Short Course

MULTIVARIATE STATISTICS IN ANIMAL SCIENCE USING R

Alberto Cesarani, Corrado Dimauro, Nicolò Pietro Paolo Macciotta,

Dipartimento di Agraria, Università degli Studi di Sassari

Athens, January 20-24 2025

Program outline

Structure of multivariate observations

- ✓ Elements of matrix algebra
- ✓ Structure of (co)variance and correlation matrices
- ✓ Eigenvectors and eigenvalues
- ✓ Examples with R

Principal Component Analysis (PCA)

- ✓ Theory
- ✓ Use of PCA to reduce the dimensionality of complex multivariate systems: practical examples
- ✓ Examples and practice with R

Program outline

Factor Analysis (FA) for studying covariance in multivariate complex systems

- ✓ Theory
- ✓ Practical examples
- ✓ Comparison between FA and PCA.
- ✓ Examples and practice with R

Canonical discriminant analysis (CDA)

- ✓ Theory
- ✓ Use of CDA to highlight differences among groups
- ✓ Examples and practice with R

Program outline

Cluster analysys (CA)

- ✓ Unsupervised learning
- ✓ Use of CA for clustering objects together
- ✓ Examples and practice with R

Recalls of matrix algebra

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

Matrix = set of numbers organized in i rows and j columns

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}_{3x3} = \begin{bmatrix} 1 & 0 & 9 \\ 3 & 15 & 7 \\ 11 & 2 & 4 \end{bmatrix}_{3x3}$$

The dimensions of a matrix are defined by the number of rows and number of columns

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} \\ a_{21} & a_{22} & \dots & a_{2j} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ij} \end{bmatrix}_{(ixj)}$$

Key elements of a matrix

- **Rows**
- **Columns**
- **Dimension**
- **Rank**
- **Determinant**
- **Inverse**

$$a_{ij}$$

$$\begin{array}{cc} & m \text{ colonne} \\ & \text{j cresce} \\ \begin{array}{c} n \text{ righe} \\ \text{i cresce} \end{array} & \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \end{array}$$

matrice n x m

Kind of matrices

Rectangular

- ✓ ex. data matrix
- ✓ rows=experimental units
- ✓ columns=variables

$$\mathbf{B}_{4,3} = \begin{bmatrix} 3 & 3 & 1 \\ 5 & 1 & 1 \\ 1 & 3 & 2 \\ 2 & 1 & 3 \end{bmatrix}$$

Squared

- ✓ ex. Covariance matrix
- ✓ Correlation matrix
- ✓ rows==columns
- ✓ Experimental units
- ✓ variables

$$\mathbf{A}_{3,3} = \begin{bmatrix} 2 & 1 & 2 \\ -4 & 3 & -1 \\ 0 & 2 & 1 \end{bmatrix}$$

Kind of matrices

Row vector

$$\mathbf{x}'_{1,n} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

Column vector

$$\mathbf{x}_{m,1} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_m \end{bmatrix}$$

Matrix: set of row and column vectors

$$\begin{bmatrix} \mathbf{a}_{.1} \dots \mathbf{a}_{.j} \dots \mathbf{a}_{.n} \end{bmatrix} = \begin{bmatrix} \mathbf{a}'_{1.} \\ \mathbf{a}'_{2.} \\ . \\ \mathbf{a}'_{m.} \end{bmatrix} \longleftarrow$$

$$\mathbf{A}_{m,n} = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ . & \dots & . & \dots & . \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix}$$

Particular kind of matrices

Diagonal matrices

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

Scalar

$$\mathbf{S} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Null

$$\mathbf{N} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Unit

$$\mathbf{U} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Identity

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{A} \cdot \mathbf{I} = \mathbf{I} \cdot \mathbf{A} = \mathbf{A}$$

Simmetrical matrices

Simmetrical matrix

$$\mathbf{S} = \begin{bmatrix} 1 & 4 & 10 & 6 \\ 4 & 2 & 7 & 5 \\ 10 & 7 & 3 & 2 \\ 6 & 5 & 2 & 4 \end{bmatrix}$$

$$\mathbf{S'} = \mathbf{S}$$

$$\mathbf{S} = \begin{bmatrix} 1 & 4 & 10 & 6 \\ & 2 & 7 & 5 \\ & & 3 & 2 \\ sym & & & 4 \end{bmatrix}$$

Trace (*tr*) of a matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

$$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + a_{33} + a_{44}$$

Trace_A=sum(diag(A))

Transposition

$\mathbf{A} = \mathbf{A'}$ or $(\mathbf{A}^{T})$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 4 & 5 & 6 \end{bmatrix}$$

$$\mathbf{A'} = \mathbf{A}^{T} \begin{bmatrix} 1 & 3 & 4 \\ 2 & 1 & 5 \\ 3 & 2 & 6 \end{bmatrix}$$

Exchange columns/rows

A_transpose=t(A)

Matrix operations

Sum

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \qquad \mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \\ a_{31} + b_{31} & a_{32} + b_{32} \end{bmatrix}$$

For matrices of equal dimensions

Multiplication of a scalar for a matrix

$$k\mathbf{A} = k \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = \begin{bmatrix} ka_{11} & ka_{12} \\ ka_{21} & ka_{22} \\ ka_{31} & ka_{32} \end{bmatrix}$$

Matrix operations: multiplication

"Along the row, down by the column………"

Column of the first matrix = rows of the second

$$\begin{bmatrix} 11 & -4 \\ 6 & 0 \\ 8 & 13 \end{bmatrix} \bullet \begin{bmatrix} 3 & 2 \\ 1 & 15 \end{bmatrix} = \begin{bmatrix} (11*3)+(-4*1) & (11*2)+(-4*15) \\ (6*3)+(0*1) & (6*2)+(0*15) \\ (8*3)+(13*1) & (8*2)+(13*15) \end{bmatrix} = \begin{bmatrix} 29 & -38 \\ 18 & 12 \\ 37 & 211 \end{bmatrix}$$

(3 x 2) x (2 x 2) (3 x 2)

result = nrows of the first, ncolumns of the second

Vector multiplications

$$\mathbf{c} = \mathbf{a}'\mathbf{b} = \begin{bmatrix} 1 & 6 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix} = 5 + 12 + 3 = 20$$

$$\mathbf{C} = \mathbf{ab}' = \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 6 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 30 & 15 \\ 2 & 12 & 6 \\ 1 & 6 & 3 \end{bmatrix}$$

The determinant of a squared matrix

- ✓ The determinant of a squared matrix is a number that characterizes the matrix
- ✓ The determinant is important for obtaining the inverse of a squared matrix
- ✓ The determinant of a squared matrix **A** is indicated as |**A**|
- ✓ Not all squared matrix can be inverted
- ✓ Only squared matrix that have the determinant different from zero can be inverted

Determinant of a squared matrix

Determinant of a squared matrix A

- ✓ Number associated to the matrix. Named as $|\mathbf{A}|$ or det($\mathbf{A}$).
- ✓ Important for obtaining the inverse of a matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$$

$$\mathbf{B} = \left[\begin{array}{cc|ccc|cc} & & a_{11} & a_{12} & a_{13} & & \\ & a_{23} & a_{21} & a_{22} & a_{23} & a_{21} & \\ a_{32} & a_{33} & a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array}\right]$$

$$\det(\mathbf{B}) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{22}a_{31}$$

$$\ldots_{2} - a_{13}a_{12}a_{31}$$

$$\mathbf{B} = \begin{bmatrix} 10 & 1 & 1 \\ 3 & 2 & 1 \\ 1 & 6 & 9 \end{bmatrix}$$

$$\det|\mathbf{B}| = \boxed{10\cdot2\cdot9} + \boxed{1\cdot1\cdot1} + \boxed{1\cdot3\cdot6} - 1\cdot3\cdot9 - 10\cdot1\cdot6 - 1\cdot2\cdot1 =$$

$$= 180 + 1 + 18 - 27 - 60 - 2 = 110$$

Rank of a squared matrix

If $|\mathbf{A}|=0$ there is not a unique **A** inverse

$|\mathbf{A}|=0$ when one or more rows (or columns) of A are a linear combination of other rows (or columns)

$$\mathbf{A}=\begin{bmatrix} 2 & -4 & 6 \\ 6 & 2 & 4 \\ 10 & 8 & 2 \end{bmatrix}$$

Third column is the difference between the **first** and the **second** columns.

Rank of a squared matrix

- The **rank** of a squared matrix n x n is the number of rows (or columns) linearly independent

- If the **rank** is equal to the number of rows (or columns) the matrix is said to be **full rank**

- If the **rank** is smaller than the number of rows (or columns) the matrix is said to be **not full rank**

- A full **rank matrix** has the determinant different from zero and has an inverse

- A not full rank matrix has the determinant equal to zero (**singular**) and cannot be inverted

Vectors and space

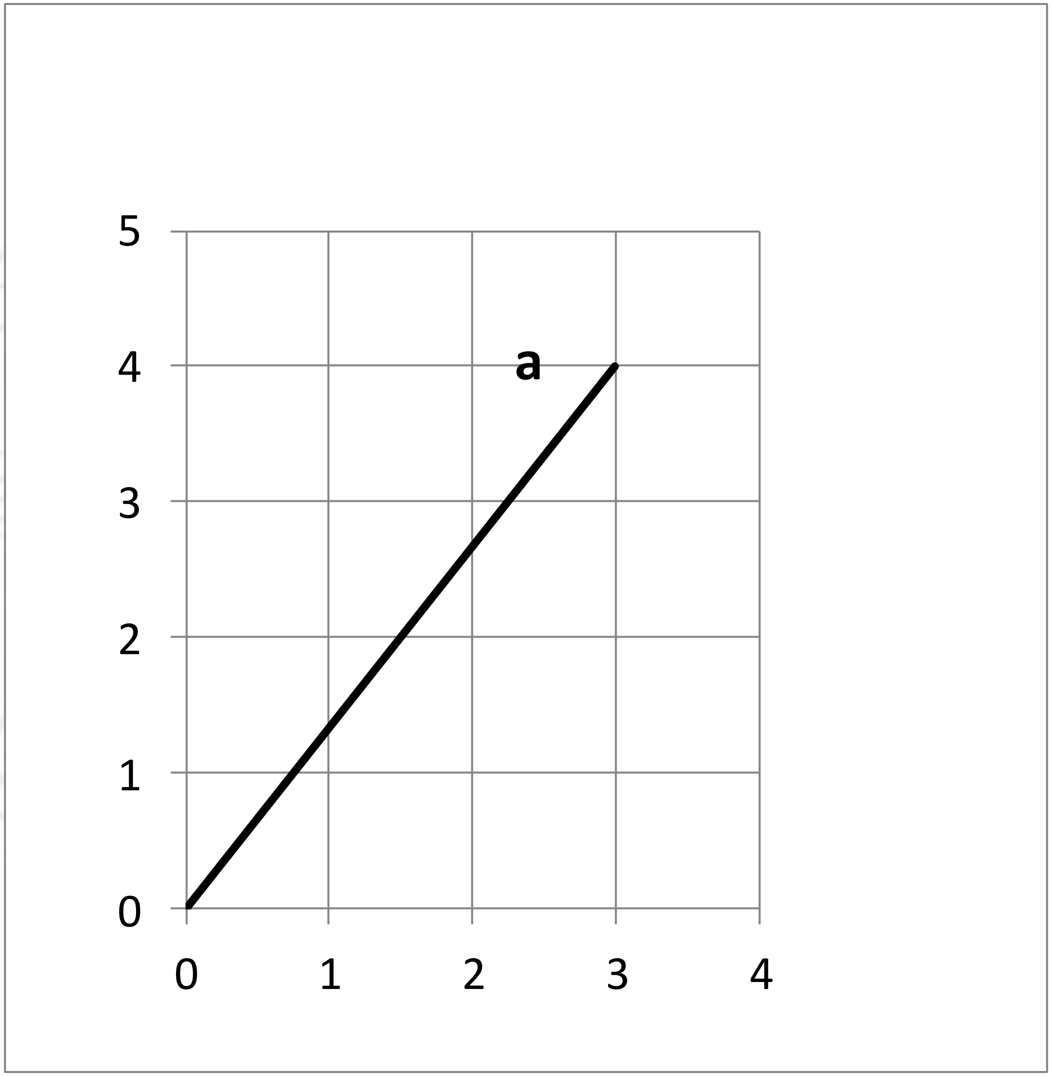
A vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \cdots\cdots \\ \cdots\cdots \\ \cdots\cdots \\ x_n \end{bmatrix}$$

defines the coordinates of a point in the Euclidean spce at n-dimensions.

The point is one of the extremes of a segment that starts in the origin of the axis

Vectors and space



Vector inner products

The inner product of two vectors with the same number of elements is the sum of the products of the corresponding elements

$$\mathbf{x}'\mathbf{y} = [x_1 \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad x_n] \begin{bmatrix} y_1 \\ \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

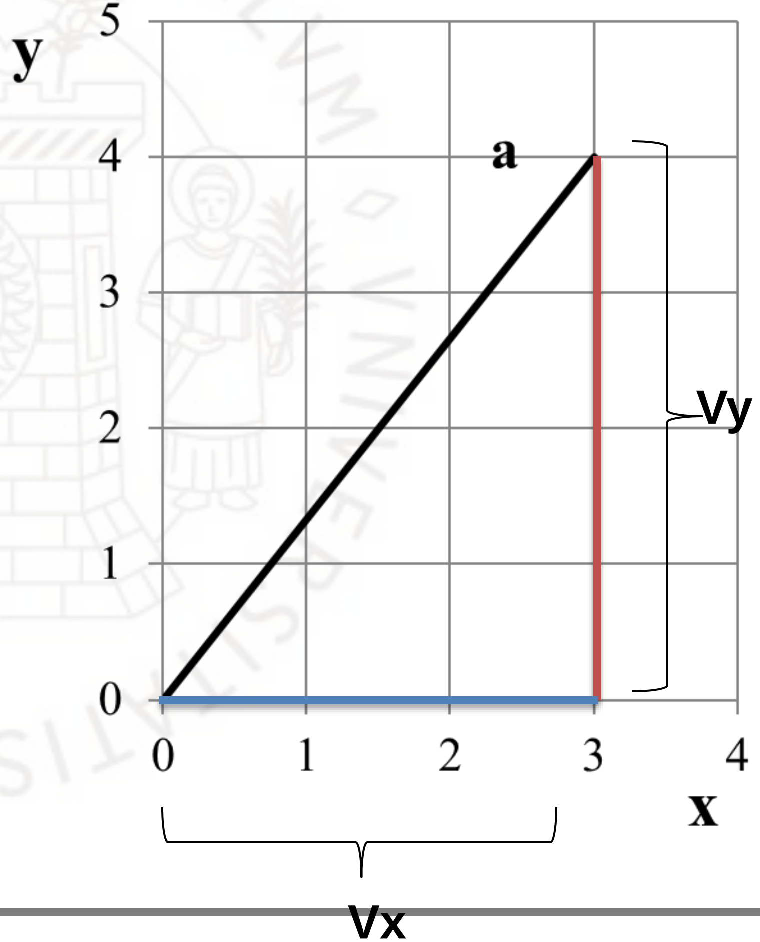
Vector inner products

The inner product of a vectors **x** by itself is the sum of squared elements of **x**

$$\mathbf{x}'\mathbf{x}=[3 \quad 2 \quad 1 \quad 4 \quad 6]\begin{bmatrix}3\\2\\1\\4\\6\end{bmatrix}=9+4+1+16+36=66$$

The inner product of **x'x** is called squared length of **x**, because it represents the squared distance from the origin of the axis of the point defined by the n elements of **x**

Vector inner product



$$\mathbf{a} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$\|\mathbf{a}\| = \sqrt{V_x^2 + V_y^2}$$

$$\|\mathbf{a}\| = \sqrt{\mathbf{a}'\mathbf{a}} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \sqrt{9+16} = 5$$

Vector normalisation

A normalised vector is divided by its length

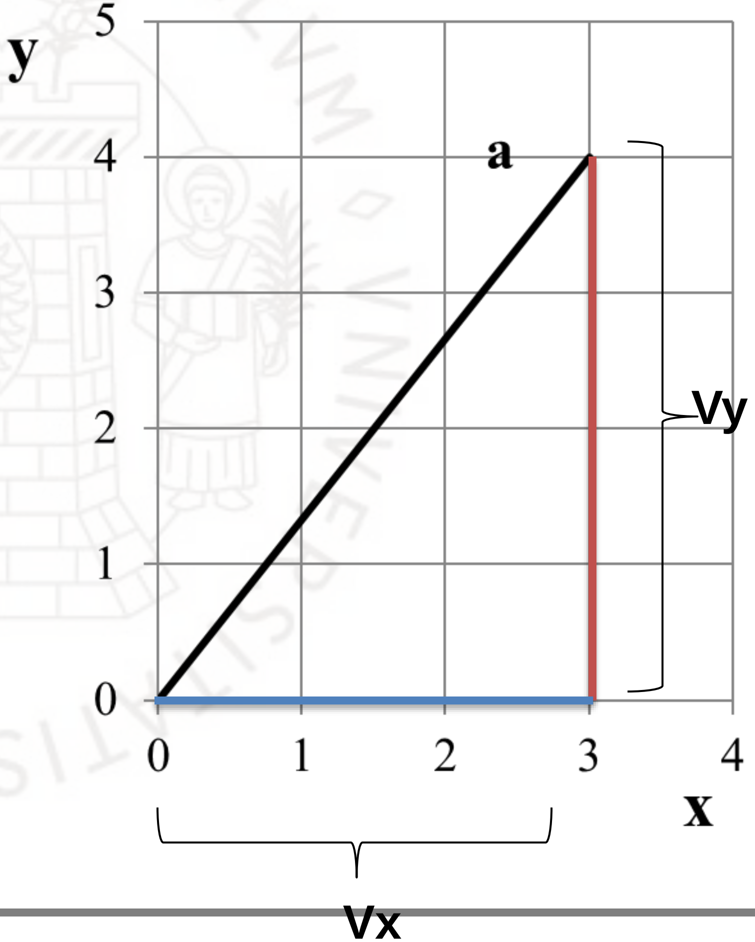
$$\mathbf{a'a} = \begin{bmatrix} 2 & 6 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix} \qquad \|\mathbf{a}\| = \sqrt{\mathbf{a'a}} = \sqrt{49} = \mathbf{7}$$

Normalised vector

$$\mathbf{n} = \begin{bmatrix} \frac{2}{7} \\ \frac{6}{7} \\ \frac{2}{7} \end{bmatrix} \qquad \mathbf{n'n} = 1$$

Vector normalisation

uniss UNIVERSITÀ DEGLI STUDI DI SASSARI agraria DIPARTIMENTO DI AGRARIA



$$\mathbf{a} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

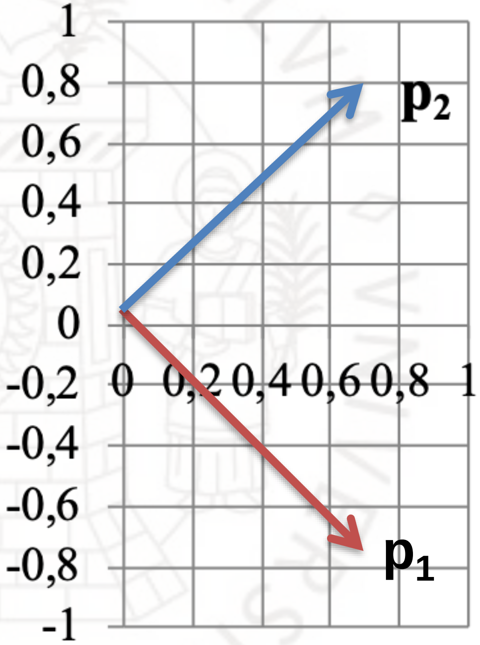
$$\|\mathbf{a}\| = \sqrt{V_x^2 + V_y^2}$$

$$\|\mathbf{a}\| = \sqrt{\mathbf{a}'\mathbf{a}} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \sqrt{9+16} = 5$$

$$\mathbf{a}_n = \frac{1}{\|\mathbf{a}\|} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \frac{1}{5} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}$$

$$\mathbf{a}_n{}'\mathbf{a}_n = 1$$

Orthogonal vectors and matrices



$$\mathbf{p}_1 = \begin{bmatrix} \frac{2}{3} \\ -\frac{5}{\sqrt{3}} \end{bmatrix} \qquad \mathbf{p}_2 = \begin{bmatrix} \frac{5}{\sqrt{3}} \\ \frac{2}{3} \end{bmatrix}$$

Orthogonal

$\mathbf{p_1'p_2=0}$

Orthonormal

$\mathbf{p_1'p_1= p_2'p_2 =1}$,

Orthogonal vectors and matrices

A squared matrix is said to be orthogonal if ***P'P=P⁻¹P=I***, *i.e. if its columns are normal and orthogonal vectors*

$$\mathbf{P'P} = \begin{bmatrix} \frac{2}{3} & -\frac{\sqrt{5}}{3} \\ \frac{\sqrt{5}}{3} & \frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} \frac{2}{3} & \frac{\sqrt{5}}{3} \\ -\frac{\sqrt{5}}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{4}{9}+\frac{5}{9} & -\frac{2\sqrt{5}}{9}+\frac{2\sqrt{5}}{9} \\ -\frac{2\sqrt{5}}{9}+\frac{2\sqrt{5}}{9} & \frac{5}{9}+\frac{4}{9} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$