EXAM HDDA PART I : RESUME PCA AND SVD		
Student Name		Deadline
Sandra Tsemo		**January 03, 2022**
January 3, 2022		2021-2022
		Lecturer: Pr. Sophie Dabo

Part 1.a : Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical method that allows the analysis and visualisation of a data set containing individuals described by several quantitative variables.

It is a statistical method for exploring so-called multivariate data (data with several variables). Each variable can be considered as a different dimension. If you have more than three variables in your dataset, it can be very difficult to visualise the data in a multidimensional "hyperspace".

PCA synthesises this information into just a few new variables called principal components. These new variables are a linear combination of the original variables. The number of principal components is less than or equal to the number of original variables.

The information contained in a data set is the total variance or inertia it contains. The purpose of PCA is to identify the directions (i.e. principal axes or principal components) along which the variation in the data is maximum. In other words, PCA reduces the dimensions of a multivariate dataset to two or three principal components, which can be visualised graphically, losing as little information as possible.

Dimension reduction is achieved by identifying the main directions, called principal components, in which the data vary. PCA assumes that the directions with the largest variances are the most important. Technically speaking, the amount of variance explained by each principal component is measured by what is called the eigenvalue. Note that PCA is particularly useful when the variables in the data set are highly correlated.

Correlation indicates that there is redundancy in the data. Because of this redundancy, PCA can be used to reduce the original variables into a smaller number of new variables. These explain most of the variance in the

original variables.

In simple terms, principal component analysis as a whole identifies hidden patterns in a dataset, then reduces the dimensions of the data by removing redundancy in the data and finally identifies correlated variables.

Part 1.b : Singular Value Decomposition (SVD)

Singular value decomposition reflects the extraction of data in the directions with the highest variances respectively. PCA is a linear model that maps m-dimensional input features to k-dimensional latent factors (k principal components). If the least significant terms are ignored, the components of least interest are eliminated, but the principal directions with the highest variances (most important information) are retained.

In the SVD, we have notions of Singular vectors and singular values who was very important.

Let be a given matrix A of size m $\times$ n. Doing an SVD means first determining the matrices AA^T and A^TA and note that in linear algebra these matrices are quite special. Indeed, they are :

1. Symmetric,
2. Square,
3. They have zero or positive eigenvalues,
4. They have the same positive eigenvalues, and the same rank r as the matrix A.

In SVD, we denote u_i as eigenvectors of AA^T and v_i as eigenvectors of A^TA. These sets of eigenvectors denoted u and v are the singular vectors of A. Both matrices have the same positive eigenvalues. The square roots of these eigenvalues are called singular values.

The principle of SVD is that any matrix A can be factored as :

$$A = UDV^T$$

Where U and V orthogonal matrices with orthonormal eigenvectors chosen from AA^T and A^TA respectively. D is a diagonal matrix with r elements

equal to the root of the positive eigenvalues of AA^T or A^TA. The diagonal elements are composed of singular values. These vectors are orthonormal because $U^TU = U$ and $V^TV = V$.

In a nutshell, the SVD is very simple. There is a matrix A. In the case of the indices computation it is a table of object-feature measure data. The objects corresponds to rows while the features corresponds to columns. Also, the matrix A is a linear operator. It maps a weights vector w in the weights space $\mathbb{R}^m$ to an indices vector q in the indices space $\mathbb{R}^n$. Here m is the number of objects and n is the number of features. A linear operator A can be represented as the product of three linear operators, $A = UDV^T$. So, an arbitrary linear operator A could be represented as the product of a rotation, scaling and rotation linear operators.

References

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman; The Elements of Statistical Learning , Section 14.5.1 of Chapter 14, page 531-541.

2. Prof. Sophie Dabo, (2022), Lecture notes of High Dimensional Data Analysis, AIMS-Senegal.