Linear algebra for applied statistics

- Linear algebra is the math of vectors and matrices.
- In statistics, the main purpose of linear algebra is to organize data and write down the manipulations we want to do to them.
- A **vector** of length n is also called an n-**tuple**, or an **ordered sequence** of length n.
- We can suppose that each data point is a **real number**. We write $\mathcal{R}$ for the set of real numbers, and $\mathcal{R}^n$ for the set of vectors of n real numbers.
- Write the US life expectancy at birth for 2011 to 2015 as $\mathbf{y} = (y_1, y_2, y_3, y_4, y_5) = (79.0, 79.1, 79.0, 79.0, 78.9)$.
- We see $\mathbf{y} \in \mathcal{R}^5$. Numerical data can always be written as a vector in $\mathcal{R}^n$ where n is the number of datapoints. Categorical data can also be written as a vector in $\mathcal{R}^n$ by assigning a number for each category.
- Note that we use a bold font for vectors, and an italic font for the **components** of the vector.

More perspectives on vectors

Question: You may or may not have seen vectors in other contexts. In physics, a vector is a quantity with magnitude and direction. How does that fit in with our definition?

Question: How can I distinguish vectors using handwriting, when a bold font is not possible?

More perspectives on vectors

Question: You may or may not have seen vectors in other contexts. In physics, a vector is a quantity with magnitude and direction. How does that fit in with our definition?

Question: How can I distinguish vectors using handwriting, when a bold font is not possible?

An underscore is a conventional handwritten alternative to a bold font, so $\mathbf{x}$ is equivalent to $\underset{\sim}{x}$. In physics and mathematics, vectors are sometimes written as $\overrightarrow{x}$, but we will not do that here.

Adding vectors and multiplying by a scalar: Example

- For a dataset, the **index** i of the component y_i of the vector y might correspond to a measurement on the ith member of a population, the outcome of the ith group in an experiment, or the ith observation out of a sequence of observations on a system. Generically, we will call i an **observational unit**, or just **unit**.
- We might want to add two quantities u_i and v_i for unit i.
- Using vector notation, if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$, $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ we define the **vector sum** $\mathbf{y} = \mathbf{u} + \mathbf{v}$ to be the **componentwise sum** $y_i = u_i + v_i$, adding up the corresponding components for each unit.
- We might also want to rescale each component by the same factor. To change a measurement y_i in inches to a new measurement z_i in mm, we rescale with the **scalar** $\alpha = 25.4$. We want $z_i = \alpha y_i$ for each i. This is written in vector notation as **multplication of a vector by a scalar**, $\mathbf{z} = \alpha \mathbf{y}$.
- Keep track of whether each object is a scalar, a vector (what is its length?) or a matrix (what are its dimensions?).

Adding vectors and multiplying by a scalar: Example

An ecologist measures the pH of ten Michigan lakes at two points in the summer. Set up vector notation to describe her data. Write a vector calculation to find the average pH in each lake.

Adding vectors and multiplying by a scalar: Example

An ecologist measures the pH of ten Michigan lakes at two points in the summer. Set up vector notation to describe her data. Write a vector calculation to find the average pH in each lake.

First, set up notation.

Let x_i be the first pH measurement in lake i, for $i \in \{1, 2, \ldots, 10\}$.

Then, $\mathbf{x} = (x_1, \ldots, x_{10})$ is the vector of the first pH measurement in each of the 10 lakes.

Let $\mathbf{y} = (y_1, \ldots, y_{10})$ be the vector of second measurements.

Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{10})$ be the average pH for each of the 10 lakes.

Adding vectors and multiplying by a scalar: Example

An ecologist measures the pH of ten Michigan lakes at two points in the summer. Set up vector notation to describe her data. Write a vector calculation to find the average pH in each lake.

First, set up notation.

Let x_i be the first pH measurement in lake i, for $i \in \{1, 2, \dots, 10\}$.

Then, $\mathbf{x} = (x_1, \dots, x_{10})$ is the vector of the first pH measurement in each of the 10 lakes.

Let $\mathbf{y} = (y_1, \dots, y_{10})$ be the vector of second measurements.

Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{10})$ be the average pH for each of the 10 lakes.

For each lake i, the mean is $\mu_i = \frac{1}{2}(x_i + y_i)$. In vector notation, this is $\boldsymbol{\mu} = \frac{1}{2}(\mathbf{x} + \mathbf{y})$.

Vectors and scalars in R

- We have seen in Chapter 1 that R has vectors. An R vector of length 1 is a scalar.
- You can check that R follows the usual mathematical rules of vector addition and multiplication by a a scalar.

```r
x <- c(1,2,3)
y <- c(4,5,6)
```

```r
x+y
```

```
## [1] 5 7 9
```

```r
3*x
```

```
## [1] 3 6 9
```

- R also allows adding a scalar to a vector

```r
x <- c(1,2,3)
```

```r
x+2
```

```
## [1] 3 4 5
```

- Mathematically, adding scalars to vectors is not allowed. Instead, we define the **vector of ones**, $\mathbf{1} = (1, 1, \ldots, 1)$, and write $\mathbf{x} + 2 \times \mathbf{1}$.

Question: Why does R break the usual rules of mathematics here?

Matrices

- Matrices provide a way to store and manipulate p quantitites for each of n units.
- An $n \times p$ matrix $\mathbb{A}$ is a numerical array with n rows and p columns,

$$\mathbb{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}.$$

- Data that have the form of a matrix are called **rectangular**.
- Many common datasets are rectangular, consisting of multiple variables collected on a groups of individual units.
- We will use blackboard bold capital letters, $\mathbb{A}$, $\mathbb{B}$, $\mathbb{X}$, $\mathbb{Z}$, etc, for matrices. We are keeping plain capital letters to use for random variables.
- We say $\mathbb{A} = [a_{ij}]_{n \times p}$ as an abbreviation for writing the full $n \times p$ matrix.

Matrix times vector multiplication

- A linear system of n equations with p unknown variables, $x_1, \ldots, x_p$ is

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1p}x_p & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2p}x_p & = & b_2 \\ \vdots & & & & & & \vdots & & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \ldots & + & a_{np}x_p & = & b_n \end{array}$$

We define matrix multiplication $\mathbb{A}\mathbf{x} = \mathbf{b}$ to match this linear system. So,

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1p} \\ a_{21} & a_{22} & \ldots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{np} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

is exactly equivalent to the collection of p linear equations above.

- Mechanically, the ith component of $\mathbb{A}\mathbf{x}$ is found by multiplying each term in the ith row of $\mathbb{A}$ with corresponding terms in the **column vector** $\mathbf{x}$, and adding these contributions. See homework for practice!

Column vectors and row vectors

- In the matrix times vector multiplication on the previous slide, the vector $\mathbf{x}$ is written in a column, as a $p \times 1$ matrix.
- We say that that $\mathbf{x}$ is a **column vector**. We interpret a vector $\mathbf{x}$ as a column vectors unless we explicitly say it is a $1 \times p$ **row vector**.
- Similarly, $\mathbf{b}$ in the previous slide is a length n column vector.
- R matches our notation: a vector in R is not a matrix, but is interpreted as a column vector for matrix times vector multiplication. R uses %*% to denote matrix multiplication.

```
x <- c(1,2)
is(x,"matrix")

## [1] FALSE

is(x,"vector")

## [1] TRUE
```

```
A <- matrix(
  c(1,0,0,-1), nrow=2)
A %*% x

##      [,1]
## [1,]    1
## [2,]   -2
```

```
xx <- matrix(x,ncol=2)

A %*% x

##      [,1]
## [1,]    1
## [2,]   -2
```

Does a system of linear equations have no solution? One solution? Many solutions?

- One linear equation in one unknown, $ax = u$, has a unique solution unless $a = 0$.
- One linear equation with two unknowns, $ax + by = u$, has a solution consisting of all points on a line in the $x - y$ plane, as long as one of a and b is nonzero.
- Two linear equations with two unknowns, $\begin{array}{ccccc} ax & + & by & = & u \\ cx & + & dy & = & v \end{array}$, have a unique solution where the lines $ax + by = u$ and $cx + dy = v$ intersect, so long as these lines are not parallel.
- Three linear equations in two unknowns will not usually have a solution—the three corresponding lines would all have to meet at a common point.
- Can you see the general pattern?

Does a system of linear equations have no solution? One solution? Many solutions? Continued...

- For three unknowns, an equation $ax + by + cz = u$ corresponds to a plane in three-dimensional space.
- Three planes will typically intersect at a single point, so three equations in three unknowns will typically have a unique solution.
- Two planes that are not parallel will meet along a line, and give a family of solutions.
- Four or more planes will typically not all meet at any point.
- In higher dimensions, we can't visualize but the pattern remains true.
- With n equations in p unknowns we expect a unique solution when $p = n$, no solution when $p < n$ and a family of solutions when $p > n$.

Using matrices to solve a system of linear equations

- We've seen how matrices can represent a system of linear equations as $\mathbb{A}\mathbf{x} = \mathbf{b}$.
- For a basic linear algebra equation $ax = b$ we would divide through by a, or equivalently multiply through by a^{-1}, to find $x = a^{-1}b$ when $a \neq 0$.
- Is there a **matrix inverse** $\mathbb{A}^{-1}$ such that $\mathbf{x} = \mathbb{A}^{-1}\mathbf{b}$?
- We will see that there is an inverse $\mathbb{A}^{-1}$ when the system of linear equations has a unique solution. Since software can compute this inverse, we can solve systems of linear equations easily. This is useful in statistics for fitting linear models to datasets. Understanding when this inverse exists, and what to do when it doesn't, will help us develop appropriate models for data analysis.
- From the previous slide, we can only expect $\mathbb{A}^{-1}$ to exist when $p = n$, in which case $\mathbb{A}$ is called a **square matrix**.

Multiplying two matrices

- Let $\mathbb{A} = [a_{ij}]_{n\times p}$ and $\mathbb{X} = [x_{ij}]_{p\times q}$ be two matrices.
- The columns of $\mathbb{X}$ can be written as $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_q$.
- $\mathbb{X}$ consists of these q columns glued together, so $\mathbb{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3\ \cdots\ \mathbf{x}_q]$.
- Here, $\mathbf{x}_j$ is a vector whose ith component is x_{ij}.
- We define the **matrix product** $\mathbb{A}\mathbb{X}$ by gluing together the matrix times vector products for each column of $\mathbb{X}$, so $\mathbb{A}\mathbb{X} = [\mathbb{A}\mathbf{x}_1\ \mathbb{A}\mathbf{x}_2\ \mathbb{A}\mathbf{x}_3\ \cdots\ \mathbb{A}\mathbf{x}_q]$.
- From this definition, we see:
  1. The (i, j) entry of $\mathbb{A}\mathbb{X}$ is found by sliding the ith row of $\mathbb{A}$ down the jth column of $\mathbb{X}$, multiplying the corresponding terms and adding them. See homework for practice!
  2. Since each product $\mathbb{A}\mathbf{x}_j$ is a vector of length n, the dimension of $\mathbb{A}\mathbb{X}$ is $n \times q$. So, the rule for the dimension of a matrix product is
  $$(n \times p) \times (p \times q) = (n \times q)$$
  3. For the matrix product to exist, the number of columns of the first matrix must equal the number of rows of the second.

A matrix product example

Question: Let $\mathbb{U} = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix}$ and $\mathbb{V} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$. Calculate $\mathbb{U}\mathbb{V}$.

We can check our working in R.

```
(U <- matrix(
   c(2,2,1,-1),2))

##      [,1] [,2]
## [1,]    2    1
## [2,]    2   -1
```

```
(V <- matrix(
   c(3,1,1,2),2))

##      [,1] [,2]
## [1,]    3    1
## [2,]    1    2
```

```
U %*% V

##      [,1] [,2]
## [1,]    7    4
## [2,]    5    0
```

Matrix multiplication is not commutative

- Scalar multiplication (i.e., the usual multiplication of two real numbers) has the **commutative** property, $uv = vu$.
- Matrix multiplication does not usually have this property, e.g.,

```
U %*% V

##      [,1] [,2]
## [1,]    7    4
## [2,]    5    0
```

```
V %*% U

##      [,1] [,2]
## [1,]    8    2
## [2,]    6   -1
```

- We are all very used to multiplication being commutative. It takes practice to get used to the fact that matrix multiplication doesn't commute.

Addition of matrices and multiplication by a scalar

- We can add matrices componentwise, and multiply them by scalars, in exactly the same way we have already done for vectors.
- If $\mathbb{A} = [a_{ij}]_{p\times q}$ and $\mathbb{B} = [b_{ij}]_{p\times q}$ then the **matrix sum** $\mathbb{A} + \mathbb{B}$ is

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ a_{21} & a_{22} & \dots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pq} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ b_{21} & b_{22} & \dots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{bmatrix} = \begin{bmatrix} (a_{11} + b_{11}) & (a_{12} + b_{12} \\ (a_{21} + b_{21}) & (a_{22} + b_{22} \\ \vdots & \vdots \\ (a_{p1} + b_{p1}) & (a_{p2} + b_{p2} \end{bmatrix}$$

Matrix multiplication does have a **distributive** property:
$\mathbb{U}(\mathbb{V} + \mathbb{W}) = \mathbb{U}\mathbb{V} + \mathbb{U}\mathbb{W}$, where matrix addition (just like vector addition) is carried out for each

The identity matrix

- The $p \times p$ **identity matrix**, $\mathbb{I}_p$, is a square matrix with 1's on the diagonal and 0's everywhere else:

$$\mathbb{I}_p = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$