CS571 Project Report
on
Component Anaysis Using PCA, NMF and Dictionary Learning

DEEPIKA
T21019-M.tech CSP

23 November 2021

Contents

1 Summary

This project is based on the component analysis of a dataset of images i.e.detection of noisy images using PCA(Principle Component Analysis),NMF(Non-Negative Matrix Factorization and Dictionary Learning Techniques. Here in this project,we generated the set of characters like alphabets and then added the noise over each alphabet sample and then we apply PCA, NMF and Dictionary Learning and then compare the results obtained from these methods.After comparing, it is observed that even PCA works fine for fewer samples data but still it is not able to bring back the original image completely but the other methods can separate the source and cause and bring back the original image.

2 Introduction

Signal processing is crucial in many data science tasks. As soon as we start handling audio files, images or even biological measurements, it is useful to know techniques to process such data. In this report, I will introduce three algorithms: Principal Components Analysis(PCA) for dimensionality reduction and feature extraction, Nonnegative Matrix Factorization (NMF) and Dictionary Learning for source separation.
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 new set of variable, 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
Non-Negative Matrix Factorization(NMF) is a state of the art feature extraction algorithm. NMF is useful when there are many attributes and the attributes are ambiguous or have weak predictability.Each feature created by NMF is a linear combination of the original attribute set. Each feature has a set of coefficients, which are a measure of the weight of each attribute on the feature. There is a separate coefficient for each numerical attribute and for each distinct value of each categorical attribute. The coefficients are all non-negative.NMF is less complex than PCA and can be applied to sparse data.
Dictionary Learning is an important problem in multiple areas, ranging from computational neuroscience, machine learning, to computer vision and image processing.Dictionary Learning also known as sparse coding. Dictionary learning applied on image patches has been shown to give good results in image processing tasks such as image completion, inpainting and denoising, as well as for supervised recognition tasks.

3 Solution

For finding out the components in a given dataset in order to overcome the problem that PCA fails to estimate the significant components ,therefore we used NMF and Dictionary Learning Technique for component analysis.By using these methods it is observed that the components which we get from NMF and Ditionary Learning are better than the components that we obtained from PCA and these components closely resemble to our original image dataset.

3.1 Assumptions

1. For genrating noisy image,Gaussian moise was added randomly with mean and variance 0,0.1 respectively.

2. Assumed the datasets as alphabet images.

In this project 4*4 images of dataset is taken and each image contains noisy sample images of approx 500 and hence total dataset contains nearly 1000 images.The components which are to be estimated are 16*16 feature vector space.

3.2 Algorithms used

The algorithms which we have used are related to PCA,NMF and dictionary Learning.

3.3 The following steps are involved in this algorithm:

1. Generate the dataset of alphabets.

2. Input dataset given by CSV formated file to the main program.

3. Read the file and convert the dataset into matrix.

4. For each alphabet 1000 noisy samples are generated(assumed Gaussian noise randomly added with mean and variance 0,0.1 respectively).

5. Rotate some of the images by some angle.

6. Combine two types of images

7. Plot some samples from noisy,rotated and combined images.
   4.Then apply PCA , NMF, dictionary Learning using sklearn decomposition.

8. Components obtained from PCA and NMF are generated and then plotted.

4 Results and analysis

Figure 1: Original images

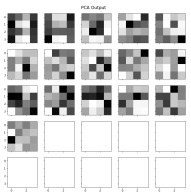

Figure 2: PCA output

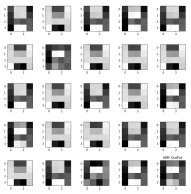

Figure 3: NMF output

5 Conclusion

From the project, I concluded that PCA,NMF and dictionary Learning are used for the component analysis.I compared the outputs obtained from these three methods then concluded that dictionary learning gives better result than PCA and NMF.

5.1 Challenges

We have to create an image dataset of 23 alphabets ,each sample is of 4*4 dimensions. During Rotation of image sapmles by a particular angle like (30,60 etc.)degree,the dimension of images changed to 5*5, de to which it suffers from some computational error beczuse we are taking the 4*4 images in our project.

6 Project Github page

Have a look at the source code,user manual, results obtained and various resources that used in the project. https://github.com/Deepikaa05/Project_PCA

7 References

1. Jolliffe, Principal Component Analysis, 2nd edition

2. "Dictionary Learning", by Tosic et al, IEEE Sig Proc Mag, March 2011

3. M.Plumbley,Algorithms for Nonnegative Independent Component Analysis,2nd,vol.28,1980