## SPIDER3 Algorithm – Online Appendix

Szymon Wojciechowski<sup>1</sup>, Szymon Wilk<sup>1</sup>, Jerzy Stefanowski<sup>1</sup>

Institute of Computing Science, Piotrowo 2, 60-965 Poznan, Poland

## 1 Experiment on Real-life Clinical Data

We performed an additional experiment that involved an additional Naive Bayes classifier (NB) and 5 real-life clinical data sets describing pediatric patients presenting to the emergency department (ED) with various complaints. In all these sets  $c_{min}$  corresponds to specialist consult,  $c_{int}$  to additional observation and finally  $c_{maj}$  to discharge. We used two-class versions of these sets in our previous study [1] and here we consider the original classification. A brief characteristic of these data sets is given in Table 1 and for a more detailed description see [1]. In the AE2 data set the numbers of examples in  $c_{int}$  and  $c_{maj}$  classes were comparable and in SP2  $c_{int}$  was the most prevalent class. Nevertheless, we decided to not keep the classes ordered according to their relevancy. Finally, we employed the 10-fold cross validation scheme repeated 5 times.

Data set	# attributes	# examples	$(c_{min},c_{int},c_{maj})$
abdominal pain (AP)	12	457	(48, 61, 348)
asthma exacerbations 1 (AE1)	32	362	(59, 140, 163)
asthma exacerbations 2 (AE2)	42	240	(21, 110, 109)
hip pain (HP)	20	412	(46, 54, 312)
scrotal pain (SP)	14	409	(56, 269, 84)

Table 1. Characteristic of clinical data sets considered in the study

## 2 Results of Experiment

Results from the additional experiment are given in Table 2 (TPR = true-positive rate). This time we focus on NB and PART-U classifiers – we decided to replace 1-NN with NB following the experience with our earlier study [1] where NB turned out to be the best performing classifier. Again, the best value for a given classifier and a data set is marked with bold font, while the second best – with italics.

The main observations are the following:

- None of the considered methods led to the best performance for both  $c_{min}$  and  $c_{int}$  classes. SPIDER2 worked better for  $c_{min}$  (it won 3 times for each of the presented classifiers), while SPIDER3 resulted in the best improvements for  $c_{int}$  (it was especially visible for PART-U where application of SPIDER3 always led to the highest TPR for this class).
- Despite differences discussed above, geometric mean (GM) of TPRs for specific classes was in most cases the highest for SPIDER3 (exceptions include the SP and HP data sets – the latter only for NB). This indicates SPIDER3 provides a more "even" performance across decision classes (and confirms our observations from artificial data).
- Performance observed for 1-NN (nor reported here) was consistent with the one for NB and PART-U with SPIDER2 leading to better performance for  $c_{min}$  and SPIDER3 resulting in best improvements for  $c_{int}$ .

		NB		]	PART-U	
Method	$c_{min}$	$c_{int}$	$c_{maj}$	$c_{min}$	$c_{int}$	$c_{maj}$
none	0.725	0.102	0.882	0.451	0.079	0.871
SP2-min+int	0.800	0.463	0.601	0.576	0.190	0.770
SP2-int+maj	0.837	0.099	0.796	0.542	0.144	0.828
SP3-default	0.791	0.372	0.679	0.560	0.328	0.650
SP3-costs	0.833	0.508	0.521	0.626	0.358	0.575
none	0.386	0.529	0.728	0.295	0.427	0.609
SP2-min+int	0.511	0.536	0.460	0.494	0.456	0.287
SP2-int+maj	0.645	0.384	0.642	0.325	0.410	0.560
SP3-default	0.471	0.583	0.569	0.362	0.610	0.412
SP3-costs	0.555	0.520	0.604	0.373	0.543	0.452
none	0.687	0.558	0.670	0.267	0.516	0.561
SP2-min+int	0.657	0.565	0.532	0.383	0.467	0.346
SP2-int+maj	0.707	0.451	0.664	0.327	0.469	0.540
SP3-default	0.697	0.695	0.529	0.453	0.724	0.325
SP3-costs	0.707	0.627	0.556	0.270	0.655	0.338
none	0.753	0.389	0.842	0.302	0.208	0.910
SP2-min+int	0.809	0.373	0.665	0.471	0.287	0.823
SP2-int+maj	0.817	0.249	0.696	0.450	0.215	0.884
SP3-default	0.775	0.319	0.771	0.441	0.293	0.818
${ m SP3-costs}$	0.828	0.284	0.711	0.468	0.323	0.778
none	0.490	0.825	0.496	0.256	0.845	0.454
SP2-min+int	0.691	0.816	0.247	0.443	0.853	0.183
SP2-int+maj	0.790	0.615	0.447	0.594	0.737	0.478
SP3-default	0.659	0.760	0.289	0.509	0.860	0.220
SP3-costs	0.703	0.700	0.316	0.506	0.855	0.224
	Method none SP2-min+int SP3-default SP3-default SP3-costs none SP2-min+int SP3-default SP3-default SP3-costs none SP2-min+int SP3-default SP3-costs none SP2-min+int SP3-default SP3-costs none SP2-min+int SP3-default SP3-costs	$\begin{array}{c c} \mbox{Method} & c_{min} \\ \mbox{none} & 0.725 \\ \mbox{SP2-min+int} & 0.800 \\ \mbox{SP2-int+maj} & 0.837 \\ \mbox{SP3-default} & 0.791 \\ \mbox{SP3-default} & 0.791 \\ \mbox{SP3-costs} & 0.833 \\ \mbox{none} & 0.386 \\ \mbox{SP2-min+int} & 0.511 \\ \mbox{SP3-default} & 0.471 \\ \mbox{SP3-costs} & 0.555 \\ \mbox{none} & 0.687 \\ \mbox{SP2-min+int} & 0.657 \\ \mbox{SP2-min+int} & 0.657 \\ \mbox{SP2-min+int} & 0.657 \\ \mbox{SP2-min+int} & 0.677 \\ \mbox{SP3-default} & 0.707 \\ \mbox{SP3-costs} & 0.707 \\ \mbox{SP3-costs} & 0.828 \\ \mbox{none} & 0.490 \\ \mbox{SP2-min+int} & 0.691 \\ \mbox{SP2-min+int} & 0.691 \\ \mbox{SP3-default} & 0.703 \\ \mbox{SP3-default} & 0.703 \\ \mbox{SP3-default} & 0.703 \\ \mbox{SP3-default} & 0.659 \\ \mbox{SP3-costs} & 0.703 \\ \end{tabular}$	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$\begin{array}{c c c c c c c c c c c c c c c c c c c $

Table 2. TPRs for specific decision classes and clinical data sets

## References

 Wilk, S., Stefanowski, J., Wojciechowski, S., Farion, K.J., Michalowski, W.: Application of preprocessing methods to imbalanced clinical data: An experimental study. In Piętka, E., Badura, P., Kawa, J., Wieclawek, W., eds.: Proc. of the 5th International Conference, ITIB 2016, Springer (2016) 503-515