e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

AN UNSUPERVISED MACHINE LEARNING ALGORITHM: PCA (PRINCIPAL COMPONENT ANALYSIS) COMPREHENSIVE REVIEW

Kshitij Patel*1, Govind Jee Tiwary*2, Kishan Kumar Pandey*3, Dr. Komal Asrani*4

*1,2,3AKTU, Computer Science And Engineering, BBDNIIT, Lucknow, UP, India.

*4Professor, Department Of Computer Science And Engineering, BBDNIIT, Lucknow, UP, India.

DOI : https://www.doi.org/10.56726/IRJMETS49457

ABSTRACT

Principal Component Analysis (PCA) is a popular unsupervised machine learning method that is well-known for simplifying complicated data. The complexities of PCA are thoroughly examined in this work, which also sheds light on its operational mechanisms, extensive applications, and mathematical underpinnings. concentrating on the many uses of Principal Component Analysis (PCA) in the field of machine learning. The paper goes into the mathematical underpinnings of PCA, such as eigen-decomposition, and its real-world uses, which range from picture reduction to multivariate data analysis. Its objective is to identify the significant information contained in the statistical data, extract it, and express it as a set of new orthogonal variables known as principal components. The pattern of similarity between the variables and the observations is then represented as spots on spot maps.

This paper offers useful comparisons between PCA and the PCA kernel, so that readers can determine which approach isbest for their particular analytical needs. This study functions as a thorough review and reveals the revolutionary potential of PCA's kernelized variant, all while imparting a good comprehension of the concept. Through the exploration, academics and practitioners will be better equipped to make educated decisions and use PCA and PCA kernel wisely in a variety of data analysis contexts.

Keywords: Machine Learning, Unsupervised Learning, Principal Component Analysis, Dimensionality Reduction, Kernel.

I. INTRODUCTION

Unsupervised learning is a fundamental concept in machine learning that has proven to be useful in revealing hidden patterns in large and complex datasets. Unsupervised learning functions independently, which makes it invaluable for a variety of data analysis and dimensionality reduction tasks. Supervised learning, on the other hand, depends on labeled data for model training. Principal Component Analysis (PCA) is a powerful and adaptable technique in the field of unsupervised learning that is essential to the toolset of academics and data scientists Developed in 1901 by Pearson, principal component analysis (PCA) has become a mainstay of unsupervised learning, providing excellent dimensionality reduction of data while preserving important information. PCA extracts the essential information from data by projecting it into a lower-dimensional subspace that is determined by its principal components. This allows the data to be subjected to additional analysis and interpretation. Its applications span diverse fields, from image and speech processing to genetics and finance.

The PCA technique tackles the problem of growing feature counts, sometimes referred to as the "curse of dimensionality."PCA intervenes as a feature extraction solution when the number of features increases and the model becomes more complex, leading to a decrease in accuracy. With little information loss, this algorithm efficiently sorts through the provided features and extracts a smaller collection. PCA essentially attempts to convert higher -dimensional data into a more controllable, lower- dimensional space with the goal of optimally capturing the essential features. In order to guarantee the use of the most appropriate method, it looks for the greatest representation of the substance of the data in higher dimensions.

PCA transcends the original set of features, crafting a new set while retaining crucial elements. By carefully selecting a subset from this new set, PCA identifies the most significant features, establishing itself as a pivotal technique in unsupervised learning for data representation and dimensionality reduction. The demand for effective feature engineering (shown in Fig. 1) has prompted a division into four crucial components: feature transformation, feature extraction, feature selection, and feature creation. Among these, feature extraction is

IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

pivotal for distilling essential patterns and information from data. Principal Component Analysis (PCA) emerges as a cornerstone in feature extraction, offering a robust approach to uncover and highlight critical aspects within the data. As the number of features grows—a phenomenon known as the "curse of dimensionality"—PCA becomes particularly valuable. It acts as a potent solution for feature extraction, navigating through the complexity of variables to derive a reduced set that encapsulates the core information of the original data.

This introduction sets the stage for an exploration of PCA's role in feature extraction, emphasizing its significance in addressing the challenges posed by increasing feature dimensions. The journey unfolds by delving into the mathematical foundations and practical applications of PCA, aiming to provide a concise yet comprehensive understanding of itsimpact on the future of feature engineering.

II. LITERATURE REVIEW

The significance of unsupervised learning in interpreting unstructured data and identifying patterns in large-scale datasets has been acknowledged. One important unsupervised learning technique that is particularly notable for compressing and displaying data in high-dimensional domains is Principal Component Analysis (PCA).

In their 2008 study, Maaten and Hinton examined the uses of PCA and noted how well it could compress high-dimensional data while preserving important information. The scalability of PCA with huge datasets, demonstrated by this work, enables it to meet the demands of contemporary research. By offering a thorough understanding of PCA's function in dimensionality reduction, olliffe (2002) makes a contribution. The study highlights how PCA can help in understanding complex dataset structures by capturing important aspects. These revelations highlight the applicability of PCA, which benefits academics working with complex datasets.

This overview of the literature emphasizes how important unsupervised learning is, especially PCA, in tackling the problemsfacing modern data-driven research. Together, the experiments demonstrate how flexible and effective PCA is for examining and evaluating huge, unstructured information.

III. A BRIEF HISTORICAL NOTE

It can be challenging to determine the exact origins of statistical approaches. Priceendorfer and Mobley (1988) pointed out that the singular value decomposition (SVD), which forms the basis of PCA, was independently discovered by Beltrami (1873) and Jordan (1874). The SVD was applied in a two-way analysis of an agricultural trial by Fisher and Mackenzie (1923).

Nonetheless, it is generally acknowledged that Pearson (1901) and Hotelling (1933) provided the initial descriptions of the method that is now known as PCA. The paper on Hotelling has two sections. The first and most significant section is included in the collection of articles compiled by Bryant and Atchley (1975), together with Pearson's work. he typical algebraic derivation provided above is similar to that proposed by Hotelling (1933), but the two works took different methodologies. Conversely, Pearson (1901) was interested in geometric optimization issues that led to PCs.

It can be challenging to determine the exact origins of statistical approaches. Priceendorfer and Mobley (1988) pointed out that the singular value decomposition (SVD), which forms the basis of PCA, was independently discovered by Beltrami (1873) and Jordan (1874). The SVD was applied in a two-way analysis of an agricultural trial by Fisher and Mackenzie (1923).

Nonetheless, it is generally acknowledged that Pearson (1901) and Hotelling (1933) provided the initial descriptions of the method that is now known as PCA. The paper on Hotelling has two sections. The first and most significant section is included in the collection of articles compiled by Bryant and Atchley (1975), together with Pearson's work. he typical algebraic derivation provided above is similar to that proposed by Hotelling (1933), but the two works took different methodologies. Conversely, Pearson (1901) was interested in geometric optimization issues that led to PCs.

He aimed to determine which lines and planes in p-dimensional space best suit a given collection of points. Considered more than 50 years before computers became widely accessible, Pearson's remarks about computing are intriguing. Although he notes that computations become "cumbersome" with four or more

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

variables, he thinks that his methods arestill highly practical and that they can be easily applied to numerical problems. Very little pertinent information appears to have been published in the 32 years that separated Pearson and Hotelling's studies, however, Rao (1964) notes that Frisch(1929) used a methodology comparable to Pearson's. Furthermore, Thurstone (1931) may have been working along similar lines as Hotelling (1933), according to a notation in Hotelling (1933). However, Thurstone's work was in a referenced paper—which is also included in Bryant and Atchley (1975)—rather than PCA. The way hotels approached PCA revealedthat it was quite different from factor analysis. The idea behind Hotelling is that a smaller fundamental set of independentvariables might determine the original p variables' value.

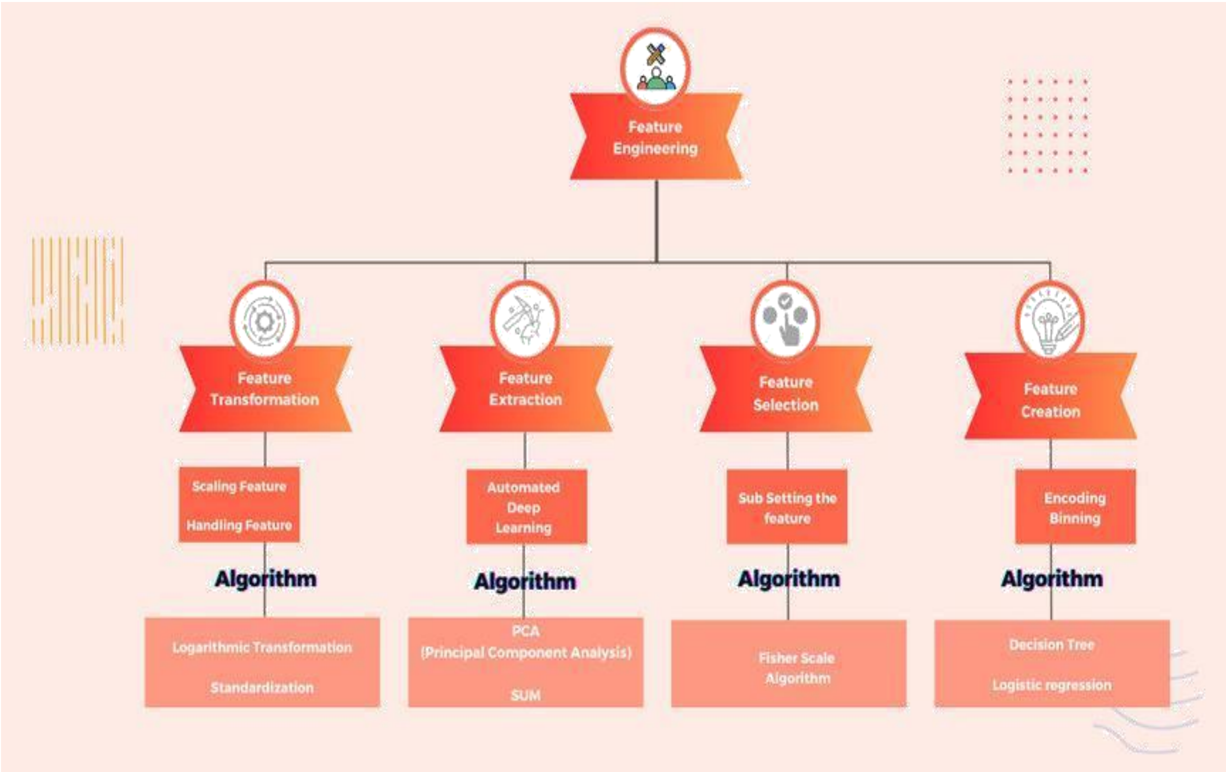

Figure 1. Feature Engineering

IV. DEFINITION

Principal Component Analysis (PCA) is a technique that aims to transform a dataset with many correlated variables into a new set of uncorrelated variables, known as principal components (PCs). PCA's primary objective is to simplify the analysis of complicated datasets by reducing their dimensionality while preserving the essential details. The first few main components capture the most significant variance found in the original dataset and are listed in order of importance. As a result, PCA enables a more effective and understandable data representation.

V. DIMENSIONALITY REDUCTION TECHNIQUES

Techniques for dimension reduction, or DR, are crucial for managing data effectively and enhancing several facets of machine learning. To enable efficient transfer learning, the main objective of these techniques is to minimize the distance in a latent space between distributions of various datasets. When Dimensionality Reduction (DR) is used, the individual device outcomes are far better than in circumstances where dimensionality is not reduced.

Benefits of dimensionality reduction techniques:

- Decreased Dimensions and Storage Space
- Time Efficiency
- Elimination of Irrelevant and Noisy Data
- Optimized Data Quality

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

- Enhanced Algorithm Efficiency and Accuracy
- Algorithms benefit from working more efficiently and achieving improved accuracy.
- Data Visualization
- Simplified Classification and Improved Performance

VI. MATHEMATICAL INTUITION

Standard deviation: The positive square root of the arithmetic mean of the squares of all the departures from the arithmetic mean is the standard deviation of a collection of observations in a series. As a result, in order to calculate the standard deviation, the arithmetic mean must first be determined, and then the deviation of each item from the mean must be squared. After adding up all of the squared deviations, divide the total by the total number of elements. With standard deviation, the algebraic symbols + and - of the deviations in the mean deviation are not removed. Therefore, a measure of dispersion with greater mathematical relevance is the standard deviation. σ is commonly used to represent the standard deviation.

Variance is the square of the standard deviation. σ ^2 represents variance.

$$\text{variance} = \sigma^2 = \frac{\Sigma (x_r - \mu)^2}{n}$$

$$\text{standard deviation} \quad \sigma = \sqrt{\frac{\Sigma (x_r - \mu)^2}{n}}$$

$$\mu = \text{mean}$$

Covariance: A useful technique for examining the relationship between two dimensions in a dataset is covariance. In contrast to variance and standard deviation, which function independently on individual dimensions, covariance sheds light on the ways in which dimensions co-vary relative to one another. There is always a comparison between two dimensions when computing covariance. For example, the covariance between A and B, B and C, and A and C can be calculated in a three-dimensional dataset (A, B, and C).

$$COV(X,Y) = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n-1}$$

Correlation: The degree of link between two variables is measured using correlation coefficients. Although there are other varieties of correlation coefficients, Pearson's is the most widely used. A correlation coefficient is known as Pearson's correlation, or Pearson's R.

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

Orthogonality: The main components are arranged such that they are perpendicular to one another. To put it briefly, the main components are separate and do not duplicate any information. In principal component analysis, every principal component (PC) isconstructed to maximize the variance it explains while adhering to the PC perpendicular criteria. In the figure PC1 is orthogonal toPC2.

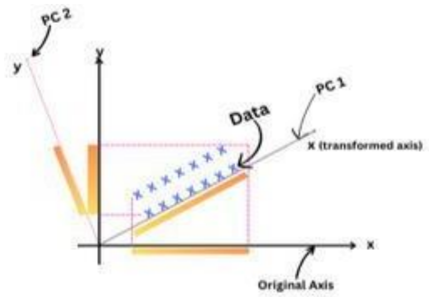

Figure 2: shows principal component

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

VII. EIGENVECTORS AND EIGENVALUES

The concepts of eigenvectors and eigenvalues are essential when discussing square matrices since they provide light on matrix structures. In unison, they constitute the eigen-decomposition, an efficacious analytical instrument that illuminates the inherent characteristics of a matrix. Eigen-decomposition is useful for some square matrices, such as correlation, covariance, and cross- product matrices, but it is not always applicable to them.

The eigen-decomposition is quite useful, particularly for analyzing matrices that are frequently encountered in statistical applications. Specifically, for matrices associated with correlation, covariance, or cross-products, the eigen-decomposition discloses crucial properties of real relevance. When working with functions that involve these matrices, this becomes very important because the eigen-decomposition is a crucial tool for determining the maximum or minimum values.

The most often used definition of an eigenvector of matrix A is a vector u that satisfies the following equation. Eigenvalues and eigenvectors can be defined in a variety of ways. $Au = \lambda u$, where λ is a scalar that is connected to the eigenvector and is known as the eigenvalue. the equation becomes $(A - \lambda I)\, u = 0$,

Where, λ is a scalar called the eigenvalue associated to the eigenvector.

VIII. PROJECTION OF DATA POINTS ON THE AXIS

By projecting data points onto these axes, represented as vectors in an R-dimensional space, PCA aims to find Principal Components (PCs). These PCs, depicted as coefficients in a new basis, reveal the most significant directions of data spread. The mathematical process involves representation and coefficient representation, aiding in data reconstruction post-compression. The figure illustrates how PCA transforms axes to capture maximum projected data compared to the original axis, providing a visual representation of variance capture. This analysis helps streamline data, making it more manageable for further exploration or modelling while preserving essential information.

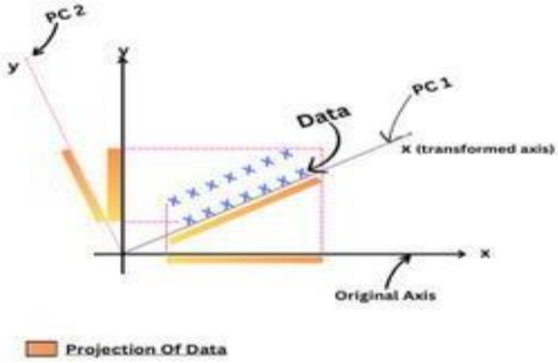

Figure 3. Projection of data points

IX. MATHEMATICAL REPRESENTATION OF PROJECTION OF DATA

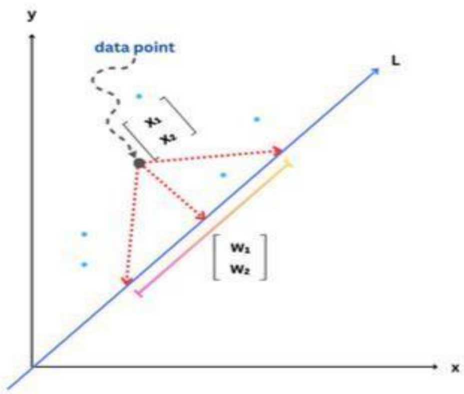

Figure 4. Projection of the data in vector form

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

formula
$\vec{\omega} \rightarrow$ unit vectar along line $L\ x \in Rd$

Projection of point on the line $= \left(\frac{(x^\top w)}{\|w\|^2}\right)w.$

A. Main goal is to find the line which have minimum reconstruction error:

Reconstruction error: The average squared distance between the original data points and their projections into the primary subspace can be used to understand the reconstruction error in primary Component Analysis (PCA). The most important information in the data is captured by the primary subspace, which is made up of the eigenvectors corresponding to the greatest eigenvalues of the covariance matrix. The disregarded subspace, on the other hand, represents the orthogonal complement and adds to the unexplained variance. The attempt to maintain the key components of the data is shown in the reduction in squared distances during the projection onto the principal subspace, highlighting the data's function in information preservation and dimensionality reduction.

$$\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

Covariance matrix: When dealing with datasets with more than two dimensions, covariance, a measure of the relationship between two variables, extends to multiple pairs. The covariance matrix, a symmetric p × p matrix (where p is the number of dimensions), encapsulates all covariances between initial variables. The entries in this matrix signify the covariances associated with all possible pairs of dimensions. Understanding the correlations between variables involves interpreting the signs of these covariances. A positive covariance indicates that the variables tend to increase or decrease together, reflecting a correlated relationship. Conversely, a negative covariance suggests that one variable increases when the other decreases, indicating an inversely correlated relationship. This insight into the sign of covariances is pivotal for discerning the directional tendencies between dimensions in a multi-dimensional dataset.

X. CONSTRUCTS THE PRINCIPAL COMPONENTS

Using a construction where the first principal component captures the most important variance, principal components are calculated to optimize variance in a dataset. Next-to-highest variances are explained by components that are uncorrelated with the ones that came before them. The original variables are then equated via this method until p primary components are found. An important part of this procedure are the eigenvalues and eigenvectors, which match the dimensions of the data in pairs. The Covariance matrix's eigenvalues quantify the variance in each component, whereas the eigenvectors show the principal components—the axis of highest variance. Understanding the most important dimensions in the dataset is streamlined by obtaining the main components in order of significance by sorting eigenvectors based on eigenvalues, highest to lowest.

After choosing principal components and creating the feature vector, the emphasis of Principal Components Analysis (PCA) moves to transforming the data along the principal component axis. The input dataset is still in terms of its initial variables, or axes, up until this stage. Reorienting the data from its original axes to those indicated by the main components is the goal here, using the feature vector obtained from the eigenvectors of the covariance matrix. The transpose of the feature vector is multiplied by the transpose of the original dataset in this transformation. By recasting the data in a new coordinate system determined by the primary components, this step effectively ends the PCA process and provides a more effective and insightful representation for additional research.

The steps below will demonstrate how the PCA approach works:

Step 1: Obtain the data set.

Step 2: Data representation in a structure.

Step 3: Standardization of data

Step 4: Z's covariance is calculated.

Step 5: Eigenvalues and eigenvectors are calculated.

Step 6: The eigenvectors are categorized.

Step 7: New characteristics or primary components are calculated.

Step 8: Remove characteristics from the new dataset that are lesssignificant or irrelevant.

XI. PRINCIPAL COMPONENT ANALYSIS LIMITATIONS

Linearity Assumption: PCA makes the assumption that the data has a linear relationship with the principal components. Other nonlinear dimensionality reduction techniques could be better suitable if the underlying data has complicated nonlinear interactions, as PCA might not be able to capture all the pertinent information.

Loss of Interpretability: Although principal component analysis (PCA) decreases the dimensionality of the data, the primary components that are produced are typically combinations of the original variables, which complicates their interpretation. When working with a large number of components, it may be difficult to comprehend the modified characteristics.

Sensitivity to Outliers: Because outliers can disproportionately affect the estimate of main components, PCA is sensitive to outliers in the data. The quality of dimensionality reduction can be impacted by outliers, which can also skew the resultant variance-covariance structure.

Choosing the Right Number of Components: Choosing how many major components to keep is a personal choice. Retaining too many components might cause overfitting or needless complexity in the data representation, while selecting too few can result in a major loss of information. It is crucial to take these restrictions into account and evaluate if PCA is appropriate for a given dataset and analytical objective. Other dimensionality reduction methods or tailored strategies could be a better fit in some situations to deal with particular issues or demands.

XII. ADVANCEMENTS IN PCA

Kernel PCA: By using a kernel function, kernel PCA extends PCA to nonlinear dimensionality reduction to create a higher-dimensional feature space map from the input. It permits the capture of intricate correlations between factors This is especially helpful when there are nonlinear frameworks.

Sparse PCA: This technique takes sparsity limitations into account into the PCA structure, encouraging the discovery of a small number of primary components. This is advantageous when handling high dimensional data, where a small number of Variables have a major role in the data structure.

Incremental PCA: This technique allows the using PCA to analyze huge datasets that don't fit into recall. It handles data sequentially or in batches. modernizes the key elements, increasing its both scalable and computationally efficient.

Online PCA: Online PCA modifies PCA to fit streaming data in which fresh observations are added all the time. By gradually updating the main components, it enables real-time These advancements effectively tackle specific challenges and meet the unique requirements of various datasets and scenarios.

XIII. COMPARE PCA ALGORITHM WITH UNSUPERVISED LEARNING ALGORITHMS

Factors	PCA	Apriori	ECLAT	Frequent Pattern Growth	K-Means
Accuracy in General	High	Moderate	Moderate	Moderate to High	High
Speed Of Learning	Fast	Moderate	Moderate	Moderate	Fast
Speed Of Classification	Fast	Moderate	Moderate	Moderate	Fast
Tolerance to Missing Values	Moderate	High	High	Low to Moderate	Low to Moderate

Tolerance to Irrelevant Attributes	High	Low to Moderate	Low	Moderate	Low to Moderate
Tolerance to Redundant Attributes	High	Low to Moderate	Low	Moderate	Low to Moderate
Tolerance to Highly Interdependent Attributes	High	Low to Moderate	Low	Moderate	Low to Moderate
Tolerance To Noise	High	Low to Moderate	Low	Moderate	Low to Moderate
Dealing with Danger Of OverfiFng	Low	Low	Low	Moderate	Moderate
Attempts for Incremental Learning	Limited	N/A	N/A	Limited	Limited
Transparency of Knowledge/Classification	Moderate	Low	Low	Low to Moderate	Low
Support Multi- Classification	Yes	No	No	No	Yes

XIV. KERNEL

A kernel is a function that computes the similarity (or distance) between two data points in a high-dimensional space. Kernels play a crucial role in kernelized algorithms, such as Support Vector Machines.

Support Vector Machines (SVMs) are a type of supervised learning algorithm used for classification and regression tasks. SVMs work by finding the hyperplane that best separates different classes in the input space. However, when the data is not linearly separable, meaning there is no straight line that can perfectly separate the classes, SVMs use a technique called kernel trick.

The kernel trick involves transforming the input data into a higher-dimensional space where it may become linearly separable. This transformation is often implicit and is performed by a kernel function. The kernel function calculates the dot product of the transformed data points without explicitly computing the transformation. Common kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid kernels.

KERNEL PCA

In the realm of dimensionality reduction, Kernel Principal Component Analysis (Kernel PCA) has emerged as a powerful tool capable of handling nonlinear data structures, thereby offering enhanced insights into high-dimensional datasets. Traditional methods like Principal Component Analysis (PCA) excel in capturing linear relationships within data but fall short when confronted with complex, nonlinear patterns that are prevalent in real-world datasets. Kernel PCA addresses this limitation by extending the PCA framework to operate in a higher-dimensional feature space via kernel functions, enabling the extraction of nonlinear principal components.

The essence of Kernel PCA lies in its ability to transform data into a space where linear separation becomes feasible, even in the presence of intricate nonlinear relationships. This transformation is achieved through the application of kernel functions, which map the original data into a higher-dimensional space where nonlinear patterns can be effectively captured. By leveraging the kernel trick, Kernel PCA facilitates the extraction of principal components that maximize the variance of the projected data while preserving essential structural information.

The essence of Kernel PCA lies in its ability to transform data into a space where linear separation becomes feasible, even in the presence of intricate nonlinear relationships. This transformation is facilitated by kernel functions, which implicitly map the original data into a higher-dimensional feature space. By extracting principal components in this transformed space, Kernel PCA enables the representation of data in a reduced-dimensional subspace while preserving essential structural information.

In this paper, we introduce Kernel PCA as a versatile tool for dimensionality reduction and nonlinear feature extraction. We discuss its theoretical foundations, algorithmic implementation, and practical applications across various domains. Additionally, we highlight current research trends and future directions, emphasizing the growing importance of Kernel PCA in modern data analysis and machine learning methodologies. Finally, we present a concise overview of Kernel PCA as an integral component of advanced data analysis pipelines.

To understand the utility of kernel PCA, particularly for clustering, observe that, while N points cannot, in general, be linearly separated in $d < N$ dimensions, they can almost always be linearly separated in dzN dimensions. That is, given N points X_i if we map them to an N-dimensional space with

$$\boldsymbol{\Phi}(\mathbf{x}_i) \text{ where } \boldsymbol{\Phi}: \mathbb{R}^d \to \mathbb{R}^N$$

A hyperplane that splits the points into random clusters is simple $\Phi(\mathbf{x})$ to build. Naturally, this produces linearly independent vectors, thus there isn't any covariance to use for directly doing eigendecomposition, unlike in linear PCA.

Rather, in kernel PCA, an arbitrary, non-trivial $(x) is 'selected,' which is never explicitly computed, opening the door to the usage of very-high dimensional $\Phi(x)$'s if we never have to evaluate the data in that space. We may construct the N-by- N kernel because we often want to avoid operating in the space, we can create the N —-by-N kernel which we shall refer to as the feature space.

$$\mathbf{AK} = \mathbf{k}(\mathbf{x}, \mathbf{y}) = (\boldsymbol{\Phi}(\mathbf{x}), \boldsymbol{\Phi}(\mathbf{y})) = \boldsymbol{\Phi}(\mathbf{x})^{\mathbf{T}} \boldsymbol{\Phi}(\mathbf{y})$$

It is a representation of the otherwise unmanageable feature space's inner product space (see matrix). We may technically describe a variant of PCA in which we never really solve the eigenvectors and eigenvalues of the covariance matrix in thestapace, thanks to the dual form that develops in the process of creating a kernel (see Kernel trick). The dot product of one converted data point with respect to all other transformed points (N points) is represented by the N-elements in each of the k columns.

The kernel-formulation of PCA is limited in that it computes our data's projections onto the principal components rather than the principal components themselves because we are never working directly in the feature space. To assess how a point in the feature $\Phi(x)$ space is projected onto the kth principal component, or Vk (where superscript k denotes the component itself, not powers of k).

$$\mathbf{V}^{\mathbf{k}^{\mathbf{T}}} \boldsymbol{\Phi}(\mathbf{x}) = \left(\sum_{\mathbf{i}=1}^{\mathbf{N}} \mathbf{a}_{\mathbf{i}}^{\mathbf{k}} \boldsymbol{\Phi}(\mathbf{x}_{\mathbf{i}}) \right)^{\mathbf{T}} \boldsymbol{\Phi}(\mathbf{x})$$

TIt should be noted $\Phi(x_i)^T \Phi(x)$ that stands for dot product, which is just the kernel K's constituent parts. It appears that the eigenvector equation must be solved in order to compute and normalize the a_i^k

$$\mathbf{N\lambda a} = \mathbf{Ka}$$

where N is the number of data points in the set, and λ and $\mathbf{a}$ are the eigenvalues and eigenvectors of K. Then to normalize the eigenvectors $\mathbf{a}^k$, we require that

$$\mathbf{1} = \left(\mathbf{V}^{\mathbf{k}}\right)^{\mathbf{T}} \mathbf{V}^{\mathbf{k}}$$

It is important to keep in mind that X may not be centered in the feature space (which we never compute directly), regardless of whether it has zero-mean in its own space. Given that centered data is necessary for a principal component analysis to be conducted effectively. We "centralize" to turn K into K′

$$\mathbf{K}' = \mathbf{K} - \mathbf{1_N K} - \mathbf{K 1_N} + \mathbf{1_N K 1_N}$$

where an N-by- N matrix with a value of 1/N for each element is indicated by 1 N. Using K, we carry out the previously mentioned kernel PCA technique.

Here, it is important to highlight one kernel PCA caution. Regarding linear PCA, the eigenvalues may be utilized to order the eigenvectors according to the extent to which each principal component accounts for the variation in the data. KPCA could benefit from this as well as data dimensionality reduction. Nonetheless, in actuality, there are instances where every data variant is same. A poor selection of kernel scale is usually the cause of this.

XV. PATTERN CLASSIFICATION FOR SYNTHETIC DATA

We look into the outcomes of our dimensionality reduction methods applied to the specialized dataset known as two-concentric- spheres. Three techniques were employed: traditional PCA, polynomial kernel PCA with a

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

degree (d) set to 5, and Gaussian kernel PCA with a standard deviation (σ) of 20. The primary objective is to transform the original 3-dimensional data into a more manageable 2-dimensional space, offering insights into the efficacy of each approach in capturing the intrinsic structure of the data.

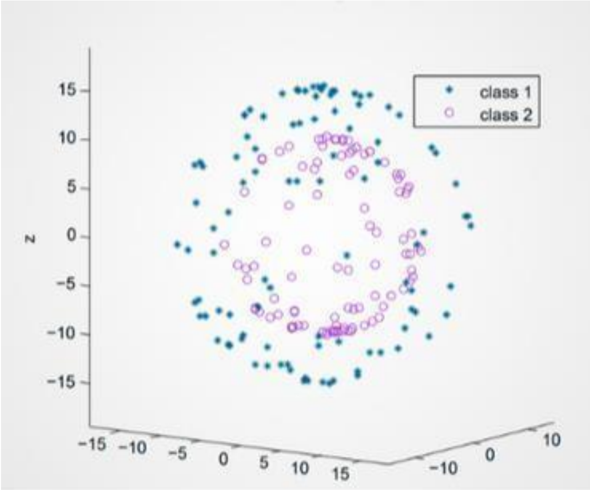

Figure 5: 3D plot of the Data Description

Figure 6 illustrates the traditional PCA results, which, regrettably, fail to provide a coherent representation of the data. Thepoints exhibit a scattered pattern, and discerning any meaningful separation between the two distinct classes is challenging.

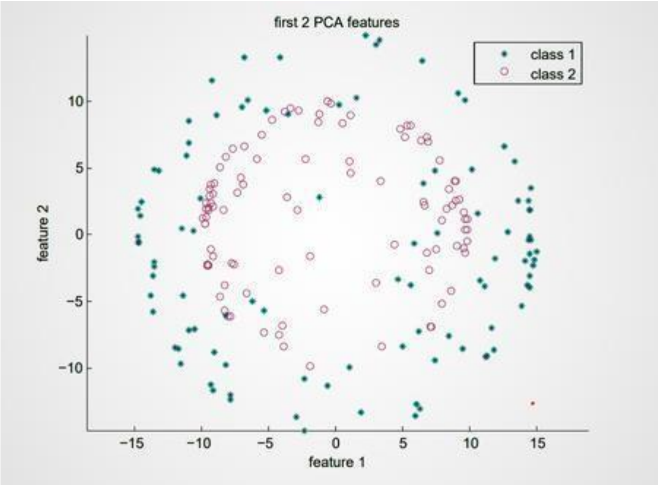

Figure 6: Traditional PCA Results

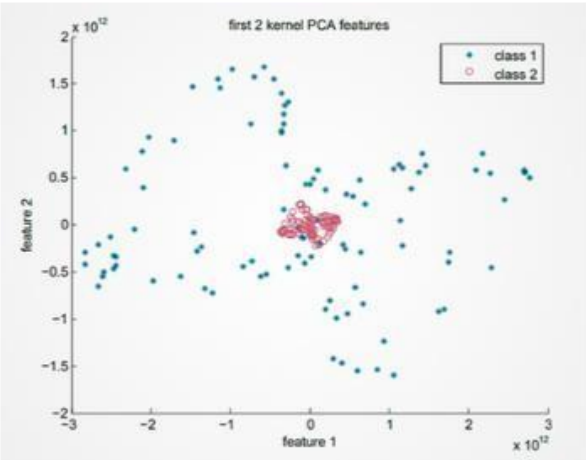

Figure 7: Polynomial Kernel PCA Results (**d**=5)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

Moving on to Figure 7, portraying the outcomes of polynomial kernel PCA with a degree (**d**) set to 5, there is an observable improvement. Class 1 data points tend to form clusters, although class 2 points still lack a clear and structured arrangement.

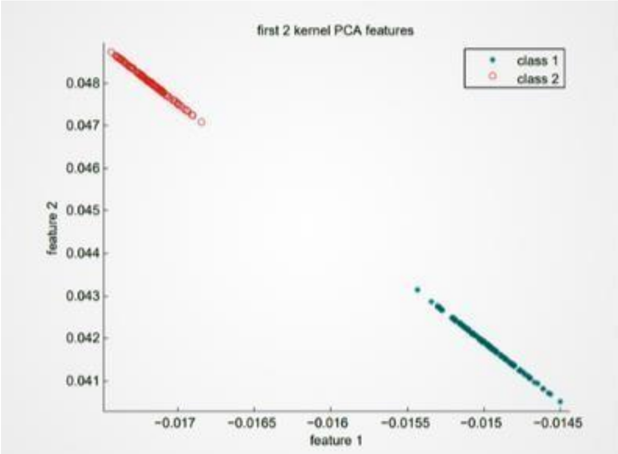

Figure 8: Gaussian Kernel PCA Results (**σ**=2)

Showcases the results of employing Gaussian kernel PCA with a standard deviation (**σ**) of 20. A notable enhancement is observed, with the two classes becoming entirely linearly separable in the new feature space. Furthermore, both features effectively capture and represent the radius information inherent in the original data.

In summary, traditional PCA falls short in revealing the underlying structure, while both polynomial and Gaussian kernel PCA show promise. Polynomial kernel PCA results in some grouping of class 1 points, while Gaussian kernel PCA achieves complete linear separability between the classes. These findings substantiate our assertions and contribute significantly to the comprehensibility of our research paper.

XVI. CONCLUSION

This research provides a succinct exploration of Principal Component Analysis (PCA) in the context of unsupervised learning.Beginning with a concise introduction and historical context, we navigated through the mathematical intricacies, algorithmic steps, and practical applications of PCA, shedding light on its strengths and limitations.

A comparative analysis juxtaposing PCA against various unsupervised learning algorithms, including frequent pattern approaches, offered valuable insights into PCA's unique contributions. Additionally, we introduced Kernel PCA as a potent extension tailored for non-linear data, exemplified by its application on the two-concentric-spheres dataset.

This study offers a brief yet comprehensive understanding of PCA in unsupervised learning, paving the way for future exploration of advanced dimensionality reduction techniques. It emphasizes PCA's significance while opening avenues for continued research into sophisticated methodologies like Kernel PCA.

XVII. REFERENCES

[1] Hodtson F, Kendall D, Tautu P, eds. Mathemematics in the Archæological and Historical Sciences.

[2] Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 25: 417-441.

[3] Hwang H, Tomiuk MA and Takane Y. 2009. Correspondence analysis, multiple correspondence analysis and recentdevelopments. In: Millsap R, Maydeu-Olivares A, eds.

[4] Jackson JE. 1991. A User's Guide to Principal Components. New York: John Wiley & Sons.

[5] Jolliffe IT. 2002. Principal Component Analysis. New York: Springer.

[6] Jordan C. 1874. Memoire´ sur les formes bilineaires´. J Math Pure Appl. 19: 35-54.

[7] Lay D. 2000. Linear Algebra and It's Applications. Addison-Wesley, New York.

[8] Mitra P and Pesaran B. 1999. "Analysis of Dynamic Brain Imaging Data". Biophysical Journal. 76: 691-708.

[9] Peres-Neto PR, Jackson DA and Somers KM. 2005. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Comput Stat Data Anal. 49: 974-997.

[10] Gower J. 1971. Statistical methods of comparing different multivariate analyses of the same data.

[11] Handbook of Quantitative Methods in Psychology. London: Pearson K. 1901. On lines and planes of closest fit to systems ofpoints in space. Philos Mag A. 6: 559-572.

[12] Quenouille M. 1956. Notes on bias and estimation. Biometrika. 43: 353-360.

[13] Saporta G and Niang N. 2006. Correspondence analysis and classification. In: Greenacre M, Blasius J, eds. MultipleCorrespondence Analysis and Related Methods. Boca Raton, FL: Chapman & Hall. 371-392.

[14] Levinson, William A. Statistical process control for real-world applications. CRC Press, 2011.

[15] Chakra Chennubhotla and Allan D. Jepson. Sparse PCA: Extracting multi-scale structure from data International Conferenceon Computer Vision, Vancouver, July 2001.

[16] Shinde, R. L., and K. G. Khadse. "Multivariate process capability using principal component analysis." Quality and Reliability Engineering International 25, no. 1 (2009): 69-77.

[17] Stojanovic, Branka, and Aleksandar Neskovic. "Impact of PCA based fingerprint compression on matching performance." In Telecommunications Forum (TELFOR), 2012 20th, pp. 693-696. IEEE, 2012.

[18] Lin, Y., & Zhang, Z. (2016). Principal component analysis (PCA). In Data mining with R: A tutorial (pp. 119-139). Springer.

[19] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (7th ed.). Springer.

[20] Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433-459.

[21] Lay, David. (2000). Linear Algebra and It's Applications. Addison-Wesley, New York. [This is a beautiful text. Chapter 7 in the second edition (p. 441-486) has an exquisite, intuitivederivation and discussion of SVD and PCA.

[22] Ganna, A. et al. Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. Science 365,eaat7693. https://doi.org/10.1126/science.aat7693 (2019).

[23] Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209.https://doi.org/10.1038/s41586-018-0579-z (2018).

[24] Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strongbackground selection. Nat. Genet. 50, 381–389. https://doi.org/10.1038/s41588-018-0059-2 (2018).

[25] Li, J. Z. et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100–1104.https://doi.org/10.1126/science.1153717 (2008).

[26] Pugach, I., Delfin, F., Gunnarsdottir, E., Kayser, M. & Stoneking, M. Genome-wide data substantiate Holocene gene flowfrom India to Australia. Proc. Natl. Acad. Sci. USA 110, 1803–1808.
https://doi.org/10.1073/pnas.1211927110 (2013).

[27] Goldfarb, S. Ma, and K. Scheinberg. Fast alternating linearization methods for minimizing the sum of two convexfunctions. Mathematical Programming, Series A, 141:349–382, 2013.