

Partial least squares (PLS) and friends

JDM and friends

2024-10-15

Contents

1	About	5
1.1	Notation	5
2	Preliminaries	7
2.1	Inner product	7
2.2	Inner product as projection	7
2.3	Matrix product	8
2.4	Column (row) space of a matrix	8
2.5	Rank of a matrix	8
2.6	Matrix as linear transform	9
2.7	Orthogonalisation	9
2.8	Orthogonal bases	9
2.9	Singular value decomposition	9

Chapter 1

About

1.1 Notation

- a, b : scalars, i.e., elements of \mathbb{R} , the set of real numbers.
- X, Y : matrices; $X \in \mathbb{R}^{n \times p}$ is an n by p matrix.
- a, b : column vectors (i.e., vectors are by default column vectors); $a \in \mathbb{R}^{n \times 1}$ has n elements, i.e., $a = (a_1, \dots, a_n)^\top$.
- $|a|$: length of vector a , i.e., $|a| = (\sum_{i=1}^n a_i^2)^{1/2}$.
- $\angle(a, b)$: angle (between -90° and 90°) between two vectors.
- \bar{a} : average of elements in a , i.e., $\bar{a} = n^{-1} \sum_{i=1}^n a_i$.
- $\text{var}(a)$: variance of elements in a , i.e., $\text{var}(a) = n^{-1} \sum_{i=1}^n (a_i - \bar{a})^2$.
- X^\top, a^\top : transpose of matrix and vectors.
- \equiv : always equivalent to, or equivalent by definition; e.g., $c \equiv a^\top b$ says that c is defined as the inner product between a and b .

Chapter 2

Preliminaries

We briefly review some basics from matrix algebra. We will focus on explaining the geometric intuition behind the concepts and how they are interconnected.

2.1 Inner product

The inner product $\mathbf{a}^\top \mathbf{b}$ results in a scalar:

$$c \equiv \mathbf{a}^\top \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos\{\angle(\mathbf{a}, \mathbf{b})\}.$$

If $|\mathbf{a}| = |\mathbf{b}| = 1$, i.e., \mathbf{a} and \mathbf{b} are *unit vectors*, we have

$$\mathbf{a}^\top \mathbf{b} = \cos\{\angle(\mathbf{a}, \mathbf{b})\},$$

which measures the affinity/similarity between \mathbf{a} and \mathbf{b} .

To illustrate the idea of affinity/similarity further, if $\check{\mathbf{a}}, \check{\mathbf{b}}$ denote the centered versions of \mathbf{a}, \mathbf{b} , then we actually have (you can work out why)

$$\text{cov}(\check{\mathbf{a}}, \check{\mathbf{b}}) \equiv \check{\mathbf{a}}^\top \check{\mathbf{b}} = \{\text{var}(\check{\mathbf{a}}) \text{var}(\check{\mathbf{b}})\}^{1/2} \cos(\check{\mathbf{a}}, \check{\mathbf{b}}).$$

That is to say $\cos(\check{\mathbf{a}}, \check{\mathbf{b}}) = \text{cor}(\check{\mathbf{a}}, \check{\mathbf{b}})$.

2.2 Inner product as projection

Consider the case where $\mathbf{a}, \mathbf{x} \in \mathbb{R}^n$ with $|\mathbf{a}| = 1$, then $\mathbf{x}^\top \mathbf{a}$ defines the length of $\hat{\mathbf{x}}$, the projection of \mathbf{x} unto the direction defined by \mathbf{a} .

2.3 Matrix product

Consider two matrices $A \in \mathbb{R}^{n \times p}$ and $B \in \mathbb{R}^{p \times r}$. Let $\mathbf{a}_1^\top, \dots, \mathbf{a}_n^\top \in \mathbb{R}^{1 \times p}$ denote the rows of A and $\mathbf{b}_1, \dots, \mathbf{b}_r \in \mathbb{R}^{p \times 1}$ denote the columns of B . Then the product AB is defined by

$$AB = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{b}_1 & \mathbf{a}_1^\top \mathbf{b}_2 & \cdots & \mathbf{a}_1^\top \mathbf{b}_r \\ \vdots & \vdots & & \vdots \\ \mathbf{a}_n^\top \mathbf{b}_1 & \mathbf{a}_n^\top \mathbf{b}_2 & \cdots & \mathbf{a}_n^\top \mathbf{b}_r \end{bmatrix} \equiv (\mathbf{a}_i^\top \mathbf{b}_j)_{ij}.$$

That is, the matrix product can be viewed as a structured way of calculating vector inner products (proximity/similarity).

In the special case where $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$ are column-centered, $\text{var}(X) = X^\top X$ and $\text{cov}(X, Y) = X^\top Y$ are the covariance matrices. That is, proximity of column vectors (variables).

2.4 Column (row) space of a matrix

For $A \in \mathbb{R}^{n \times p}$, the column space of A is defined as

$$\text{Col}(A) = \{ \mathbf{v} \in \mathbb{R}^{n \times 1} \mid \mathbf{v} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^{p \times 1} \}.$$

Let $\mathbf{a}_1, \dots, \mathbf{a}_p$ denote the p columns of A . The above definition states that $\text{Col}(A)$ contains all vectors that can be written as $\mathbf{v} = x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n$, i.e., linear combinations of the column vectors of A .

The column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ therefore define a linear space $\text{Col}(A)$, i.e., a collection of vectors whose arbitrary linear combinations are also in that collection. We refer to $\mathbf{a}_1, \dots, \mathbf{a}_n$ as the *basis* of $\text{Col}(A)$; we also say that the basis vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ *span* the space $\text{Col}(A)$.

2.5 Rank of a matrix

Sometimes a basis can contain redundant basis vectors. For example, consider a matrix $A = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ where $\mathbf{a}_2 = 0.4\mathbf{a}_1 + 0.6\mathbf{a}_3$. Then, in terms of defining the column space $\text{Col}(A)$, \mathbf{a}_2 is redundant. In this case, we say that A is not of *full rank*, or that A is *rank deficient*. The rank of A , denoted by $\text{rank}(A)$, is defined as the number of *linearly independent* column vectors. (It is also equivalent to the number of linearly independent *row* vectors.) The value of $\text{rank}(A)$ gives us a good indicator of the “true dimensionality” of A .

In biology, suppose we have a gene expression matrix $X \in \mathbb{R}^{n \times p}$ with normalized values. Then X not having full rank means that the expression level of some gene can be perfectly predicted as the linear combinations of other genes.

We can safely exclude this redundant gene from analysis without causing any information loss. In reality, however, due to observational noise, we can rarely have a strictly rank deficient gene expression matrix. Instead, it is often useful to assume that the observed gene expression X is an additive combination of signal and noise, i.e., $X = A + E$, where A is a low-rank signal matrix and E is a full rank noise matrix.

2.6 Matrix as linear transform

With the definition of column and row spaces in mind, we can view a matrix $A \in \mathbb{R}^{n \times p}$ as defining a linear transform from $\text{Col}(A)$ to $\text{Row}(A)$, i.e., from \mathbb{R}^p to \mathbb{R}^n . Taking $\mathbf{x} \in \mathbb{R}^{p \times 1}$, $A\mathbf{x}$ defines a vector that lives in \mathbb{R}^n .

For $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times K}$.

2.7 Orthogonalisation

In data analysis

2.8 Orthogonal bases

Suppose $A \in \mathbb{R}^{n \times p}$ and $r \equiv \text{rank}(A) < p < n$. Then we can find a subset of r linearly independent column vectors $\mathbf{a}_{i(1)}, \dots, \mathbf{a}_{i(r)}$, with $i(1), \dots, i(r) \in \{1, \dots, p\}$, forming a basis for $\text{Col}(A)$. In general, any set of r linearly independent vectors living in $\text{Col}(A)$ can serve as a (non-redundant) basis of $\text{Col}(A)$. It is often more convenient to find an *orthogonal* basis: when we plot a 3D space, we usually try to render 3 vectors that are perpendicular to each other, e.g., $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$.

2.9 Singular value decomposition