Criteria for choosing the number of dimensions in a principal component analysis: An empirical assessment

Renata B. Silva[1], Daniel de Oliveira[2], Davi P. Santos[3], Lucio F. D. Santos[4], Rodrigo E. Wilson[1], and Marcos Bedo[1]

[1]Fluminense Federal University (INFES/UFF), Brazil

{renatabarbosa, rodrigoerthal, marcosbedo}@id.uff.br

[2]Fluminense Federal University (IC/UFF), Brazil

danielcmo@ic.uff.br

[3]University of São Paulo (ICMC/USP), Brazil

davips@icmc.usp.br

[4]Federal Institute of North of Minas Gerais (IFNMG), Brazil

lucio.santos@ifnmg.edu.br

Abstract. *Principal component analysis (PCA) is an efficient model for the optimization problem of finding d' axes of a subspace $\mathbb{R}^{d'} \subseteq \mathbb{R}^d$ so that the mean squared distances from a given set $\mathcal{R}$ of points to the axes are minimal. Despite being steadily employed since 1901 in different scenarios, e.g., mechanics, PCA has become an important link in machine learning chained tasks, such as feature learning and AutoML designs. A frequent yet open issue that arises from supervised-based problems is how many PCA axes are required for the performance of machine learning constructs to be tuned. Accordingly, we investigate the behavior of six independent and uncoupled criteria for estimating the number of PCA axes, namely Scree-Plot %, Scree Plot Gap, Kaiser-Guttman, Broken-Stick, ρ-Score, and 2D. In total, we evaluate the performance of those approaches in 20 high dimensional datasets by using (i) four different classifiers, and (ii) a hypothesis test upon the reported F-Measures. Results indicate Broken-Stick and Scree-Plot % criteria consistently outperformed the competitors regarding supervised-based tasks, whereas estimators Kaiser-Guttman and Scree-Plot Gap delivered poor performances in the same scenarios.*

1. Introduction

Principal component analysis (PCA) is a widely adopted model for dimensionality reduction[1], a pre-processing step related to machine learning tasks [Pearson 1901, Aggarwal 2015]. Such a step is particularly relevant for supervised-driven problems, in which the *curse of dimensionality* [Pestov 2008] may disrupt the learning bias of certain classifiers, *e.g.*, Naïve-Bayes (NB), Instance-based Learning (IbL), Decision-Tree (DT), and Multi-Layer Perceptron (MLP), as well severely degraded their computational performance [Aggarwal 2015, James et al. 2013]. Formally, given a dataset $\mathcal{R} \subset \mathbb{R}^d$, PCA

[1]The most relevant *dimensions* for a particular set of points are the most prominent data *features*. Accordingly, we use the terms dimensions and features alternately.

enables finding the d' orthogonal axes of a subspace $\mathbb{R}^{d'} \subseteq \mathbb{R}^d$ so that the mean squared distances from elements in $\mathcal{R}$ to the axes are minimal.

A common yet open issue that arises in practice is distinguishing *relevant* and *non-relevant* axes so that data are reduced to a proper subspace [Pestov 2008, Aggarwal 2015]. Unlike previous approaches that investigate the relationship between d' and co-variance patterns within artificial data [Jackson 1993, Neto et al. 2005], we focus on examining distinct criteria for choosing the number of PCA axes whose performance is assessed by different classifiers. Figure 1 highlights the challenges of exhaustively estimating number d' in supervised UCI dataset `WINE`[2] regarding two wrapped classifiers: *(i)* labeling performance is not monotonic with d', and *(ii)* individual maxima are overfitting-prone.

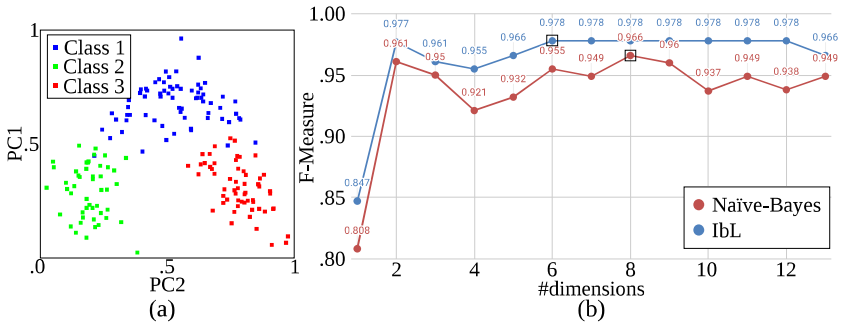


Figure 1. PCA reduction and labeling of `WINE` dataset.

In this study, we investigate the behavior of six global, distinct, and classifier-unwrapped criteria for choosing the number of dimensions in PCA reductions for labeling problems, namely *(i)* graphical-based estimators Scree-Plot %, Scree Plot Gap, and 2D; *(ii)* statistical-based indicators Broken-Stick, and Kaiser-Guttman; and *(iii)* intrinsic dimension-based criterion ρ-Score. We compared those criteria in the labeling of 20 datasets with four classifiers (NB, IbL, DT, and MLP), and results indicate estimators Broken-Stick and Scree-Plot % surpassed the competitors, while indicators Kaiser-Guttman and Scree-Plot Gap performed modestly. Such outcomes provide indications to devise the tuning of PCA-based pieces within AutoML designs.

This remainder of this paper is organized as follows. Section 2 discusses the estimators for the number of PCA dimensions, while Section 3 describes the material and methods. Sections 4 and 5 provide the experimental comparison and conclude the study.

2. Preliminaries

PCA axes can be found as the uncorrelated coefficients calculated from discrete data features. In a nutshell, reducing a dataset $\mathcal{R} \subset \mathbb{R}^d$ by PCA into a d'-dimensional representation, $d' \leq d$, is a sequence of six sequential steps, namely: *(i)* scale each $\mathcal{R}$ feature to the $[0, 1]$ interval, *(ii)* calculate means $\mu_{i,i\in[1,d]}$ for every $\mathcal{R}$ feature, *(iii)* subtract means $\mu_{i,i\in[1,d]}$ from each $\mathcal{R}$ element, *(iv)* calculate co-variance matrix $\mathcal{C}_{d\times d}$ from $\mathcal{R}$ entries, *(v)* obtain both $\mathcal{C}$ eigenvalues and eigenvectors, and *(vi)* calculate the cross product of $\mathcal{R}$ entries and d' eingenvectors related to the d' *highest and descending-sorted* eigenvalues.

[2]Data links at `github.com/Renata-Barbosa/cpca`

Existing criteria for estimating the d' value can be divided into *(i)* graphical, *(ii)* statistical, and *(iii)* intrinsic dimension-based approaches [Neto et al. 2005, Pestov 2008]. Representative approaches of the first group include the following rules:

Scree-Plot % (*SP-%*). The estimator assumes a few eigenvalues concentrate the largest part of data variance, and calculates d' so that a *percentage* of variance is kept. Figure 2(a) illustrates the *SP-%* rule for covering 70% of the area under the `WINE` eigenvalues.

Scree-Plot Gap (*SP-G*). The criterion uses a greedy search for finding the largest variance difference between two consecutive pairs of scale-normalized eigenvalues [Zhu and Ghodsi 2006]. The intersection of both pairs is returned as d'. Figure 2(b) shows the search on `WINE` where line segments represent the scaled differences.

Plane visualization (*2D*). This approach is a baseline rule that sets $d' = 2$ so that data can be visualized in a Euclidean plane. Figure 1(a) shows that PCA reduction for set `WINE`.

Statistical criteria examine whether a set of eigenvalues is larger than an expected value drawn from a known data distribution. Approaches of that category include:

Kaiser-Guttman (*KG*). This rule retrieves every eigenvalue greater than 1.0 aiming at retaining shared data variance [Guttman 1954, Neto et al. 2005].

Broken-Stick (*B-St*). If joint variance is randomly distributed within d axes, then eigenvalues are supposed to follow the *Broken-Stick distribution* [Legendre and Legendre 2012]. For a set of eigenvalues e_i, Broken-Stick entries are given by $bs_i = \left(\frac{\sum_{j=1}^{d} e_j}{d}\right) / \left(\sum_{j=1}^{d-i+1} \frac{1}{d+1-j}\right)$. *B-St* returns the last scaled eigenvalue that deviates from the random distribution, *i.e.*, $d' = k \mid \left(e_{k+1}/\sum_{i=1}^{d} e_i \leq bs_{k+1} \wedge bs_1 \leq e_1/\sum_{i=1}^{d} e_i\right) \vee \left(e_{k+1}/\sum_{i=1}^{d} e_i \geq bs_{k+1} \wedge bs_1 \geq e_1/\sum_{i=1}^{d} e_i\right)$. Figure 2(c) depicts the *B-St* rule for set `WINE`.

Finally, intrinsic dimension-based criteria estimate possible correlations embedded within data features. A stable rule for obtaining such a value is as follows.

Rho-Score (ρ-*sct*). This estimator approximates data intrinsic dimension by using its distance distribution [Pestov 2008]. Formally, let $\mathcal{R} \subset \mathbb{R}^d$ be a dataset and $\delta : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ be the Euclidean distance, the ρ-*sct* criterion estimates d' as in Eq. 1.

$$d' = \left\lceil \frac{\mu_{\mathcal{R}}^2}{2\sigma_{\mathcal{R}}^2} \right\rceil ; \mu_{\mathcal{R}} = \frac{1}{2|\mathcal{R}|} \sum_{r_i, r_j \in \mathcal{R}} \delta(r_i, r_j); \sigma_{\mathcal{R}} = \sqrt{\frac{1}{|\mathcal{R}| - 1} \sum_{r_i, r_j \in \mathcal{R}} \left(\frac{\delta(r_i, r_j) - \mu_{\mathcal{R}}}{2}\right)^2} \quad (1)$$

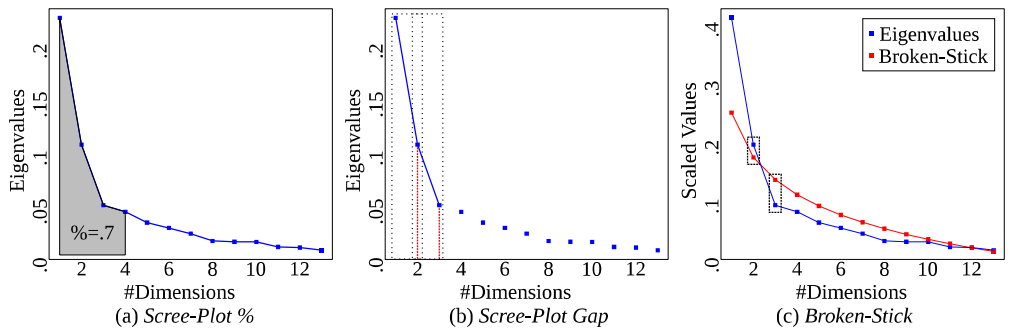


Figure 2. Choosing the number of dimensions d' in a PCA reduction of `WINE`.

3. Material and Methods

Estimation criteria implementation. We implement PCA, and the six reviewed estimation criteria from scratch in R^2 by taking advantage of the spectral decomposition *eigen* routine available at R-core package. Graphical-based criteria are constructed upon sorted eigenvalues produced by the *eigen* routine, as well as the Kaiser-Guttman rule. An equi-width histogram H with d positions is employed as the underlying data structure for the *B-St* implementation. Each histogram position corresponds to a theoretical descending-sorted eigenvalue that follows a scale-normalized Broken-Stick distribution. Accordingly, the *B-St* estimation is carried out by a linear search over H for finding the first position that deviates from the eigenvalues' distribution. Finally, we implemented the Euclidean distance, as well as both mean and variance Welford's one-pass calculation for efficiently obtaining the intrinsic dimension returned by the ρ*-sct* rule.

Non-parametric and pairwise hypothesis tests. We adopt pairwise tests for assessing which criteria are suitable for estimating the number of dimensions in PCA reductions of labeled sets. The two-tailored Wicoxon test is convenient for such a pairwise comparison because the null hypothesis is that results produced by two competing estimators are drawn from the same distribution, whereas the alternative hypothesis is that outputs are from different distributions [Wilcoxon 1992]. We model the results of PCA criteria as the quality values measured after a classifier 10-folds cross-validation, *e.g.*, F-Measure [Aggarwal 2015]. Therefore, a pair of criteria is compared by its paired list of label-driven results, being that pairwise list sorted by the absolute differences between the two juxtaposed outputs. An incremental *rank* ranging from one to the number of observations is assigned to each position of the sorted list. Ranks of positions where the second rule outperforms the first are multiplied by -1, which generates two rank groups aggregated by a *ranking sum*. Such a sum adjusted by the number of observations produces a z-score, which is employed in the analysis of Wilcoxon's null hypothesis.

4. Experiments

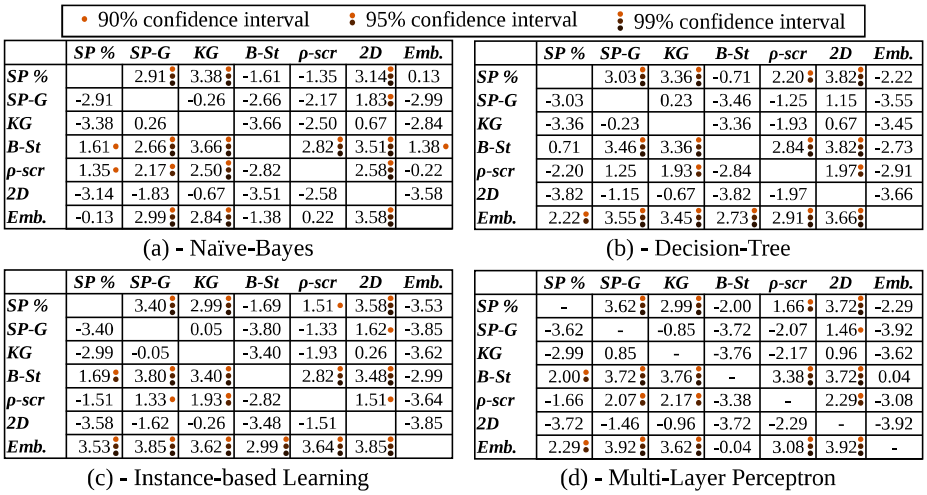
We evaluate the performance of criteria *SP-%*, *SP-G*, *2D*, *KG*, *B-St*, and ρ*-sct* in association with four different classifiers NB, IbL, DT, and MLP in 20 medium to high dimensional labeled datasets ($\mathcal{R}$) from UCI, MILD, GBDI and QTDU repositories[2]. In particular, we relied on both Weka v3.8.4 and R v3.6.1 to set up the classifiers as follows: *(i)* DT with binary splits and early pruning, *(ii)* MLP with one hidden layer of $\sqrt{d}$ neurons, and *(iii)* IbL with Euclidean distance and one neighbor. Evaluations were conducted in a KUbuntu machine 19.10 with an Intel $i5$ processor, 8GB RAM, and a 1TB disk.

We tuned the *SP-%* criterion by evaluating its performance with parameters $\% = \{.65, .7, .75\}$, being the setup *SP-%* $= .7$ the tuning with highest F-Measures, on average. Table 1 details the F-Measure reached by each criterion for every evaluation scenario, where $Emb.$ lines stand for data original dimensionality d. Results reinforce no monotonic relationship can be drawn from F-Measure and the number of dimensions (d').

Next, we compare the competing criteria through a set of pairwise Wilcoxon's tests grouped by classifiers. Figure 3 shows the z-scores obtained for each pairwise comparison (**line *vs.* column**), and highlights the cases in which the null hypothesis was re-

Table 1. F-Measures for data reduced by PCA with distinct criteria.

Crit.	d'	F-Measure NB	DT	IbL	MLP	$\mathcal{R}$	d'	F-Measure NB	DT	IbL	MLP	$\mathcal{R}$	d'	F-Measure NB	DT	IbL	MLP	$\mathcal{R}$
SP-%	26	.855	.858	.965	.936	MNIST (784)	15	.668	.788	.795	.870	CNAE2 (128)	19	.829	.711	.839	.870	ISOLET (617)
SP-G	2	.373	.424	.379	.430		3	.468	.689	.715	.689		2	.322	.276	.257	.311	
KG	13	.799	.852	.947	.880		1	.326	.446	.408	.416		6	.654	.608	.666	.698	
B-St	47	.870	.854	.972	.951		7	.557	.747	.780	.797		28	.871	.722	.865	.912	
ρ-sct	27	.856	.858	.973	.934		3	.468	.689	.715	.689		17	.815	.698	.829	.859	
2D	2	.373	.424	.379	.430		2	.420	.573	.586	.559		2	.322	.276	.257	.311	
Emb.	784	.686	.860	.965	.883		128	.743	.797	.830	.883		617	.841	.823	.896	.873	
SP-%	3	.321	.909	.901	.921	OZONE (72)	17	.715	.870	.879	.896	SPAM (57)	25	.504	.631	.670	.659	FOX (230)
SP-G	2	.321	.321	.885	.898		2	.765	.809	.794	.812		2	.543	.498	.572	.537	
KG	2	.321	.321	.885	.898		1	.610	.813	.768	.806		1	.505	.469	.558	.542	
B-St	3	.321	.909	.901	.921		2	.765	.809	.794	.812		28	.503	.626	.669	.658	
ρ-sct	4	.901	.921	.911	.921		2	.765	.809	.794	.812		1	.505	.469	.558	.542	
2D	2	.321	.321	.885	.898		2	.765	.809	.794	.812		2	.543	.498	.572	.537	
Emb.	72	.751	.922	.914	.914		57	.792	.931	.894	.901		230	.525	.704	.670	.641	
SP-%	571	.877	.886	.700	.763	GISETTE (5000)	14	.936	.875	.970	.957	MFEAT (649)	12	.743	.876	.873	.875	ULCER_SET (400)
SP-G	2	.696	.699	.624	.698		2	.596	.621	.554	.635		3	.761	.822	.803	.296	
KG	28	.892	.908	.965	.977		11	.929	.877	.965	.947		1	.181	.181	.725	.181	
B-St	517	.879	.885	.719	.971		26	.952	.871	.973	.973		23	.616	.875	.871	.893	
ρ-sct	68	.870	.888	.954	.976		23	.949	.870	.974	.970		1	.181	.181	.725	.181	
2D	2	.696	.699	.624	.698		2	.596	.621	.554	.635		2	.242	.271	.764	.227	
Emb.	5000	.916	.938	.958	.898		649	.954	.949	.978	.960		400	.280	.895	.882	.821	
SP-%	4	.724	.784	.755	.508	SMART (561)	37	.957	.938	.976	.879	SEMEION (264)	2	.507	.789	.792	.603	GAS_DRIFT (128)
SP-G	2	.341	.629	.577	.370		2	.316	.316	.838	.316		2	.507	.789	.792	.603	
KG	2	.341	.629	.577	.370		12	.948	.939	.971	.974		1	.174	.394	.302	.264	
B-St	13	.828	.855	.885	.899		27	.957	.932	.972	.982		4	.609	.957	.983	.828	
ρ-sct	4	.724	.784	.755	.508		65	.961	.935	.964	.963		1	.174	.394	.302	.264	
2D	2	.341	.629	.577	.370		2	.316	.316	.838	.316		2	.507	.789	.792	.603	
Emb.	561	.674	.930	.960	.783		264	.926	.943	.977	.975		128	.595	.978	.995	.938	
SP-%	13	.270	.272	.293	.100	MAMMOSET (1111)	35	.656	.907	.968	.928	UJIN (520)	29	.565	.685	.712	.744	TIGER (230)
SP-G	2	.069	.070	.253	.068		2	.200	.610	.615	.279		3	.571	.683	.664	.660	
KG	4	.081	.085	.250	.084		2	.200	.610	.615	.279		1	.474	.578	.555	.582	
B-St	36	.308	.269	.301	.121		34	.652	.905	.967	.923		27	.495	.698	.711	.743	
ρ-sct	12	.266	.112	.287	.095		17	.087	.182	.280	.178		1	.474	.578	.555	.582	
2D	2	.069	.070	.253	.068		2	.020	.610	.615	.279		2	.536	.647	.643	.661	
Emb.	1111	.104	.284	.310	.133		520	.508	.895	.973	.836		230	.608	.803	.719	.675	
SP-%	4	.410	.581	.667	.546	LIBRAS (90)	4	.508	.553	.706	.262	GRAMPACIAL (300)	11	.964	.783	.955	.965	PLANT (192)
SP-G	4	.410	.581	.667	.546		2	.262	.512	.611	.262		2	.366	.306	.322	.304	
KG	1	.181	.240	.355	.191		2	.262	.512	.611	.262		1	.073	.072	.073	.041	
B-St	8	.724	.665	.848	.747		11	.516	.572	.711	.516		14	.968	.789	.974	.975	
ρ-sct	7	.697	.638	.826	.727		4	.508	.553	.706	.262		9	.949	.786	.957	.961	
2D	2	.174	.346	.362	.278		2	.262	.512	.611	.262		2	.366	.306	.322	.304	
Emb.	90	.631	.694	.856	.801		300	.218	.759	.734	.770		192	.970	.743	.987	.993	
SP-%	243	.577	.660	.500	.573	MADELON (400)	122	.912	.899	.925	.923	QSAR (1024)	4	.955	.949	.921	.949	WINE (13)
SP-G	2	.606	.615	.580	.573		3	.319	.882	.881	.319		2	.977	.949	.961	.961	
KG	1	.599	.607	.544	.579		9	.891	.897	.907	.898		1	.966	.938	.949	.972	
B-St	3	.634	.649	.670	.625		54	.909	.905	.924	.917		2	.977	.949	.961	.961	
ρ-sct	404	.569	.637	.503	.547		16	.908	.912	.919	.915		1	.966	.938	.949	.972	
2D	2	.606	.615	.580	.573		2	.319	.319	.867	.319		2	.977	.949	.961	.961	
Emb.	400	.595	.390	.542	.657		1024	.837	.920	.930	.922		13	.966	.938	.949	.972	



	SP %	*SP-G*	*KG*	*B-St*	*ρ-scr*	*2D*	*Emb.*
SP %		2.91⋮	3.38⋮	-1.61	-1.35	3.14⋮	0.13
SP-G	-2.91		-0.26	-2.66	-2.17	1.83⋮	-2.99
KG	-3.38	0.26		-3.66	-2.50	0.67	-2.84
B-St	1.61•	2.66⋮	3.66⋮		2.82⋮	3.51⋮	1.38•
ρ-scr	1.35•	2.17⋮	2.50⋮	-2.82		2.58⋮	-0.22
2D	-3.14	-1.83	-0.67	-3.51	-2.58		-3.58
Emb.	-0.13	2.99⋮	2.84⋮	-1.38	0.22	3.58⋮	

(a) - Naïve-Bayes

	SP %	*SP-G*	*KG*	*B-St*	*ρ-scr*	*2D*	*Emb.*
SP %		3.03⋮	3.36⋮	-0.71	2.20⋮	3.82⋮	-2.22
SP-G	-3.03		0.23	-3.46	-1.25	1.15	-3.55
KG	-3.36	-0.23		-3.36	-1.93	0.67	-3.45
B-St	0.71	3.46⋮	3.36⋮		2.84⋮	3.82⋮	-2.73
ρ-scr	-2.20	1.25	1.93⋮	-2.84		1.97•	-2.91
2D	-3.82	-1.15	-0.67	-3.82	-1.97		-3.66
Emb.	2.22⋮	3.55⋮	3.45⋮	2.73⋮	2.91⋮	3.66⋮	

(b) - Decision-Tree

	SP %	*SP-G*	*KG*	*B-St*	*ρ-scr*	*2D*	*Emb.*
SP %		3.40⋮	2.99⋮	-1.69	1.51•	3.58⋮	-3.53
SP-G	-3.40		0.05	-3.80	-1.33	1.62•	-3.85
KG	-2.99	-0.05		-3.40	-1.93	0.26	-3.62
B-St	1.69⋮	3.80⋮	3.40⋮		2.82⋮	3.48⋮	-2.99
ρ-scr	-1.51	1.33•	1.93⋮	-2.82		1.51•	-3.64
2D	-3.58	-1.62	-0.26	-3.48	-1.51		-3.85
Emb.	3.53⋮	3.85⋮	3.62⋮	2.99⋮	3.64⋮	3.85⋮	

(c) - Instance-based Learning

	SP %	*SP-G*	*KG*	*B-St*	*ρ-scr*	*2D*	*Emb.*
SP %	-	3.62⋮	2.99⋮	-2.00	1.66⋮	3.72⋮	-2.29
SP-G	-3.62	-	-0.85	-3.72	-2.07	1.46•	-3.92
KG	-2.99	0.85	-	-3.76	-2.17	0.96	-3.62
B-St	2.00⋮	3.72⋮	3.76⋮	-	3.38⋮	3.72⋮	0.04
ρ-scr	-1.66	2.07⋮	2.17⋮	-3.38	-	2.29⋮	-3.08
2D	-3.72	-1.46	-0.96	-3.72	-2.29	-	-3.92
Emb.	2.29⋮	3.92⋮	3.62⋮	-0.04	3.08⋮	3.92⋮	-

(d) - Multi-Layer Perceptron

Figure 3. Pairwise Wilcoxon's tests (lines *vs.* columns) grouped by classifiers.

jected and the corresponding level of confidence. Results indicate *B-St* rule outperformed every competitor (including the original data dimensionality) for a confidence level of at

least 90% regarding NB and MLP classification, whereas *SP-%* also outperformed estimators *SP-G*, *KG*, and *2D*. A similar result was observed for classifiers IbL and DT in which *B-St* dominated *SP-G*, *KG*, ρ*-sct*, and *2D* within significance levels. In those scenarios, *SP-%* also outperformed *SP-G*, *KG*, *2D*, and ρ*-sct* with statistical significance. Such findings pinpoint *both criteria B-St and SP-% are suitable rules* for choosing the number of dimensions in a PCA reduction. Notice, however, *B-St is a parameterless estimator that may be preferable to SP-%* whenever adjusting variance area $\%$ is unpractical.

A counterpart discovery is estimators *KG and SP-G performed poorly* in comparison to other criteria regarding labeling-driven tasks. In particular, *KG* did not outperform any competitor, including *SP-G*, and baseline rules *2D* and $Emb.$. Lastly, ρ*-sct* criterion showed an intermediary performance, which indicates there may be a relationship between the intrinsic dimension and the number of PCA axes, but not strong as a correlation to be spotted by the experimental supervised evaluation we carried out.

5. Conclusions and Future Work

This study has discussed global criteria for finding the number of dimensions in a PCA reduction of labeled datasets. Since estimators are based on distinct theoretical grounds, we examine their performance from an experimental perspective regarding the biases of different classifiers. Results indicate *B-St* and *SP-%* are suitable rules for estimating the number of PCA axes, whereas *KG* and *SP-G* shall be avoided in the reductions. Such outcomes enable devising the tuning of PCA-based pieces of AutoMLs in future work.

Acknowledgments. The study was supported by FAPERJ and CEPID-CeMEAI/FAPESP (Grants 2013/07375-0 – 2019/01735-0).

References

Aggarwal, C. (2015). *Data mining: The textbook*. Springer.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2):149–161.

Jackson, D. A. (1993). Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology*, 74(8):2204–2214.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to Statistical Learning*, volume 112. Springer.

Legendre, P. and Legendre, L. F. (2012). *Numerical Ecology*. Elsevier.

Neto, P., Jackson, D., and Somers, K. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *C. Stat.*, 49(4):974–997.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Phil. Magazine and J. of Science*, 2(11):559–572.

Pestov, V. (2008). An axiomatic approach to intrinsic dimension of a dataset. *Neural Networks*, 21(2-3):204–213.

Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in statistics*, pages 196–202. Springer.

Zhu, M. and Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Stat. & Data Analysis*, 51(2):918–930.