The algebra & geometry of vectors & matrices

2 Jan 09

AMS 225

ex. measure height, weight of n individuals (p = 2)

\[ \overline{\mathbf{x}} = \begin{bmatrix} \mathbf{h}^\top \\ \mathbf{w}^\top \end{bmatrix} \]

key multivariate

insight: can either think of this as n points in \( p = 2 \)-space

or as n vectors in \( \mathbb{R}^2 \)-space:

this scatterplot is our usual way of visualizing this dataset
(Fisher was particularly good at thinking geometrically in this way:)

but there is another useful way to visualize the data: 

\[ h = 2 \]

\[
\begin{pmatrix}
69 & 145 \\
72 & 170 \\
70 & 150 \\
\end{pmatrix}
\]

2 points in 3-space

\( \text{obs} \)

\( \text{wt} \)

\( \text{ht} \)

50 100 150 200

50 100 150 200

\( \text{obs}_2 \)
as \( a \to h_t + \theta \)

\( w_t \) becomes affine multiply of each other (e.g., \( w_t = 2.1 h_t \))

is they become perfectly correlated, turns out that \( \cos(\theta, h_t^*, w_t^*) = \cos \theta \), \( \theta \) = angle between \( h_t \& \) \( w_t, h_t^*, w_t^* \) appropriately normalized, makes sense:

angle near 0 \( \leftrightarrow \) cosine (correlation) near 1

the variables carry nearly the same (linear) information
To exactly come to rest at $(0,0)$, struck along line $a_2 = \frac{1}{2}a$. 
and of force $F_2 = \frac{1}{2}F_1$.

The direction of the initial velocity vector should be taken at least $90^\circ$ from the direction of the force $F_1$.
multiplication of vector by scalar. 0

geometry

if you keep the direction the same & double the force the ball will end up at (4, 2), so it makes sense to define \( 2 \cdot (1, 2) = (4, 2) \)

and in general \( c \cdot \mathbf{a} = c \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} c a_1 \\ c a_2 \\ \vdots \end{bmatrix} \) (algebra)

vector addition

geometry:

(3, 2 1/2) in old coordinates

(0, 1 1/2) in new coordinates

new origin at (2, 1)
having got the ball to \((2,1)\), if you now redefine \(g\) at what used to be \((2,1)\) & strike the ball again so that it ends up at \((1,1\frac{1}{2})\) in new coordinates, this will clearly be at \((2+1, 1+1\frac{1}{2}) = (3, 2\frac{1}{2})\) in old coordinates, so \(\left(\frac{2}{1}\right) + \left(\frac{1}{1\frac{1}{2}}\right) = \left(\frac{3}{2}\right)\) and more generally (algebra) if
\[
\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}
\]
then
\[
\begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix}
\]
NB doesn't make sense to all vectors of different length.
vector subtraction is just vector addition combined with multiplication by a scalar: \( \vec{a} - \vec{b} = \vec{a} + (-1)\vec{b} \).

Vector multiplication (geometry) begins with observation that there's something special about two vectors being perpendicular to each other:

people chose the convention \( \vec{a} \perp \vec{b} \) to mean simple function of them is 0 and then looked for such a function:
The simplest thing that works is to notice that \((-1^{\frac{1}{2}})(2) + (3)(1) = 0\), which leads to the suggestion that if \(a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}\) and \(b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}\) then
\[
\text{"times" } a \sim \sum_{i=1}^{n} a_i b_i \cdot \text{ By convention people only allow this to make sense if the dimensions of the two vectors match in a particular way, } a' = (a_1, \ldots, a_n) \text{ and } b, \text{ can be "multiplied" in the order } a' \sim b, \text{ because the cols of } a' \text{ matches the rows of } b.
\[ X = \left[ \begin{array}{c|c|c|c} \vdots & \vdots & \vdots & \vdots \\ x_1 & x_2 & \cdots & x_p \\ \end{array} \right] \]

In a table with \( i \) rows and \( p \) columns.

Vectors of some dimension (\( n \))

For \( i \leq p \)

The same reason.

For some matrix of coefficients, we have

\[ \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \end{bmatrix} = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \end{bmatrix} \]

For some \( f \) and \( g \)
All the previous operations now readily generalize:

1. \( cX = \left[ \begin{array}{ccc} cX_{11} & \cdots & cX_{1p} \\ \vdots & \ddots & \vdots \\ cX_{n1} & \cdots & cX_{np} \end{array} \right] \)

2. If \( A = (a_{ij}) \) and \( B = (b_{ij}) \),
   then \( (A + B) = \left[ \begin{array}{ccc} a_{11} + b_{11} & \cdots & a_{1p} + b_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{np} + b_{np} \end{array} \right] \)

3. If \( A = (a_{ij}) \) and \( B = (b_{ij}) \),
   then \( A \cdot B \) makes sense (but not \( B \cdot A \))
and is given by $AB = (a_{ij} \cdot b_{kj})$

$$[\sum_{k=1}^{\infty} a_{1k} b_{k1}, \ldots, \sum_{k=1}^{\infty} a_{1k} b_{k2}, \ldots, \sum_{k=1}^{\infty} a_{ik} b_{k1}, \ldots, \sum_{k=1}^{\infty} a_{ik} b_{k2}]$$

what about division? division by a scalar $c \neq 0$ is just multiplication by $c^{-1}$: $\frac{x}{c} = c^{-1}x$.

For this idea (division = multiplication by an inverse) to generalize, need to extend idea of inverse to matrices.
with scalars \( a = b^{-1} \) means \( a \cdot b = 1 \)

the analogue with matrices would be

\[
A = B^{-1} \rightarrow A \cdot B = 1 = B \cdot A
\]

Awkward to have \( B^{-1} \) and \( B \) be of different dimensions; simplest way forward is to define 'inverse' only for matrices with \( \# \text{rows} = \# \text{cols} \)

such matrices are square & the \( B^{-1} \) would become

\[
A = B^{-1} \rightarrow A \cdot B = 1
\]

Easiest & most natural matrix version.
of scalar 1 is \( I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \).

The \( n \times n \) identity matrix; this motivates another natural definition: any matrix with 0 entries off the diagonal is a diagonal matrix:

\[
I = \text{diag}(1), \quad \text{where } I = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.
\]

\[n\]
def: The inverse of a square matrix \( A \) is the matrix \( A^{-1} \) (if it exists) satisfying

\[
AA^{-1} = A^{-1}A = I.
\]
(can show easily that if \( A^{-1} \) exist it's unique). So: why might \( A^{-1} \) not exist? A: Scalar division by 0 is not defined, so matrices that are "analogous to 0" in some sense will not have inverses; when is a matrix analogous to 0?

\[
\text{ex. } (n=2) \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad A^{-1} = \begin{pmatrix} e & f \\ g & h \end{pmatrix}
\]

\[
AA^{-1} = \begin{pmatrix} ae + bg & uf + bh \\ ce + dg & cf + dh \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\]
Also want

$$A^{-1}A = \begin{pmatrix} ae + cf & de + df \\ bg + ch & bg + dh \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(Linear)

Each of these amounts to 4 equations in 4 unknowns, readily solved:

sauternes 53> maple

```
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
\$\$\$
This doesn't make sense if \(ad - bc \neq 0\).

So \(2 \times 2\) matrices \((a\ b)
\begin{pmatrix}
c & d
\end{pmatrix}
\) with
\[ad - bc = 0\]
are "analogous to 0".

What's special about such matrices?

Ex: \(A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\) leave \(a, b, c\)

free \& solve for \(d\) in \(ad - bc = 0\),
yielding \(d = \frac{bc}{a}\) and \(A^2 = \begin{pmatrix} a & b \\ c & \frac{bc}{a} \end{pmatrix}\).

But this can be expressed as

\[A = \begin{pmatrix} a & (\frac{b}{a})a \\ c & (\frac{b}{a})c \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & \frac{bc}{a} \end{pmatrix}\] for

some \(k\), i.e., 2\text{nd}\ col. of \(A\) is scalar
Multiple of 1st col: e.g. \[ \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \]
def. two vectors \[ a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \] and \[ b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \]
are linearly dependent if \( b = \delta a \) for some \( \delta \neq 0 \) (special case of non-trivial solution).

when the columns of \( A \) are linearly dependent it's not possible to use them as the basis of a coordinate system in \( \mathbb{R}^n \); here \( \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \) only span a 1-dimensional subspace of \( \mathbb{R}^2 \).
A collection of vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_n \) is linearly independent if none of them can be written as a linear combination of finitely many other vectors in the collection. Related def: The \underline{rank} of a matrix \( A \) is the maximal number of linearly independent columns of \( A \) (when \( A \) is viewed as vectors); similarly, for \underline{row rank}, \( \text{rank} \) of \( A \).
The problem with \( \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \) and

indeed only \( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) with \( ad - bc = 0 \)
is that the matrix is of linear

2 but only has rank 1;

conjecture (turns out to be true):

only \( A_n \) with rank \( (A) < n \)
does not have an inverse (is

not invertible).

\[ \text{Def: For a} \]

\[ 2 \times 2 \text{ matrix } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \]

\( ad - bc \) is the determinant of

\( A \), written \( |A| \).
A.2.3 Determinants and cofactors

Definition The determinant of a square matrix $A$ is defined as

$$|A| = \sum (-1)^{\pi} a_{1\tau(1)} \cdots a_{p\tau(p)},$$

where the summation is taken over all permutations $\tau$ of $(1, 2, \ldots, p)$, and $|\tau|$ equals $+1$ or $-1$, depending on whether $\tau$ can be written as the product of an even or odd number of transpositions.

For $p=2$,

$$|A| = a_{11}a_{22} - a_{12}a_{21}.$$ 

Definition The cofactor of $a_{ij}$ is defined by $(-1)^{i+j}$ times the minor of $a_{ij}$, where the minor of $a_{ij}$ is the value of the determinant obtained after deleting the $i$th row and the $j$th column of $A$.

\[ \text{area} = \text{det} A \]
(This extends readily to \( n > 2 \)).

**Def:** A square matrix \( A \) is **nonsingular** (invertible) if \( |A| \neq 0 \).

**Fact:** \( A^{-1} \) exists if \( A \) is nonsingular.

Some special square matrices:

- Some matrices (e.g., correlation, covariance) are symmetric:
  - \( A \) is symmetric if \( A = A' \)
  - or equivalently if \( a_{ij} = a_{ji} \) for all \( i \) and \( j \).
  - If not true, \( A \) is **asymmetric**.

\( A' \) is the \( p \times q \) matrix whose rows are the cols of \( A \) (vice versa).
(2) Sometimes useful to have square notation for a matrix whose entries are all 1; N x N call
the N x N matrix of this type

\[ \mathbf{I}_n = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \]

(3) Sometimes a matrix will have all 0 entries above (below) the diagonal; such matrices are
called lower (upper) triangular.

(4) The null matrix is \( \mathbf{0} \), i.e., (all 0's).
The trace of a square matrix $A$ is defined to be the sum of its diagonal values:

$$\text{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$ This summing can be interpreted as another measure of the size of $A$, as will be seen below.

Vectors have both length & direction, by Pythagoras, the length of $\vec{a} = (a_1, a_2)$ is

$$\sqrt{a_1^2 + a_2^2} = \| \vec{a} \|,$$

the (Euclidean) norm of $\vec{a}$. More about the dot product.
and more generally if \( \mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \),

then \( \| \mathbf{a} \| = \sqrt{\sum_{i=1}^{n} a_i^2} = (\mathbf{a} \cdot \mathbf{a})^{1/2} \).

By the same token,

\[ \| \mathbf{b} - \mathbf{a} \| = \| \mathbf{a} - \mathbf{b} \| = \left( \sum_{i=1}^{n} (a_i - b_i)^2 \right)^{1/2} \]

Geometric interpretation of matrix multiplication is broader.

**Q:** How do you change from one coordinate system to another?

**A:** Let's look at the line \((\mathbb{R}^2)\) first.
Temperature in Fahrenheit & Centigrade

\[ ^\circ F = \frac{9}{5} (^\circ C) + 32 \leftrightarrow ^\circ C = \frac{5}{9} (^\circ F - 32) \]

Simplified coordinate system changes are linear:

\[ x^* = a x + b \]

\( b \) represents translation of origin (left or right along number line);

\( a \) represents rescaling (making distance from 0 to 1 bigger or smaller).
what about in $\mathbb{R}^2$?

Translation is easy:

1. New coordinates (shifting if original)
   \[ x = x' + a \]

2. Old coordinates
   \[ x' = x - a \]

Regarding can occur in 2 different ways (horizontal & vertical)

Ex. Here
\[ (7, 1) \rightarrow (2x_1, 7_2) \]

\[ = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \]

\[ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \]

So regarding involves multiplication by a diagonal matrix
In 3 dimensions there is a third possible operation: rotating (counterclockwise, say) through an angle \( \theta \).
This time: algebra & geometry of vectors & matrices

Next time: multivariate normal distribution

Scalar $c$

Vector $a$

Proof reader's symbol for bold face

\[
\cos \theta = \frac{c}{r}
\]

\[
\begin{align*}
\theta & \quad \text{corr} \\
0 & \quad 1 \\
90^\circ & \quad 0
\end{align*}
\]
2. \[ f\left(\left(\frac{a_1}{a_2}, \frac{b_1}{b_2}\right)\right) = 0 \]

when
\[ a_1 \neq 0 \]

3. \[ a = b^{-1} \] if and only if \[ a \cdot b = b \cdot a = 1 \]

\[ \text{area of parallelogram} = \text{area of parallelogram} \]
\[ 1 + \begin{pmatrix} 1 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \ldots & 1 \end{pmatrix} = \sum_{n=1}^{\infty} \begin{pmatrix} 3 \\ 1 \ 2 \\ 0 \ 2 \ 1 \end{pmatrix} \]