ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 2

The Dimensionality Reduction Problem: A Comprehensive Exploration of Disjoint Principal Component Analysis (DPCA) and Disjoint Multiple Correspondence Analysis (DMCA)

Mario Fordellone[(1)]

(1) Department of Mental, Physical Health and Preventive Medicine, Medical Statistics Unit, University of Campania Luigi Vanvitelli

CORRESPONDING AUTHOR: Mario Fordellone, Università degli Studi della Campania Luigi Vanvitelli. E-mail: mario.fordellone@unicampania.it

SUMMARY

This paper delves into the realm of advanced data analysis, focusing on two powerful dimensionality reduction methods: Disjoint Principal Component Analysis (DPCA) and Disjoint Multiple Correspondence Analysis (DMCA). Methodological marvels in their own right, these approaches are scrutinized for their unique properties and applications across diverse domains. We navigate through the intricacies of their algorithms and explore how they unveil patterns within complex datasets. The comparative analysis highlights the strengths and weaknesses of DPCA and DMCA, shedding light on their distinct contributions to the analytical landscape. This paper serves as a comprehensive guide for researchers and analysts seeking deeper insights into these cutting-edge techniques for dimensional reduction.

Keywords: dimensionality reduction model; multivariate analysis; principal component analysis; multiple corrispondence analysis.

INTRODUCTION

In the era of big data, large and massive data sets are increasingly common that often include a progressive increase of the measurements and the number of variables used is always bigger. For this reason, the research of new statistical approaches to reduce the number of variables considerably while still retaining much of the information in the original data set is always under way. We know that in the field of dimensionality reduction methods, a variety of very known techniques have been proposed [1]. The most used and cited methods are surely principal component analysis [2] and factorial analysis [3] for quantitative data, and multiple correspondence analysis [4] for categorical data. Nevertheless, one of the most crucial topics related to these methods is the interpretation of components (i.e., factors) that define the latent subspace.

For example, in the case of PCA and FA, the main issue is related to the fact that each principal component (PC) typically is a linear combination of all manifest (i.e., observed) variables (MVs). In particular, for each PC all loadings are typically nonzero, even though only few MVs are relevant for the corresponding PC. This makes it often difficult to interpret the derived PCs, i.e., to understand what are the variables that really define each factor. In the specialized literature, several extensions of PCA have been proposed to specify subsets of MVs that most explain PCs. A very known approach consists to proceed is to artificially set the loadings with absolute values smaller than a threshold to zero, although Cadima and Jolliffe [5] consider this thresholding approach potentially misleading and subjective. Alternative and more statistically rigorous procedures for enhancing the interpretation are based on postprocessing methods such as rotations [6]. Nevertheless, the rotation procedures do not generate loadings exactly or

DOI: 10.54103/2282-0930/22513
Accepted: 19th March 2024

© 2024 Fordellone

close to zero and, then, thresholding is still required. Although, we note that in the context of FA, standard errors of rotated loadings are available, and thus, the evaluation of small loadings can be facilitated by using this inferential information. In Tibshirani et al. [7], a regularization of PCA is proposed to solve the sparsity problem. This approach consists in shrinking loadings toward zero by maximizing the explained variance of PCs penalized to shrink and select nonzero loadings. Also, Jolliffe et al. [8], propose an example of sparse principal component analysis (SPCA), in which the sparse loading matrix (i.e., namely with very few nonzero loadings) is obtained by using a simple least absolute shrinkage and selection operator (LASSO)-based approach. A probabilistic formulation of the SPCA approach is proposed by Guan and Dy [9]. Moreover, Shen and Huang [10] propose a SPCA via a regularized singular value decomposition (SVD) approach. Conversely, d'Aspremont et al. [11] propose another extension of SPCA: the direct sparse PCA (DSPCA), which reformulates the problem directly incorporating a sparsity criterion in the PCA.

However, even though several extensions of PCA have been proposed, they do not necessarily provide a simpler PC interpretation, since some MVs may still load on several PCs leaving the problem unresolved.

On the other hand, approaches similar to those used in PCA framework have been used in the FA. An example is confirmatory factor analysis (CFA) proposed by Jöreskog [12], where all the relationships between MVs and factors are studied and only few relationships between MVs and factors are specified by associating each MV to a single factor inducing disjoint classes of MVs. Obviously, in this way, the interpretation is greatly simplified since factors are exactly explained by a subset of MVs only. However, a drawback of CFA is that the assignment of a MV to a factor is based on the a priori knowledge of the researcher, which is not often guaranteed in the empirical cases.

An important method to solve the interpretability problem is the disjoint principal component analysis (DPCA) model introduced by Ferrara et al. [13] which is a particular case of the clustering and disjoint principal component analysis (CDPCA) model proposed by Vichi and Saporta [14], focusing only on the classification of MVs. Note that here, components/factors are formed by disjoint classes of MVs automatically identified instead that a priori fixed. In the last work, Ferrara et al. [15] propose a probabilistic approach of DPCA, named probabilistic disjoint principal component analysis (PDPCA).

This work aims to explore from a methodological point of view how key factors emerge during the process of dimensionality reduction for categorical data in the medical field, influencing the final representation of crucial information for clinical research. By closely examining the selection and combination of categorical variables in our specific medical context, we aim to identify determining factors that can significantly impact the understanding of relationships between different medical conditions, treatments, or responses to therapies. This approach not only provides valuable support for scholars and data analysts but also contributes to enhancing confidence in the insights extracted from the data, fostering a clear and transparent understanding of information. In summary, our research aims to improve interpretability in the dimensionality reduction of categorical data in the field of clinical research, offering an analytical and methodological framework that can be applied in medical contexts to achieve clearer and more meaningful results.

BACKGROUND

Let $X = [x_{ij}]$ $(i = 1, \ldots, n;\ j = 1, \ldots, J)$ be a $n \times J$ data matrix containing the measurements of J variables on n objects. Without loss of generality, after a location and scale transformation, we assume that all the variables are centred. For better understanding the algebraic proofs of the manuscript, the reader can refer Trefethen, L. N. and Bau, D. [16].

Principal component analysis

Principal component analysis [2] is generally seen as the orthogonal linear transformation of a set of J correlated variables, in matrix X, into a set of H (where $1 \leq H \leq J$ and $n > J$) principal components (PCs). Given a $J \times H$ loadings matrix A (i.e., factorial weights matrix), the $n \times H$ scores matrix Y can be written as

$$Y = XA \tag{1}$$

such that $\max_A \{tr(\Sigma_Y)\}$ subject to the constraint $A'A = I_H$ that implies $\Sigma_Y = diag(\sigma^2_{y_1}, \ldots, \sigma^2_{y_H})$. With standardized data PCs are such that $\Sigma_Y = I_H$.

Another formalization of PCA is the reconstruction of data matrix. In particular, the PCA model for reconstructing data is

$$X = YA' + E \tag{2}$$

where E is the $n \times J$ error matrix. Substituting Equation (1) in Equation (2) we obtain

$$X = XAA' + E \tag{3}$$

Proof 1

It is proved that the LSE problem of model (3), i.e., $\min_A \|X - XAA'\|^2$ is equivalent to maximize $tr(\Sigma_Y)$, i.e.,

$$tr(\Sigma_Y) = tr\left(n^{-1}Y'Y\right) = n^{-1}\left\|Y^2\right\| = n^{-1}\|XA\|^2 \tag{4}$$

therefore, it corresponds to compute PCs. Moreover, let us recall that the following decomposition holds for any orthogonal matrix A:

$$\|X\|^2 = \|X - XAA'\|^2 + \|XAA'\|^2 \quad (5)$$

in fact, let start from $\|X\|^2$ then add and subtract XAA', we obtain

$$\begin{aligned}\|X\|^2 &= \|X - XAA' + XAA'\|^2 \\ &= tr\begin{pmatrix}((X - XAA') + XAA')' \\ ((X - XAA') + XAA')\end{pmatrix} \\ &= \|X - XAA'\|^2 + \|XAA'\|^2 \\ &+2tr\,(X - XAA')'\,(XAA')) \end{aligned} \quad (6)$$

where the double product is null since $tr\,(X'XAA') - tr\,(AA'X'XAA') = 0$, because A is orthogonal. Therefore, minimising $\|X - XAA'\|^2$ corresponds to maximise $\|XAA'\|^2$ since $\|X\|^2$ is constant as the orthogonal matrix A varies. Now, it is easy to show that

$$\begin{aligned}\|XAA'\|^2 &= tr\,(AA'X'XAA') = tr\,(A'X'XAA'A) \\ &= tr\,(A'X'XA) = \|XA\|^2 = ntr\,(\Sigma_Y)\end{aligned} \quad (7)$$

and therefore if $\|XAA'\|^2$ is maximised also $ntr\,(\Sigma_Y)$ is maximised. In other words, minimize the error resulting in Equation 3 corresponds to maximize the variance of the principal components matrix Y. This is the objective function of the PCA algorithm.

Probabilistic principal component analysis

Probabilistic principal component analysis (PPCA) [17], is a probabilistic formulation of PCA. In particular, just recall the model formalization of PCA shown in Equation (2) and assume the following hypothesis:

i. $y_i \sim N_H\,(0, \Sigma_Y)$, where $\Sigma_Y = I_H$;
ii. $e_i \sim N_J\,(0, \Sigma_E)$, where $\Sigma_E = \sigma^2 I_J$;
iii. $Cov\,(e_i, y_i) = \Sigma_{EY} = 0$

Thus, like factorial analysis (FA), PCs are defined independent, standardized, Gaussian, and a mutual independence between Y and E is assumed. Statistically, these hypotheses imply the following covariance matrix structure of X:

$$\begin{aligned}\Sigma_X &= n^{-1}X'X = n^{-1}AY'YA' + E'E \\ &= A\Sigma_Y A' + \Sigma_E = AA' + \sigma^2 I_J\end{aligned} \quad (8)$$

and, consequently, the Gaussian distribution of data $x_i \sim N_J\left(0, AA' + \sigma^2 I_J\right)$. A similar form of the covariance matrix is specified in FA, which differs from PPCA only in the more general specification of $\Sigma_E = diag\left(\sigma_1^2, \ldots, \sigma_J^2\right)$, which is not necessarily based on a isotropic error covariance as in PPCA. This modification leads to significant differences in the behavior of the two methods [18].

The ML estimate of Σ_X (i.e., the estimation of A and σ^2 can be obtained by the standard EM algorithm [19]. We can define the log-likelihood function as follows:

$$l\left(A, \sigma^2 \mid X\right) = -\frac{n}{2}\begin{Bmatrix} Jln\,(2\pi) + \ln\left(AA' + \sigma^2 I_J\right) \\ +tr\left[\left(AA' + \sigma^2 I_J\right)^{-1} S\right]\end{Bmatrix} \quad (9)$$

where $S = n^{-1}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)'$ is the observed sample covariance matrix with μ supposed known and estimated by the sample mean.

Proof 2

It is proved that the ML estimators of A and σ^2 for the isotropic error model correspond to the PCA solution. A formal, short and easy proof depends on the two following results:

a) $AA' + \sigma^2 I_J = AA' + \sigma^2 I_H \sigma^{2(J-H)} = \left(1 + \sigma^2\right)^H \sigma^{(J-H)}$

b) $\left(AA' + \sigma^2 I_J\right)^{-1} = \sigma^{-2} I_J - A\left[\sigma^2\left(\sigma^2 I_H + AA'\right)\right]^{-1}$
$A' = \sigma^{-2} I_J - (\sigma^2\left(\sigma^2 + 1)AA'\right)^{-1}$

thus substituting a) and b) in Equation (9), the log-likelihood function can be written as:

$$l\left(A, \sigma^2 \mid X\right) = -\frac{n}{2}\begin{Bmatrix}\left[\ln\left(1 + \sigma^2\right)^H \sigma^{(J-H)}\right] + tr\left[\frac{S}{\sigma^2}\right] \\ -(\sigma^2\left(\sigma^2 + 1\right)\right)^{-1} tr\,(AA'S)\end{Bmatrix} + C \quad (10)$$

where C is a constant not depending on both A and σ^2. In this way, it can be directly observed that the ML estimate of A (i.e., $\hat{A}$) is equal to the LS estimate. In particular, to maximize (10), we need to maximize

$$tr(AA'S) = tr(A'SA) \quad (11)$$

and considering the spectral decomposition of S show below:

$$S = ULU' \quad (12)$$

where U is orthogonal matrix which columns are eigenvectors of S and L is diagonal matrix which elements are the corresponding eigenvalues.

The solution is given by the H eigenvectors $U_{(H)}$ corresponding to the largest H eigenvalues (reported in the diagonal matrix $L_{(H)}$) of the covariance matrix S, i.e., $\hat{A} = U_{(H)}$. However, since $\Sigma_Y = I_H$, to reconstruct the matrix X according model (2), PCs have to be scaled for their variance and the variance of the error term. Then,

$$\hat{A} = U_{(H)}\left(L_{(H)} - \sigma^2 I_H\right)^{1/2} \quad (13)$$

For estimating σ^2 we need to set the derivative of the log-likelihood function with respect to σ^2 equal to zero as shown in Equation (14)

$$\frac{\partial l}{\partial \sigma^2} = K\sigma^4\left(\sigma^2+1\right) + (J-H)\sigma^2\left(\sigma^2+1\right)^2 - tr(S)\left(\sigma^2+1\right)^2 + \left(2\sigma^2+1\right)tr(AA'S) = 0 \quad (14)$$

Thus, the solution is given by

$$\hat{\sigma}^2 = \frac{tr(S) - tr(A'SA)}{(J-H)} \quad (15)$$

Disjoint principal component analysis

The disjoint principal component analysis (DPCA) model can be formally written as the PCA model (2) where some constraints on the loading matrix A are imposed [14], following the idea of SEM, which allows researchers to model LVs through disjoint classes of correlated MVs [20]. In particular, the following constraints are defined:

iv. $\sum_{j=1}^{J} a_{jh}^2 = 1, \quad h = 1,\ldots, H$

v. $\sum_{j=1}^{J}\left(a_{jh}a_{jr}\right)^2 = 0, \quad h = 1,\ldots, H-1;\ r = h+1,\ldots, H;$

vi. $\sum_{h=1}^{H} a_{jh}^2 > 0, \quad j = 1,\ldots, J.$

The constraints iv-vi imply:

c) A is column-orthonormal, i.e. A'A = I;
d) each row of A has at most a single loading for a LV, i.e. a MV can contribute only to a single LV;
e) from (d) a partition of MVs is induced and each LV is represented as a linear combination of a single class of variables.

Moreover, the loading matrix A can be re-parameterized as the product of two matrices as follows:

$$A = BV \quad (16)$$

where $V = [v_{jh}]$ is a $J \times H$ binary and row stochastic matrix defining a partition of variables into H classes identifying H PCs, with $v_{jh} = 1$, if the jth variable belong to h^{th} class, $v_{jh} = 0$, otherwise; B is a $J \times J$ diagonal matrix weighting MVs. In this way, constraints iv-vi become $\sum_{j=1}^{j} v_{jh}b_j^2 = 1$; $\sum_{h=1}^{H}\sum_{j=1}^{J} v_{jh}b_j^2 = H$, and the DPCA is can be specified as follows:

$$X = YV'B + E \quad (17)$$

where Y is a linear combination defined as Y = XBV. Thus, model (14) can be expressed as

$$X = XBVV'B + E \quad (18)$$

such that

1) $V = \left[v_{jk} : \forall v_{jk} \in \{0,1\}\right]$ (binary);

2) $V1_k = 1_J$ (row stochastic);

3) $B = diag(b_1, \ldots b_J)$ (diagonal);

4) $V'BBV = I_k$ (orthonormal);

Proof 3

Note that in FA framework the Bartlett's weighted LS score, which takes the following form

$$Y = X\Sigma_E^{-1}BV\left(V'B\Sigma_E^{-1}BV\right)^{-1} \quad (19)$$

it is reduced to Y=XBV when an isotropic error is specified. In fact

$$\begin{aligned} Y &= X(\sigma^2 I_J)^{-1}BV\left(V'B(\sigma^2 I_J)^{-1}BV\right)^{-1} = \\ &= X\sigma^2 BV\sigma^2\left(V'BBV\right)^{-1} = \\ &= XBV\left(V'BBV\right)^{-1} = \\ &= XBV \end{aligned} \quad (20)$$

The LS estimators of the models (17) and (18) are the optimal solutions of the following quadratic problem with respect to unknown parameters B and V:

$$\min_{B,V}\left\|X - XBVV'B\right\|^2 \quad (21)$$

such that constraints 1) - 4) are satisfied.

Proof 4

It is interesting to note the following decomposition:

$$\|X\|^2 = \|X - XBVV'B\|^2 + \|XBVV'B\|^2 \quad (22)$$

The proof of the decomposition is given by

$$\|X\|^2 = \|X - XBVV'B\|^2 + \|XBVV'B\|^2 + 2\,tr\left[\left(X - XBVV'B\right)'\left(XBVV'B\right)\right] \quad (23)$$

thus,

$$tr\left[\left(X - XBVV'B\right)'\left(XBVV'B\right)\right] = 0 \quad (24)$$

In fact,

$$tr\left[\left(X - XBVV'B\right)'\left(XBVV'B\right)\right] = \\ = tr\left(X'XBVV'B\right) - tr\left(BVV'BX'XBVV'B\right) = \\ = tr\left(V'BX'XBV\right) - tr\left(V'BX'XBVV'BBV\right) = 0 \quad (25)$$

since in the second member of the previous equation $V'BBV = I_k$. From decomposition (22), by minimizing equation (21), the second term of the right-hand side of (22) is maximized.

Moreover, the second term of the (22) can be written as

$$\begin{aligned} \|XBVV'B\|^2 &= tr\left(BVV'BX'XBVV'B\right) \\ &= tr\left(V'BX'XBV(V'BBV)\right) = \\ &= tr\left(V'BX'XBV\right) = XBV^2 \\ &= n\ tr\left(\Sigma_Y\right) = tr\left(Y'Y\right) = \\ &= tr\left(V'B\Sigma_X BV\right) = \sum_{k=1}^{K} v_k' B\Sigma_X B v_k \end{aligned} \quad (26)$$

therefore, the minimization problem shown in Equation (21) corresponds to maximize the total variance of the PCs (26). Then, DPCA model is a constrained formulation of the PCA model, where the loading matrix has the form A=BV, with B and V satisfying constraints 1) - 4), and the solution of Equations (21) and (26) can be find through a constrained Alternating Least Squares (ALS) algorithm.

Probabilistic disjoint principal component analysis

The probabilistic disjoint principal component analysis (PDPCA) is an isotropic error model that joins the features of PPCA and DPCA [15]. PDPCA model is defined by the DPCA model in shown in Equations (17) and (18), subject to the constraints defined in 1) - 4), in which we consider the PPCA assumptions i-iii. For these properties, the PDPCA model produces the following covariance matrix structure of X:

$$\Sigma_X = \frac{1}{n}X'X = \frac{1}{n}BV(Y'Y)V'B + \Sigma_E = BV\Sigma_Y V'B + \sigma^2 I_J \quad (27)$$

with the related Gaussian distribution $x_i \sim N_J(0, BVV'B + \sigma^2 I_J$, since $\Sigma_Y = I_H$. Let $(x_1, \ldots x_n)$ be a sample of i.i.d. J dimensional observations, where $x_i \sim N_J\left(0, BVV'B + \sigma^2 I_J\right)$, the corresponding log-likelihood function can be formulated as

$$l\left(V, B, \sigma^2 \mid X\right) = -\frac{n}{2}\left\{\begin{array}{l} \ln\left|BV\Sigma_Y V'B + \sigma^2 I_J\right| \\ +tr\left(\left(BV\Sigma_Y V'B + \sigma^2 I_J\right)^{-1} S\right) \\ +Jln\left(2\pi\right) \end{array}\right\} \quad (28)$$

By the maximizing the log-likelihood function shown in Equation (28) through an expectation-maximization (EM) algorithm [21], subject to the constraints defined in 1) - 4) and under PPCA assumptions i-iii, we obtain the following ML estimator:

the ML estimator of b_h is

$$\hat{b}_k = {}_hU_{(1)}\left({}_hL_{(1)} - \sigma^2 I_H\right)^{1/2} \quad (29)$$

where ${}_hU_{(1)}$ and ${}_hL_{(1)}$ respectively are the eigenvector and the corresponding largest eigenvalue of matrix Σ_{X_h}. The estimates of V is obtained by assigning each variable to the class that most increases the log-likelihood, i.e.,

$$\begin{aligned} \hat{v}_{jh} = 1 \text{ if } & l\left(V, B_h, \sigma^2\right) \\ &= max\left\{l\left(V, B_m, \sigma^2\right) : m = 1, \ldots, H\right\} \end{aligned} \quad (30)$$

$\hat{v}_{jh} = 0$ otherwise.

Finally, the ML estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{J-H}\sum_{h=1}^{H}\left(tr(\Sigma_{X_h}) - \frac{1}{n}y_h' y_h\right) \quad (31)$$

that is the average of the loss corresponding to the H classes.

Disjoint multiple correspondence analysis

The disjoint multiple correspondence analysis (DMCA) model is a particular case of the disjoint principal component analysis (DPCA) introduced in the subsection 2.3. In fact, the DMCA model can be considered as the DPCA applied to a categorical data matrix appropriately centred as $X = J^{1/2}J\Psi L^{1/2}$. Where J is the number of qualitative variables; $\Psi = \left[\Psi_1, \ldots, \Psi_J\right]$ is the binary block matrix formed by J indicator binary matrices Ψ_J with elements Ψ_{i,j_m}=1 if the ith observation has assumed category m for variable J, $\psi_{i,j_m} = 0$ otherwise; $L = diag\left(\Psi' 1_N\right)$; $J = I_N - N^{-1}1_N 1_N'$ is the idempotent centring matrix with.

Therefore, to introduce the DMCA model we can consider PCA model shown in Equation (3) and the re-parameterization of the loading matrix given by A = BV:

$$J^{1/2}J\Psi L^{1/2} = J^{1/2}J\Psi L^{1/2}BVV'B + E \quad (32)$$

such that

1) $V = \left[v_{jk} : \forall v_{jk} \in \{0, 1\}\right]$ (binary);
2) $V1_K = 1_J$ (row stochastic);
3) $B = diag\left(b_1, \ldots, b_J\right)$ (diagonal);
4) $V'BBV = I_K$ (orthonormal).

The LS estimators of the models (32) are the optimal solutions of the following quadratic problem with respect to unknown parameters B and V:

$$\min_{B,V} \left\| J^{1/2} J \Psi L^{1/2} - J^{1/2} J \Psi L^{1/2} B V V' B \right\|^2 \qquad (33)$$

such that constraints 1) - 4) are satisfied. Then, fixed the number of factor *H*, the maximization of (33) can be solved by using ALS algorithm.

DISCUSSION

In this work we analysed the methodological properties of the dimensionality reduction approaches in the case of continuous and categorical data. The two approaches discussed in this work are Disjoint Principal Component Analysis (DPCA) and Disjoint Multiple Correspondence Analysis (DMCA), two statistical methods that could find interesting applications in the clinical field.

In clinical genetics, DPCA could be a valuable tool for delving into complex genetic data. It enables the identification of specific variance patterns within subsets of genes or genetic markers. This application is particularly beneficial for uncovering genetic associations related to specific clinical conditions or responses to various treatments. Similarly, in biomarker research, DPCA offers a means to separate the variance linked to different categories of biomarkers. This capability aids in identifying patterns that hold clinical significance, providing valuable insights for diagnostic and prognostic purposes. Medical imaging, such as data obtained from magnetic resonance or computed tomography scans, stands to benefit from DPCA. This method can assist in pinpointing specific patterns within different regions of medical images, contributing significantly to the early diagnosis of various pathologies.

Turning our attention to DMCA, its application in epidemiological analysis is noteworthy. When dealing with categorical clinical data, such as classifying diseases into distinct categories, DMCA proves useful. It helps in identifying specific risk factors associated with particular health conditions, aiding in the development of targeted preventive measures. Moreover, in lifestyle habits studies, DMCA serves as a valuable analytical tool. By examining the relationships between various categorical variables, such as tobacco consumption, physical activity, and diet, DMCA contributes to a more nuanced understanding of their impacts on health, providing crucial information for personalized interventions. In the assessment of patients' quality of life, DMCA plays a significant role. Analysing data related to categorical variables like symptoms, emotional impact, and overall satisfaction, DMCA provides a comprehensive view of health status. This holistic understanding can guide healthcare professionals in tailoring treatment plans and interventions to improve patients' overall well-being.

In conclusion, both DPCA and DMCA offer valuable insights for clinical research and analysis. Their applications in clinical genetics, biomarker research, medical imaging, epidemiological analysis, lifestyle habits studies, and quality of life assessment showcase their versatility in addressing diverse aspects of healthcare and medical research. The choice between these approaches depends on the specific characteristics of the data and the research goals in the clinical context.

REFERENCES

1. Yan S, Xu D, Zhang B, Zhang H, Yang Q, Lin S. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. IEEE Trans Pattern Anal Mach Intell. 2007;29:40-51.
2. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1901;2(11):559-72.
3. Spearman C. "General Intelligence," Objectively Determined and Measured. The American Journal of Psychology. 1904;15(2):201-92.
4. Michael Greenacre JB. Multiple Correspondence Analysis and Related Methods (1st ed.): Chapman and Hall/CRC. ; 2006.
5. Cadima J, Jolliffe IT. Loading and correlations in the interpretation of principle compenents. Journal of Applied Statistics. 1995;22(2):203-14.
6. Widaman K. Common Factor Analysis Versus Principal Component Analysis: Differential Bias in Representing Model Parameters? Multivariate Behavioral Research. 1993;28:263-311.
7. Hastie T, Tibshirani R, Wainwright M. Statistical Learning with Sparsity: The Lasso and Generalizations2015. 1-337 p.
8. Jolliffe I, Trendafilov N, Uddin M. A Modified Principal Component Technique Based on the LASSO. Journal of Computational and Graphical Statistics. 2003;12.
9. Guan Y, Dy J. Sparse Probabilistic Principal Component Analysis. Journal of Machine Learning Research - Proceedings Track. 2009;5:185-92.
10. Shen H, Huang JZ. Sparse principal component analysis via regularized low rank matrix approximation. Journal of Multivariate Analysis. 2008;99(6):1015-34.
11. d'Aspremont A, El Ghaoui L, Jordan MI, Lanckriet GRG. A direct formulation for sparse PCA using semidefinite programming. Siam Rev. 2007;49(3):434-48.
12. Jöreskog K. A General Approach to Confirmatory Factor Analysis. Psychometrika. 1969;34:183-202.
13. Ferrara C, Martella F, Vichi M. Dimensions of Well-Being and Their Statistical Measurements. Stud Theor Appl Stat. 2016:85-99.

14. Vichi M, Saporta G. Clustering and Disjoint Principal Component Analysis. Computational Statistics & Data Analysis. 2009;53:3194-208.
15. Ferrara C, Martella F, Vichi M. Probabilistic Disjoint Principal Component Analysis. Multivariate Behavioral Research. 2018;54:1-15.
16. 16. Lloyd N. Trefethen DB. Numerical linear algebra: Society for Industrial and Applied Mathematics; 2022. xvi + 370 p.
17. Tipping ME, Bishop CM. Probabilistic Principal Component Analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1999;61.
18. Tipping ME, Bishop CM. Mixtures of Probabilistic Principal Component Analyzers. Neural Computation. 1999;11:443-82.
19. Meng X-L, van Dyk D. The EM Algorithm–An Old Folk-Song Sung to a Fast New Tune. Journal of the Royal Statistical Society Series B (Methodological). 1997;59(3):511-67.
20. Fordellone M, Vichi M. Structural Equation Modeling and simultaneous clustering through the Partial Least Squares algorithm2018.
21. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B (Methodological). 1977;39(1):1-38.