

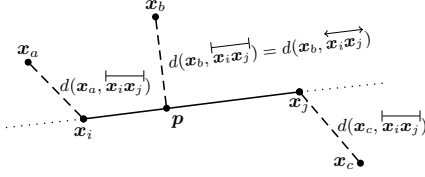
# A CHEAPER RECTIFIED-NEAREST-FEATURE-LINE-SEGMENT CLASSIFIER BASED ON SAFE POINTS

Mauricio Orozco-Alzate<sup>1</sup> and Manuele Bicego<sup>2</sup>

<sup>1</sup>Universidad Nacional de Colombia - Sede Manizales, Colombia

<sup>2</sup>University of Verona, Italy

## Rectified-Nearest-Feature-Line-Segment (RNFLS) classifier



- RNFLS [1] **improves** over NFL [2] by solving two drawbacks of the latter: *interpolation* and *extrapolation* inaccuracies.
- Segmentation: Distances on the extrapolating part of the feature line are replaced with the distance to the nearest endpoint.
- Rectification: Remove feature lines segments crossing the territory of other classes. **Very costly!**
- Degenerated lines are also considered. So, RNFLS includes 1-NN as a special case.

## Typification according to the 5-nearest neighbors

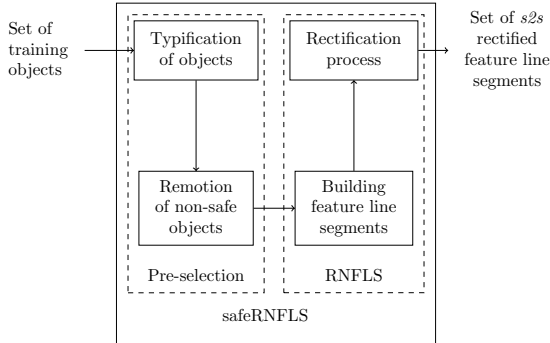
According to the proportion  $\{Same\ class\}:\{Different\ class\}$  among its 5-nearest neighbors, a point  $x$  is categorized as [3]:

- **safe (s)** if 5:0 or 4:1;
- **borderline (b)** if 3:2 or 2:3;
- **rare (r)** if 1:4 but, only if its nearest neighbor from the same class has, in turn, a proportion of either 0:5 or 1:4. Otherwise,  $x$  is  $b$  [4];
- **outlier (o)** if 0:5.

## Proposed typification of feature line segments

- We propose to categorize each feature line segment according to the types of its endpoints: **s2s, s2b, s2r, s2o, b2b, b2r, b2o, r2r, r2o, o2o**.
- The most preserved category after the rectification process is **s2s**. In addition, most of the class labels are assigned by **s2s** feature line segments.
- **Hypothesis**: removing of all non-safe examples, prior the building of the feature line segments, allows to avoid computations without significantly deteriorating the classification performance of the original RNFLS.

## safeRNFLS: the cheaper proposal



## Classification accuracies (20 rep., 50-50 random train-test)

Dataset	(a) 1-NN vs. safeNN		(b) RNFLS vs. safeRNFLS		(c) NFL vs. safeNFL	
	1-NN	safeNN	RNFLS	safeRNFLS	NFL	safeNFL
Hepatitis	92.25±0.95	87.75±1.16	91.50±0.99	91.38±0.99	93.62±0.86	93.00±0.90
Iris	93.40±0.64	93.67±0.63	94.87±0.57	94.80±0.57	87.07±0.87	87.47±0.85
Pima	70.29±0.52	<u>72.93±0.51</u>	74.14±0.50	74.44±0.50	68.05±0.53	68.31±0.53
Wine	94.33±0.55	94.27±0.55	95.45±0.49	95.34±0.50	95.73±0.48	95.62±0.49
Liver	59.83±0.83	58.32±0.84	63.67±0.82	62.86±0.82	61.16±0.83	61.04±0.83
Ionosphere	<u>84.49±0.61</u>	77.50±0.70	90.43±0.50	89.38±0.52	83.89±0.62	83.38±0.63
WDBC	94.88±0.29	<u>95.61±0.27</u>	96.47±0.24	96.53±0.24	94.77±0.29	94.88±0.29
WPBC	65.46±1.08	<u>75.36±0.98</u>	72.99±1.01	74.33±0.99	72.16±1.02	71.75±1.02
Glass	<u>66.40±1.02</u>	58.41±1.07	68.36±1.01	67.10±1.02	<u>62.90±1.04</u>	60.28±1.06
Gastro	52.11±1.81	49.08±1.81	<u>55.66±1.8</u>	45.53±1.81	<u>58.55±1.79</u>	51.97±1.81

## Execution times (in seconds) and percentage of savings

Dataset	(a) 1-NN vs. safeNN			(b) RNFLS vs. safeRNFLS			(c) NFL vs. safeNFL		
	1-NN	safeNN	Saving	RNFLS	safeRNFLS	Saving	NFL	safeNFL	Saving
Hepatitis	0.03	0.02	7.27%	0.44	0.40	8.53%	0.49	0.46	6.78%
Iris	0.14	0.10	26.20%	1.56	1.38	11.56%	1.41	1.26	10.56%
Pima	2.03	0.91	55.11%	127.74	96.19	24.70%	334.62	134.60	59.78%
Wine	0.11	0.08	21.17%	1.74	1.69	2.90%	2.59	2.08	19.85%
Liver	0.34	0.11	69.49%	4.71	1.72	63.54%	28.85	2.82	90.22%
Ionosphere	0.35	0.27	21.35%	17.62	14.77	16.12%	33.67	27.08	19.59%
WDBC	0.91	0.87	4.21%	101.81	98.03	3.71%	135.29	117.02	13.51%
WPBC	0.11	0.06	45.63%	2.64	1.78	32.54%	6.24	1.21	80.53%
Glass	0.13	0.05	61.60%	1.08	0.42	61.12%	3.55	0.88	75.31%
Gastro	0.02	0.005	78.90%	0.06	0.01	81.14%	0.34	0.004	98.89%

## Conclusions

- **s2s** feature line segments are typically the ones providing the class label assignments for the RNFLS classifier; see the exhaustive study in our conference paper.
- **safeRNFLS** is, in general, **not significantly different** from RNFLS but **cheaper** (saved computations and execution times, in most cases, are outstanding).
- Safe variants of 1-NN and NFL were also studied; see results for **safeNN** and **safeNFL** respectively.
- **safeRNFLS** is not recommended for complicated compositions along with very sparse representations (few examples in very high-dimensional feature spaces); c.f. **Gastro**.

## References

- [1] Hao Du and Yan Qiu Chen. "Rectified nearest feature line segment for pattern classification". In: *Pattern Recognition* 40.5 (2007), pp. 1486–1497.
- [2] Stan Z. Li and Juwei Lu. "Face Recognition Using the Nearest Feature Line Method". In: *IEEE Transactions on Neural Networks* 10.2 (1999), pp. 439–443.
- [3] Krystyna Napierala and Jerzy Stefanowski. "Types of minority class examples and their influence on learning classifiers from imbalanced data". In: *Journal of Intelligent Information Systems* 46.3 (2016), pp. 563–597.
- [4] José A. Sáez, Bartosz Krawczyk, and Michał Woźniak. "Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets". In: *Pattern Recognition* 57 (2016), pp. 164–178.