1 Introduction

In this document, you will see different interpretations and meanings to certain mathematical quantities. These interpretations are crucial to understand numerical results from software like SAS, Minitab, Mathematica, Matlab that perform multivariate techniques such as Principal Component Analysis (PCA), Canonical Correlation Analysis (CCA), Correspondence Analysis (CA), and Canonical Correspondence Analysis (CCPA). These techniques are based on eigenvector-eigenvalue decomposition of certain matrices.

To understand the numerical results when we perform these multivariate techniques, it is necessary to have some background in an upper level undergraduate course in linear algebra covering some contents like dot product and distance, basis and coordinates of a vector, eigenvectors and eigenvalues, singular value decomposition (svd), quadratic forms and ellipsoids, metrics and projections. We will briefly go over some of these ideas here in this document.

2 A Data Matrix

We have taken a small data matrix here to demonstrate some important ideas in matrix algebra/linear algebra. However, the same ideas apply to large data sets and computer software like SAS, Minitab, and Mathematica can easily handle all such computations for large matrices as well.

Consider the following 3 by 2 (three rows, each size 2 and two columns, each size 3) matrix **Z**.

$$\mathbf{Z} = \begin{pmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{pmatrix}$$

We look at it in two different ways: A set of two 3-component vectors in R^3 (a 3D space) and set of three 2-component vectors in R^2 (a 2D plane). Here is how we may represent them.

$$\text{Column Cloud} : \left\{ \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} \right\}$$

and

$$\text{Row Cloud} : \left\{ \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \end{pmatrix} \right\}$$

Each row vector represents an observation measured on two correlated characteristics on an experimental unit called a subject. The column variables in general are measured in different unit of measurements.

2.1 A Geometric View of a Data Matrix

Our ability to visualize these clouds geometrically is limited to a 2D plane or to a 3D space. However, such geometric intuition we gain is valid in higher dimensions also.

The row cloud is also called observation space or subject space. When we plot the rows of a matrix, we label them with some symbols or subject IDs using a coloring scheme. When we plot the columns of a matrix, we label them with column IDs and also, we join an arrow from the center of the cloud to the terminal point of that vector. Such schemes of plotting may help in understanding some relationships between the variables or similarities pattern among subjects.

2.2 Centering a Data Matrix

The row cloud and the column cloud each have a mean vector which may call the center. The mean vector of all the observations or the row vectors is a meaningful center in applications. The individual mean of each column variable make the mean row vector. In some applications, we may be interested in the mean of all the column variables (the center in the column cloud) also. We move the center of the row cloud to the origin. This process is called the centering a data matrix like $\mathbf{Z}$ which we will explain as shown below.

$$\left\{ \mathbf{Z} = \begin{pmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{pmatrix}, \bar{\mathbf{Z}} = \begin{pmatrix} 2 & 3 \\ 2 & 3 \\ 2 & 3 \end{pmatrix}, \mathbf{Z}_c = \begin{pmatrix} 2 & -2 \\ -3 & 0 \\ 1 & 2 \end{pmatrix} \right\}$$

We first find the mean row vector which is $\{2, 3\}$ in our example. Note that the quantity 2 is the mean of the first column of $\mathbf{Z}$ and the quantity 3 is the mean of the second column of $\mathbf{Z}$. Then we create the mean matrix $\bar{\mathbf{Z}}$, same size as $\mathbf{Z}$, by repeating the mean values as many times as required to make it compatible with $\mathbf{Z}$ such that we can subtract the mean row vector from each row vector. In the example above, $\mathbf{Z}_c$ is the matrix with its mean row vector shifted to the origin. Check that the mean of each column vector is zero in $\mathbf{Z}_c$.

3 Dot Product

The **dot product** is perhaps the most frequently used operation in matrix/linear algebra. This operation is valid on two equal size vectors. This is represented by $\mathbf{x}_1 \cdot \mathbf{x}_2$ and computed as $\sum \mathbf{x}_{1i} \times \mathbf{x}_{2i}$, the sum of products of respective pairs in the two vectors. It is usually written as $\mathbf{x}_1'\mathbf{x}_2$ in matrix language without using "$\cdot$" notation.

Length or Norm: The quantity $\mathbf{x}_1 \cdot \mathbf{x}_1$ is called the squared length of $\mathbf{x}_1$. The $\sqrt{\mathbf{x}_1 \cdot \mathbf{x}_1}$ is called the length or the euclidian norm a vector. As we see later, the norm of a vector is related to the standard deviation of the measurements in a vector when we center the vector to have mean zero. This quantity is directly affected by the unit of measurements. Check that the norm is about about 5 units for the first column of **Z** and it is about 6 units for the second column of **Z**. Also these norms for the columns of $\mathbf{Z}_c$ are about 3.7 and 2.8 respectively. The vectors with length/norm 1 are called **unit vectors**.

Let $\mathbf{d} = \mathbf{r}_1 - \mathbf{r}_2$ represent the difference vector. The length of this vector is called the euclidian distance between the two vectors. If $\mathbf{r}_2$ is the origin, then the length of **d** is the distance from the origin. The values of **d** for row vectors may be of interest to researchers who like to study patterns in a row cloud.

In geometric representation, the points with bigger norm in a subject space fall farther away from the center and the vectors with bigger norm in a column cloud will have longer arrows from the center. Also, the distance between the vectors in these clouds is an important measure to study the patterns and correlations.

Correlation: The quantity $\mathbf{x}_1 \cdot \mathbf{x}_2$ is related to the angle between the two vectors and also, it is related to Person correlation coefficient, a measure of linear relationship between the two vectors. Usually this quantity is of main interest in a column cloud. Check that it is about -2 units between the two columns in $\mathbf{Z}_c$. The negative sign indicates that the angle is obtuse and the correlation is negative. Again, just like length, this quantity is also affected by scales of measurements.

Perpendicular: If $\mathbf{x}_1 \cdot \mathbf{x}_2 = 0$, we say that the two vectors are uncorrelated or perpendicular. Sometimes, the word orthogonal is used as a synonym with the word perpendicular. Check that the vector $\mathbf{1}_n$ and any column in $\mathbf{Z}_c$ are perpendicular to each other. This is a geometric meaning for centering or moving the center of the row cloud to the origin.

A square matrix is called an **orthogonal** if its rows are pair-wise perpendicular unit vectors and its column vectors are also pair-wise perpendicular unit vectors.

3.1 Matrix Multiplication

Multiplication of a matrix **Z** and a vector **b** is represents as **x= Zb**. See it below how it can be computed and interpreted.

$$\left\{ \begin{pmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix} = \mathbf{x} = \begin{pmatrix} 17 \\ -1 \\ 17 \end{pmatrix} \right\}$$

$$\left\{ (4) \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix} + (1) \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 17 \\ -1 \\ 17 \end{pmatrix} \right\}$$

- In the first interpretation, we can view **x** as a result of the dot product of **b** with the row vectors of $\mathbf{Z}$. In this view, we see how $\mathbf{b}$ is correlated with the row vectors or what sorts of angle it is making with row vectors of $\mathbf{Z}$.

- In the second interpretation, we can view $\mathbf{x}$ as a linear combination of the column vectors in $\mathbf{Z}$ with combination coefficients (weights) provided by $\mathbf{b}$.

- In the third interpretation, we view $\mathbf{x}$ as an image of $\mathbf{b}$ when we consider $\mathbf{Z}$ as a matrix transformation (linear function). In this interpretation of $\mathbf{Z}$ as a function, the matrix has the domain (in our example, R^2) and co-domain (in our example, R^3). In this sense, $\mathbf{Z}$ transforms the vectors in the domain into a vector in the co-domain. In some applications, we will be looking for the vectors in the codomain that have the maximum distance from the origin.

Multiply two Matrices: If $\mathbf{A}$ is an n by q matrix and $\mathbf{B}$ is an q by r matrix, the result AB of their multiplication is an n by r matrix $\mathbf{X}$ defined only if the number of columns q in $\mathbf{A}$ is equal to the number of rows q in $\mathbf{B}$. This definition of matrix multiplication involves again the dot product operation between the rows of $\mathbf{A}$ and the columns of $\mathbf{B}$. In the example below, we will demonstrate how to do this and interpret the results in two important special cases of $\mathbf{A}$ and $\mathbf{B}$.

$$\left\{ \mathbf{Z}_c'\mathbf{Z}_c = \begin{pmatrix} 14 & -2 \\ -2 & 8 \end{pmatrix}, \mathbf{Z}_c\mathbf{Z}_c' = \begin{pmatrix} 8 & -6 & -2 \\ -6 & 9 & -3 \\ -2 & -3 & 5 \end{pmatrix} \right\}$$

- In the first product, $\mathbf{Z}_c'\mathbf{Z}_c$ we have carried out the dot products between the columns of $\mathbf{Z}_c$. For example, the quantity 16 is the squared distance of the first column of $\mathbf{Z}_c$ from the origin in R^3. It is easy to see why this product is a square symmetric matrix. This product summarizes the distances within and between the column vectors of $\mathbf{Z}_c$. This matrix may also be called as the variance-covariance type matrix.

- In the second product, $\mathbf{Z}_c\mathbf{Z}_c'$ we have carried out the dot products between the rows of $\mathbf{Z}_c$. For example, the quantity -6 is the dot product of the first row of $\mathbf{Z}_c$ and the second row of $\mathbf{Z}_c$. Note that this product is also a square symmetric matrix.

- The sum of all the diagonal numbers in $\mathbf{Z}_c'\mathbf{Z}_c$ or in $\mathbf{Z}_c\mathbf{Z}_c'$ (called the Trace) is an important number in many applications. This summarizes a total variation that is present in all the column variables. In applications, one objective is to see if this variation can be captured in one or two independent linear combinations of the column variables.

For any given rectangular matrix $\mathbf{Z}_c$ we can always produce these two types ($\mathbf{Z}_c'\mathbf{Z}_c$ and $\mathbf{Z}_c\mathbf{Z}_c'$) of square symmetric matrices. These two matrices belong another set of special matrices called non-negative definite matrices which play an important role in multivariate techniques that we will discuss in this document. In fact, when $\mathbf{Z}_c$ has the full column rank, the $\mathbf{Z}_c'\mathbf{Z}_c$ type matrices are called positive definite and $\mathbf{Z}_c\mathbf{Z}_c'$ type matrices are called positive semi-definite.

For the un-centered matrix $\mathbf{Z}$ also, with full column rank, the product $\mathbf{ZZ}'$ is positive semi-definite and $\mathbf{Z}'\mathbf{Z}$ is positive definite.

4 Basis and Coordinates

Linear independence is an important property of a set of vectors like the row cloud or the column cloud. We say that a set of vectors , say, the columns of $\mathbf{Z}$ is linearly dependent if a linear combination $\mathbf{Zb}$ is the zero vector for some non-zero vector $\mathbf{b}$. In other words, linear dependence implies that some vector in the set is a linear combination some other vectors in the set. In our example, we can show that the row cloud is a dependent set and the column cloud is an independent set. Ideally speaking, we like our column variables to form an independent set.

The **rank** is the number of independent vectors in a set is called the rank of that set. The rank of a matrix is the same as the rank of column cloud or the rank of the row cloud. It is a well known result that both the clouds will have the same rank. The rank of a matrix perhaps single most important idea that connects several issues related to matrix in applications.

A square matrix is called **non-singular** (invertible) if its row cloud (and hence column cloud) form an independent set. That is, it has full rank. Otherwise we call the matrix as singular or degenerate.

Basis and coordinates: A basis is simply a set of certain number of linearly independent vectors. For example, any set S = $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ of three independent 3 by 1 vectors of real values form a basis for R^3. That is, any arbitrary vector in R^3 can be expressed uniquely as a linear combination of these three vectors. In general, the columns of a non-singular matrix $\mathbf{E}$ represent a basis for $\mathbb{R}^n$. Then

$$\mathbf{x} = \mathbf{E}\mathbf{y}, \quad or, \quad \mathbf{E}^{-1}\mathbf{x} = \mathbf{y}.$$

Here we say $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ in the basis $\mathbf{E}$.

In our example, the columns of $\mathbf{Z}_c'\mathbf{Z}_c$ form a basis for R^2 where as the columns of $\mathbf{Z}$ or the columns of $\mathbf{Z}_c\mathbf{Z}_c'$ cannot form a basis for R^3.

Standard basis: Suppose $\mathbf{E}$ is the identity matrix, we call the corresponding basis the standard basis. In this case, the coordinate vector of $\mathbf{x}$ is the vector itself. In applications, we collect data as standard coordinates and find coordinates in some other basis that makes it easier to understand and explain the complexities hidden in the data.

Orthonormal Basis: When $\mathbf{E}$ is an orthogonal matrix, $\mathbf{E}^{-1} = \mathbf{E}'$, the columns of $\mathbf{E}$ form an orthonormal basis. When $\mathbf{E}$ is an orthogonal matrix, we see that $\mathbf{E}'\mathbf{x} = \mathbf{y}$. That is, the coordinate of $\mathbf{x}$ is a set of dot products of $\mathbf{x}$ with the basis vectors.

The standard basis is an orthonormal basis. Finding some other orthonormal basis is a major goal in many multivariate techniques. The orthogonal matrices are also called rotation matrices. In other words, non-trivial orthonormal bases are some rotations of standard basis. Check that the the columns and the rows of the following matrix form an orthonormal basis for R^2.

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

4.1 Eigen Basis and SVD

Eigenvector and Eigenvalue: An non-zero vector **b** is an eigenvector associated with a square matrix, say **C**, if it is a solution of the linear system of equations $\mathbf{Cb} = \lambda\mathbf{b}$ for some real number λ. Here we call the scalar quantity λ as the corresponding eigenvalue of **b**. In other words, the eigenvectors of **C** are those non-zero vectors that do not change their directions when **C** acts on them. They only get magnified or shrunk by the eigenvalue. Note that if **b** is an eigenvector with eigenvalue λ, then for any non-zero real number a, $a \times \mathbf{b}$ is also an eigenvector with eigenvalue λ. in this sense, we say eigenvectors are unique upto scalar multiplications. All software like SAS and Mathematica generate eigenvectors that have unit lengths.

There are some issues related to its existence, uniqueness, non-negativity, orthogonality etc. However, we list below two important results when **C** is a symmetric matrix of the type $\mathbf{ZZ'}$ or $\mathbf{Z'Z}$.

- All the eigenvalues are non-negative real numbers. If all the columns are independent (full rank), then all the eigenvalues are positive. If all the eigenvalues are positive, we say that the square matrix is *positive definite*. Otherwise these types of symmetric matrices are *positive semi-definite*. In our case, $\mathbf{ZZ'}$ is a positive semi-definite and $\mathbf{Z'Z}$ is positive definite when we assume rank is q.

- if all eigenvalues are distinct, then all the corresponding eigenvectors are uniquely determined (upto to scalar multiplications) and are all orthogonal to each other. Hence they can be converted to an orthonormal basis. This property is true in general for any symmetric matrix. If eigenvalues are not distinct, then there are infinitely many possibilities to form an orthonormal basis.

EVD: Assume **C** is positive semi definite definite symmetric matrix of the type $\mathbf{ZZ'}$. Then we can factor **C** , an orthogonal matrix **P** and a diagonal matrix Λ, as

$$\mathbf{CP} = \mathbf{P}\Lambda, \quad equivalently, \quad \mathbf{C} = \mathbf{P}\Lambda\mathbf{P'}.$$

This is known as Eigenvalue-Eigenvector Decomposition (EVD) of **C**. Here only the first q eigenvalues are positive and the remaining eigenvalues are all zeroes in the diagonal of Λ. All the n columns of $\mathbf{P}$ form an orthonormal basis (*eigen basis*) for all of $\mathbb{R}^n$ space, but only the first q columns (say,$\mathbf{P}_q$) corresponding to the q positive eigenvalues form an orthonormal basis for the column space of $\mathbf{Z}$.

Many software reduce this EVD using only the first q columns of $\mathbf{P}$ and the first q positive eigenvalues in Λ. Hence we state the short version as

$$\mathbf{C} = \mathbf{P}_q \Lambda_q \mathbf{P}_q'.$$

The short version generally do not affect the results in many applications.

EVD Assume that $\mathbf{C}$ is a q by q positive definite symmetric matrix of the type $\mathbf{Z}'\mathbf{Z}$. Then we can factor $\mathbf{C}$ with an orthogonal matrix $\mathbf{Q}$ and a diagonal matrix Λ_q of positive values, as

$$\mathbf{C}\mathbf{Q} = \mathbf{Q}\Lambda_q, \quad equivalently, \quad \mathbf{C} = \mathbf{Q}\Lambda_q\mathbf{Q}'.$$

The columns of $\mathbf{Q}$ form an orthonormal eigen basis for the row space of $\mathbf{Z}$ (and for all of $\mathbb{R}^q$.)

Same Eigenvalues Both the EVDs ($\mathbf{Z}\mathbf{Z}'$ type and $\mathbf{Z}'\mathbf{Z}$ type) share the same positive eigenvalues.

SVD: As we saw, for any given n by q rectangular matrix $\mathbf{Z}$ there are two associated symmetric matrices $\mathbf{Z}\mathbf{Z}'$ and $\mathbf{Z}'\mathbf{Z}$. Combining the two associated EVD's, we can write $\mathbf{Z}$ as

$$\mathbf{Z} = \mathbf{P}\Lambda^{1/2}\mathbf{Q}', \text{using the full version of } \mathbf{Z}\mathbf{Z}',$$

or

$$\mathbf{Z} = \mathbf{P}_q\Lambda_q^{1/2}\mathbf{Q}', \text{using the short version of } \mathbf{Z}\mathbf{Z}'.$$

Here in the middle factor $\Lambda^{1/2}$, the diagonal values are known as the singular values of $\mathbf{Z}$ and the whole factorization is known as the Singular Value Decomposition (SVD) of a rectangular matrix $\mathbf{Z}$. The columns of $\mathbf{P}$ are called left singular vectors of $\mathbf{Z}$ (eigenvectors of $\mathbf{Z}\mathbf{Z}'$) and the columns of $\mathbf{Q}$ are called the right singular vectors of |Zm (eigenvectors of $\mathbf{Z}'\mathbf{Z}$.)

An Example

$$SVD(\mathbf{Z}_c) : \left\{ \mathbf{P} = \begin{pmatrix} -0.653 & -0.491 \\ 0.751 & -0.32 \\ -0.099 & 0.81 \end{pmatrix}, \Lambda^{1/2} = \begin{pmatrix} 3.822 & 0. \\ 0. & 2.719 \end{pmatrix}, \mathbf{Q} = \begin{pmatrix} -0.957092 & 0.289784 \\ 0.289784 & 0.957092 \end{pmatrix} \right\}$$

$$EVD(\mathbf{Z_c Z_c'}) : \left\{ \mathbf{P} = \begin{pmatrix} -0.653 & -0.491 \\ 0.751 & -0.32 \\ -0.099 & 0.81 \end{pmatrix}, \Lambda = \begin{pmatrix} 14.6077 & 0. \\ 0. & 7.39296 \end{pmatrix} \right\}$$

$$EVD(\mathbf{Z_c' Z_c}) : \left\{ \mathbf{Q} = \begin{pmatrix} -0.957092 & 0.289784 \\ 0.289784 & 0.957092 \end{pmatrix}, \Lambda = \begin{pmatrix} 14.6077 & 0. \\ 0. & 7.39296 \end{pmatrix} \right\}$$

Comments: The SVD of a matrix is used to solve many different problems in applications. We can interpret these results in many ways. We will list a few important interpretations.

- Check the measure of total variation the SVD analyzes, the trace of the matrices $\mathbf{Z_c Z_c'}$or $\mathbf{Z_c' Z}$. It is about 22 units. This is also equal to the trace of Λ, the sum of eigenvalues. The first eigenvalue of 14.7 tells us that about 67 percent of the total variation is explained by the first eigenvector (either the first column of **P** or the first column of **Q**.

  The columns of **P** form a new basis for the column cloud and the columns of **Q** form a new basis for the row cloud.

- Next we apply the ideas of coordinates and basis that we talked about earlier: **x**= **Py**. Let the first column of $\mathbf{Z_c}$ be **x**. Then we can check that **x**= **Py** where **y** equals the first column of $\Lambda^{1/2}\mathbf{Q}'$. In other words,

$$\left\{ \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.652521 & -0.490799 \\ 0.751304 & -0.3197 \\ -0.0987837 & 0.810499 \end{pmatrix} \cdot \begin{pmatrix} -3.65774 \\ 0.788002 \end{pmatrix} \right\}$$

Column Cloud: Here the quantity -3.66 is the coordinate of **x** (an arrow in the column cloud) with respect to the first axis (the first column of **P** and 0.79 is the coordinate of **x** with respect to the second axis. To another way to look at it, the quantity -3.66 is the dot product of **x** with the first column of **P** and the quantity 0.79 is the dot product of **x** with the second column of **P**. That is, **x** is negatively correlated with the first axis and positively correlated with the second axis.

Similarly, the vector $(1.10747, 2.6026)'$ contains the coordinates of the second column of $\mathbf{Z_c}$ in basis **P**. Note that $\{-3.66, 1.11\}$ are the coordinates of the columns of $\mathbf{Z_c}$ with respect to the first axis in the column cloud. The sum of squares of these two values is the first eigenvalue.

Some researchers pay attention to the values in a basis vector. These values are called factor loadings. Remember that the coordinate of **x** with a axis is the dot product with these loadings. Hence the magnitude and the sign of these loadings can be used used in determining which loadings play a dominant role in overall correlation of **x** with that axis.