

Can Principal Component Analysis be Applied in Real Time to Reduce the Dimension of Human Motion Signals?

Vittorio Lippi^{*} Giacomo Ceccarelli[†]

(^{*}) PERCRO Scuola Superiore Sant'Anna, Italy

([†]) Università di Pisa, Italy

E-mail: v.lippi@sssup.it, giacomo.ceccarelli@df.unipi.it

Abstract

Principal Component Analysis (PCA) is a usual method in multivariate analysis to reduce data dimensionality. PCA relies on the definition of a linear transformation of the data through an orthonormal matrix that is computed on the basis of the dataset itself. In this work we discuss the application of PCA on a set of human motion data and the cross validation of the result. The cross validation procedure simulates the application of the transformation on real time data. The PCA proved to be suitable to analyze data in real time and showed some interesting behavior on the data used as cross validation.

1. Introduction

1.1 Overview

Data dimensionality is an usual issue in multivariate analysis, and Principal Component Analysis (PCA) is a very popular method to perform it. PCA is a method to convert a set of observations of possibly correlated variables into a set of values of linearly independent variables called principal components using a transformation described by an orthogonal matrix. After the principal component produced have the same dimensionality of the original dataset. The transformation is defined in such a way that the first principal component has as high a variance as possible, and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. The Data reduction is obtained cutting off a number of principal components, the last ones, bearing less variance. While in most of the work PCA is exploited as a first data filtering on the whole dataset involved into the performed analysis, in this work we are interested in using

it on online data. This means that the data over which the PCA is applied are not known when the transformation defining it is computed. This method provided good results for gesture recognition in conjunction with hidden Markov models and neural networks[5] where the data dimensionality reduction is crucial to avoid the overfitting problems connected with the proliferation of parameters[9]. Cross validation is the natural approach when the procedure is thought to be applied on data not available when building up the transformation[4]

1.2 PCA Terms and Symbols

In this section we briefly present some of the terms used in the paper through a quick exposition of the PCA method. For the sake of simplicity we will avoid demonstrations and details about the numerical methods to be applied to implement the steps of the algorithm for a detailed for which several dissertation are available e.g. [2][8][3]. The dataset consist in a table of data, let it be the matrix X , whose rows represent samples and whose N_C columns represent the sampled variables x_i . As a first step we compute the Z-score of each x_i , subtracting the estimated average and normalizing it dividing it by its variance (actually the normalization is not always used), we call this the X^Z . The actual PCA consist then in finding a linear transformation, represented by the matrix C to obtain the new table $P = X^Z C$ with the following properties:

- the column p_i of P , called *principal components* are the linearly independent and ordered by decreasing variance i.e. $var(p_i) > var(p_{i+1})$
- C is an orthonormal matrix;

Each component p_i is a linear combination of the original x_k through the values of the i^{th} column of C . the transformation is reversible, hence non destructive

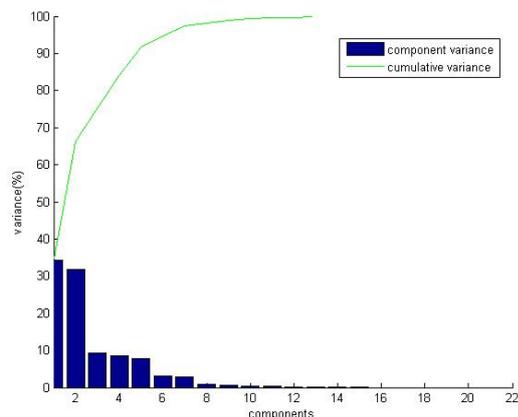


Figure 1: Distribution of variance over the principal components for juggling, the whole set of available data has been exploited.

respect to the information. To perform the reduction of data dimensionality just the first N principal components are retained and the remaining $N_c - N$ are dropped, saving the highest possible of the variance. The number N should be chosen on the basis of a test, *the stopping criterion*, that could be defined in different ways, according to the nature of data and the aim of the analysis.

2 Dataset

We performed the experiment over two datasets: the first consist into the hands position(x,y and z) of a juggler performing the *3 ball cascade* together with the positions of the juggled balls (15 variables); the second consists in the position of wrists, elbows and shoulders of a rower using SPRINT rowing simulator system[7], together with the two angles describing the position of the oars (22 variables). We sampled data with the VICON [1]optical tracking system at $100Hz$ sampling rate. Notice that the data used in this work represent a continuous function of time sampled at constant frequency. The variance captured into the first components takes in account mainly the trend of the function[3]. The movements are limited and, although not strictly periodic, repetitive: this allows to use the average of the training set data as an estimation of the average of the real time data. The figures 1 and 2 show the distribution of variance over the principal components for rowing and juggling data: notice how, especially for the rowing, the variance is concentrated into the first components, this reveals a strong linear correlation between the original variables.

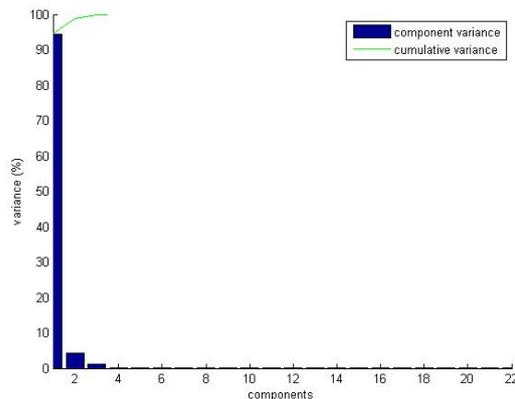


Figure 2: Distribution of variance over the principal components for rowing, the whole set of available data has been exploited.

3 Cross Validation

To perform the cross validation the dataset is divided into two subsets. Adopting a terminology borrowed by machine learning we call *training set* the set used to produce the parameters of the model used for the PCA i.e. the transformation matrix, the average to be subtracted from data and the variance of the input variables to be normalized, and *validation set* the subset of data used to test the performance of the obtained parameters in the analysis of unseen data. The training set is produced sampling randomly the decided number of samples from the dataset, the remaining samples are used as validation set. To check the impact of the number of samples N_s used in building the transformation we experimented training set of different sizes. The procedure is repeated N_{reps} times for each N_s taken into account.

Since we are interested into the amount of variance retained over the validation set, for each repetition of the experiment we compute the cumulative of the variance over the number of principal components

$$CV_k = \frac{\sum_{i=0}^k \hat{var}(p_i^v)}{\sum_{i=0}^{N_c} \hat{var}(p_i^v)} \cdot 100 \quad (1)$$

where with p_i^v we addressed the variables obtained multiplying the C matrix to the validation set samples, that are not strictly the principal components and with $\hat{var}()$ the estimated variance. This value tells us how much variance has been retained over the training set, it is expressed as a percentage of the total variance. by definition CV_0 is always equal to 0% and CV_{N_c} is always 100%. Calling CV_k^r the CV_k obtained with the r^{th} rep-

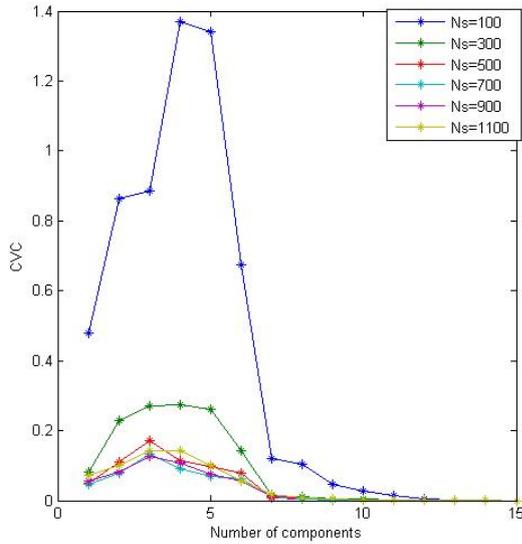


Figure 3: VCV over the juggling data for different N_s . Notice the EQT at the 7th component

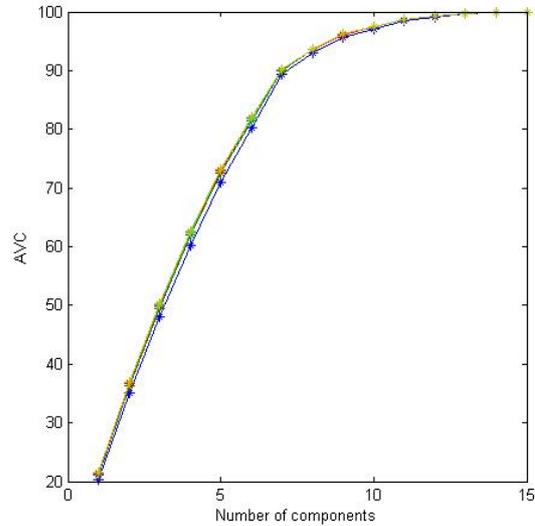


Figure 4: ACV over the juggling data for different N_s . the results are difficult to be distinguished

etition of the validation experiment we may define the average (ACV) and the variance (VCV) of CV as:

$$ACV_k = \frac{\sum_{r=1}^{N_{reps}} CV_k^r}{N_{reps}} \quad (2)$$

$$VCV_k = \frac{\sum_{r=1}^{N_{reps}} (CV_k^r - ACV_k)^2}{N_{reps} - 1} \quad (3)$$

Once a threshold for the retained performance has been defined with a stop criterion, the VCV gives a measure of the expected variation of it over the real time data.

4 Results

The test has been performed using a Matlab[®] script, the C matrix has been computed through the PRINCOMP function [6]. We tested the values for N_s of 100, 300, 500, 700, 900 and 1100 over 2328 samples. We repeated the experiment 10000 times. The ACV is in practice independent from the number of samples as displayed in figures 4 and 6.

For the first components VCV increases with N to reach a maximum, then starts to decrease as shown in figures 3 and 5. There is an $N < N_c$ for which the VCV is in practice zero. This happened at the third component for the rowing (with an $ACV > 95\%$) for all the values of N_s we tried) and at the seventh component for the juggling (with an $ACV > 90\%$) for all the values of N_s we tried). Since a small VCV indicates a robust choice of N (implying a small expected

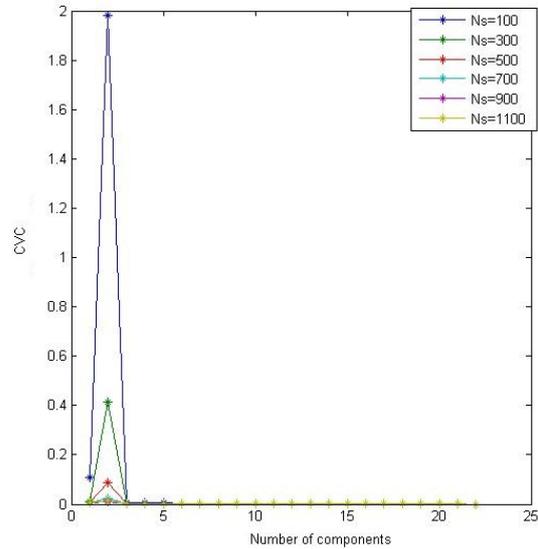


Figure 5: VCV over the rowing data for different N_s . Notice the EQT at the 3th component

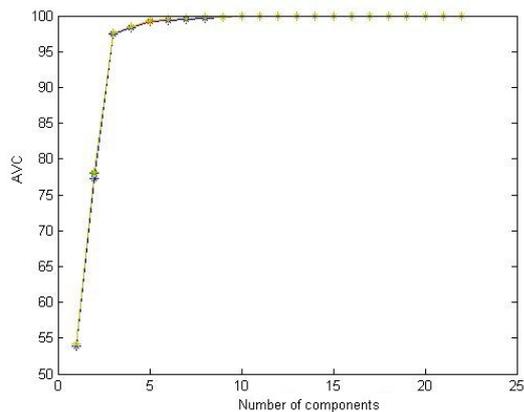


Figure 6: ACV over the rowing data for different N_s . the results are difficult to be distinguished

variation of the CV over the real time data) we call this threshold the empirical quality threshold EQT . It is interesting to note that, as showed in figures 3 and 5, the EQT appears to be independent from N_s .

5. Conclusions and Future Work

5.1 Main Conclusions

The cross validation proved that the PCA is suitable to be applied on real time samples not available while computing the transformation matrix. An interesting result has been the presence of a threshold beyond which the empirical variance cumulative variance over the cross validation dataset (VCV) was in practice null, while the average of cumulative variance over the cross validation dataset (ACV) was still increasing with the number of components accounted. We call this threshold the empirical quality threshold (EQT). A very interesting empirical result consist in the fact that, although the number of data into the *training set* decreases in general the VCV, the EQT is practically independent from it. It is to be noticed also that the (ACV) is barely independent from the number of samples into the dataset in the analyzed cases, this can be thought as a first test of consistency of the procedure applied over the data. The *recipe* for using PCA for online data analysis can be hence formalized in three points:

- check the ACV, it is expected to be roughly the same on the *validation set* and on the *training set*;
- perform a stopping criterion over the *validation set*;
- check the VCV, take a value over the EQT to expect a better robustness;

5.2 Future Work

Our next step over this topic will be to evaluate the proposed procedure under a probabilistic point of view assuming a known distribution for data. In particular, since the core of PCA is the diagonalization of the correlation matrix (computed over the training set) the variance of the correlation matrix estimation could give a formal indication on the bias over online data. Besides we are interested to extend the proposed experiment over a wider set of human motion data.

6 Acknowledgements

This work was supported by EU SKILLS Integrated Project.

References

- [1] O.m.g vicon system <http://www.vicon.com>.
- [2] H. Abdi and L. J. Williams. Principal component analysis, 2010.
- [3] J. E. Jackson. *A user's Guide to Principal Components*. John Wiley & Sons, Inc., 2004.
- [4] W. J. Krzanowski. Efficient cross-validation of principal components. *Statistics and Computing*, 6:177–177, 1996. 10.1007/BF00162529.
- [5] V. Lippi, E. Ruffaldi., C. A. Avizzano, and M. Bergamasco. Recognition of hand gestures tracked by a dataglove: Exploiting hidden markov models discriminative training and environment description to improve recognition performance. In *5th International Workshop on Artificial Neural Networks and Intelligent Information Processing*, 2009.
- [6] The Math Works, Inc., 24 Prime Park Way, Natick, MA 01760-1500. *Statistics Toolbox User's Guide*.
- [7] E. Ruffaldi, A. Filippeschi, A. Frisoli, C. A. Avizzano, B. Bardy, and D. Gopher. Sprint: a training system for skills transfer in rowing. In T. Gutierrez and E. Sanchez, editors, *SKILLS09 International Conference on Multimodal Interfaces for Skills Transfer. Bilbao, Spain.*, 2009.
- [8] L. I. Smith. A tutorial on principal component analysis.
- [9] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16(2):264–280, 1971.