How to Perform Principal Component Analysis on a Dataset of Automobile MPG

***Dr.Renuka sagar,phD(IOT and ML)***

***Dept of ECE,***

***Ballari Institute of Technology and Management,Ballari***

***Email:renukasagar83@gmail.com***

***Abstract-*** **A car's average age can be anywhere from a few to 17 years. There are numerous methods for calculating MPG, one of the key elements of the vehicle. In this study, fuel consumption is calculated using principal component analysis. PCA is a versatile tool that is most frequently used for multivariate data processing and visualisation. Finding the variables that affect an automobile's MPG is the major goal of this investigation. A statistical method known as principal component analysis (PCA) identifies the linear combinations of the original variables that are most beneficial. PCA is employed in a variety of industries, including marketing, engineering, chemistry, and also biology. Principal components analysis (PCA) is a statistical method for reducing dimensions while retaining as much variance as possible within a given data set. The first principal components produced by PCA are those that capture the most information or variance in the data set, i.e. the first few components represent the majority of explained variance.**

Index terms-primary components, PCA, object,dataset,data analysis

1. NOMENCLATURE

Principal analysis (PCA) is used to reduce the of a data set while maintaining as much variability as possible within the data set. Transformations to the using the technique produce new feature . Principal components are linear combinations (orthogonal transformations) of initial predictors within the . The basic idea of PCA is to project data onto the two-dimensional plane, to make necessary relationships visible.

1. Introduction

Although automobiles have been around for a while, the first one was created in 1769. That occurred more than three centuries ago. The modern automobile is distinctive and special due to its many diverse features. A car's average age can be anywhere from a few to 17 years. There are numerous methods for calculating MPG, one

of the key elements of the vehicle. Principal component analysis is one of the most often used approaches to calculating MPG. In this study, fuel consumption is calculated using principal component analysis. A linear algebraic method called principal component analysis (PCA) can be used to decrease the of a set of features. The singular value decomposition and

other linear combinations of features can be discovered using PCA, which is especially

helpful for exploratory data analysis and . PCA is a versatile tool that is most frequently used for multivariate data processing and . Each car's mileage over a certain time period is recorded in the automotive MPG. For every car in the , there are a of 30 variables. Finding the variables that affect an automobile's MPG is the major goal of this investigation. A statistical method known as principal component analysis (PCA) identifies the linear combinations of the original variables that are most beneficial. It is employed in a variety of industries, including marketing, engineering, chemistry, and also biology In order to uncover intriguing patterns and insights, a mathematical methodology known as principal component analysis is used in this study on a of automotive gasoline usage.

* 1. Data

component analysis (PCA) is a common technique for reducing the number of dimensions for a given data collection. It can be applied to lower the number of variables while maintaining the same level of information in a data set. PCA is an iterative procedure that can be used to reduce the of a data set. Utilizing Python's PCA function is a nice place to begin. The data set's PCA is determined by the PCA function. PCA is an iterative procedure that can be used to reduce the of a data set.

2.2. Steps involved in Principal Component Analysis of autompg dataset.

A method for extracting sets of orthogonal (uncorrelated) variables is principal component analysis (PCA), which can subsequently be used to reduce the of a data set. A common method in machine learning is PCA. A data set's can be decreased using PCA, a linear treatment that extracts orthogonal variables.

To determine PCA, use the vehicle mpg

Step 1: The data set is initially converted into a fresh set of variables[1] that stand in for the main data elements. The new main components are then projected with the data. To achieve this, the dot product of each data point with its matching primary component is calculated. The objective of this phase is to align the data using a feature vector format and the [3] of a matrix from its original axis to that represented by the principal components (hence principal component analysis).

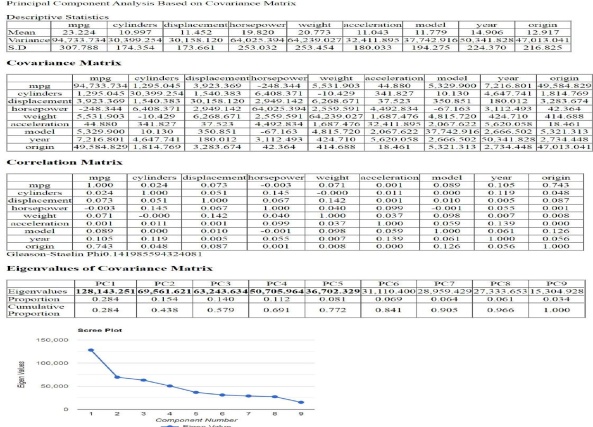
The data input sets still correspond to the original axes because in the preceding steps, with the exception of , no modifications were done to the data; only principal components were chosen, and a feature vector was created (i.e., the original variables). Realignment is accomplished by multiplying the feature vector's transpose by the input data set's transpose. The direction of the between specific feature variables () within a data set is used by PCA[2]. It produces a matrix that reflects the of the link between each pair of feature variables in our .

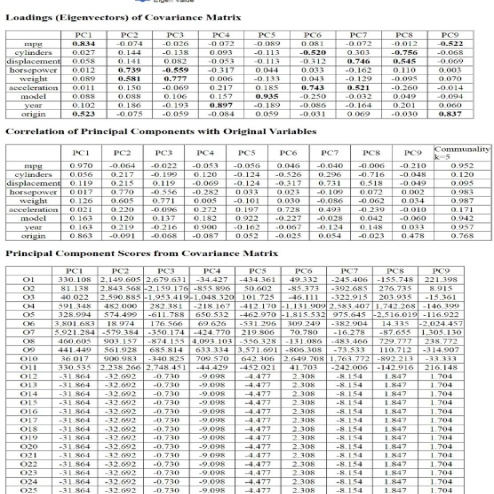
Step 3: Standardize the values of all the feature variables to a single scale before calculating the . However, in the traditional computations for PCA, an ML vehicle fitted data set is naturally as a side consequence of constructing a matrix. The subtraction has no influence on the variance of the data itself. By using linear combinations ( on zero mean) to the features with the biggest variance, PCA calculates the first principal component of the feature/variable set, which after some mathematical calculations results in estimations for (Z1) for the initial major elements. The first principal component produced by PCA, as its name suggests, has the most information or the highest variation in a data set.

Step4: Using the second main component (PCA2) (feature1, feature2). Main components reduce a big number of features to a small number of principal components in order to noise for very large (where the number of dimensions may surpass 100 distinct variables).

A feature vector of shape (feature1, feature2)[4] is extremely similar to a vector of shape (first principal component (PCA1), second principal component (PCA2)[5] in the little two-dimensional example above, therefore employing PCA doesn't really help much. The total number of variables that must be examined is decreased in this way. Large scale transformations are By performing transformations, large portions of the entire data set are efficiently compressed into multiple feature columns. With each additional component, less information is contributed.

1. Results





1. Conclusion

A common statistical method for reducing dimensions while retaining as much variance as possible within a given data set is principal component analysis (PCA) (Pearson, 1901; , 1933; , 2008). Principal analysis (PCA) is simply a straightforward reduction technique that converts the columns of the into a new feature group (PCs). We are interested in the features of the data, and PCA, which is closely related to factor analysis, frequently leads to comparable results. first principal components produced by PCA, as the name implies, are those that capture the most information or variance in the data set. The typical setting for PCA as a tool for data analysis and exploration involves a data. An observational data set containing P distinct variables for each of N entities or individuals and P different variables for each of N entities or individuals constitutes the typical context for PCA as an analysis tool for data exploration. Here, we employ some PCA criteria to select the best subsets of the p output variables—the main variables. Imagine running PCA on a with hundreds of variables and seeing that the first few components represent the majority of the explained variance, the NxP data matrix column choices, X, that capture the most variance. The primary components of a data set are these two linear combinations. First, I a PCA class and call fit transform just on X. to transform X to obtain the new set of X's components, then compute the weights of the primary components all at once. .decomposition module offers the PCA object that can be simply mapped and converts data to the components, i.e., it a PCA object from and changes the data in accordance with the computed components. When the principal component input features are entered, the PCA object's Inverse Transform method returns the original data. A lesser quantity of information, but more nuance, is added to the data set with each subsequent component.

V.Acknowledgment

I am Associate professor in the department of Electronics and communication Engineering. Earned how to manipulate and complicated, high-volume, high-dimensional data from a variety of data sources. Strong desire to put machine learning technologies to use, build artificial intelligence systems to automate company processes, conduct statistical research, and use data mining techniques in using test findings to fine-tune statistical analyses thorough knowledge of machine learning algorithms; the development of analytical techniques to new methods for processing data and information.

**Ethical Statement for Environmental Science and Pollution Research**

Hereby, **Dr Renuka Sagar** consciously assure that for the manuscript **How to Perform Principal Component Analysis on a Dataset of Automobile MPG** the following is fulfilled:

1) This material is the my own original work, which has not been previously published elsewhere.

2) The paper is not currently being considered for publication elsewhere.

3) The paper reflects the authors' own research and analysis in a truthful and complete manner.

4) The paper properly credits the meaningful contributions of co-authors and co-researchers.

5) The results are appropriately placed in the context of prior and existing research.

I agree with the above statements and declare that this submission follows the policies **Environmental Science and Pollution Research** as outlined in the Guide for Authors and in the Ethical Statement.

Date:21.9.2022

Corresponding author’s signature:



consent to participate/publish

# Title of manuscript: How to Perform Principal Component Analysis on a Dataset of Automobile MPG.

Journal: **Environmental Science and Pollution Research**

I, the undersigned, give my consent for the publication of identifiable details, which can include photograph(s) and/or videos and/or case history and/or details within the text (“Material”) to be published in the above Journal and Article. I confirm that I have seen and been given the opportunity to read both the Material and the Article (as attached) to be published. I have discussed this consent form with **Dr.Renuka sagar,** who is an author of this paper.

I understand that all journals may be available in both print and on the internet, and will be available to a broader audience through marketing channels and other third parties.

Therefore, anyone can read material published in the Journal. I understand that readers may include not only medical professionals and scholarly researchers but also journalists and general members of the public.

**study participant name : Dr.Renuka sagar** (please print)

**Signed by (name) :**



NOTE: If the patient/study participant is a minor (i.e. less than 18 years of age), or is unable to provide informed consent for publication, this must be signed by their parent or legal guardian.

NOTE: If the patient/study participant is deceased, this must be signed by their next of kin.

# Date 21.09.22

**Signed**



Relationship to patient/study participant, if applicable: self

**Author name DR. Renuka sagar** (please print)

# Date 21.9.22

**Signed**



**Declaration of interests**

**Dr Renuka sagar** declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



**Declaration of interests**

**Dr Renuka sagar** declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Dr Renuka sagar** declare the following financial interests/personal relationships which may be considered as potential competing interests:



VI.References

1. https://en.wikipedia.org/wiki/Principal\_component\_analysis
2. https://www.sae.org/publications/technical-papers/content/2019-26-0359/
3. https://towardsdatascience.com/the-mathematics-behind-principal-component-analysis-fff2d7f4b643
4. https://www.mathworks.com/help/stats/pca.html
5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792409/
6. https://www.polarsparc.com/xhtml/PCA.html
7. https://muthu.co/mathematics-of-principal-component-analysis/
8. http://rasbt.github.io/mlxtend/user\_guide/feature\_extraction/PrincipalComponentAnalysis/