Subspace Determination Through Local Intrinsic Dimensional Decomposition

Becker, Ruben; Hafnaoui, Imane; Houle, Michael E.; Li, Pan; Zimek, Arthur

Published in:
Similarity Search and Applications - 12th International Conference, SISAP 2019, Newark, NJ, USA, October 2-4, 2019, Proceedings

DOI:
10.1007/978-3-030-32047-8_25

Publication date:
2019

Document version:
Accepted manuscript

Citation for pulished version (APA):
Becker, R., Hafnaoui, I., Houle, M. E., Li, P., & Zimek, A. (2019). Subspace Determination Through Local Intrinsic Dimensional Decomposition. In G. Amato, C. Gennaro, V. Oria, & M. Radovanovic (Eds.), *Similarity Search and Applications - 12th International Conference, SISAP 2019, Newark, NJ, USA, October 2-4, 2019, Proceedings* (pp. 281-289). Springer. Lecture Notes in Computer Science Vol. 11807 https://doi.org/10.1007/978-3-030-32047-8_25

Subspace Determination through Local Intrinsic Dimensional Decomposition

Ruben Becker[1], Imane Hafnaoui[2], Michael E. Houle[3], Pan Li[4], and Arthur Zimek[5]

[1] Gran Sasso Science Institute, L'Aquila, Italy, ruben.becker@gssi.it
[2] Ecole Polytechnique de Montréal, Montréal, Canada, imane.hafnaoui@polymtl.ca
[3] National Institute of Informatics, Tokyo, Japan, meh@nii.ac.jp
[4] Saarland Informatics Campus, Saarbrücken, Germany, panli1989@gmail.com
[5] Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark, zimek@imada.sdu.dk

Abstract. Axis-aligned subspace clustering generally entails searching through enormous numbers of subspaces (feature combinations) and evaluation of cluster quality within each subspace. In this paper, we tackle the problem of identifying subsets of features with the most significant contribution to the formation of the local neighborhood surrounding a given data point. For each point, the recently-proposed Local Intrinsic Dimension (LID) model is used in identifying the axis directions along which features have the greatest local discriminability, or equivalently, the fewest number of components of LID that capture the local complexity of the data. In this paper, we develop an estimator of LID along axis projections, and provide preliminary evidence that this LID decomposition can indicate axis-aligned data subspaces that support the formation of clusters.

Keywords: Intrinsic dimensionality · Estimation · Subspace

1 Introduction

In data mining, machine learning, and other areas of AI, we are often faced with datasets that contain many more attributes than needed, or that can even be helpful for tasks such as clustering or classification. Problems associated with such high dimensional data are for example the concentration effect of distances [6, 9] or irrelevant features [13, 26]. For clustering [17, 23] and outlier detection [26], researchers have made use of various techniques to identify relevant subspaces, as defined by subsets of features that are informative for a particular task. Examples of how relevant subspaces can be determined for individual clusters or outliers include local density estimation in a systematic search through candidate subspaces, or the adaptation of distance measures based on the distribution within local neighborhoods. For sufficiently tight local neighborhoods, the underlying local data manifold can be regarded as approaching a linear form [21], an assumption that further justifies the determination of locally relevant features for subspace determination.

In this paper, we present a novel technique for the identification of subsets of features with the most significant contribution to the formation of the local neighborhood surrounding a given data point, using the recently introduced Local Intrinsic Dimensionality (LID) [10, 11] model. LID is a distributional form of intrinsic dimensional modeling in which the volume of a ball of radius r is taken to be the probability measure associated with its interior, denoted by $F(r)$. The function F can be regarded as the cumulative distribution function (cdf) of an underlying distribution of distances. Theoretical properties of LID in multivariate analysis have been studied recently [12]. LID has also seen practical applications in such areas as similarity search [7], dependency analysis [20], and deep learning [18, 19].

To make use of the LID model to identify locally-discriminative features, we develop an estimator of LID decomposed along axis projections that compensates for the bias introduced during projection. We also provide preliminary experimental evidence that LID decomposition can indicate axis-aligned data subspaces that support the formation of clusters, by implementing a simple two-stage technique whereby points are first assigned to relevant subspaces, and then clustered. As the relevant features can be different for each cluster, feature relevance is assessed cluster-wise or even point-wise (as the clusters are not known in advance). It is not our intent here to propose a complete subspace clustering strategy; rather the goal in this preliminary investigation is to provide some guidance as to how subspace identification could be done as an independent, initial step as part of a larger clustering strategy.

In Section 2, we give some preliminaries on intrinsic dimensionality, before discussing LID decomposition and its estimation. In Section 3, to illustrate how LID decomposition could be used within subspace clustering, we propose as an example a simple method using LID to determine eligible subspaces within which DBSCAN is used for clustering. In this preliminary version, only a brief summary of the experimentation is given; more details can be found in [4]. We conclude the paper in Section 4.

Preliminaries on ID. Let $X \in \mathbb{R}^m$ be an m-variate random variable, let $F : \mathbb{R}^m \to \mathbb{R}$ be its joint probability distribution, and let $\|\cdot\|$ denote an arbitrary norm. The ID of F at a non-zero point x is defined as follows.

Definition 1 ([12]). *Let $x \in \mathbb{R}^m_{\neq 0}$ such that $F(x) \neq 0$. Assume that the partial derivatives $\frac{\partial f}{\partial x_i}(x)$ at x exist for all $i \in [m] = \{1, \ldots, m\}$, the* ID *of F at x is defined as $\mathrm{ID}_F(x) := x^T \nabla F(x)/F(x)$.*

It is well-known that, under suitable mild continuity assumptions, the ID of F at x is equivalent to both the indiscriminability and the intrinsic dimensionality of F at x, see [12, Theorem 1]. Local intrinsic dimensionalities have also been shown to satisfy the following useful decomposition rule.

Theorem 1 ([12]). *Let $x \in \mathbb{R}^m_{\neq 0}$ and let $I \subseteq \mathbb{R}$ with $0 \in I$ be an open interval such that F is non-zero and its partial derivatives exist and are continuous at $(1+\varepsilon)x$ for all $\varepsilon \in I$. Assume that $x_i \neq 0$ for each $i \in [m]$. Then $\mathrm{ID}_F(x) = \sum_{i=1}^m \mathrm{ID}_{F_{i,x}}(x_i)$, where $F_{i,x}(t) := F(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_m)$ for $i \in [m]$.*

2 Decomposed LID Estimation

Definition and Properties. We now define $N_\delta := \{x \in \mathbb{R}^m : 0 < \|x\|_\infty < \delta\}$, and assume that F is non-zero and that its partial derivatives exist and are continuous at every $x \in N_\delta$. Then, for every $x \in N_\delta$, there is an interval I with $0 \in I$ such that F is non-zero and its partial derivatives exist and are continuous at $(1+\varepsilon)x$ for every $\varepsilon \in I$. Following [12], we define $\mathrm{ID}^*_F := \lim_{x\to 0, \|x\|_\infty \le \delta} \mathrm{ID}_F(x)$ as the ***local intrinsic dimensionality of*** F.

Definition 2. *Let I_δ be the 'hollow' open interval $(-\delta,\delta) \setminus \{0\}$. For $x \in N_\delta$, we define the functions $F_{i,x} : I_\delta \to \mathbb{R}$ and $g_i : I_\delta \times {I_\delta}^{m-1} \to \mathbb{R}$ as $F_{i,x}(t) := F(x_1,\dots,x_{i-1},t,x_{i+1},\dots,x_m)$ and $g_i(t,x_{-i}) := t \cdot F'_{i,x}(t)/F_{i,x}(t)$, where $x_{-i} = (x_1,\dots,x_{i-1},x_{i+1},\dots,x_m) \in {I_\delta}^{m-1}$ for some $x \in N_\delta$.*

Using the Moore-Osgood theorem to interchange the order of limits, we obtain a decomposition rule for LID. For the precise statement of the Moore-Osgood theorem, see for example [16].

Theorem 2. *Assume that for every $i \in [m]$, it holds that (1) $\lim_{t\to 0} g_i(t,y)$ exists for every $y \in {I_\delta}^{m-1}$, (2) $\lim_{y\to 0} g_i(t,y)$ exists for every $t \in I_\delta$, and (3) at least one of the two limits exists uniformly. Then the limits $\mathrm{ID}^*_{F,i} := \lim_{x\to 0} x_i \cdot F'_{i,x}(x_i)/F_{i,x}(x_i)$ exist for all $i \in [m]$, and thus*

$$\mathrm{ID}^*_F = \sum_{i=1}^{m} \mathrm{ID}^*_{F,i} = \sum_{i=1}^{m} \lim_{x\to 0} \frac{x_i \cdot F'_{i,x}(x_i)}{F_{i,x}(x_i)} = \sum_{i=1}^{m} \lim_{y\to 0}\lim_{t\to 0} g_i(t,y). \tag{1}$$

We refer to $\mathrm{ID}^*_{F,i}$ as the *local intrinsic dimensionality of F in direction i.*

Estimating $\mathrm{ID}^*_{F,i}$. Now let $\phi : \mathbb{R} \to \mathbb{R}$ be a univariate funcction and assume that $\mathrm{ID}^*_\phi := \mathrm{ID}_\phi(0) = \lim_{t\to 0} t \cdot \phi'(t)/\phi(t)$ exists. We note that Theorems 2 and 3 in [11] yields that, as w approaches 0, it holds that $\phi(t) \approx \phi(w) \cdot (t/w)^{\mathrm{ID}^*_\phi}$. Moreover, differentiating this quantity yields $(\phi(w)/w) \cdot \mathrm{ID}^*_\phi \cdot (t/w)^{\mathrm{ID}^*_\phi - 1}$ as an approximation of $\phi'(t)$. We now apply this observation to the estimation of $\mathrm{ID}^*_{F,i}$ for some $i \in [m]$. Let us fix some $x \in \mathbb{R}^m_{\neq 0}$ and let us denote $\mathrm{ID}^*_i := \mathrm{ID}^*_{F_{i,x}}$ for $i \in [m]$. Given $p^{(1)},\dots,p^{(k)} \in \mathbb{R}^m$ following the joint distribution F, we are now in a position to state the log-likelihood function for the parameter ID^*_i under the observations $p^{(1)},\dots,p^{(k)}$. Assume that we associate a weight $\omega(p_i^{(j)})$ to the projection $p_i^{(j)}$ of each observation $p^{(j)}$ — for the standard unweighted case of the log-likelihood function, all weights are set to 1. We may regard these weights as assigning a-priori likelihoods to the observations, by which an individual observation $p_i^{(j)}$ is accounted as having occurred $\omega(p_i^{(j)})$-many times. The weighted log-likelihood function can then be derived as

$$\mathcal{L}(\mathrm{ID}^*_i : p^{(1)},\dots,p^{(k)}) = \sum_{j=1}^{k} \omega(p_i^{(j)}) \cdot \log\Big(\frac{F_{i,x}(w)}{w} \cdot \mathrm{ID}^*_i \cdot \big(\frac{|p_i^{(j)}|}{w}\big)^{\mathrm{ID}^*_i - 1}\Big).$$

We are now interested in the parameter ID_i^* that maximizes $\mathcal{L}(\mathrm{ID}_i^* : p^{(1)}, \dots, p^{(k)})$. For this purpose, we form its derivative w.r.t. ID_i^* and set it to zero. A straightforward derivation shows that the likelihood is maximized at

$$\widehat{\mathrm{ID}_i^*} = \Big(- \frac{1}{\sum_{j=1}^k \omega(p_i^{(j)})} \sum_{j=1}^k \omega(p_i^{(j)}) \log\big(\frac{|p_i^{(j)}|}{w}\big)\Big)^{-1}, \tag{2}$$

which has the form of a weighted variant of the Hill estimator with threshold w.

Note that we have now developed an estimator for ID_i^*. Assuming however, that for a reference point $x_0 \in \mathbb{R}^m$, the considered neighborhood from which the points $p^{(1)}, \dots, p^{(k)}$ are chosen is sufficiently small, it is reasonable to use the same estimator for $\mathrm{ID}_{F,i}^*$ as well, as the outer limit in (1) can be neglected.

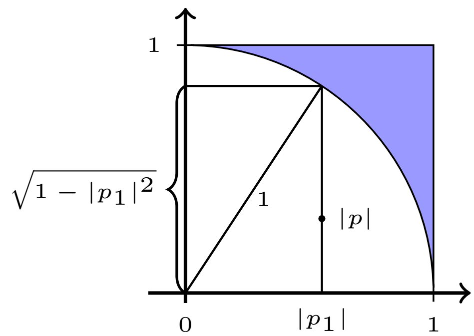

Fig. 1: When considering a circular neighborhood, points p with projections $|p_1|$ close to one are much less likely than points with small projections, since the blue region is not accounted for. Such a neighborhood can however still be employed, by associating a weight $\omega(p_1)$ with p that is proportional to 1 over the length of the line segment that contains all points with this projection $|p_1|$.

Neighborhood Weighting. In the previous subsection, we have developed an estimator for $\mathrm{ID}_{F,i}^*$; however, we have not yet stated how to determine a neighborhood for x_0. This turns out to be a delicate question, for which the use of observation weighting will become essential.

Note that the estimator for $\mathrm{ID}_{F,i}^*$ that we developed above assumes that neighborhood points $p^{(j)}$ with projections $|p_j|$ stem from the interval $[0, w]$. If we pick a 'box neighborhood' of x_0 consisting of the k closest points to x_0 with respect to the L_∞ norm (defined as $\|v\|_\infty := \max\{|v_i| : i \in [m]\}$ for $v \in \mathbb{R}^m$), the points p with projections $|p_i|$ close to zero are equally likely to be neighbors as points with projections close to one. This is however, not the case if we pick the neighborhood as the k closest points with respect to the Euclidean norm. In this case, points p with projections $|p_i|$ close to zero will be much more likely to be neighbors than points with $|p_i|$ close to one. As the Euclidean norm is however much more common in practical applications, due to its rotational invariance, we would still like to be able to handle this situation as well. In order to compensate for the bias that results from the fact that points with large projections are less likely than points with small projection, we will use the weighting scheme introduced in the previous subsection. When estimating $\mathrm{ID}_{F,i}^*$, an observation p with projection $|p_i|$ must be weighted according to the ratio of the volume of the $m-1$-dimensional sphere with radius $(1 - |p_i|^2)^{1/2}$ on the one hand, and the volume of its bounding hypercube on the other. This leads to the definition of weights $\omega(p_i) := 1/(1-|p_i|^2)^{(m-1)/2}$ for the case of the Euclidean norm. See Figure 1 for an illustration of the 2D-case.

Verification of $\mathrm{ID}^*_F = \sum_{i=1}^m \mathrm{ID}^*_{F,i}$. In this paragraph we report on an experiment that aims at verifying the equation $\mathrm{ID}^*_F = \sum_{i=1}^m \mathrm{ID}^*_{F,i}$ from Theorem 2 for the case of a uniform distribution in a space equipped with the Euclidean distance metric. For the purpose of estimating ID^*_F, we use the MLE (Hill) estimator proposed in [3] (`hill_distances`). We compare its output value $\widehat{\mathrm{ID}^*_F}$ on a hyperspherical neighborhood of radius 1 with the sum $\sum_{i=1}^m \widehat{\mathrm{ID}^*_i}$, where we consider two different ways of obtaining the estimates $\widehat{\mathrm{ID}^*_i}$. In the first case (`sum_hill_projections`), we pick a (unit-)hypercubical neighborhood, while in the second case (`sum_w_hill_projections`), we use the weighted estimator for the hyperspherical neighborhood compensating for bias using weights. In our experiment, we create neighborhoods of 100 points for dimensions $m = 2, 4, 8, \ldots, 1024$. Note that in this example of a uniform distribution in m dimensions, the true LID value is m. The experiments show that the two decomposition-based estimators, when summed over all components, do match the total intrinsic dimensionality m, as does the MLE estimator, see Figure 2.

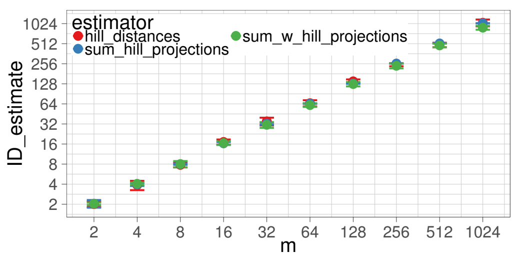

Fig. 2: Results for the three estimators. Errorbars denote 95% confidence intervals. Every measurement is an average of 5 runs.

3 Subspace Clustering Based on LID Decomposition

We now consider some of the issues surrounding the use of LID-decomposition ranking to support subspace clustering. It is not our intent here to propose a single full subspace clustering strategy; rather, the goal is to provide some guidance as to how subspace identification could be done as an independent, preliminary step as part of a larger clustering strategy.

We rely on the LID decomposition to determine relevant attributes for the cluster to which the neighborhood of q belongs. The subspace dimensionality of a point q is determined by searching for attributes with low ID estimates. A common way of doing this is by locating a gap in the sequence of LID estimates that best separates relevant attributes from irrelevant ones, much in the same way as a projective basis is found in PCA decompositions through gaps in the sequence of eigenvalues or variances. We track the relative difference in ID from attributes with low ID to high ID and fix the cut-off that determines the subspace dimensionality at the attribute that exhibits the highest relative difference.

Cluster Membership. To better define the local subspace preference vectors, we propose an additional refinement step. We use a sample of data points $\tilde{\mathcal{X}}$ to build a profile from their subspace preference vectors $\mathcal{P} = \{\mathcal{S}(x) \mid x \in \tilde{\mathcal{X}}\}$. The local subspace preference is refined by determining the membership of points $\mathcal{M}$ to the collected subspace profiles. Given the ordered attributes vector $\mathcal{O}(q)$, $\mathcal{M}(q)$ is selected as the subspace whose attributes are present in the first elements of $\mathcal{O}(q)$. Inside a subspace, points with preference towards that subspace are clustered using a traditional algorithm such as DBSCAN [8].

Experimental Evaluation

Besides the recall, we rely on three other metrics that are widely used in the literature to measure the performance of clustering techniques, namely the Adjusted Rand-Index (ARI) [15], the Normalized Mutual Information (NMI) [24], and the Adjusted Mutual Information (AMI) [25].

Synthetic Data. We synthetically generated three datasets (T1, T2, T3) with 30, 50, and 100 attributes, respectively, each consisting of 5 standard Gaussian clusters with each attribute value from a given cluster generated according to $\mathcal{N}(c, r)$, with c and r having been selected uniformly at random from $[-1, 1]$ and $(0, 0.2]$, respectively. For T1 and T2, each cluster was generated in its own *distinct* subspace (with no attributes in common between clusters). For the purpose of studying the resilience of the approach to noise, the data was augmented with attributes whose values were drawn uniformly at random from $[-1, 1]$. T3 was generated from T2 by adding 50 additional attributes with uniform noise. The details are summarized in Table 1. Table 2 summarizes the clustering performance for these datasets comparing our approach against DiSH [1] and CLIQUE [2]. We chose DiSH as it also relies on a point-wise determination of relevant attributes (essentially comparing the spread of distances of nearest neighbors in all attributes) and could be seen as closely related to our approach. In addition, we test against the classical method CLIQUE, as it is arguably the best-known subspace clustering method. In most cases, our approach shows a superior performance.

Table 1: Synth. Datasets: Description.

	d	$\|\|\mathcal{S}\|\|$	Noisy $\mathcal{A}_i$
T1	30	$\{5, 5, 5, 5, 5\}$	5
T2	50	$\{3, 5, 7, 7, 11\}$	17
T3	100	$\{3, 5, 7, 7, 11\}$	67

Table 2: Synth. Datasets: Results.

		NMI	AMI	ARI	Recall
T1	DiSH	0.535	0.362	0.264	0.582
	CLIQUE	0.431	0.275	0.303	0.635
	LID-DBSCAN	**0.801**	**0.734**	**0.803**	**0.726**
T2	DiSH	0.568	0.396	0.532	0.7
	CLIQUE	0.644	0.473	0.568	**0.78**
	LID-DBSCAN	**0.779**	**0.695**	**0.716**	0.765
T3	DiSH	0.570	0.397	0.412	0.702
	CLIQUE	0.644	0.473	0.568	**0.78**
	LID-DBSCAN	0**.749**	**0.671**	**0.699**	0.76

Manifold Data. For the purpose of further validating the efficiency of the approach to detect significant subspaces on more complex datasets, we relied on the manifold generator proposed in [22] and generated manifolds of differing distributions in different dimensions. We compared the performance of the LID decomposition approach with the one of DiSH with respect to two different metrics (RNIA and ARR) that are generally used to judge clustering algorithms. With respect to both metrics, LID decomposition outperforms DiSH for each of the datasets considered, particularly for D4 (the set with highest average manifold dimension). We refer the reader to the full version for details.

4 Conclusion

Using decomposed LID as a new primitive for estimating the local relevance of a feature, future work could explore more refined subspace clustering approaches. Clustering approaches can be tailored to this new primitive but presumably many existing subspace clustering methods could be adapted to using the new primitive instead of conventional building blocks such as density-estimates, analysis of variance, or distance distributions. Beyond subspace clustering, many more applications can be envisioned, for example in subspace outlier detection [26] or in subspace similarity search [5, 14].

Variance-based measures of feature relevance, such as those underlying PCA and its variants, have an advantage over LID in that sample variances decompose perfectly across the coordinates within a Euclidean space. However, although the theoretical values within an LID decomposition are guaranteed to be additive, their estimates are not. Although the experimental results shown in Figure 2 indicate for the case of uniform distributions that MLE estimates for decomposed LID do sum to the overall LID estimate within reasonable tolerances, it is not clear how well additivity is conserved for real data. Since the additivity of estimators for LID decomposition may depend significantly on their accuracy, future research in this area could benefit from the further development of LID estimators of good convergence properties.

Acknowledgments M. E. Houle was supported by JSPS Kakenhi Kiban (B) Research Grant 18H03296.

References

1. Achtert, E., Böhm, C., Kriegel, H.P., Kröger, P., Müller-Gorman, I., Zimek, A.: Detection and visualization of subspace cluster hierarchies. In: Proc. DASFAA. pp. 152–163 (2007)
2. Agrawal, R., Gehrke, J., Gunopulos, D., Raghavan, P.: Automatic subspace clustering of high dimensional data for data mining applications. In: Proc. SIGMOD. pp. 94–105 (1998)
3. Amsaleg, L., Chelly, O., Furon, T., Girard, S., Houle, M.E., Kawarabayashi, K., Nett, M.: Estimating local intrinsic dimensionality. In: Proc. KDD. pp. 29–38 (2015)
4. Becker, R., Hafnaoui, I., Houle, M.E., Li, P., Zimek, A.: Subspace determination through local intrinsic dimensional decomposition: Theory and experimentation. arXiv e-prints arXiv:1907.06771 (2019)

5. Bernecker, T., Emrich, T., Graf, F., Kriegel, H.P., Kröger, P., Renz, M., Schubert, E., Zimek, A.: Subspace similarity search: Efficient k-NN queries in arbitrary subspaces. In: Proc. SSDBM. pp. 555–564 (2010)
6. Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U.: When is "nearest neighbor" meaningful? In: Proc. ICDT. pp. 217–235 (1999)
7. Casanova, G., Englmeier, E., Houle, M., Kröger, P., Nett, M., Schubert, E., Zimek, A.: Dimensional testing for reverse k-nearest neighbor search. PVLDB **10**(7), 769–780 (2017)
8. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proc. KDD. pp. 226–231 (1996)
9. François, D., Wertz, V., Verleysen, M.: The concentration of fractional distances. IEEE TKDE **19**(7), 873–886 (2007)
10. Houle, M.E.: Dimensionality, discriminability, density and distance distributions. In: Proc. ICDM Workshops. pp. 468–473 (2013)
11. Houle, M.E.: Local intrinsic dimensionality I: an extreme-value-theoretic foundation for similarity applications. In: Proc. SISAP. pp. 64–79 (2017)
12. Houle, M.E.: Local intrinsic dimensionality II: multivariate analysis and distributional support. In: Proc. SISAP. pp. 80–95 (2017)
13. Houle, M.E., Kriegel, H.P., Kröger, P., Schubert, E., Zimek, A.: Can shared-neighbor distances defeat the curse of dimensionality? In: Proc. SSDBM. pp. 482–500 (2010)
14. Houle, M.E., Ma, X., Oria, V., Sun, J.: Efficient algorithms for similarity search in axis-aligned subspaces. In: Proc. SISAP. pp. 1–12 (2014)
15. Hubert, L., Arabie, P.: Comparing partitions. J. Classif. **2**(1), 193–218 (1985)
16. Kadelburg, Z., Marjanović, M.: Interchanging two limits. The Teaching of Mathematics **8**(1), 15–29 (2005)
17. Kriegel, H.P., Kröger, P., Zimek, A.: Clustering high dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM TKDD **3**(1), 1–58 (2009)
18. Ma, X., Li, B., Wang, Y., Erfani, S.M., Wijewickrema, S.N.R., Schoenebeck, G., Song, D., Houle, M.E., Bailey, J.: Characterizing adversarial subspaces using local intrinsic dimensionality. In: Proc. ICLR. pp. 1–15 (2018)
19. Ma, X., Wang, Y., Houle, M.E., Zhou, S., Erfani, S.M., Xia, S., Wijewickrema, S.N.R., Bailey, J.: Dimensionality-driven learning with noisy labels. In: Proc. ICML. pp. 3361–3370 (2018)
20. Romano, S., Chelly, O., Nguyen, V., Bailey, J., Houle, M.E.: Measuring dependency via intrinsic dimensionality. In: ICPR16. pp. 1207–1212 (Dec 2016)
21. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**, 2323–2326 (2000)
22. Rozza, A., Lombardi, G., Ceruti, C., Casiraghi, E., Campadelli, P.: Novel high intrinsic dimensionality estimators. Machine learning **89**(1-2), 37–65 (2012)
23. Sim, K., Gopalkrishnan, V., Zimek, A., Cong, G.: A survey on enhanced subspace clustering. Data Min. Knowl. Disc. **26**(2), 332–397 (2013)
24. Strehl, A., Ghosh, J.: Cluster ensembles – a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. **3**, 583–617 (2002)
25. Vinh, N.X., Epps, J., Bailey, J.: Information theoretic measures for clustering comparison: Variants, properties, normalization and correction for chance. J. Mach. Learn. Res. **11**, 2837–2854 (2010)
26. Zimek, A., Schubert, E., Kriegel, H.P.: A survey on unsupervised outlier detection in high-dimensional numerical data. Stat. Anal. Data Min. **5**(5), 363–387 (2012)