# PCA

*by* Pca Pca

**Submission date:** 22-Nov-2021 11:54PM (UTC-0600)
**Submission ID:** 1710927265
**File name:** AlgorithmsD.docx (35.88K)
**Word count:** 2314
**Character count:** 13528

# ALGORITHM

By Student's Name

Course Code

Professor's Name

University

City, State

Date

# The Use of an Algorithm to Solve a Specific Business Problem.

## Definition of an algorithm

An algorithm is a defined set of instructions often applied in problem-solving or executing activities in specialized areas. An algorithmic business is based on algorithms that describe its operations and offer client services in the business sector. In contrast to huge data, algorithms specify actions and provide tangible value. Certain corporations produce enormous volumes of data, analyze it, and then delegate decision-making to individuals (Balachandran and Prasad, 2017). Then some firms make several decisions independently of the human intellect, including customer service and pricing. Numerous businesses utilize algorithms to solve problems. For instance, e-commerce websites (Alibaba and Amazon) rely heavily on algorithmic business models for pricing, online proof of identity, stock control, and seller matching. The organizations, as mentioned earlier, use a dynamic pricing method that enables merchants to adjust their prices in response to market changes automatically. Additionally, they consider customer wants, rival pricing and conduct large-scale valuation tests and simulations to make more informed strategic business choices.

**Company of focus**: *Amazon*

**Business problem to focus on**: *Need for online identity verification (Face recognition)*

## Problem definition

Ensuring that an e-commerce platform visitor who joins up is a legitimate buyer, the e-commerce platform must make some determination about the visitor. This process ensures that fake accounts or bots are not created, resulting in lost revenue (particularly when purchasing items using cash on delivery (COD)). To confirm that the identification of a person fits the one that is intended to be, such as the one entered during the first sign-up approach on the website, authentication systems are an essential step (Roy et al., 2017). This process is also known as the "KYC (Know your customer)" process. Hence to curb this issue, a facial recognition algorithm must be embedded in the code. This algorithm will identify the person from their image, which can be found on ID, Driver's license, or passport submitted during the sign-up process. Some features like eyes, lips, shape, etc. extract from the image or video source to identify the person's identity.

## Meaning of verifying identity

Verifying identification is imperative. It shows that a genuine person is behind an activity. The user is who they purport to be, avoiding another individual from undertaking clear procedures on our behalf without authority and a person from inventing fake identities or engaging in fraudulent activity

Identity verification services are available in a variety of forms and via a variety of mechanisms. Identification verification can be conducted in various ways based on the medium via which it is executed and the method used to carry out the validation.

An individual perfectly represents themselves in an organization (commercial office, branch, store), institution (employment service, tax agency) and achieves their identity records (ID, passport) simultaneously (Barkadehi et al., 2018). The agent can authenticate that the image on the document coincides with the individual who is providing it. In the conventional face-to-face method, a replica of the document is created, a form is completed, and inspections are performed in preparation for the authorized user to continue with the necessary procedures.

**Algorithm of focus**

**Eigenface (PCA)**

Eigenfaces is a face detection and recognition method that uses singular value decomposition to identify faces. The principal component analysis (PCA) is a statistical technique that is applied to reduce the dimensionality of a dataset (Anggo and Arapu, 2018). Eigenfaces are used to remove certain less significant aspects from a picture and retain just the most critical and required image characteristics. Eigenfaces reduce dimensionality with having important features of the image. When dimensionality reduced then quality and space occupied will also reduce because of losing information (less important).

**History of Eigenface**

The quest for a low-dimensional description of facial pictures was the starting point for the eigenface technique. Sirovich and Kirby (1987) demonstrated that PCA might be used in a series of facial photos to provide a set of basic characteristics for further investigation (Zhang and Turk, 2008). A technique to face recognition proposed by Turk and Pentland in 1991 employs discretization and linear algebra ideas to identify faces (Zhang and Turk, 2008). The fact that this technique is theoretically less costly and easy to configure led to its adoption in various applications at the time, including handwriting assessment, lip-reading, diagnostic picture assessment, and other similar tasks.In 1901, Pearson invented the method of PCA (Principal Component Analysis), which is used to reduce the dimensionality of a dataset.

It reduces dimensionality by using EigenVectors and Eigenvalues. It projects a trained model or data into a compact feature map using these methods.

Let's take a closer look at the algorithm itself (from a face recognition perspective).

To rebuild images from the training sample, these foundation images, also termed as eigen pictures, could be concatenated linearly.The training set contains M photos, the principal component analysis might be used to create a basic set of N images, where N = M. If the training set contains M images, PCA could develop a mathematical formulation of N images N = M.

The inaccuracy in the reconstruction is minimized by expanding the number of eigen images; nonetheless, the amount of eigen pictures required is less than the integer M. In the case of producing a set of N eigenfaces for a training sample of M face pictures, it is possible to claim that every face image could be composed of "proportions" of every K eigenfaces or "features."

**How it works**

First, an input image from the database is taken. a large dataset for training for getting a more accurate result. After taking the image, the image is classified using an image classifier thereafter, a single layer neural networks in our classifier. In this classifier, a 2D image to vector is made (If image size is pxq then we make a vector of pqx1). Then we are doing the process of feature extraction.

**Algorithm:**

In Eigenfaces we are using Principal Component Analysis . From 2D image we get to 1D vector it called feature vector.

PCA performs as given below

Let $P=\{P_1,P_2,\ldots,P_n\}$ is random vector.

1. Find mean

$$m = \frac{1}{n}\sum_{i=1}^{n} P_i$$

2. Find Covariance Matrix

$$S = \frac{1}{n}\sum_{i=1}^{n}(P_i-m)(P_i-m)^T$$

3. Find eigenvector $\lambda_i$ and eigenvalue $b_i$ of S : $Sb_i = \lambda_i b_i$ ,i=1,2,…,n
4. Sort all eigenvectors by their value and select the first k largest eigenvectors. So the new observed vector is given by

$$v = W^T(P - m)$$

where $W = \{b_1,b_2,\ldots,b_k\}$

Algorithm 5

From using PCA, we get eigenvector and dimensionality smaller than original dimensionality (pqx1). From the classifier, we get if the image is matching or not.

The potential areas of face recognition technology include:

Surveillance is carried out automatically intending to identify individuals and what they do. Face recognition technology can be integrated into current closed-circuit television (CCTV) systems to search for missing kids or other lost individuals and follow suspicious wanted offenders (Lawson et al., 2018). Inquiries into picture databases, scanning image databases of motorists, benefit claimants, and locating persons in massive news images and video archives, as well as checking the Facebook digital platform, are all possible.

Multimedia settings with human cognitive connections are being developed. Face recognition technology could be employed at the airport entrance gate or other areas where travelers are subjected to random inspections to screen them for further examination. Additionally, the platform can be used in slot machines where the design process of betting floors integrates camera systems at face height with perfect lighting. The framework could be used for face image scans for personal identification and to take photographs to create a comprehensive exhibition for future watchlists, proof of identity, and verification tasks, among other applications.

**Pros and cons**

**Pros**

**Removes Correlated Features**

This method makes it quicker and more efficient to discover correlations between features since finding correlations explicitly among thousands of characteristics is practically difficult.

**Minimizes Overfitting**

It is most common when there are multiple parameters in a dataset that overfitting arises. Consequently, by lowering the number of features, this kind of method assists in solving the problem of overfitting.

**Improves Visualization**
When dealing with large amounts of data, it is not easy to comprehend and interpret the information. This method reduces a large amount of high-dimensional input to a small amount of low-dimensional data, which can then be readily displayed. Using 2D Scree Plot, we can determine which Principal Components cause the most variation and have

the greatest influence when matched to different Principal Components (Wagner et al., 2018). Even the most basic IRIS dataset is four-dimensional, making it difficult to visualize. We can use this kind of method to convert it to a 2-dimensional representation for easier display.

**Cons**

**Independent variables are less interpretable:**
The original features would be transformed into Principal Components when they have been implemented. Principal Components are the weighted sum of the initial characteristics. Unlike original features, Principal Components are inaccessible and not easily decipherable as they might be.

**Data standardization**

Prior to doing Principal Component Analysis, it is necessary to standardize the data. Normalize your data before employing PCA; otherwise, PCA will not identify the ideal Principal Components. For example, if a feature set contains data represented in different measures, the variance scale in the test dataset is quite large compared to the feature functionality.

If PCA is used to such feature functionality, the resulting loadings for components with high variation will similarly be substantial, as will the total loadings for all variables. Thus, principal components would be skewed in favor of characteristics with large volatility, resulting in false positives when analyzing data.

Furthermore, to achieve standardization, all categorical characteristics must first be turned into nominal attributes before PCA can be implemented.

**Information Loss**
Although Principal Components attempt to capture the greatest variation among the elements in a set, if the quantity of Principal Components is not carefully chosen, this could lose out on some information as opposed to the earlier list of selected features.

**Aspect of business addressed by this kind of Algorithm**

**Customer service**

With this kind of algorithm, the business is able to distinguish original clients from fake clients hence adjusting the way of treating customers/clients with given priorities.

**Record keeping**

With this kind of algorithm, record keeping which is the biggest aspect of any given business is achieved since records with verified identities are given much priorities hence making the business to run smoothly and also helps in record manipulation at any given scenario

**Alternative of PCA**

**The following is the other alternative of PCA**

**Factor analysis**

When applied to a collection of measured variables, factor analysis is often used to detect or validate the latent component structure of the variables. Latent factors are not seen and cannot be assessed directly; nonetheless, they are thought to be responsible for the values we notice on the assessed or indicator parameters.  The second use of factor analysis reduces the number of variables that can properly assess or express the latent structural model.

Factor analysis is a method that is based on models. Mapping the links between observed data, hidden components, and error is at the heart of what it does. Factor analysis is often more constant across samples due to the identification of erro (Bandalos and Finney, 2018)r.  The potential of factor analysis to acknowledge individual item variability (also known as item error variance) is critical in discerning it from principal component analysis (PCA), which deems all variance evenly. Several assumptions, including continuity, multifactorial normality, and homogeneity of variance, are used in factor analysis.

**Pros and cons of Factor analysis**

**Pros**

The following are some of the benefits of factor analysis:

Detection of groupings of inter-related variables helps to determine how the analysis is connected to the environment. Direct analysis may reveal hidden aspects or concepts, and factor analysis may be employed to uncover these dimensions or constructs. It is possible to operate both actual and perceived qualities. Reduce the number of variables by merging two or more independent variables to form a single component (also known as factorization). The use of dimensions allows for more freedom in naming. It is not incredibly tough to do, it is economical, and it is quite precise.

**Cons**

For the technique to be helpful, the researchers must develop a comprehensive and accurate collection of product features. If significant attributes are overlooked, the procedure's relevance is lowered proportionally. It's also worth mentioning that naming the components might be challenging since numerous traits can be strongly associated for no obvious reason at the same time.

Algorithm 8

# References

Anggo, M. and Arapu, L., 2018, June. Face recognition using fisherface method. In *Journal of Physics: Conference Series* (Vol. 1028, No. 1, p. 012119). IOP Publishing. Retrieved from https://iopscience.iop.org/article/10.1088/1742-6596/1028/1/012119/meta

Balachandran, B.M. and Prasad, S., 2017. Challenges and benefits of deploying big data analytics in the cloud for business intelligence. *Procedia Computer Science*, *112*, pp.1112-1122. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050917314953

Bandalos, D.L. and Finney, S.J., 2018. Factor analysis: Exploratory and confirmatory. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 98-122). Routledge. Retrieved from https://www.taylorfrancis.com/chapters/edit/10.4324/9781315755649-8/factor-analysis-deborah-bandalos-sara-finney

Barkadehi, M.H., Nilashi, M., Ibrahim, O., Fardi, A.Z. and Samad, S., 2018. Authentication systems: A literature review and classification. *Telematics and Informatics*, *35*(5), pp.1491-1511. Retrieved from https://www.sciencedirect.com/science/article/pii/S0736585318301400

Lawson, T., Rogerson, R. and Barnacle, M., 2018. A comparison between the cost effectiveness of CCTV and improved street lighting as a means of crime reduction. *Computers, Environment and Urban Systems*, *68*, pp.17-25. Retrieved from https://www.sciencedirect.com/science/article/pii/S0198971516304240

Roy, A., Memon, N. and Ross, A., 2017. Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems. *IEEE Transactions on Information Forensics and Security*, *12*(9), pp.2013-2025. Retrieved from https://ieeexplore.ieee.org/abstract/document/7893784/

Wagner Filho, J.A., Rey, M.F., Freitas, C.M. and Nedel, L., 2018, March. Immersive visualization of abstract information: An evaluation on dimensionally-reduced data scatterplots. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 483-490). IEEE. Retrieved from https://ieeexplore.ieee.org/abstract/document/8447558/

Zhang, S. and Turk, M., 2008. Eigenfaces. *Scholarpedia*, *3*(9), p.4244. Retrieved from http://www.scholarpedia.org/article/Eigenfaces

PCA

ORIGINALITY REPORT

| 4% | 2% | 1% | 4% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to University of Birmingham<br>Student Paper | 1% |
| 2 | Submitted to University of Lincoln<br>Student Paper | 1% |
| 3 | scholar.sun.ac.za<br>Internet Source | 1% |
| 4 | link.springer.com<br>Internet Source | 1% |
| 5 | www.coursehero.com<br>Internet Source | 1% |

| | | | |
|---|---|---|---|
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | On | | |