Some matrix revision

Ability to do and understand simple matrix algebra is central to understanding the theory of statistical modelling and inference. These notes revise what is essential for this course. We will be concerned only with real matrices.

1 Matrices and vectors

- A n dimensional **vector** is an ordered array of n numbers. We will write $\mathbf{v}$ for the whole vector and v_i for its i^{th} element.
- An $r \times c$ **matrix** is a 2 dimensional array of numbers, with r rows, c columns and rc elements in total. We will write $\mathbf{M}$ for the whole matrix, and M_{ij} for the element in row i, column j. The indices are *always* row followed by column.

The following give examples of a 3 vector and a 3×2 matrix:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v3 \end{pmatrix} = \begin{pmatrix} \pi \\ -2 \\ 5.6 \end{pmatrix}, \qquad \mathbf{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \\ M_{31} & M_{32} \end{pmatrix} = \begin{pmatrix} -2 & 4.3 \\ 0.5 & 45 \\ 0 & 0.1 \end{pmatrix}$$

Notice that by default a vector is treated as a *column vector*, so that an n vector is an $n \times 1$ matrix. A matrix for which the number of rows is equal to the number of columns is called a *square matrix*.

2 Partitioning matrices

It is sometimes necessary to partition matrices into sub matrices. For example, I might want to partition 3×5 matrix $\mathbf{A}$ into $[\mathbf{A}_0 : \mathbf{A}_1]$ where $\mathbf{A}_0$ has 2 columns and $\mathbf{A}_1$ has 3. Here is exactly what this means

$$\mathbf{A} = [\mathbf{A}_0 : \mathbf{A}_1] = \left(\begin{array}{cc|ccc} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ A_{21} & A_{22} & A_{23} & A_{24} & A_{25} \\ A_{31} & A_{32} & A_{33} & A_{34} & A_{35} \end{array}\right) \text{ so } \mathbf{A}_0 = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{pmatrix} \text{ \& } \mathbf{A}_1 = \begin{pmatrix} A_{13} & A_{14} & A_{15} \\ A_{23} & A_{24} & A_{25} \\ A_{33} & A_{34} & A_{35} \end{pmatrix}$$

Of course we can also partition row-wise.

3 Products

- The *inner product* of two n vectors $\mathbf{a}$ and $\mathbf{b}$ is $< \mathbf{a}, \mathbf{b} >= \sum_i^n a_i b_i$.
- The product of $r \times n$ matrix $\mathbf{A}$ and $n \times c$ matrix $\mathbf{B}$ is the $r \times c$ matrix $\mathbf{C}$ with elements $C_{ij} = \sum_k^n A_{ik} B_{kj}$. Notice that this only exists when the number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$. Basically C_{ij} is the inner product of row i of $\mathbf{A}$ with column j of $\mathbf{B}$. It is very important to have a mental picture of this. The following example should help…

$$\begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ . & C_{32} = 8 & . & . & . \\ . & . & . & . & . \end{pmatrix} = \begin{pmatrix} . & . & . \\ . & . & . \\ A_{31} = 1 & A_{32} = 3 & A_{33} = 2 \\ . & . & . \end{pmatrix} \begin{pmatrix} . & B_{12} = 3 & . & . & . \\ . & B_{22} = 1 & . & . & . \\ . & B_{32} = 1 & . & . & . \end{pmatrix}$$

- Note the special case that the product of a matrix and a vector is always a vector.

4 Transposition

The transpose of $n \times m$ matrix $\mathbf{A}$ is the $m \times n$ matrix $\mathbf{A}^\mathsf{T}$ whose i^{th} row(column) is given by the i^{th} column(row) of $\mathbf{A}$. Here is an example

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 8 \\ 1 & 0 \end{pmatrix} \Rightarrow \mathbf{A}^\mathsf{T} = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 8 & 0 \end{pmatrix}.$$

A *symmetric* matrix is one for which $\mathbf{A} = \mathbf{A}^\mathsf{T}$. The inner product of vectors $\mathbf{a}$ and $\mathbf{b}$ can be written as $\mathbf{a}^\mathsf{T}\mathbf{b}$. It is essential to know one easily proven identity: $(\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$.

5 Euclidean norm of a vector

The squared Euclidean norm of a vector, $\mathbf{v}$, is just its squared length, i.e. the sum of squares of its elements $\sum_i v_i^2$, which is also its inner product with itself. It can be written $\|\mathbf{v}\|^2$ and we have

$$\|\mathbf{v}\|^2 = \mathbf{v}^\mathsf{T}\mathbf{v} = \sum_i v_i^2.$$

6 Matrix rank

The **rank** of an $r \times c$ matrix, $\mathbf{A}$, is the number of linearly independent rows or columns it has. A set of rows(columns) is linearly independent if none can be made as a linear combination of the others. $\text{rank}(\mathbf{A}) \leq \min(r, c)$. Equivalently, the rank of a matrix is the number of non-zero eigenvalues it possesses. The rank of the product of two matrices is at most the smaller of their two individual ranks. $\mathbf{A}$ has *full rank* if $\text{rank}(\mathbf{A}) = \min(r, c)$.

7 Matrix inversion

Let $\mathbf{A}$ be an $n \times n$ matrix with full rank (i.e. $\text{rank}(\mathbf{A}) = n$). Then there exists a matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, where $\mathbf{I}$ is the $n \times n$ identity matrix, i.e.

$$I_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

$\mathbf{A}^{-1}$ is the *inverse* of $\mathbf{A}$. Notice that it *only* exists for $\mathbf{A}$ square and full rank (so obviously these notes assume these conditions whenever an inverse is written). In R, `solve(A)` will produce the inverse of `A`.

8 Orthogonal matrices

An orthogonal matrix $\mathbf{Q}$ is a matrix for which $\mathbf{Q}^{-1} = \mathbf{Q}^\mathsf{T}$. That is $\mathbf{Q}\mathbf{Q}^\mathsf{T} = \mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}$. Clearly $\mathbf{Q}$ must be square, but we will sometimes be interested in non-square matrices made up of just some columns of an orthogonal matrix. For example, $\mathbf{U}$ might be the first $p < n$ columns of $n \times n$ orthogonal matrix $\mathbf{Q}$, in which case $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}_p$, the $p \times p$ indentity matrix, but $\mathbf{U}\mathbf{U}^\mathsf{T}$ will not produce an identity matrix (since $\mathbf{U}$ does not have orthogonal *rows*).

Consider the (squared) Euclidean length of the vector $\mathbf{Qy}$, where $\mathbf{y}$ is any vector. We have

$$\|\mathbf{Qy}\|^2 = \mathbf{y}^\mathsf{T}\mathbf{Q}^\mathsf{T}\mathbf{Q}\mathbf{y} = \mathbf{y}^\mathsf{T}\mathbf{I}\mathbf{y} = \mathbf{y}^\mathsf{T}\mathbf{y} = \|\mathbf{y}\|^2.$$

So $\mathbf{Qy}$ has the same length as $\mathbf{y}$. i.e. $\mathbf{Qy}$ can be viewed as a rotation and or reflection of $\mathbf{y}$. The product of two $n \times n$ orthogonal matrices is an $n \times n$ orthogonal matrix (the result of two successive rotation/reflections is a rotation/reflection).

9 QR decomposition of a matrix

Any real $r \times c$ matrix $\mathbf{A}$ can always be decomposed into the product of an orthogonal $r \times r$ matrix $\boldsymbol{\mathcal{Q}}$ and an upper triangular matrix, as follows

$$\mathbf{A} = \boldsymbol{\mathcal{Q}} \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix}$$

where $\mathbf{R}$ is $c \times c$ upper triangular (meaning that $R_{ij} = 0$ if $i > j$). Here $\mathbf{0}$ denotes an $(r - c) \times c$ matrix of zeroes.

Here is an example of a QR decomposition

$$\begin{pmatrix} 1 & \sqrt{1/2} \\ 0 & 1 \\ 1 & \sqrt{2} \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} & -\sqrt{1/10} & \sqrt{2/5} \\ 0 & \sqrt{4/5} & \sqrt{1/5} \\ \sqrt{1/2} & \sqrt{1/10} & -\sqrt{2/5} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 3/2 \\ 0 & \sqrt{5/4} \\ 0 & 0 \end{pmatrix}$$

This course only requires you to know what a QR decomposition is, **not** how it is computed. However, if you are curious, it works like this. Given any two vectors $\mathbf{a}$ and $\mathbf{b}$, of the same length, it is easy to write down a simple symmetric orthogonal matrix, known as a *Housholder* matrix, such that $\mathbf{a} = \mathbf{H}\mathbf{b}^*$. Armed with this fact, we can construct the QR decomposition. First find $\mathbf{H}_1$ to rotate the first column of $\mathbf{A}$ to a vector containing a single non-zero entry in its first row, and zeroes elsewhere. $\mathbf{H}_1$ will also rotate all the other columns of $\mathbf{A}$, but we don't care about that. Now create $\mathbf{H}_2$ which is a rotation that leaves the first element of any vector unchanged, but rotates the second column of $\mathbf{H}_1\mathbf{A}$ so that it has non-zeroes only in its first two elements. Clearly $\mathbf{H}_2$ has no effect on the first column of $\mathbf{H}_1\mathbf{A}$, since it can't change it's length, and leaves the only element contributing to that length unchanged! Continuing in this way for c steps we can always reduce $\mathbf{A}$ to upper triangular form, accumulating $\boldsymbol{\mathcal{Q}}^\mathsf{T}$ on the way. At the end $\boldsymbol{\mathcal{Q}} = \mathbf{H}_1\mathbf{H_2}\cdots\mathbf{H}_c$. Here is how the process works for the above example.

$$\begin{aligned} \mathbf{H}_1\mathbf{A} &= \begin{pmatrix} \sqrt{1/2} & 0 & \sqrt{1/2} \\ 0 & -1 & 0 \\ \sqrt{1/2} & 0 & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} 1 & \sqrt{1/2} \\ 0 & 1 \\ 1 & \sqrt{2} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 3/2 \\ 0 & -1 \\ 0 & -1/2 \end{pmatrix} \\ \mathbf{H}_2\mathbf{H}_1\mathbf{A} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\sqrt{4/5} & -\sqrt{1/5} \\ 0 & -\sqrt{1/5} & \sqrt{4/5} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 3/2 \\ 0 & -1 \\ 0 & -1/2 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 3/2 \\ 0 & \sqrt{5/4} \\ 0 & 0 \end{pmatrix} \end{aligned}$$

In other courses you may have covered the Gram-Schmidt process for producing a QR decomposition, but this is not as stable as the Householder method so is not used for practical numerical computation.

10 Choleski decomposition of a matrix: a matrix 'square root'

Positive definite matrices are the 'positive real numbers' of matrix algebra. They have particular computational advantages and occur frequently in statistics, because covariance matrices are usually positive definite (and always positive semi-definite). To see why matrix square roots might be useful, consider the following.

Example Generating multivariate normal random variables. There exist very quick and reliable methods for simulating i.i.d. $N(0, 1)$ random deviates, but suppose that $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vectors are required. Clearly we can generate vectors $\mathbf{z}$ from $N(\mathbf{0}, \mathbf{I})$. If we could find a matrix $\mathbf{R}$ such that $\mathbf{R}^\mathsf{T}\mathbf{R} = \boldsymbol{\Sigma}$, then $\mathbf{y} \equiv \mathbf{R}^\mathsf{T}\mathbf{z} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, because the covariance matrix of $\mathbf{y}$ is $\mathbf{R}^\mathsf{T}\mathbf{I}\mathbf{R} = \mathbf{R}^\mathsf{T}\mathbf{R} = \boldsymbol{\Sigma}$ and $\mathbb{E}(\mathbf{y}) = \mathbb{E}(\mathbf{R}^\mathsf{T}\mathbf{z} + \boldsymbol{\mu}) = \boldsymbol{\mu}$.

In general the square root of a positive definite matrix is not uniquely defined, but there is a unique *upper triangular* square root of any positive definite matrix: its *Choleski factor*. The algorithm for finding the Choleski factor is easily derived. Consider a 4×4 example first. The defining matrix equation is

$$\begin{pmatrix} R_{11} & 0 & 0 & 0 \\ R_{12} & R_{22} & 0 & 0 \\ R_{13} & R_{23} & R_{33} & 0 \\ R_{14} & R_{24} & R_{34} & R_{44} \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ 0 & R_{22} & R_{23} & R_{24} \\ 0 & 0 & R_{33} & R_{34} \\ 0 & 0 & 0 & R_{44} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{12} & A_{22} & A_{23} & A_{24} \\ A_{13} & A_{23} & A_{33} & A_{34} \\ A_{14} & A_{24} & A_{34} & A_{44} \end{pmatrix}.$$

*You *really* don't need to know this for this course, but $\mathbf{H} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^\mathsf{T}/\|\mathbf{u}\|^2$ where $\mathbf{u} = \mathbf{b} - \mathbf{a}$

If the component equations of this expression are written out and solved in the right order, then each contains only one unknown, as the following illustrates (unknowns are in bold):

$$\begin{aligned}
A_{11} &= \mathbf{R_{11}}^2 \\
A_{12} &= R_{11}\mathbf{R_{12}} \\
A_{13} &= R_{11}\mathbf{R_{13}} \\
A_{14} &= R_{11}\mathbf{R_{14}} \\
A_{22} &= R_{12}^2 + \mathbf{R_{22}}^2 \\
A_{23} &= R_{12}R_{13} + R_{22}\mathbf{R_{23}} \\
&. \\
&.
\end{aligned}$$

Generalising to the $n \times n$ case, and using the convention that $\sum_{k=1}^{0} x_i \equiv 0$, we have

$$R_{ii} = \sqrt{A_{ii} - \sum_{k=1}^{i-1} R_{ki}^2}, \text{ and } R_{ij} = \frac{A_{ij} - \sum_{k=1}^{i-1} R_{ki}R_{kj}}{R_{ii}}, \quad j > i.$$

Working through these equations in row order, from row one, and starting each row from its leading diagonal component, ensures that all right-hand-side quantities are known at each step. Choleski decomposition requires $n^3/3$ flops and n square roots. In R it is performed by function `chol`[†].

[†]Actually, numerical analysts do not consider the Choleski factor to be a square root in the strict sense, because of the transpose in $\mathbf{A} = \mathbf{R}^\mathsf{T}\mathbf{R}$.