHC— can be very beneficial.

References

[1] A. Carli et al. "Nonlinear mappings for generative kernels on latent variable models," in: *Proc. Int. Conf. on Pattern Recognition.* 2010, 2134–2137.

[2] N. Lopes and B. Ribeiro. "Incremental Hypersphere Classifier (IHC)". In: *Machine Learning for Adaptive Many-Core Machines - A Practical Approach.* Vol. 7. Studies in Big Data. Cham: Springer International Publishing, 2015. Chap. 6, pp. 107–123.

[3] C. Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* 2019. URL: `https://christophm.github.io/interpretable-ml-book/`.

[4] M. Orozco-Alzate, S. Baldo, and M. Bicego. "Relation, Transition and Comparison Between the Adaptive Nearest Neighbor Rule and the Hypersphere Classifier". In: *ICIAP 2019 Proceedings.* Vol. 11751. LNCS. GIRPR/IAPR. Trento, Italy: Springer, 2019, pp. 141–151.

[5] J. Wang, P. Neskovic, and L. N. Cooper. "Improving nearest neighbor rule with a simple adaptive distance measure". In: *Pattern Recognition Letters* 28.2 (2007), pp. 207 –213.