***MATHEMATICS***

***END SEM PROJECT***

***GROUP 5***

**ACKNOWLEDGEMENT**

We are highly indebted to ***AMRITA VISHWA VIDYHAPEETAM*** for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

We would like to express my gratitude towards our professor **Dr . M Ranjith Kumar, Assistant Professor (Sr. Gr), Department of Maths, ASE, Chennai** for their kind co-operation and encouragement which help me in completion of this project.

ABSTRACT

Large datasets are increasingly common and are often difficult to interpret. Principal component analysis (PCA) is a technique for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. It does so by creating new uncorrelated variables that successively maximize variance. Finding such new variables, the principal components, reduces to solving an eigenvalue/eigenvector problem, and the new variables are defined by the dataset at hand, not a priori, hence making PCA an adaptive data analysis technique. This article will begin by introducing the basic ideas of PCA, discussing what it can and cannot do. It will then describe some variants of PCA and their application.

***INTRODUCTION***

The central idea of *PRINCIPAL COMPONENT ANALYSIS (PCA)* is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

WHAT ARE PRINCIPAL COMPONENTS?

A principal component is a normalized linear combination of the original predictors in a data set. In image above, *PC1* and *PC2* are the principal components.

FIRST PRINCIPAL COMPONENTS:

First principal component is a linear combination of original predictor variables which captures the maximum variance in the data set.

SECOND PRINCIPAL COMPONENTS:

The second principal component is the direction which maximizes variance among all directions orthogonal to the first.

***STEPS TO CALCULATE PCA :***

1. Find the mean vector.
2. Assemble all the data samples

in a mean adjusted matrix.

1. Create the covariance matrix.
2. Compute the Eigen vectors and

Eigen values.

1. Compute the basis vectors.
2. Represent each sample as a

linear combination of basis

vectors.

***USE OF PCA :***

The most important use of PCA is to represent a multivariate data table as smaller **set** of variables (summary indices) in order to observe trends, jumps, clusters and outliers. This overview may uncover the relationships between observations and variables, and among the variables.

*The goals of PCA are to*

1. extract the most important information from the data table;
2. compress the size of the data set by keeping only this important information;
3. simplify the description of the data

set

1. Analyze the structure of the

observations and the variables.

1. Compress the data, by reducing the number of dimensions, without much loss of information.
2. This technique used in image

compression.

***ADVANTAGES OF PRINCIPAL COMPONENT ANALYSIS***

1. Removes Correlated Features

2. Improves Algorithm Performance

3. Reduces Overfitting

4. Improves Visualization

***DISADVANTAGES OF PRINCIPAL COMPONENT ANALYSIS***

1. Independent variables become less interpretable

2. Data standardization is must before PCA

3. Information Loss

***APPLICATIOMS:***

**The applications of Principal Component Analysis (PCA) are:**

* Spike-triggered covariance analysis in Neuroscience
* Image Compression
* Facial Recognition
* Other applications like Medical Data correlation

***A TEAM WORK OF***

  

*POORANI PRANAV RAJ YAGNITHA*

*AYSWARIYA*



*SARRVESH SIDESH SUNDAR*