

# Basic methods for data analysis

Philipp Renz

January 4, 2022



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Linear algebra: Basics</b>	<b>3</b>
2.1	Matrix multiplication . . . . .	3
2.1.1	Vector-vector products . . . . .	4
2.1.2	Matrix-vector multiplication . . . . .	4
2.1.3	Matrix-matrix multiplication . . . . .	5
<b>3</b>	<b>Principal component analysis</b>	<b>7</b>
3.1	Variance maximization . . . . .	7



# Chapter 1

## Introduction



## Chapter 2

# Linear algebra: Basics

In data analysis we will often represent our data as matrices. Consequently many algorithms are described using the concept of matrices and their properties which could be broadly called the study of linear algebra. Here I don't want to give an axiomatic introduction but rather provide some basic results with the prerequisite that the reader already knows some basics about vectors and matrices. We will denote vectors as bold lowercase letters

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \quad (2.1)$$

where  $a_i$  is the  $i^{th}$  entry of the vector  $\mathbf{a}$ .

Matrices are denoted by bold uppercase letters

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}. \quad (2.2)$$

Depending on the context it is useful to view a matrix as a concatenation of either row or column vectors.

$$\mathbf{A} = \begin{pmatrix} - & \mathbf{a}_1^r & - \\ & \vdots & \\ - & \mathbf{a}_m^r & - \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{a}_1^c & \dots & \mathbf{a}_n^c \\ | & & | \end{pmatrix}. \quad (2.3)$$

We made an explicit distinction between row- and column vectors using superscripts here but these will often be left out for better readability.

## 2.1 Matrix multiplication

Given two matrices

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in R^{m \times n}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & \dots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{np} \end{pmatrix} \in R^{n \times p} \quad (2.4)$$

their product  $\mathbf{C} = \mathbf{AB}$  is of shape  $(m \times p)$  and its entries are

$$c_{ik} = \sum_{j=1}^n a_{ij} b_{jk} \quad (2.5)$$

The resulting matrix  $\mathbf{C}$  has a shape of  $(m \times p) \leftarrow (m \times n)(n \times p)$ . The inner dimension  $n$  is "eliminated" by summing over it. One can also see that the sum moves horizontally/rowwise over the entries of  $\mathbf{A}$  and down the columns of  $\mathbf{B}$

Often we are interested in a special case in which one of the matrices reduces to a vector, that means it has only one column or one row. We define a column vector as a shorthand notation for a matrix  $\mathbf{B} \in \mathbb{R}^{n \times 1}$  with only one column and drop the column index:

$$\mathbf{B} = \begin{pmatrix} b_{11} \\ \vdots \\ b_{n1} \end{pmatrix} \hat{=} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \mathbf{b} \quad (2.6)$$

Implicitly when we talk of a vector  $\mathbf{b}$  with  $n$  elements we mean a matrix  $\mathbf{B} \in \mathbb{R}^{n \times 1}$ . When we want to talk about a matrix with only one row we write this a transposed vector

$$\mathbf{b}^T = (b_1 \dots b_n) \hat{=} (b_{11} \dots b_{1n}) \quad (2.7)$$

### 2.1.1 Vector-vector products

Two vectors can be multiplied with each other in two different ways. The scalar or inner product is defined as

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \sum_{i=1}^n a_i b_i \quad (2.8)$$

This can be seen as a multiplication of matrices with shapes  $(1 \times m)(m \times 1) \rightarrow (1 \times 1)$  and results in a scalar value.

The outer product between two vectors is

$$\mathbf{C} = \mathbf{a} \mathbf{b}^T = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} (b_1 \dots b_n) = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \dots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \dots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m b_1 & a_m b_2 & \dots & a_m b_n \end{pmatrix} \quad (2.9)$$

This means that every column is a multiple of  $\mathbf{a}$  and every row a multiple of  $\mathbf{b}$ . We see that the entry  $c_{ij} = a_i b_j$ .

### 2.1.2 Matrix-vector multiplication

The product of a matrix and vector  $\mathbf{c} = \mathbf{A} \mathbf{b}$  again is a vector. If  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{b} \in \mathbb{R}^n$  their product will be of shape  $(m \times 1)$ , or  $\mathbf{c} \in \mathbb{R}^m$ .

#### Linear combination of columns

One can view matrix-vector product,  $\mathbf{c} = \mathbf{A} \mathbf{b}$ , as calculating a linear combination of the columns of  $\mathbf{A}$  weighted by the entries of  $\mathbf{b}$ :

$$\mathbf{A} \mathbf{b} = \begin{pmatrix} \sum_{j=1}^n a_{1j} b_j \\ \sum_{j=1}^n a_{2j} b_j \\ \vdots \\ \sum_{j=1}^n a_{mj} b_j \end{pmatrix} = \begin{pmatrix} \textcolor{red}{a}_{11} b_1 + \textcolor{blue}{a}_{12} b_2 + \dots + \textcolor{green}{a}_{1n} b_n \\ \textcolor{red}{a}_{21} b_1 + \textcolor{blue}{a}_{22} b_2 + \dots + \textcolor{green}{a}_{2n} b_n \\ \vdots \\ \textcolor{red}{a}_{m1} b_1 + \textcolor{blue}{a}_{m2} b_2 + \dots + \textcolor{green}{a}_{mn} b_n \end{pmatrix} = \begin{pmatrix} \textcolor{red}{|} \textcolor{red}{a}_1 b_1 + \textcolor{blue}{|} \textcolor{blue}{a}_2 b_2 + \dots + \textcolor{green}{|} \textcolor{green}{a}_n b_n \\ \textcolor{red}{|} \textcolor{red}{a}_1 b_1 + \textcolor{blue}{|} \textcolor{blue}{a}_2 b_2 + \dots + \textcolor{green}{|} \textcolor{green}{a}_n b_n \\ \vdots \\ \textcolor{red}{|} \textcolor{red}{a}_1 b_1 + \textcolor{blue}{|} \textcolor{blue}{a}_2 b_2 + \dots + \textcolor{green}{|} \textcolor{green}{a}_n b_n \end{pmatrix} \quad (2.10)$$

$$= \textcolor{red}{a}_1 b_1 + \textcolor{blue}{a}_2 b_2 + \dots + \textcolor{green}{a}_n b_n = \sum_{i=1}^n \mathbf{a}_i b_i \quad (2.11)$$

One prominent example where this view is useful is that if the columns of  $\mathbf{A}$  form a basis. Then a vector  $\mathbf{c}$  can be written as  $\mathbf{c} = \mathbf{A} \mathbf{c}_A$ , where  $\mathbf{c}_A$  are the coordinates of  $\mathbf{c}$  with respect to the basis  $\mathbf{A}$ . Equivalently if the columns of  $\mathbf{B}$  form another basis then

$$\mathbf{c} = \mathbf{A} \mathbf{c}_A = \mathbf{B} \mathbf{c}_B. \quad (2.12)$$

From this equation one can easily get the prescription of calculating the coordinates wrt. to  $\mathbf{A}$  given those wrt.  $\mathbf{B}$ ,

$$\mathbf{c}_A = \mathbf{A}^{-1} \mathbf{B} \mathbf{c}_B. \quad (2.13)$$



**Dot product**

Alternatively we can view a matrix vector product  $\mathbf{A}\mathbf{b}$  as calculating the dot products of  $\mathbf{b}$  with each row of  $\mathbf{A}$ ,

$$\mathbf{A}\mathbf{b} = \begin{pmatrix} - & \mathbf{a}_1 & - \\ & \vdots & \\ - & \mathbf{a}_m & - \end{pmatrix} \begin{pmatrix} | \\ \mathbf{b} \\ | \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{b} \\ \vdots \\ \mathbf{a}_m \cdot \mathbf{b} \end{pmatrix} \quad (2.14)$$

**2.1.3 Matrix-matrix multiplication**

Now that we've looked at some special cases of matrix multiplication we can go back to the general case  $\mathbf{C} = \mathbf{A}\mathbf{B}$ .

Recalling the definition of matrix multiplication

$$c_{ik} = \sum_{j=1}^n a_{ij}b_{jk} \quad (2.15)$$

we can look at the index  $k$  and see that the  $k$ -th column of  $\mathbf{C}$  is formed by the matrix vector product of  $\mathbf{A}$  and the  $k$ -th column vector of  $\mathbf{B}$ ,

$$\mathbf{C} = \mathbf{A}\mathbf{B} = \mathbf{A} \begin{pmatrix} | & & | \\ \mathbf{b}_1 & \dots & \mathbf{b}_p \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{A}\mathbf{b}_1 & \dots & \mathbf{A}\mathbf{b}_p \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{c}_1 & \dots & \mathbf{c}_p \\ | & & | \end{pmatrix}. \quad (2.16)$$

Breaking matrix-matrix multiplication down to multiple matrix vector operations allows us to view it as calculating multiple linear combinations of the columns of  $\mathbf{A}$  or calculating the all inner products between the rows of  $\mathbf{A}$  and the columns of  $\mathbf{B}$ ,

$$\mathbf{C} = \begin{pmatrix} - & \mathbf{a}_1 & - \\ & \vdots & \\ - & \mathbf{a}_m & - \end{pmatrix} \begin{pmatrix} | & & | \\ \mathbf{b}_1 & \dots & \mathbf{b}_p \\ | & & | \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \mathbf{a}_1 \cdot \mathbf{b}_2 & \dots & \mathbf{a}_1 \cdot \mathbf{b}_p \\ \mathbf{a}_2 \cdot \mathbf{b}_1 & \mathbf{a}_2 \cdot \mathbf{b}_2 & \dots & \mathbf{a}_2 \cdot \mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m \cdot \mathbf{b}_1 & \mathbf{a}_m \cdot \mathbf{b}_2 & \dots & \mathbf{a}_m \cdot \mathbf{b}_p \end{pmatrix}. \quad (2.17)$$

From this we can read that  $c_{ik} = \mathbf{a}_i \cdot \mathbf{b}_k$ .

**Sum of outer products**

Another view that is often useful is to write a matrix-matrix multiplication as a sum of outer products. Looking again at our basic definition

$$c_{ik} = \sum_{j=1}^n a_{ij}b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + a_{i3}b_{3k} \quad (2.18)$$

we can see that for example the term  $a_{i2}b_{2k}$  is the  $ik$ -th entry of the dot-product between the second column vector of  $\mathbf{A}$  and the second row vector of  $\mathbf{B}$ . Thus we can deduce that the matrix  $\mathbf{C}$  is a sum of outer-products,

$$\mathbf{C} = \mathbf{A}\mathbf{B} = \begin{pmatrix} | & & | \\ \mathbf{a}_1 & \dots & \mathbf{a}_n \\ | & & | \end{pmatrix} \begin{pmatrix} - & \mathbf{b}_1 & - \\ & \vdots & \\ - & \mathbf{b}_n & - \end{pmatrix} = \sum_{i=1}^n \mathbf{a}_i \mathbf{b}_i^T. \quad (2.19)$$



## Chapter 3

# Principal component analysis

**Motivating example** There is a park and we are given the locations of the trees in it given two numbers. It appears that they are approximately planted in a line with slight deviations. However we would like to find a single direction which tells us the most about the location of tree. In the figure it is evident that if we can get a very good estimate of where trees are, by knowing how far to go along the red line. The (blue) distance orthogonal to it is of less importance. If we encoded each point by it's red and blue distance we would find that the red ones vary a lot more than the blue ones. Thus we will define the most important direction as the one along which the data has the highest variance.

### 3.1 Variance maximization

Given a vector  $\mathbf{v}$  with unit length  $\|\mathbf{v}\| = 1$  we can calculate the coordinate wrt. to it via a dot product,

$$y_i = \mathbf{x}_i \cdot \mathbf{v}, \quad (3.1)$$

yielding a scalar for each data-point. The variance of the values  $\{y_1, \dots, y_n\}$  is

$$\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \quad (3.2)$$

$$= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i \mathbf{v})^2 - \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{v} \right)^2 = \quad \text{def. of } y_i \quad (3.3)$$

$$= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^m x_{ij} v_j x_{ik} v_k - \mathbf{v}^T \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{v} = \quad \text{write dot product as sum} \quad (3.4)$$

$$= \frac{1}{n} \sum_{j=1}^m \sum_{k=1}^m v_j \sum_{i=1}^n \mathbf{x}_{ij}^T \mathbf{x}_{ik} v_k - \mathbf{v}^T \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{v} = \quad \text{rearrange} \quad (3.5)$$

$$= \frac{1}{n} \sum_{j=1}^m \sum_{k=1}^m v_j [\mathbf{X}^T \mathbf{X}]_{jk} v_k - \mathbf{v}^T \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{v} = \quad \text{replace sum by m.m.} \quad (3.6)$$

$$= \frac{1}{n} \sum_{j=1}^m \sum_{k=1}^m v_j [\mathbf{X}^T \mathbf{X}]_{jk} v_k - \mathbf{v}^T \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{v} = \quad \text{replace sum by m.m.} \quad (3.7)$$

$$= \frac{1}{n} \mathbf{v}^T \mathbf{X}^T \mathbf{X} \mathbf{v} - \mathbf{v}^T \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{v} = \quad \text{replace sums by m.m.} \quad (3.8)$$

$$= \mathbf{v}^T \left( \frac{1}{n} \mathbf{X}^T \mathbf{X} - \boldsymbol{\mu} \boldsymbol{\mu}^T \right) \mathbf{v} = \quad \text{rearrange} \quad (3.9)$$

$$= \mathbf{v}^T \mathbf{C} \mathbf{v} \quad \text{def. of covariance} \quad (3.10)$$

Since  $\mathbf{C}$  is real and symmetric we can express it using it's eigenvalues and eigenvectors,

$$\mathbf{C} = \sum_{j=1}^m \lambda_j \mathbf{w}_j \mathbf{w}_j^T. \quad (3.11)$$

Since  $(w_1, \dots, w_m)$  form an orthonormal basis we can also write  $\mathbf{v}$  as linear combination of them,

$$\mathbf{v} = \sum_{i=1}^m \alpha_i \mathbf{w}_i. \quad (3.12)$$

Thus we shift problem of finding the entries of  $\mathbf{v}$  to finding the  $\alpha_i$  that maximize the variance. We can write our objective as

$$\mathbf{v}^T \mathbf{C} \mathbf{v} = \sum_{i=1}^m \alpha_i \mathbf{w}_i^T \sum_{j=1}^m \mathbf{w}_j \mathbf{w}_j^T \lambda_j \sum_{k=1}^m \alpha_k \mathbf{w}_k^T \quad (3.13)$$

$$= \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \alpha_i \alpha_k \lambda_j \mathbf{w}_i^T \mathbf{w}_j \mathbf{w}_j^T \mathbf{w}_k \quad (3.14)$$

$$= \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \alpha_i \alpha_k \lambda_j \delta_{ij} \delta_{jk} \quad (3.15)$$

$$= \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j \lambda_j \delta_{ij} \quad \text{sum over } k \quad (3.16)$$

$$= \sum_{i=1}^m \alpha_i \alpha_i \lambda_i \quad \text{sum over } j \quad (3.17)$$

$$= \sum_{i=1}^m \alpha_i^2 \lambda_i. \quad (3.18)$$

Since  $\|\mathbf{v}\| = 1$  we know that  $\sum_{i=1}^m \alpha_i^2 = 1$ . The above result gives us a weighted average which is maximal when  $\alpha_1 = 1$  since  $\lambda_1$  is (one of) the largest eigenvalue(s). Thus the direction in which the variance of our data is maximal is just the direction of the eigenvector with the largest eigenvalue.

We now want to find the second most important direction

**Problem:** Why do we want the directions to be orthogonal.