Matrix algebra

Patrick Breheny

January 20

Introduction

- The mathematics of multiple regression revolves around ordering and keeping track of large arrays of numbers and solving systems of equations
- The mathematical constructs that allow us to do this in a relatively simple and straightforward was are called *matrices*, and the tools describing their manipulations are the tools of *matrix algebra*
- The goal of this class is to introduce the important definitions, results, operations, and concepts that we will be use constantly in the rest of the course
- Note: You may not see why some of these definitions/results are important right away, but we will use them in the course and I think you'll benefit from having them all in one set of notes

Matrices and vectors

- A matrix is a collection of numbers arranged in a rectangular array of *rows* and *columns*, such as

$$\begin{bmatrix} 3 & 2 \\ 4 & -1 \\ -1 & 2 \end{bmatrix}$$

- A matrix with r rows and c columns is said to be an $r \times c$ matrix (*e.g.*, the matrix above is a 3×2 matrix)
- In the degenerate case where a matrix has just a single row or column, it is said to be a *vector*, such as

$$\begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Matrix and vector notation

- Vectors and matrices are denoted in lower- and upper-case boldface, respectively (*e.g.*, x is a scalar, $\mathbf{x}$ is a vector, and $\mathbf{X}$ is a matrix)
- By convention, vectors are *column vectors* – *i.e.*, a vector of n numbers is an $n \times 1$ matrix, not a $1 \times n$ matrix
- It is worth pointing out that the use of upper/lower case in matrix algebra differs from that in probability, where upper/lower case is used to distinguish random variables from fixed quantities
- Thus, in this course, a vector will be denoted $\mathbf{y}$ regardless of whether it is fixed or random; if the distinction is important and not clear from context, random/fixed vectors will be distinguished in some other way

Matrix and vector notation (cont'd)

- The ijth element of a matrix $\mathbf{M}$ is denoted by M_{ij} or $(\mathbf{M})_{ij}$; *e.g.*, letting $\mathbf{M}$ denote the matrix on the third slide of today's lecture, $M_{11} = 3$, $(\mathbf{M})_{32} = 2$, and so on
- Similarly, the jth element of a vector $\mathbf{v}$ is denoted v_j; *e.g.*, letting $\mathbf{v}$ denote the vector on the same slide, $v_1 = 3$

Transposition

- It is often useful to switch the rows and columns of a matrix around
- The resulting matrix is called the *transpose* of the original matrix, and denoted with a superscript ^{T} or an apostrophe $'$

$$\mathbf{M} = \begin{bmatrix} 3 & 2 \\ 4 & -1 \\ -1 & 2 \end{bmatrix} \qquad \mathbf{M}^T = \begin{bmatrix} 3 & 4 & -1 \\ 2 & -1 & 2 \end{bmatrix}$$

- Note that $M_{ij} = M_{ji}^T$, and that if $\mathbf{M}$ is an $r \times c$ matrix, $\mathbf{M}^T$ is a $c \times r$ matrix

Addition

- There are two kinds of addition operations for matrices
- The first is *scalar addition*:

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

- The other kind is *matrix addition*:

$$\mathbf{M}+\mathbf{M}=\begin{bmatrix} 3 & 2 \\ 4 & -1 \\ -1 & 2 \end{bmatrix}+\begin{bmatrix} 3 & 2 \\ 4 & -1 \\ -1 & 2 \end{bmatrix}=\begin{bmatrix} 6 & 4 \\ 8 & -2 \\ -2 & 4 \end{bmatrix}$$

- Formally, $(\mathbf{A}+\mathbf{B})_{ij}=A_{ij}+B_{ij}$
- Note that only matrices of the same dimension can be added to each other – it makes no sense to add a 4×5 matrix to a 2×9 matrix

Scalar multiplication

- There are also two common kinds of multiplication for matrices
- The first is *scalar multiplication*:

$$4\mathbf{M} = 4\begin{bmatrix} 3 & 2 \\ 4 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 12 & 8 \\ 16 & -4 \\ -4 & 8 \end{bmatrix}$$

- Formally, $(c\mathbf{M})_{ij} = cM_{ij}$

Vector multiplication

- The other kind is *matrix multiplication*
- To get an idea of how matrix multiplication works, let's start with the simpler case of vector multiplication
- Suppose $\mathbf{u}$ and $\mathbf{v}$ are two $n \times 1$ vectors; their product is defined as

$$\mathbf{u}^T\mathbf{v} = \sum_j u_j v_j$$

$$\begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 6 - 2 = 4$$

- Specifically, this is called their *inner product* (there is also an *outer product*, but don't worry about that right now)

Matrix multiplication

- Now, the product of two matrices, $\mathbf{AB}$, is defined by taking all the (inner) products of $\mathbf{A}$'s rows with $\mathbf{B}$'s columns:

$$(\mathbf{AB})_{ik} = \sum_k A_{ij} B_{jk}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 4 & -1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 12 & 9 \end{bmatrix}$$

- Note that
  - Matrix multiplication is only defined if the number of columns of $\mathbf{A}$ matches the number of rows of $\mathbf{B}$
  - If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times p$ matrix, then $\mathbf{AB}$ is an $m \times p$ matrix

Elementary vs. matrix algebra

- Most aspects of elementary algebra carry over to matrix algebra:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \qquad (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$
$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \qquad \mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$
$$k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$$

- One important exception, however, is that $\mathbf{AB} \neq \mathbf{BA}$
- The order of matrix multiplication matters, and we must remember to, for instance, "left multiply" both sides of an equation by a matrix $\mathbf{M}$ to preserve equality
- This is tricky at first, but you get used to it

Restating regression with matrices

- We can write out simple linear regression in terms of vectors and matrices as follows:

$$\mathbf{y} = \alpha + \beta\mathbf{x} + \boldsymbol{\epsilon}$$

- Now, let $\boldsymbol{\beta} = (\alpha \; \beta)$ and $\mathbf{X}$ be an $n \times 2$ matrix whose first column is a n-dimensional vector of all 1s and whose second column is $\mathbf{x}$
- With this notation, we can write the regression equation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- Note that all the dimensions are consistent: this is an important check to perform in general to make sure that the matrix equations you're writing down make sense

Square and symmetric matrices

- There are a number of special kinds of matrices that are important to know about, because certain theorems only apply to certain kinds of matrices
- For example, in the special case where a matrix has the same numbers of rows and columns, it is said to be *square*
- If $\mathbf{A}^T = \mathbf{A}$, the matrix is said to be *symmetric*

$$\text{Symmetric: } \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \qquad \text{Not symmetric: } \begin{bmatrix} 3 & 2 \\ 0 & -1 \end{bmatrix}$$

- Note that a matrix cannot be symmetric unless it is square

Diagonal and identity matrices

- The elements A_{ii} of a matrix are called its *diagonal entries*; a matrix for which $A_{ij} = 0$ for all $i \neq j$ is said to be a *diagonal matrix*:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

- Consider in particular the following diagonal matrix:

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

Diagonal and identity matrices (cont'd)

- Note that this matrix has the interesting property that $(\mathbf{AI})_{ij} = A_{ij}$ for all i, j
- In other words, $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$
- Because of this property, $\mathbf{I}$ is referred to as the *identity matrix*
- Note: some authors use $\mathbf{I}_k$ to mean the $k \times k$ identity matrix; I will simply use $\mathbf{I}$ with the understanding that $\mathbf{I}$ is whatever dimension it needs to be in order to be *conformable* (for its dimensions to match the other matrices in the equation)

Some other matrices

- Some other notations which are commonly used are $\mathbf{1}$, the vectors of 1s, and $\mathbf{0}$, the vector of zeros:

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- Unfortunately, there is no "capital zero"; matrices of zeros or ones are usually represented by explicitly specifying their dimensions, as in $\mathbf{0}_{2\times 2}$ or $\mathbf{1}_{4\times 5}$
- The vector $\mathbf{e}_j$ is also useful: it has element $e_j = 1$ and $e_k = 0$ for all other elements:

$$\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

The matrix inverse

- Suppose $\mathbf{Ax} = \mathbf{B}$ and we want to solve for $\mathbf{x}$ ... can we "divide" by $\mathbf{A}$?
- Sort of – by multiplying both sides by the *inverse* of $\mathbf{A}$
- If a matrix $\mathbf{A}^{-1}$ satisfies $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, then $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$
- If we have $\mathbf{A}^{-1}$, then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{B}$ in the equation above
- Note that, by this definition, only square matrices can have inverses; this can be extended by defining a *left inverse* and *right inverse*, but we're not going to need them in this class

Orthogonality

- If two vectors $\mathbf{u}$ and $\mathbf{v}$ satisfy $\mathbf{u}^T\mathbf{v} = 0$, they are said to be *orthogonal* to each other
- If all the columns and rows of a matrix $\mathbf{A}$ are orthogonal to each other and satisfy $\mathbf{a}^T\mathbf{a} = 1$, then $\mathbf{A}$ (transposed) can serve as its own inverse: $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$
- In this case, the matrix $\mathbf{A}$ is said to be an *orthogonal matrix*
- If a matrix $\mathbf{X}$ is not square, then it is possible that $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ but $\mathbf{X}\mathbf{X}^T \neq \mathbf{I}$; in this case, the matrix is said to be *column orthogonal*, although in statistics it is common to refer to these matrices as orthogonal also
- A somewhat related definition is that a matrix is said to be *idempotent* if $\mathbf{A}\mathbf{A} = \mathbf{A}$

One and only one inverse?

- Does every matrix have one and only one inverse?
- Well, if a matrix has an inverse, it is said to be *invertible*, and yes, all invertible matrices have exactly one, unique inverse
- However, not every matrix is invertible
- For example, there are no values of a, b, c, and d that satisfy

$$\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Linear dependence

- Why didn't the matrix on the previous slide have an inverse?
- Well, there were four equations and four unknowns, but some of those equations contradicted each other
- The term for this situation is *linear dependence*

Linear combinations and linear independence

- If you have a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then you can form new vectors from *linear combinations* of the old vectors: $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$
- A collection of vectors is said to be *linearly independent* if none of them can be written as a linear combination of the others; if it can, then they are linearly dependent
- This is the key to whether a matrix is invertible or not: a matrix $\mathbf{A}$ is invertible if and only if its columns (or rows) are linearly independent
- Note that the columns of our earlier matrix were not linearly independent, since $2(2\ 1) = (4\ 2)$

Rank

- The *rank* of a matrix is the number of linearly independent columns (or rows) it has
- If they're all linearly independent, then the matrix is said to be of *full rank*
- This is a very important concept in regression, because we'll have to take inverses to solve for the regression coefficients $\boldsymbol{\beta}$, and if our matrix $\mathbf{X}$ that contains all the explanatory variables is not full rank, then we won't be able to solve for $\boldsymbol{\beta}$

Additional helpful identities

- $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$

Expectation and variance

- As statisticians, we need to deal with vectors of random quantities like $\mathbf{y}$ and $\boldsymbol{\epsilon}$, so we need to define the expectation and variance of random vectors
- If $\mathbf{u} = (U_1 \ U_2 \ \cdots U_n)$ is a vector of random variables, then $\mathrm{E}(\mathbf{u}) = (\mathrm{E}(U_1) \ \mathrm{E}(U_2) \ \cdots \mathrm{E}(U_n))$
- Meanwhile, $\mathrm{Var}\mathbf{u}$ is an $n \times n$ matrix with elements

$$(\mathrm{Var}(\mathbf{u}))_{ij} = \mathrm{E}\left\{(U_i - \mu_i)(U_j - \mu_j)\right\}$$

  where $\mu_i = \mathrm{E}(U_i)$
- The matrix $\mathrm{Var}(\mathbf{u})$ is referred to as the *variance-covariance matrix* of $\mathbf{u}$

Positive definite

- With scalars, we know that x^2 is always positive (well, at least, non-negative)
- The equivalent notion for matrices is that $\mathbf{X}^T\mathbf{X}$ is said to be a “positive definite” matrix
- To be more precise, if $\mathbf{X}$ is full rank, then $\mathbf{X}^T\mathbf{X}$ is *positive definite*; if $\mathbf{X}$ is *singular* (not of full rank), then $\mathbf{X}^T\mathbf{X}$ is said to be *positive semidefinite*
- This comes up from time to time in statistics; for example, a covariance matrix is always positive definite (or possibly positive semidefinite if, for example, some random variables have variance 0)

Matrix calculus

- Last but not least, we also need just a little bit of matrix calculus
- Specifically, the derivative of a scalar with respect to a vector is defined as

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_m} \end{bmatrix}$$

- The derivative of a vector $\mathbf{y}$ with respect to a vector $\mathbf{x}$ has columns $\frac{\partial y_1}{\partial \mathbf{x}}, \frac{\partial y_2}{\partial \mathbf{x}}, \ldots$ (*i.e.*, if $\mathbf{y}$ is an n-dimensional vector and $\mathbf{x}$ is an m-dimensional vector, $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ is an $m \times n$ matrix)

Linear and quadratic forms

- Two kinds of scalars that appear often are linear and quadratic forms
- Derivatives of linear forms:

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{a}^T\mathbf{x} = \mathbf{a}$$
$$\frac{\partial}{\partial \mathbf{x}}\mathbf{A}^T\mathbf{x} = \mathbf{A}$$

- Derivatives of quadratic forms:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{x} &= (\mathbf{A} + \mathbf{A}^T)\mathbf{x} \\ &= 2\mathbf{A}\mathbf{x} \quad \text{if } \mathbf{A} \text{ is symmetric}\end{aligned}$$

Chain and product rules

- Lastly, we need the chain rule and the product rule:

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \frac{\partial \mathbf{z}}{\partial \mathbf{y}}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{z}^T \mathbf{y}) = \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \mathbf{z}$$

- It is important to note that

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} \neq \frac{\partial \mathbf{z}}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{z}^T \mathbf{y}) \neq \mathbf{y} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} + \mathbf{z} \frac{\partial \mathbf{y}}{\partial \mathbf{x}}$$