ISSN 2959-6157

Exploring Principal Component Analysis: A Comprehensive Survey

Ruixin Yuan

Chengdu Shude High School International Department, Chengdu, 610000, China

Abstract

In many research fields, high-dimensional data poses significant difficulties for experimental algorithms and simulations. The pro-post of Principal Component Analysis (PCA) provides an important preprocessing idea for the study of large sample experiments. PCA achieves information refinement by constructing an uncorrelated variable. Based on basic and generalized ideas of PCA, we illustrate the feasibility and flexibility of its application in a wide range of involved fields. It demonstrates that PCA is a worthy method not only in different fields but also in specific applications as a classical theory that can be further investigated in depth according to the characteristics of the data.

Keywords: principal component analysis, preprocessing method, dimension reduction

1 Introduction

Principal component analysis (PCA) is the most common way to reduce the data dimension. The main idea of PCA is to reduce the dimension of data and simultaneously minimize the loss of information. The work of principal component analysis is to find a set of mutually orthogonal coordinate axes sequentially from the original space, and the choice of new coordinate axes is closely related to the data. The first new axis is chosen in the direction in which initial data has the largest variance, then the second new axis, which is orthogonal to the first axis and has the largest variance. We can get the K axis to include most of the variance to reduce the dimension of data.

PCA was proposed by Karl Pearson in 1901 and used to reduce the dimension of data and help understand and analyze the data. In the 1980s, with the development of computer technology, PCA was popularized and applied in image processing [6], and its advantages in data reduction and feature extraction began to appear. Up to now, PCA has played an important role in many areas.

In Section 2, we review the basic concept of PCA. Section 3 concludes the valuable research of PCA in finance and face recognition areas. Finally, the conclusion is presented in Section 4.

In this paper, we denote X^T by the transport of vector X. Using $E_p(\xi)$ to express the mean value (expected value) for random variable ξ. R^p denotes the set of all p-dimensional real-valued vectors.

2 Theory chapters

2.1 Correlation of statistical concepts

Random variable and random vector

A random variable is a variable whose value is randomly selected; for example, when through the coin, suppose that the probabilities of each result are equal, we can use random variable ξ to quantization results:

$$\xi('' \text{ face upward}'') := 1,\ \xi('' \text{ face downward}'') := 2. \quad (1)$$

Denote ω_i; i = 1, . . . , 6, show simple random sampling, the results of a random event, these sample constitute sample space $\Omega = \{\omega_i\}_{i=1}^{6}$

$$\xi : \Omega \to \{1, . . . , 6\},\ \xi(\omega_i) = i, i = 1, . . . , 6. \quad (2)$$

The distribution of random variables is:

$$P_{\omega_i} := \frac{1}{6},\ i = 1, . . . , 6. \quad (3)$$

A random vector is a vector whose components are all random variables. A random vector can express the results of throwing the coin. Suppose we throw the coin n times, the quantization of each result is random X_j, j = 1,...,n, we can use an-dimensions vector $\mathbf{X} = (X_1,..., X_n)^T$, a random variable is a special random vector.

The characteristics of the sample and overall (discrete type)

For the random vector **X**, suppose that the support of **X** is $R_x := \{\mathbf{a}_1, \mathbf{a}_2, . . . \}$, the distribution of corresponding R_x is:

$$P = \{p_1, p_2, . . . \}. \quad (4)$$

For any $\mathbf{a}_j$, j=1,2,..., the probability of the value is denoted as $p_j \geq 0$, and the basic condition is $\Sigma_j p_j = 1$.

Mean is the weighted average of random variable values under corresponding probabilities; the computing method is:

$$EP(\mathbf{X}) := \sum_j pj\ \mathbf{a}j. \quad (5)$$

The mean value is a basic and important characteristic of a random variable that shows the average standard. The probability of a value is the weight.

Denote VarP by the variance operator. Variance is a non-negative value used to measure the disorder level of random variables; the computing method is:

$$VarP(\mathbf{X}) := E[(\mathbf{X} - Ep(\mathbf{X})2)] = \sum_j pj(\mathbf{a}j - EP(\mathbf{X}))2 . \quad (6)$$

The square root of variance var (**X**) is called standard deviation.

When we have a sample {**X**1, . . . , **X**N }, we can use the method below to calculate the mean and variance:

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_i, \quad \sigma_j^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{X}_{ij} - \mu_j\right)^2 \quad (7)$$

Where μj is the jth component of **μ**, the equivalent probability 1/N of each sample point is assumed.

2.2 The definition of principal component

Let

$$\mathbf{Y} := (Y1, ..., Yp)T \in Rp$$

be a realization of some random data. There are two steps to reduce the dimension of the **P** random variables.

1. Find the principal component

Let **α1** = (α11,...,α 1p) ∈ Rp be a vector of constants. Find the maximum

variance of $\alpha_1^T Y$, which is the linear transformation of element **Y**

$$\alpha_1^T Y = \sum_{i=1} \alpha\ 1iYi = \alpha\ 11\ Y1 + \cdots + \alpha 1pYp \in R, \quad (8)$$

which is called the first principal component of **Y**.

2. Find the other principal components

Find $\alpha_2^T Y$, which is uncorrelated with $\alpha 1^T Y$, and calculate the maximum

variance. This is the second principal component. Repeat these steps to find all

principal components $\alpha_k^T Y$, k = 1,...,k.

3. Maximization of variance

The variance maximization problem could be reformulated as follows:

$$\underset{\alpha \in}{m}\overset{ax}{Rp}\ Var(\alpha TY) = \underset{\alpha \in}{m}\overset{ax}{Rp}\ \alpha T\ \Sigma\alpha, \quad (9)$$

Where Σ is the covariance matrix of **Y**.

Then, use the Lagrange multiplier method to solve the equation. We can get the

value of V and λ. Take the derivative of the following formula.

$$L(\mathbf{e},\lambda) = \mathbf{e}T\ \mathbf{YY}\ T\mathbf{e} - \lambda(\mathbf{e}T\mathbf{e} - 1) \quad (10)$$

We can get:

$$\partial\mathbf{e} = 2\mathbf{YY}\ \mathbf{e} - 2\lambda\mathbf{e} = 0,\ \partial\lambda = -\mathbf{ee} + 1 = 0 \quad (11)$$

Solve it can get:

$$\mathbf{YY}\ T \cdot \mathbf{ee} = \lambda\mathbf{e} \quad (12)$$

In the above equations, **e** is called the eigenvector (corresponds to λ) of **Y**, and also we can get:

$$\mathbf{YY}\ T \cdot \mathbf{e} = \lambda\mathbf{ee}T\ \mathbf{YY}\ T \cdot \mathbf{e} = \lambda\mathbf{e} = \lambda \quad (13)$$

So, we can prove that λ is also an eigenvalue of **e**.

3 Recently research

3.1 The application of PCA in finance

Since data in the financial area has a strong linear relationship, PCA is widely applied in the financial area. It is always used to analyze financial market data and control the allocation of assets. In optimizing investment portfolios, using PCA can identify the relationship between different properties, helping investors build a risk-diversified portfolio. In risk management, PCA can be used to identify the main risk factors in the financial market, helping financial institutions recognize the essential risks. PCA can analyze the price of a stock, using time series data such as interest rate and exchange rate to capture the trend and period of the market. In the following part, we will introduce the recent application of PCA in the fiance area in detail.

3.1.1 Predict stock price

1. Combine NeuroEvolution and PCA

In the stock market, determining the trading signal that can achieve the visitor's aim is a very popular research direction. This is a complicated task. Recently, there has been some research using deep learning methods in the financial area to find suitable entry and exit points in the stock market. In 2018, Nadkarni and Neves [9] proposed an approach combining a NeuroEvolution of Augmenting Topologies (NEAT), with the technique PCA, which is used for reducing data dimension. This method considers the earnings, daily profits, investing risk, and investment time. They also test the performance of this algorithm. They used the different markets' daily volume and prices (open, high, low, and adjusted close) from 27/03/2006 to 13/04/2017 to train and test the implemented system (80% data used for training, 20% for testing). They measure the performance of NEAT and PCA by using four indicators, including returns obtained (ROR), risk-return ratio (RR), mean daily profits (MDP), and returns obtained by day in the market.

(ROR/day). Also considered is the time that the capital invested in the market (Market Days) and the maximum drawdown (MDD) of the returns. They select eight representative stocks from the eight highest market capitalization sectors of the S&P 500 index for testing. The results showed that the implemented system is robust as it performs better in Buying and Holding in most tested stocks.

2. Combine PCA, Discrete Wavelet Transform, and XGBoost

Researchers continue to find more models that can be used to determine the trading point in financial markets. In 2019, Nobre and Neves [10] proposed a new model that combines Principal Component Analysis (PCA), Discrete Wavelet Transform (DWT), Extreme Gradient Boosting (XGBoost), and Multi-Objective Optimization Genetic Algorithm (MOO-GA). They used PCA to reduce the dimension of the input financial dataset and DWT to reduce the noise of each feature. The processed dataset was then fed into the XGBoost binary classifier, the hyperparameters optimized by MOO-GA. They analyze the effectiveness of PCA, where the reduction of data dimension can better identify patterns in the processed data, resulting in a better generalization ability and higher accuracy of the XGBoost binary classifier. The results show that PCA improves the system's performance, especially at high oscillating periods.

3. Scaled PCA

In 2022, Huang et al. [2] proposed Scaled PCA. This method scales each predictor according to its ability to predict the target before prediction. Generally, Scaled PCA has two steps. First, a target prediction regression for each predictor's lagged value is performed to assess its predictive power. Factors were then extracted from these scaled predictors using the PCA method: underweight variables with weak predictive power and overweight variables with strong predictive power. The experiments show that scaled PCA outperforms compared with PCA because PCA treats all standardized predictors equally and completely ignores the target information; when one predictor is noisier than the others, it inevitably affects the factor weights disproportionately, but Scaled PCA takes into account the target in the downscaling process. In their article, they mentioned the application of Scaled PCA to predict the stock market in terms of investor sentiment. Also, an application to predict inflation with a large panel of macro variables. While the PCA does not show significant pre-predictive power, the Scaled PCA shows significant predictive power within and outside the 1- to 12-month forecast range.

3.1.2 Analysis of the structure of financial markets

In 2011, Fenn et al. [1] used random matrix theory and PCA to analyze the relationship in the financial market. They use random matrix theory to demonstrate that the correlation matrices of asset price changes contain a structure incompatible with uncorrelated random price change. Then they use

PCA to testify that a fraction of components could interpret a large fraction of the market variability. They study the relevance between the asset price time series and principal components to characterize the time-evolving relationships between the different assets. They also found that the relationship between different markets increased after the 2007-2008 liquidity and credit crisis.

In 2014, Lin, Shang, and Zhou [7] also used PCA in their research on financial market structure. They measure the multiscale behavior and interactions between stock markets by using multifractal detrended cross-correlation analysis (MF-DXA: combine MF-DFA and DCCA). They get many conclusions. The market within one geographic region has a higher level of correlation. Two financial crises (the global financial crisis and the Asian financial crisis) led to an increase in stock market correlations. Market correlation was stronger during the global financial crisis than during the Asian financial crisis. In general, the cross-correlation increased during the crisis. In normal times, the US stock market plays an important role. In times of crisis, the US stock market also retains an important role. Meanwhile, the influence of China's stock market is rising.

3.1.3 Summary

There is a large amount of data in the financial area, and a large part belongs to time series data. PCA has many advantages for reducing the dimension of such data. The accuracy of predicting the price trend of financial products will be improved, and financial markets can be analyzed more quickly and clearly. In conclusion, the application of PCA greatly helps researchers in the financial area, and many researchers in the financial industry are using PCA to reduce the dimension of data to help them work (Table 1).

Table 1: Related research of PCA in the financial area

Year	Author	Application Areas	View
2009	Hubert M et al. [3]	Analysis of financial data	Robust PCA proposed
2018	Nadkarni J et al. [9]	Trade in financial markets	NeuoEvolution and PCA

2019	Nobre J et al. [10]	Trade in financial markets	PCA in discrete wavelet transform
2021	Zheng L et al. [15]	Predict share price	PCA hybrid prediction model
2022	Huang D et al. [2]	Forecast market return	Scaled PCA proposed

3.2 The application of PCA in face recognition

Nowadays, there are a lot of scenarios where we use face recognition. Face recognition always includes these key procedures: (1) image acquisition using camera- capture equipment. (2) face detection: using a classifier to locate the face part of the image. (3) face preprocessing: adjust the detected face image (including size, angle, color, etc.) (4) feature extraction: face image contains a large number of pixels (features), using dimension reduction algorithm to extract important features (5) face mating and recognition: compared

Extracted features with the features in the database to complete face recognition. (6) output results. The (4) is a key procedure of face recognition, PCA, and some related algorithms always used to extract important features in a large number of features, reduce the computational load in the recognition process, and improve the recognition performance.

PCA still has great potential for improvement due to its recognition speed, accuracy, and cost. So, based on PCA, the researchers optimized from different angles and then proposed kernel PCA [11], 2DPCA [13], and other algorithms. The following sections will introduce some of the most influential PCA optimization algorithms and their application.

3.2.1 Theory

1. Kernel PCA

PCA is suitable for processing linear data, but it is not effective for nonlinear data. So, in 1998, Scholkop et al. [11] proposed kernel PCA, a algorithm that processes nonlinear data.

2. 2D PCA

Since PCA could not capture the simplest invariance, some scientists proposed independent component analysis (ICA) and used kernel PCA to extract face features. These two methods outperformed the classical eigenfaces method. However, when doing computation, both kernel PCA and ICA are more expensive than PCA. So, in 2004, Yang et al. [13] put forward 2DPCA. Compared with PCA, 2DPCA directly uses 2D image matrices and does not translate them into a 1D vector, which better preserves the original data and reduces the amount of computation (the size of the image covariance matrix using 2DPCA is much smaller).

However, 2DPCA has a key shortage: it needs more coefficients to represent images than PCA. Researchers all used PCA after 2DPCA to solve this problem. In 2005, Zhang et al. [14] proposed an algorithm (2D)2 PCA to solve this problem. The main difference between 2DPCA and (2D)2 PCA is that 2DPCA works only in the row direction of the face images, while (2D)2 PCA works in both row and column directions. Since (2D)2 PCA analyzes both row and column directions, the algorithm is particularly suitable for analyzing images with important information in both row and column directions.

3. Sparse PCA

It is difficult to interpret the results because all components after using PCA are linear combinations of the original variables. Subsequently, in 2006, Zou et al. [18] proposed a new method, sparse PCA. They used elastic nets (lasso), a promising variable selection technique, to generate modified principal components with sparse loadings. SPCA is based on the fact that PCA can be written as a regression-type optimization problem with a quadratic penalty; the lasso penalty (via the elastic net) is directly integrated into the regression criterion to obtain an improved PCA under sparse loadings.

4. Robust PCA

Since PCA is very sensitive to outliers, it has facilitated the development of robust techniques. Researchers developed an improved robust PCA method applicable to skewed data. In 2009, Hubert et al. [3] proposed robust PCA. ROBPCA combines projection pursuit (PP) and robust covariance estimation, which is well suited for analyzing high-dimensional data and has been applied to multivariate calibration and classification.

3.2.2 Application

1. Kernel PCA

Using kernel PCA, we can map the low-dimensional data to a high-dimensional space, showing the nonlinear relationships in the original data. Kernel PCA has high flexibility. We can choose different kernel functions to meet different data and needs when using kernel PCA. Scholkop et al. [11] experimentally examined the performance of nonlinear PCA. Experiments on character recognition. They used kernel PCA to extract nonlinear principal components from a database of handwritten characters.

The results show that the nonlinear principal components have higher recognition accuracy than the corresponding number of linear principal components. The performance of the nonlinear components is improved more than that of the linear components by using more components. In conclusion, kernel PCA only requires solving an eigenvalue problem without nonlinear optimization compared to other feature extraction techniques. Therefore, kernel PCA can be applied to all areas where traditional PCA and nonlinear extensions make sense.

2. 2D PCA

Yang et al. [13] used three well-known face image databases (ORL, AR, Yale) experimentally to test face recognition performance using 2DPCA. They used the ORL database to evaluate the performance of 2DPCA under conditions carrying different poses and sample sizes. The AR database was used to test the system's performance under different facial expressions, illumination conditions, and occlusions. The Yale database was used to test the system's performance regarding facial expression and illumination changes. In the experiments based on the ORL database, they compared 2DPCA with Fisherfaces, ICA, and kernel eigenfaces, and the results show that 2DPCA performs better in almost all conditions. In experiments based on the AR database, the results show that 2DPCA has higher recognition accuracy and faster feature extraction speed than PCA under light variation conditions. The experiments based on the Yale database show that 2DPCA outperforms PCA, ICA, and kernel eigenfaces in terms of recognition rate. The results of all tests (based on ORL, AR, and Yale databases) show that 2DPCA has more advantages than PCA: it extracts image features directly, and 2DPCA has higher accuracy and faster computation speed. Since 2DPCA can directly process two-dimensional data, retain more information, and have fast computation speed, high accuracy,

At low cost, the algorithm is particularly suitable for processing and analyzing two-dimensional data such as images.

After Zhang et al. [14] proposed (2D)2 PCA in 2005, they conducted evaluation experiments on PCA, 2DPCA, alternative 2DPCA, and (2D)2 PCA based on two well-known face databases, ORL and FERET. Experiments based on the ORL database show that (2D)2 PCA is more accurate than PCA, and (2D)2 PCA has the shortest runtime. Experiments based on the FERET database compare the ability of PCA, 2DPCA, and (2D)2 PCA to represent face images at similar compression ratios. The results show that (2D)2 PCA generates higher-quality images than the other two methods with the same storage capacity.

3. Sparse PCA

In the article by Zou in 2006 [18], they introduce the methodological details of SPCA. An effective SPCA model fitting algorithm is proposed to derive suitable expressions for correcting the variance explained by the principal components. They also considered special cases of the SPCA algorithms that effectively handle gene expression arrays. They illustrate the proposed method with real data and simulation examples. The traditional lasso has some limitations [17], so they proposed the elastic net. The elastic net penalty is a convex combination of the ridge and lasso penalty. In this article, they compared SPCA with lasso constraints in PCA and SCoTLASS [4]. Compared to using lasso constraints in PCA, SCoTLASS successfully derives sparse loading but is not computationally efficient. When the spar- sity (lasso) penalty term disappears, accurate PCA results are obtained using SPCA. The sparse structure of the SPCA control makes the loading more flexible. SPCA has several advantages, including computational efficiency, high explained variance, and the ability to identify important variables.

4. Robust PCA

This algorithm consists of six steps. First, if the number of variables is much larger than the observation, use singular value decomposition to reduce the dimension for original data without losing information. Next, choose the coverage α between 1/2 and 1 (the default value is α = 0.75). Set h = [αn] (denote [.] the integer part). The value of coverage α determines the robustness and effectiveness of this method. The robust is stronger and less accurate when α is smaller. Then, outliers will be calculated based on the definition proposed by Stahel (1981)and Dono ho and Gasko (1992). After that, all data will be projected onto the k-dimensional subspace spanned by the first K eigenvectors of the robust cover- ance estimator obtained in Step 3 to reduce the data dimension. Subsequently, the orthogonal distances to the subspace for each observation are calculated, and an improved robust subspace estimate is obtained based on these distances. Finally, the weighted minimum covariance determinant (MCD) calculates the robust center and covariance matrix in the k-dimension subspace. The eigenvectors of this robust covariance matrix are the final principal components. In conclusion, experiments applying ROBPCA to both real and simulated data

This shows that the method accurately estimates the PCA subspace and distinguishes between regular observations and outliers.

3.2.3 Summary

The development of face recognition technology is synchronized with the improvement and optimization of PCA.

The most common improvement of PCA is improved the speed and accuracy of face recognition. In addition, there are also improved algorithms that reduce the required input data, reduce costs, and so on. In conclusion, PCA is widely used in the area of facial recognition. There are a lot of relevant studies (Table 2).

Table 2: Related research of PCA in the face recognition area

Year	Author	Application Areas	View
1998	Scholkopfsch B et.al[11]	Nonlinear feature extraction	Nonlinear component analysis
1999	Tipping M et al. [12]	Probabilistic dimension reduction	Probabilistic principal component analysis
2002	Kim K et al. [5]	Extract facial feature	Apply kernel principal component analysis
2004	Yang J et al. [13]	Image projection technique	Two-dimensional PCA
2005	Daoqiang Zhang et al. [14]	Face representation and recognition	(2D)2PCA
2006	Zou H et al. [18]	Data processing and dimensionality reduction	Sparse PCA
2009	Hubert M et al. [3]	Analysis high-dimensional data	Robust PCA
2011	Mohammed A et al. [8]	Human face recognition algorithm	Bidirectional two-dimensional PCA
2018	Zhu Y et al. [16]	Face feature representation	Improved PCA

4 Conclusion

With the development of related computer technology and the continuous improvement of theories, the application of PCA has become more and more widespread in recent years, mainly focused on the areas that need to process large amounts of data. In the financial area, PCA is used to predict the price trend of financial products and analyze the structure of financial markets. The experiments show that in the investment, the return of portfolios that use PCA is significantly better than other portfolios. Using PCA helps researchers to quickly and analyze the financial market structure. The application of PCA in face recognition has improved the speed and accuracy of face recognition and reduced costs. At the same time, more improved algorithms (2D PCA/kernel PCA/Scaled PCA, etc.) have further improved face recognition techniques.

References

[1] Daniel J Fenn, Mason A Porter, Stacy Williams, Mark McDonald, Neil F Johnson, and Nick S Jones. Temporal evolution of financial-market correlations. Physical Review E, 84(2):026109, 2011.

[2] Dashan Huang, Fuwei Jiang, Kunpeng Li, Guoshi Tong, and Guofu Zhou. Scaled pca: A new approach to dimension reduction. Management Science, 68(3):1678–1695, 2022.

[3] Mia Hubert, Peter Rousseeuw, and Tim Verdonck. Robust PCA for skewed data and its outlier map. Computational Statistics & Data Analysis, 53(6):2264–2274, 2009.

[4] Ian T Jolliffe, Nickolay T Trendafilov, and Mudassir Uddin. A modified principal component technique based on the lasso. Journal of Computational and Graphical Statistics, 12(3):531–547, 2003.

[5] Kwang In Kim, Keechul Jung, and Hang Joon Kim. Face recognition using kernel principal component analysis. IEEE Signal Processing Letters, 9(2):40–42, 2002.

[6] Valero Laparra, Gustavo Camps-Valls, and Jesu ́s Malo. PCA Gaussian- ization for image processing. In IEEE International Conference on Image Processing, 2010.

[7] Aijing Lin, Pengjian Shang, and Huachun Zhou. Cross-correlations and structures of stock markets based on multiscale mf-dxa and pca. Nonlinear Dynamics, 78:485–494, 2014.

[8] Abdul Adeel Mohammed, Rashid Minhas, QM Jonathan Wu, and Maher A Sid-Ahmed. Human face recognition is based on multidimensional PCA and extreme learning machines. Pattern recognition, 44(10-11):2588–2597, 2011.

[9] Jo˜ao Nadkarni and Rui Ferreira Neves. Combining neuroevolution and principal component analysis to trade in the financial markets. Expert Systems with Applications, 103:184–195, 2018.

[10] Jo˜ao Nobre and Rui Ferreira Neves. Combining principal component analysis, discrete wavelet transform, and boost to trade in the financial markets. Expert Systems with Applications, 125:181–194, 2019.

[11] Bernhard Sch¨olkopf, Alexander Smola, and Klaus-Robert M¨uller. Nonlin- ear component analysis as a kernel eigenvalue problem. Neural computation, 10(5):1299–1319, 1998.
[12] Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. Journal of the Royal Statistical Society Series B: Statistical Methodology, 61(3):611–622, 1999.
[13] Jian Yang, David Zhang, Alejandro F Frangi, and Jing-yu Yang. Two- Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Transactions Transactions on pattern analysis and MachineMachine intelligenceintelligence, 26(1):131–137, 2004.
[14] Daoqiang Zhang and Zhi-Hua Zhou. (2d) 2pca: Two-directional two- dimensional pca for efficient face representation and recognition. Neurocomputing, 69(1-3):224–231, 2005.
[15] Linyu Zheng and Hongmei He. Share price prediction of aerospace-relevant companies with recurrent neural networks based on PCA. Expert Systems with Applications, 183:115384, 2021.
[16] Yani Zhu, Chaoyang Zhu, and Xiaoxin Li. Improved principal component analysis and linear regression classification for face recognition. Signal Processing, 145:175–182, 2018.
[17] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2):301–320, 2005.
[18] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. Journal of computational and graphical statistics, 15(2):265–286, 2006.