ECON207 Session 4

Introduction to Matrix Algebra

Anthony Tay

This Version: 17 Sep 2024

Session 4

- Introduction to Matrix Algebra
  - Different types of matrices
  - Matrix operations
  - Partitioned matrices
  - Vectors and matrices of random variables
  - Principal Component Analysis

Matrix Definition

Session 4.1

- Matrix Definitions and Operations
  - Types of matrices, notation
  - Additions, scalar multiplications, Hadamard product, matrix multiplication, transposition

Matrix Definition

rows
$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$
columns

Matrix with m rows, n columns

- **Order** or **Dimension** $m \times n$
- If $m = n$: **square matrix**
- If $m > 1$ and $n = 1$: **column vector**
- If $m = 1$ and $n > 1$: **row vector**
- If $m = 1$ and $n = 1$: **scalar**

a_{ij} is the (i, j)th **element** or **term** of the matrix

Other notational conventions:

- $(a_{ij})_{m \times n}$ refers to a $m \times n$ matrix with (i, j)th element a_{ij}
- $(A)_{ij}$ refers to the (i, j)th element of the matrix A

Matrix Operations (Add, Hadamard Prod, Scalar Mult.)

Given two matrices A and B of the same dimensions:

- Equality: $A = B \iff (A)_{ij} = (B)_{ij}$ for all $i = 1, \ldots, m;\ j = 1, \ldots, n$
- Addition: $(A+B)_{ij} = (A)_{ij} + (B)_{ij}$ for all $i = 1, \ldots, m;\ j = 1, \ldots, n$
  - Matrix addition is element-by-element addition
- Hadamard Product: $(A \odot B)_{ij} = (A)_{ij}(B)_{ij}$
  - Hadamard Product is element-by-element multiplication
  - Alternative notations for Hadamard Product: $A \circ B,\ A * B$

For any matrix A and any scalar $\alpha \in \mathbb{R}$

- Scalar Multiplication: $(\alpha A)_{ij} = (A\alpha)_{ij} = \alpha(A)_{ij}$ for all $i = 1, \ldots, m;\ j = 1, \ldots, n$

Matrix Operations (Add, Hadamard Prod, Scalar Mult.)

Examples: If $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$, then

- $A + B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{bmatrix}$
- $A \odot B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & a_{13}b_{13} \\ a_{21}b_{21} & a_{22}b_{22} & a_{23}b_{23} \end{bmatrix}$
- $\alpha A = A\alpha = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} \end{bmatrix}$
- $A - B = A + (-1)B = \begin{bmatrix} a_{11}-b_{11} & a_{12}-b_{12} & a_{13}-b_{13} \\ a_{21}-b_{21} & a_{22}-b_{22} & a_{23}-b_{23} \end{bmatrix}$

Matrix Operations (Add, Hadamard Prod, Scalar Mult.)

The following should be obvious:

- $(A+B)+C = A+(B+C)\,,\ (A \odot B) \odot C = A \odot (B \odot C)$
- $A+B = B+A\,,\ A \odot B = B \odot A$
- $A \odot (B+C) = A \odot B + A \odot C$
- $\alpha(A+B) = \alpha A + \alpha B\,,\ (\alpha+\beta)A = \alpha A + \beta A$

Matrix Operations (Transposition)

- Matrix Transpose of A, denoted A^{T}, is defined by $(A^{\mathrm{T}})_{ij} = A_{ji}$

e.g., if $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$, then $A^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$

e.g., if $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, then $x^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$

We often write column vectors as $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ to save space

Sometimes a matrix transpose is written as A' instead of A^{T}

Matrix Operations (Transposition)

Clearly

- $(A+B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}}$
- $(A \odot B)^{\mathrm{T}} = A^{\mathrm{T}} \odot B^{\mathrm{T}}$
- $(\alpha A)^{\mathrm{T}} = \alpha A^{\mathrm{T}}$
- $(A^{\mathrm{T}})^{\mathrm{T}} = A$

Definition: A square matrix is **symmetric** if $(A)_{ij} = (A)_{ji}$, i.e., $A^{\mathrm{T}} = A$

e.g., $\begin{bmatrix} 1 & 3 & 2 \\ 3 & 4 & 6 \\ 2 & 6 & 3 \end{bmatrix}$ is symmetric, $\begin{bmatrix} 1 & 3 & 2 \\ 7 & 4 & 6 \\ 2 & 6 & 3 \end{bmatrix}$ is not

Matrix Operations (Matrix Multiplication)

Matrix Multiplication/Product: For any $m \times n$ matrix A and $n \times p$ matrix B, we have

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \,.$$

i.e., (i,j)th element of AB is the sum of the product of the elements of the ith row of A with the corresponding elements in the jth column of B

For example,

- $(AB)_{11} = \sum_{k=1}^{n} a_{1k} b_{k1} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} + \cdots + a_{1n}b_{n1}$
- $(AB)_{23} = \sum_{k=1}^{n} a_{2k} b_{k3} = a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} + \cdots + a_{2n}b_{n3}$

Matrix Operations

For the product of a 3×3 matrix and a 3×2 matrix, we have

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & \bullet \\ \bullet & \bullet \\ \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{3} a_{1k}b_{k1} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ \bullet & \bullet \\ \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{3} a_{1k}b_{k1} & \sum_{k=1}^{3} a_{1k}b_{k2} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & \bullet \\ \bullet & \bullet \end{bmatrix}$$

and so on.

Matrix Operations (Matrix Multiplication)

If $A = \begin{bmatrix} 2 & 8 \\ 3 & 0 \\ 5 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 4 & 7 \\ 6 & 9 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 3 & 4 \\ 6 & 2 & 5 \end{bmatrix}$ then

$$AB = \begin{bmatrix} 2 & 8 \\ 3 & 0 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 6 & 9 \end{bmatrix} = \begin{bmatrix} 2\cdot 4 + 8 \cdot 6 & 2 \cdot 7 + 8 \cdot 9 \\ 3 \cdot 4 + 0 \cdot 6 & 3 \cdot 7 + 0 \cdot 9 \\ 5 \cdot 4 + 1 \cdot 6 & 5 \cdot 7 + 1 \cdot 9 \end{bmatrix} = \begin{bmatrix} 56 & 86 \\ 12 & 21 \\ 26 & 44 \end{bmatrix}.$$

- A and B "conformable" for the product AB requires no.cols(A) $=$ no.rows(B)
- Even if A and B are conformable for AB, the product BA might not be possible
- Even if AB and BA are possible, they may not be equal (might not even be the same dimensions)

Matrix Operation (Matrix Multiplication)

```
A = matrix(c(2,8,3,0,5,1),
        nrow=3, byrow=T)
A
```

```
     [,1] [,2]
[1,]    2    8
[2,]    3    0
[3,]    5    1
```

```
B = matrix(c(4,7,6,9),
        nrow=2, byrow=T)
B
```

```
     [,1] [,2]
[1,]    4    7
[2,]    6    9
```

```
C = matrix(c(1,3,4,6,2,5),
        nrow=2, byrow=T)
C
```

```
     [,1] [,2] [,3]
[1,]    1    3    4
[2,]    6    2    5
```

```
A %*% B
```

```
     [,1] [,2]
[1,]   56   86
[2,]   12   21
[3,]   26   44
```

```
A %*% C
```

```
     [,1] [,2] [,3]
[1,]   50   22   48
[2,]    3    9   12
[3,]   11   17   25
```

```
B %*% A
```

```
Error in B %*% A: non-conformable arguments
```

```
C %*% A
```

```
     [,1] [,2]
[1,]   31   12
[2,]   43   53
```


Matrix Operations (Matrix Multiplication)

Easy to show

- $(AB)C = A(BC)$ if A is $m \times n$, B is $n \times p$ and C is $p \times q$
- $A(B + C) = AB + AC$ if A is $m \times n$, B and C are $n \times p$
- $(A + B)C = AC + BC$ if A and B are $m \times n$ and C is $n \times p$

Proof of $(AB)C = A(BC)$:

$$((AB)C)_{ij} = \sum_{k=1}^{p}((AB))_{ik}(C)_{kj} = \sum_{k=1}^{p}\left(\sum_{r=1}^{n}(A)_{ir}(B)_{rk}\right)(C)_{kj}$$
$$= \sum_{r=1}^{n}(A)_{ir}\left(\sum_{k=1}^{p}(B)_{rk}(C)_{kj}\right) = \sum_{r=1}^{n}(A)_{ir}(BC)_{rj} = (A(BC))_{ij}$$

Matrix Operations

In some respect, matrix multiplication behaves quite differently from multiplication of numbers:

e.g.,

- $\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
- $\begin{bmatrix} 1 & b \\ -\frac{1}{b} & -1 \end{bmatrix}\begin{bmatrix} 1 & b \\ -\frac{1}{b} & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

A matrix with all elements 0 is the “zero matrix” $0_{m \times n}$

Sometimes subscripts left out

- $A0 = 0$
- $0B = 0$
- But $AB = 0$ does not imply $A = 0$ or $B = 0$
- Possible for $A \neq 0$, yet $A^2 = AA = 0$

Matrix Operations (Matrix Multiplication)

It is possible for $Ab = Ac$, yet $b \neq c$

e.g.,

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$$

There is an important special case where $Ab = Ac \implies b = c$

We'll come to it later

Important to understand $Ab = Ac$ does not imply $b = c$ in general

Relationship between Matrix Multiplication and Transpose

Suppose A is $m \times n$ and B is $n \times p$, then

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}$$

Proof: We have

$$\begin{aligned}((AB)^{\mathrm{T}})_{ij} = (AB)_{ji} &= \sum_{k=1}^{n} (A)_{jk}(B)_{ki} \\ &= \sum_{k=1}^{n} (A^{\mathrm{T}})_{kj}(B^{\mathrm{T}})_{ik} \\ &= \sum_{k=1}^{n} (B^{\mathrm{T}})_{ik}(A^{\mathrm{T}})_{kj} = (B^{\mathrm{T}}A^{\mathrm{T}})_{ij}\end{aligned}$$

The Identity Matrix

The **identity matrix** I_n is the $n \times n$ matrix such that

$$(I_n)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \text{for} \quad i, j = 1, \dots, n$$

That is,

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

If A is $m \times n$ then

$$AI_n = A \quad \text{and} \quad I_m A = A$$

Diagonal, Upper and Lower Triangular Matrices

The identity matrix is an example of a diagonal matrix

A **diagonal matrix** D is a square matrix such that $(D)_{ij} = 0$ for all $i \neq j$

- It doesn't matter what the diagonal elements $(D)_{ii}$ are
- Diagonal matrices are often written diag$(d_1, d_2, \dots, d_n)$
- The identity matrix is diag$(1, 1, \dots, 1)$

A **lower triangular matrix** L is a square matrix such that $(L)_{ij} = 0$ for all $i < j$

An **upper triangular matrix** U is a square matrix such that $(U)_{ij} = 0$ for all $i > j$

Diagonal, Upper and Lower Triangular Matrices

diagonal / lower triangular / upper triangular

$$D = \begin{bmatrix} * & 0 & \cdots & 0 & 0 \\ 0 & * & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & * & 0 \\ 0 & 0 & \cdots & 0 & * \end{bmatrix} \quad L = \begin{bmatrix} * & 0 & \cdots & 0 & 0 \\ * & * & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & \cdots & * & 0 \\ * & * & \cdots & * & * \end{bmatrix} \quad U = \begin{bmatrix} * & * & \cdots & * & * \\ 0 & * & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & * & * \\ 0 & 0 & \cdots & 0 & * \end{bmatrix}$$

where * means any value, including 0

Important examples of matrix products

Example: The general linear system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

can be written as $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ or $Ax = b$

Often the problem is: given A and b, want to find x so that equation holds

Important examples of matrix products

If $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$, then

Inner Product: $x^{\mathrm{T}}x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i^2$

The norm of a vector x is defined as $\|x\| = \sqrt{x^{\mathrm{T}}x}$

Outer Product: $xx^{\mathrm{T}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2 \end{bmatrix}$

Important examples of matrix products

If $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$, and $A = (a_{ij})_{n \times n}$, then $x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}$

For the case when $n = 3$:

$$\begin{aligned} x^T A x &= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} x_1a_{11} + x_2a_{21} + x_3a_{31} & x_1a_{12} + x_2a_{22} + x_3a_{32} & x_1a_{13} + x_2a_{23} + x_3a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= x_1^2a_{11} + x_2^2a_{22} + x_3^2a_{33} + x_1x_2(a_{12} + a_{21}) + x_1x_3(a_{13} + a_{31}) + x_2x_3(a_{23} + a_{32}) \end{aligned}$$

When A is symmetric, $x^{\mathrm{T}}Ax$ is called a **quadratic form**

Session 4.2

Session 4.2 Comments on the Inner Product

- Geometric understanding of the inner product
- Norms and angles between vectors

Comments on the Inner Product

In some contexts, a "vector" is simply an ordered list of numbers with no shape

- e.g., an general n-vector $x = (x_1, x_2, \dots, x_n)$
- e.g., a specific 4-vector $(1, 5, 3, 2)$

Inner product / Scalar Product / Dot Product $u \cdot v$ or $\langle u, v \rangle$

$$u \cdot v = (u_1, u_2, \dots, u_n) \cdot (v_1, v_2, \dots, v_n) = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

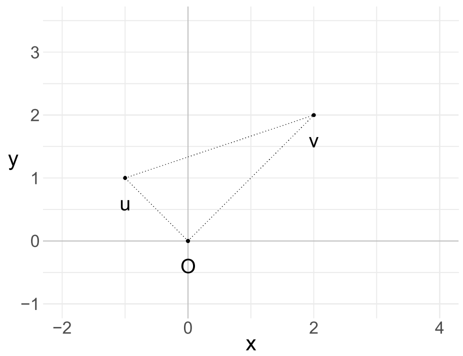
In matrix algebra, we organize vectors into rows or columns

If u and v are columns vectors, the inner product is $u^{\mathrm{T}} v$

Comments on the Inner Product

We can think of vectors as points in space

e.g. for vectors in $\mathbb{R}^2$ such as $u = (u_1, u_2) = (-1, 1)$, $v = (v_1, v_2) = (2, 2)$



- The **norm** of v is the distance from $O = (0, 0)$ to $v = (2, 2)$

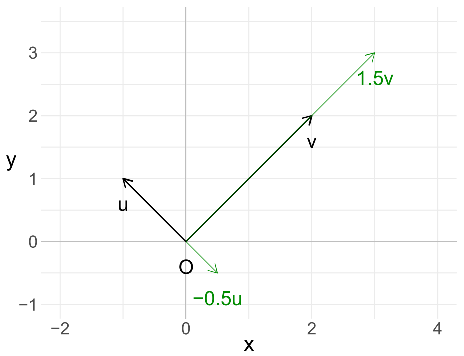
$$\|v\| = \sqrt{v_1^2 + v_2^2} = \sqrt{2}$$

(Pythagoras's Theorem)

- distance from u to v is

$$\sqrt{(v_1 - u_1)^2 + (v_2 - u_2)^2} = \|v - u\| = \sqrt{10}$$

Comments on the Inner Product



Can view a vector v as an "arrow" from O to v

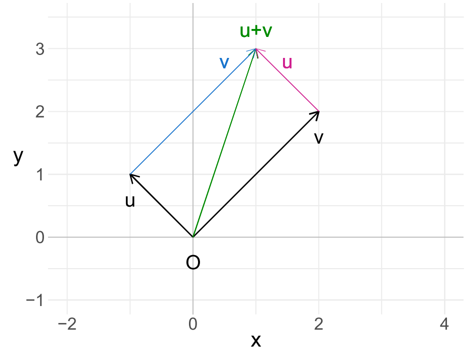
- $\|v\|$ is length of arrow representing v
- But use "norm" instead of "length"

Scalar Multiplication αv

- Stretches a vector if $|\alpha| > 1)$
- Shrinks a vector if $|\alpha| < 1$
- Returns the vector to the origin if $\alpha = 0$
- flips the direction of the vector if $\alpha < 0$

The vector $\dfrac{v}{\|v\|}$ has unit norm

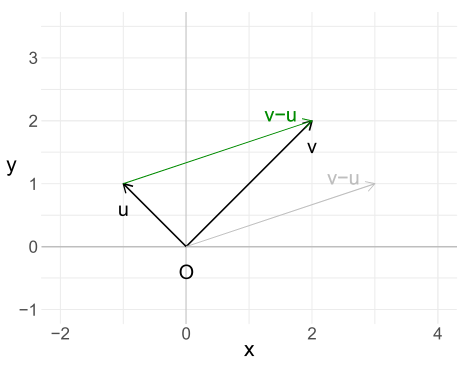
Comments on the Inner Product



Vector Addition:

- $u + v$ is the diagonal, starting at O, of the parallelogram formed by u and v
- When thinking of vectors as "arrows", the starting position is irrelevant
- The black arrow v and the blue arrow are the same vector
- The red arrow u and the red arrow are the same vector

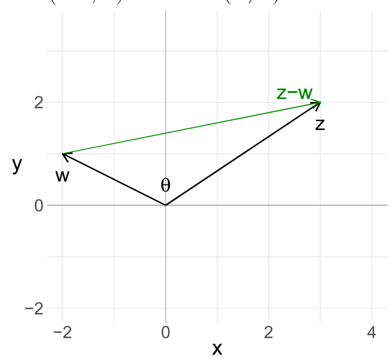
Comments on the Inner Product



- Since
$$u + (v - u) = v$$
the vector $v - u$ is represented by the arrow from u to v
- It is also represented by the arrow from the origin to the point
$$(3, 1) = (2, 2) - (-1, 1)$$

Comments on the Inner Product

e.g., $w = (-2, 1)$ and $z = (3, 2)$



Interpretation of the dot product of two different vectors:

If θ is the angle formed at the origin by two vectors w and z, then
$$w \cdot z = \|w\|\,\|z\|\,\cos\theta$$
i.e.,
$$\frac{w}{\|w\|} \cdot \frac{z}{\|z\|} = \cos\theta$$
i.e., the dot product of two unit vectors gives the cosine of the angle formed by the two vectors at the origin

Comments on the Inner Product

Proof using the Cosine Rule:
$$\|z - w\|^2 = \|z\|^2 + \|w\|^2 - 2\,\|z\|\,\|w\|\cos\theta$$
(NB: Pythagoras's Theorem is when $\theta = \pi/2$, so $\cos\theta = 0$)

Converting to inner products (we'll take the vectors to be column vectors)
$$\begin{aligned}(z-w)^{\mathrm{T}}(z-w) &= (z^{\mathrm{T}} - w^{\mathrm{T}})(z-w)\\ &= z^{\mathrm{T}}z + w^{\mathrm{T}}w - z^{\mathrm{T}}w - w^{\mathrm{T}}z\\ &= z^{\mathrm{T}}z + w^{\mathrm{T}}w - 2w^{\mathrm{T}}z \quad (\text{since } z^{\mathrm{T}}w = w^{\mathrm{T}}z)\end{aligned}$$
Comparing the two, we have $w \cdot z = \|z\|\,\|w\|\cos\theta$

Comments on the Inner Product

The Cosine Rule also gives the **Triangle Inequality**

Since $-1 \le \cos\theta \le 1$, we have
$$\begin{aligned}\|z-w\|^2 &= \|z\|^2 + \|w\|^2 - 2\,\|z\|\,\|w\|\cos\theta\\ &\le \|z\|^2 + \|w\|^2 + 2\,\|z\|\,\|w\|\\ &= (\|z\| + \|w\|)^2\end{aligned}$$
$$\|z - w\| \le \|z\| + \|w\|$$
Length of one side of a triangle is less than the sum of the lengths of the other two sides

Comments on the Inner Product

All of this extends to n-vectors

- If $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$, then
  - $\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^{\mathrm{T}}x}$ is the "distance" from the origin to x
  - $\|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$ is the "distance" between x and y

But what is "distance" in n-dimensions? We can show $\|x - y\|$ satisfies the three main properies of distance:

- it is non-negative (obvious)
- it is symmetric, i.e., distance between x and y is the same as distance between y and x (also obvious)
- it satisfies the triangle inequality

Comments on the Inner Product

The key is the **Cauchy-Schwarz Inequality**: For any two $n \times 1$ vectors u and v, we have

$$|u^{\mathrm{T}}v| \leq \|u\|\|v\|$$

Equality holds only if $u = \alpha v$

Proof:

- If $u = 0_{n\times 1}$ or $v = 0_{n\times 1}$, then the CS-Inequality holds trivially with equality
- $u = \alpha v$, then

$$|u^{\mathrm{T}}v| = |(\alpha v)^{\mathrm{T}}v| = |\alpha|\|v\|^2 \quad \text{and} \quad \|u\|\|v\| = \|\alpha v\|\|v\| = |\alpha|\|v\|^2$$

  so the CS-Inequality holds (with equality)

Comments on the Inner Product

- If $u \neq \alpha v$ for any α. Then

$$0 < (u - \alpha v)^{\mathrm{T}}(u - \alpha v) = u^{\mathrm{T}}u - \alpha v^{\mathrm{T}}u - \alpha u^{\mathrm{T}}v + \alpha^2 v^{\mathrm{T}}v = u^{\mathrm{T}}u - 2\alpha u^{\mathrm{T}}v + \alpha^2 v^{\mathrm{T}}v$$

  This inequality holds for all α. At the particular values $\alpha = u^{\mathrm{T}}v/v^{\mathrm{T}}v$, we have

$$0 < u^{\mathrm{T}}u - 2\frac{u^{\mathrm{T}}v}{v^{\mathrm{T}}v}u^{\mathrm{T}}v + \frac{(u^{\mathrm{T}}v)^2}{(v^{\mathrm{T}}v)^2}v^{\mathrm{T}}v = u^{\mathrm{T}}u - \frac{(u^{\mathrm{T}}v)^2}{v^{\mathrm{T}}v}$$

$$\implies \quad (u^{\mathrm{T}}u)(v^{\mathrm{T}}v) < (u^{\mathrm{T}}v)^2$$

  Taking square roots gives the result $|u^{\mathrm{T}}v| \leq \|u\|\|v\|$

Comments on the Inner Product

Cauchy-Schwarz Inequality implies the Triangle inequality

Let x, y, z be any three $n \times 1$ vectors and let $u = x - y$ and $v = y - z$. Then

$$\begin{aligned}
\|x - z\|^2 &= \|x - y + y - z\|^2 \\
&= (x - y + y - z)^{\mathrm{T}}(x - y + y - z) \\
&= (x - y)^{\mathrm{T}}(x - y) + 2(x - y)^{\mathrm{T}}(y - z) + (y - z)^{\mathrm{T}}(y - z) \\
&\leq \|x - y\|^2 + \|y - z\|^2 + 2|(x - y)^{\mathrm{T}}(y - z)| \\
&\leq \|x - y\|^2 + \|y - z\|^2 + 2\|x - y\|\,\|y - z\| = (\|x - y\| + \|y - z\|)^2
\end{aligned}$$

Taking square roots gives $\|x - z\| \leq \|x - y\| + \|y - z\|$

Comments on the Inner Product

- We can treat $\|x - y\|$ as "distance" from x to y even in n-dimensions
  - even though we can't literally measure this distance with a ruler
- We can treat $\|x\|$ as "distance" from origin to x ("norm")

Furthermore, since $|x^{\mathrm{T}}y| \leq \|x\|\|y\|$ implies $-1 \leq \frac{x^{\mathrm{T}}y}{\|x\|\|y\|} \leq 1$

We define the "angle" between x and y to be θ such that

$$\cos\theta = \frac{x^{\mathrm{T}}y}{\|x\|\|y\|}$$

If $x^{\mathrm{T}}y = 0$ we say that x and y are **orthogonal** (general n-dimensional version of "perpendicular")

Session 4.3

Session 4.3 Inverse Matrices

- Left-, right-, and "two-sided" inverse matrix
- Properties of the inverse matrix
- Determinants

Less emphasis on methods for computing inverses and determinants

The Inverse Matrix

Let A be an $m \times n$ matrix

- The $n \times m$ matrix B is a **left-inverse** of A if $BA = I_n$
- The $n \times m$ matrix C is a **right-inverse** of A if $AC = I_m$.

Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 0.2 & 0.4 \\ 2 & -0.2 & -0.4 \end{bmatrix}$

B is a left-inverse of A (or A is the right-inverse of B) since

$$BA = \begin{bmatrix} -1 & 0.2 & 0.4 \\ 2 & -0.2 & -0.4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} -1+0.4+1.6 & -1+0.2+0.8 \\ 2-0.4-1.6 & 2-0.2-0.8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The Inverse Matrix

B is not a right-inverse of A since $AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.2 & 0.4 \\ 0 & 0.4 & 0.8 \end{bmatrix}$

In fact A has no right-inverse

Suppose $AC = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ then we have

$$2b + e = 1 \text{ and } 4b + 2e = 0$$

which is a contradiction

The Inverse Matrix

A has left- and right-inverse only if it is square, and they will be the same matrix, i.e.,

- If A is $n \times n$, and $BA = AC = I_n$, then it must be that $B = C$

$$BA = I_n \implies BAC = I_n C \implies BI_n = C \implies B = C$$

Then $B = C$ is the "two-sided inverse", or simply the **inverse** of A, denoted A^{-1}

The inverse of a $n \times n$ matrix A, *if it exists*, is the unique matrix A^{-1} such that

$$A^{-1}A = I_n = AA^{-1}$$

The Inverse Matrix

We emphasize

- A has a (two-sided) inverse only if it is square
- But not all square matrices have an inverse

Examples:

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \quad \text{is} \quad A^{-1} = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$$

Verify by direct multiplication:

$$A^{-1}A = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The Inverse Matrix

The matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ has no inverse

Proof: Suppose

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

This implies $a + 2c = 1$ but $2a + 4c = 2(a+2c) = 0$ which gives a contradiction

The Inverse Matrix

Consider the system of n equations in n unknowns

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \implies \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \vdots & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \text{ or } Ax = b$$

If A has an inverse, then the unique solution to this system is

$$Ax = b \iff x = A^{-1}b$$

- $Ax = b \Rightarrow A^{-1}Ax = A^{-1}b \Rightarrow x = A^{-1}b$
- $x = A^{-1}b \Rightarrow Ax = AA^{-1}b = b$ so $x = A^{-1}b$ is indeed a solution

The Inverse Matrix

Example: consider the system

$$\begin{aligned} x_1 + 3x_2 &= 1 \\ 2x_1 + 4x_2 &= 3 \end{aligned} \quad \text{or} \quad \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{or} \quad Ax = b$$

We saw earlier that the inverse of $A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ is $A^{-1} = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$

The unique solution is

$$A^{-1}b = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5/2 \\ -1/2 \end{bmatrix}$$

The Inverse Matrix

(Warning) The same argument doesn't quite hold for left-inverses

Suppose the system is $Ax = b$ where A is $m \times n$, $m < n$, with left-inverse A_{left}^{-1}

Pre-multiplying both side of $Ax = b$ by A_{left}^{-1} gives

$$A_{left}^{-1}Ax = A_{left}^{-1}b \implies x = A_{left}^{-1}b$$

However, when we check if $x = A_{left}^{-1}b$ is a solution, we get

$$Ax = AA_{left}^{-1}b$$

which *may or may not* be equal to b, since $AA_{left}^{-1} \neq I_m$

- If $AA_{left}^{-1}b = b$, there is a unique solution and you have found it
- If $AA_{left}^{-1}b \neq b$, there is no solution

The Inverse Matrix

Example: consider the systems

$$\text{(i)} \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \qquad \text{and} \qquad \text{(ii)} \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

which we write as (i) $Ax = b$ and (ii) $Ax = c$ respectively

The left inverse of $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix}$ is $A_{left}^{-1} = \begin{bmatrix} -4/9 & 5/9 & 1/9 \\ 5/9 & -4/9 & 1/9 \end{bmatrix}$ (verify!)

You can verify that

- $AA_{left}^{-1}b = b$ (despite $AA_{left}^{-1} \neq I_3$) so $A_{left}^{-1}b$ is a unique solution to (i)
- $AA_{left}^{-1}c \neq c$ so $A_{left}^{-1}c$ is not a solution to (ii)

The Inverse Matrix

The inverse of an arbitrary 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, if it exists, is

$$\mathit{Memorize\ this!} \rightarrow \quad A^{-1} = \frac{1}{\det(A)}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad \text{where } \det(A) = a_{11}a_{22} - a_{12}a_{21}$$

- $\det(A)$ is the **determinant** of the 2×2 matrix A
- the inverse exists only if det(A) $\neq 0$
- If inverse of A does not exist, we say that A is **singular**
- If inverse exists, we say that A is **non-singular**
- An alternative notation for $\det(A)$ is $|A|$

The Inverse Matrix

The inverse of $A = \begin{bmatrix} 1 & 4 \\ 5 & 6 \end{bmatrix}$ is

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} 6 & -4 \\ -5 & 1 \end{bmatrix} = -\frac{1}{14} \begin{bmatrix} 6 & -4 \\ -5 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{7} & \frac{2}{7} \\ \frac{5}{14} & -\frac{1}{14} \end{bmatrix}.$$

The determinant of $B = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$ is $\det(B) = 1 \cdot 6 - 2 \cdot 3 = 0$, so B is singular

When will det$(A) = 0$?

- if one or both rows or columns are all zero, or
- if one row is a multiple of the other, or
- if one column is a multiple of the other

The Inverse Matrix

See readings for

- general formula for the determinant and inverse matrix for general $n \times n$ matrix
- algorithmic approach to calculating inverses of general $n \times n$ matrix
- deeper understanding of determinant and inverse matrix

Generally speaking, determinant will be zero (and the inverse will not exist) if

- if one or more rows or columns of the matrix are all zero
- if one column is a multiple of another
- if one column is exactly a linear combination of the others

If det$(A) \neq 0$, then A is "full rank", and the inverse exists

The Inverse Matrix

A few additional results:

- The inverse of a diagonal matrix diag$(d_1, \dots, d_n)$ is diag$(d_1^{-1}, \dots d_n^{-1})$
- If A is $n \times n$ and non-singular, then $(A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}$

$$\begin{aligned} AA^{-1} = I_n \;&\Rightarrow\; (A^{-1})^{\mathrm{T}} A^{\mathrm{T}} = I_n \\ &\Rightarrow\; (A^{-1})^{\mathrm{T}} A^{\mathrm{T}} (A^{\mathrm{T}})^{-1} = I (A^{\mathrm{T}})^{-1} \;\Rightarrow\; (A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1} \end{aligned}$$

- A and B are both $n \times n$ and non-singular, then $(AB)^{-1} = B^{-1}A^{-1}$.

$$B^{-1}A^{-1}AB = ABB^{-1}A^{-1} = I \text{ implies } B^{-1}A^{-1} = (AB)^{-1}$$

The Inverse Matrix

- The inverse of a non-singular symmetric matrix is symmetric (exercise)
- If X is $n \times k$ with $n > k$ and $Xc = 0_{n \times 1} \iff c = 0_{k \times 1}$, then

$$X^{\mathrm{T}}X \text{ is non-singular}$$

Elaboration: If X is $n \times k$ with $n > k$ and c is $k \times 1$, then

$$Xc = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} c_1 x_{11} + c_2 x_{12} + \cdots + c_k x_{1k} \\ c_1 x_{21} + c_2 x_{22} + \cdots + c_k x_{2k} \\ \vdots \\ c_1 x_{n1} + c_2 x_{n2} + \cdots + c_k x_{nk} \end{bmatrix}$$

$$= c_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} + c_2 \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix} + \cdots + c_k \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{bmatrix} = c_1 X_{*1} + c_2 X_{*2} + \cdots c_k X_{*k}$$

The Inverse Matrix

$c = 0_{k\times1} \implies Xc = 0_{n\times1}$ always hold. When would $c \neq 0_{k\times1}$ yet $Xc = 0_{n\times1}$?

Suppose $c_i \neq 0$ for some i, yet $Xc = 0_{n\times1}$. Then we can write

$$c_i X_{*i} = c_1 X_{*1} + \cdots + c_{i-1} X_{*(i-1)} + c_{i+1} X_{*(i+1)} + \cdots + c_k X_{*k}$$
$$\Rightarrow \quad X_{*i} = d_1 X_{*1} + \cdots + d_{i-1} X_{*(i-1)} + d_{i+1} X_{*(i+1)} + \cdots + d_k X_{*k} \text{ where } d_j = c_j / c_i$$

- if all the $d_j = 0$, $j \neq i$, then $X_{*i} = 0_{n\times1}$
- if exactly one $d_j \neq 0$, $j \neq i$, then $X_{*i} = d_j X_{*j}$, i.e., one column is a multiple of another
- if two or more $d_j \neq 0$, then X_{*i} is a linear combination of some of the other columns

Agenda | Definitions and Matrix Operations | Inner Product | Inverse Matrices | Partitioned Matrices | Vectors of RVs | PCA | Roadmap

The Inverse Matrix

- If there is a vector $c \neq 0_{k\times1}$ such that $Xc = 0_{n\times1}$, we say that the columns of X are "linearly dependent"
- If X is a data matrix (one column per variable) whose columns are "linearly dependent", we also say that there is "perfect collinearity" in X
- If $Xc = 0_{n\times1} \iff c = 0_{k\times1}$, then the columns of X are "linearly independent"
- We also say the "X" has full column rank
- For more on matrix rank, please see readings

If the columns of X are linearly independent, i.e., $Xc = 0_{n\times1} \iff c = 0_{k\times1}$, then

$$(X^{\mathrm{T}}X)^{-1} \text{ exists}$$

Agenda | Definitions and Matrix Operations | Inner Product | Inverse Matrices | Partitioned Matrices | Vectors of RVs | PCA | Roadmap

Session 4.4

Session 4.4 Partitioned Matrices

- Partitioned or block matrices
- Addition, multiplication and transpose of partitioned matrices

Agenda | Definitions and Matrix Operations | Inner Product | Inverse Matrices | Partitioned Matrices | Vectors of RVs | PCA | Roadmap

Partitioned Matrices

We can partition contents of an $m \times n$ matrix into blocks of submatrices, e.g.,

$$A = \begin{bmatrix} 1 & 3 & 2 & 6 \\ 2 & 8 & 2 & 1 \\ 3 & 1 & 2 & 4 \\ 4 & 2 & 1 & 3 \\ 3 & 1 & 1 & 7 \end{bmatrix} = \left[\begin{array}{c|ccc} 1 & 3 & 2 & 6 \\ 2 & 8 & 2 & 1 \\ \hline 3 & 1 & 2 & 4 \\ 4 & 2 & 1 & 3 \\ 3 & 1 & 1 & 7 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where $A_{11} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $A_{21} = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$, $A_{12} = \begin{bmatrix} 3 & 2 & 6 \\ 8 & 2 & 1 \end{bmatrix}$ and $A_{22} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 1 & 3 \\ 1 & 1 & 7 \end{bmatrix}$

Partitioned Matrices

- Partitioned matrices are often called **block matrices**
- Many ways of partitioning any given matrix, e.g.,

$$A=\begin{bmatrix}1&3&2&6\\2&8&2&1\\3&1&2&4\\4&2&1&3\\3&1&1&7\end{bmatrix}=\left[\begin{array}{cc|cc}1&3&2&6\\2&8&2&1\\3&1&2&4\\\hline 4&2&1&3\\3&1&1&7\end{array}\right].$$

Main point of this section: as long as the matrices are appropriately partitioned, we can add / multiply partitioned matrices as though the blocks were elements

Partitioned Matrices

Addition of Partitioned Matrices If A and B are two $m\times n$ matrices A and B partitioned as:

$$A=\begin{bmatrix}\underbrace{A_{11}}_{m_1\times n_1}&\underbrace{A_{12}}_{m_1\times n_2}\\\underbrace{A_{21}}_{m_2\times n_1}&\underbrace{A_{22}}_{m_2\times n_2}\end{bmatrix}\quad\text{and}\quad B=\begin{bmatrix}\underbrace{B_{11}}_{m_1\times n_1}&\underbrace{B_{12}}_{m_1\times n_2}\\\underbrace{B_{21}}_{m_2\times n_1}&\underbrace{B_{22}}_{m_2\times n_2}\end{bmatrix}$$

where $n_1+n_2=n$ and $m_1+m_2=m$, then

$$A+B=\begin{bmatrix}\underbrace{A_{11}+B_{11}}_{m_1\times n_1}&\underbrace{A_{12}+B_{12}}_{m_1\times n_2}\\\underbrace{A_{21}+B_{21}}_{m_2\times n_1}&\underbrace{A_{22}+B_{22}}_{m_2\times n_2}\end{bmatrix}$$

Partitioned Matrices

Multiplication of Partitioned Matrices. If A and B are $m\times p$ and $p\times n$ respectively, and partitioned as:

$$A=\begin{bmatrix}\underbrace{A_{11}}_{m_1\times p_1}&\underbrace{A_{12}}_{m_1\times p_2}\\\underbrace{A_{21}}_{m_2\times p_1}&\underbrace{A_{22}}_{m_2\times p_2}\end{bmatrix}\quad\text{and}\quad B=\begin{bmatrix}\underbrace{B_{11}}_{p_1\times n_1}&\underbrace{B_{12}}_{p_1\times n_2}\\\underbrace{B_{21}}_{p_2\times n_1}&\underbrace{B_{22}}_{p_2\times n_2}\end{bmatrix}$$

then

$$AB=\begin{bmatrix}\underbrace{A_{11}}_{m_1\times p_1}&\underbrace{A_{12}}_{m_1\times p_2}\\\underbrace{A_{21}}_{m_2\times p_1}&\underbrace{A_{22}}_{m_2\times p_2}\end{bmatrix}\begin{bmatrix}\underbrace{B_{11}}_{p_1\times n_1}&\underbrace{B_{12}}_{p_1\times n_2}\\\underbrace{B_{21}}_{p_2\times n_1}&\underbrace{B_{22}}_{p_2\times n_2}\end{bmatrix}=\begin{bmatrix}\underbrace{A_{11}B_{11}+A_{12}B_{21}}_{m_1\times n_1}&\underbrace{A_{11}B_{12}+A_{12}B_{22}}_{m_1\times n_2}\\\underbrace{A_{21}B_{11}+A_{22}B_{21}}_{m_2\times n_1}&\underbrace{A_{21}B_{12}+A_{22}B_{22}}_{m_2\times n_2}\end{bmatrix}$$

Partitioned Matrices

Transposition of Partitioned Matrices: We have

$$A=\begin{bmatrix}\underbrace{A_{11}}_{m_1\times n_1}&\underbrace{A_{12}}_{m_1\times n_2}\\\underbrace{A_{21}}_{m_2\times n_1}&\underbrace{A_{22}}_{m_2\times n_2}\end{bmatrix}\quad\Rightarrow\quad A^{\mathrm{T}}=\begin{bmatrix}\underbrace{A_{11}^{\mathrm{T}}}_{n_1\times m_1}&\underbrace{A_{21}^{\mathrm{T}}}_{n_1\times m_2}\\\underbrace{A_{12}^{\mathrm{T}}}_{n_2\times m_1}&\underbrace{A_{22}^{\mathrm{T}}}_{n_2\times m_2}\end{bmatrix}$$

e.g., If X is an $n\times k$ data matrix partitioned into columns, then

$$X=\begin{bmatrix}x_{11}&x_{12}&\cdots&x_{1k}\\x_{21}&x_{22}&\cdots&x_{2k}\\\vdots&\vdots&\ddots&\vdots\\x_{n1}&x_{n2}&\cdots&x_{nk}\end{bmatrix}=\begin{bmatrix}X_{*1}&X_{*2}&\cdots&X_{*k}\end{bmatrix}\Longrightarrow X^{\mathrm{T}}=\begin{bmatrix}X_{*1}^{\mathrm{T}}\\X_{*2}^{\mathrm{T}}\\\vdots\\X_{*k}^{\mathrm{T}}\end{bmatrix}$$

X_{*i} is the column vector of all N observations of variable i

Partitioned Matrices

$$Xc = \begin{bmatrix} X_{*1} & X_{*2} & \cdots & X_{*k} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = c_1 X_{*1} + c_2 X_{*2} + \cdots + c_k X_{*k}$$

$$X^{\mathrm{T}} X = \begin{bmatrix} X_{*1}^{\mathrm{T}} \\ X_{*2}^{\mathrm{T}} \\ \vdots \\ X_{*k}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} X_{*1} & X_{*2} & \cdots & X_{*k} \end{bmatrix} = \begin{bmatrix} X_{*1}^{\mathrm{T}} X_{*1} & X_{*1}^{\mathrm{T}} X_{*2} & \cdots & X_{*1}^{\mathrm{T}} X_{*k} \\ X_{*2}^{\mathrm{T}} X_{*1} & X_{*2}^{\mathrm{T}} X_{*2} & \cdots & X_{*2}^{\mathrm{T}} X_{*k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{*k}^{\mathrm{T}} X_{*1} & X_{*k}^{\mathrm{T}} X_{*2} & \cdots & X_{*k}^{\mathrm{T}} X_{*k} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{N} x_{i1}^2 & \sum_{i=1}^{N} x_{i1} x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1} x_{ik} \\ \sum_{i=1}^{N} x_{i2} x_{i1} & \sum_{i=1}^{N} x_{i2}^2 & \cdots & \sum_{i=1}^{N} x_{i2} x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} x_{ik} x_{i1} & \sum_{i=1}^{N} x_{ik} x_{i2} & \cdots & \sum_{i=1}^{N} x_{ik}^2 \end{bmatrix}$$

Partitioned Matrices

If we partition the data matrix X into rows, i.e.,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} = \begin{bmatrix} X_{1*} \\ X_{2*} \\ \vdots \\ X_{n*} \end{bmatrix}$$

where X_{j*} is the row vector containing the jth obs of all variables

then

$$X^{\mathrm{T}} X = \begin{bmatrix} X_{1*}^{\mathrm{T}} & X_{2*}^{\mathrm{T}} & \cdots & X_{n*}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} X_{1*} \\ X_{2*} \\ \vdots \\ X_{n*} \end{bmatrix}$$

$$= X_{1*}^{\mathrm{T}} X_{1*} + X_{2*}^{\mathrm{T}} x_{2*} + \cdots + X_{n*}^{\mathrm{T}} X_{n*} = \underbrace{\sum_{i=1}^{n} X_{i*}^{\mathrm{T}} X_{i*}}_{\text{sum of } n\ k \times k \text{ matrices}}$$

Partitioned Matrices

If $A = \begin{bmatrix} \underbrace{A_{11}}_{m_1 \times m_1} & \underbrace{A_{12}}_{m_1 \times m_2} \\ \underbrace{A_{21}}_{m_2 \times m_1} & \underbrace{A_{22}}_{m_2 \times m_2} \end{bmatrix}$ and non-singular, then

$$A^{-1} = \begin{bmatrix} (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} & -(A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} A_{12} A_{22}^{-1} \\ -A_{22}^{-1} A_{21} (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} & A_{22}^{-1} + A_{22}^{-1} A_{21} (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} A_{12} A_{22}^{-1} \end{bmatrix}$$

You can verify this by direct multiplication, to show that

$$A^{-1} A = \begin{bmatrix} I_{m_1} & 0_{m_1 \times m_2} \\ 0_{m_2 \times m_1} & I_{m_2} \end{bmatrix}$$

Partitioned Matrices

Yet another type of matrix product is the *Kronecker product*

Kronecker product, denoted $\otimes$, of an $m \times n$ matrix A with a $p \times q$ matrix B is the $mp \times nq$ block matrix formed by multiplying each element of A by the entire B matrix

For example

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \left[\begin{array}{cc|cc|cc} a_{11} & 0 & a_{12} & 0 & a_{13} & 0 \\ 0 & a_{11} & 0 & a_{12} & 0 & a_{13} \\ \hline a_{11} & 0 & a_{12} & 0 & a_{13} & 0 \\ 0 & a_{11} & 0 & a_{12} & 0 & a_{13} \end{array}\right]$$

Session 4.5

Session 4.5 Vectors of Random Variables

- Expectations
- Variance-covariance matrices

Vectors and Matrices of Random Variables

Matrix algebra helps in organizing large numbers of random variable, especially their expectations and variances and covariances

If x is a $m \times 1$ vector of random variables $x = [X_1 \quad X_2 \quad \dots \quad X_m]^{\mathrm{T}}$, then we **define**

$$E(x) = [E(X_1) \quad E(X_2) \quad \dots \quad E(X_m)]^{\mathrm{T}}$$

If X is a matrix $m \times n$ matrix of random variables, then

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} \Leftrightarrow E(X) = \begin{bmatrix} E(X_{11}) & E(X_{12}) & \dots & E(X_{1n}) \\ E(X_{21}) & E(X_{22}) & \dots & E(X_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{m1}) & E(X_{m2}) & \dots & E(X_{mn}) \end{bmatrix}$$

Vectors and Matrices of Random Variables

Let x be a $m \times 1$ vector of random variables. Let

$$\tilde{x} = x - E(x) = \begin{bmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \\ \vdots \\ X_m - E(X_m) \end{bmatrix} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_m \end{bmatrix}$$

Then the **variance-covariance matrix** of x, denoted $Var(x)$, is defined as

$$Var(x) = E((x - E(x))(x - E(x))^{\mathrm{T}}) = E(\tilde{x}\tilde{x}^{\mathrm{T}})$$

Vectors and Matrices of Random Variables

$$\begin{aligned} Var(x) &= E(\tilde{x}\tilde{x}^{\mathrm{T}}) = E((x - E[x])(x - E[x])^{\mathrm{T}}) \\ &= E\begin{bmatrix} \tilde{X}_1^2 & \tilde{X}_1\tilde{X}_2 & \dots & \tilde{X}_1\tilde{X}_m \\ \tilde{X}_2\tilde{X}_1 & \tilde{X}_2^2 & \dots & \tilde{X}_2\tilde{X}_m \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{X}_m\tilde{X}_1 & \tilde{X}_m\tilde{X}_2 & \dots & \tilde{X}_m\tilde{X}_m \end{bmatrix} \\ &= \begin{bmatrix} E(\tilde{X}_1^2) & E(\tilde{X}_1\tilde{X}_2) & \dots & E(\tilde{X}_1\tilde{X}_m) \\ E(\tilde{X}_2\tilde{X}_1) & E(\tilde{X}_2^2) & \dots & E(\tilde{X}_2\tilde{X}_m) \\ \vdots & \vdots & \ddots & \vdots \\ E(\tilde{X}_m\tilde{X}_1) & E(\tilde{X}_m\tilde{X}_2) & \dots & E(\tilde{X}_m\tilde{X}_m) \end{bmatrix} \\ &= \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_m) \\ Cov(X_1, X_2) & Var(X_2) & \dots & Cov(X_2, X_m) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_1, X_m) & Cov(X_2, X_m) & \dots & Var(X_m) \end{bmatrix} \end{aligned}$$

Vectors and Matrices of Random Variables

Recall that if X is a (univariate) random variable, then

- $E(aX+b) = aE(X)+b$
- $Var(aX+b) = a^2\,Var(X)$
- $Var(X) = E(X^2) - E(X)^2$

We have matrix analogues of these results: Suppose x is an $m\times 1$ vector of random variables, $A=(a_{ij})_{km}$ is a $k\times m$ matrix of constants and b is a $k\times 1$ vector of constants. Then

- $E(Ax+b) = AE(x)+b$
- $Var(Ax+b) = A\,Var(x)A^{\mathrm{T}}$
- $Var(x) = E(xx^{\mathrm{T}}) - E(x)E(x)^{\mathrm{T}}$

Vectors and Matrices of Random Variables

Proof of $E(Ax+b) = AE(x)+b$:

The ith element of the $k\times 1$ vector $Ax+b$ is $\sum_{j=1}^{m}(a_{ij}X_j+b_i)$. The expectation of this term is

$$E\left(\sum_{j=1}^{m}(a_{ij}X_j+b_i)\right) = \sum_{j=1}^{m} a_{ij}E(X_j)+b_i$$

which is the ith element of the vector $AE(x)+b$.

Proof of $Var(Ax+b) = A\,Var(x)A^{\mathrm{T}}$:

Since $Ax+b-E(Ax+b) = A(x-E(x)) = A\tilde{x}$, we have

$$\begin{aligned} Var(Ax+b) &= E((A\tilde{x})(A\tilde{x})^{\mathrm{T}}) = E(A\tilde{x}\tilde{x}^{\mathrm{T}}A^{\mathrm{T}}) = AE(\tilde{x}\tilde{x}^{\mathrm{T}})A^{\mathrm{T}} \\ &= A\;Var(x)A^{\mathrm{T}}\,. \end{aligned}$$

Proof of $Var(x) = E(xx^{\mathrm{T}}) - E(x)E(x)^{\mathrm{T}}$: Exercise!

Vectors and Matrices of Random Variables

Let x be $m\times 1$ vector of random variables, c be $m\times 1$ non-zero vector of constants

- Obviously the variance-covariance matrix of x is symmetric
- Consider the linear combination $c^{\mathrm{T}}x$ (this is a now a single random variable). We have
$$Var(c^{\mathrm{T}}x) = c^{\mathrm{T}}\,Var(x)c \geq 0 \;\text{ for all }\; c\neq 0_{m\times 1}$$
- $Var(c^{\mathrm{T}}x)$ cannot be negative since it is a variance
- If $Var(c^{\mathrm{T}}x)=0$ then either
  - one of the random variables is not actually random, or
  - one of the random variables is just a multiple of the other
  - one of the random variables is a linear combination of two or more of the other random variables

Digression on Symmetric Matrices

A $m\times m$ symmetric (and square) matrix A is **positive definite** if

$$c^{\mathrm{T}}Ac > 0 \;\text{ for all }\; c\neq 0_{m\times 1}$$

It is **positive semidefinite** if $c^{\mathrm{T}}Ac \geq 0$ for all $c\neq 0_{m\times 1}$. Similar definitions for **negative definiteness** and **negative semidefiniteness**

- Variance covariance matrices $Var(x)$ are **positive semidefinite**
- If the random variables in x are not linearly dependent, then $Var(x)$ is **positive definite**
- Another example: suppose the columns of a $n\times k$ data matrix X are linearly independent, i.e.,
$$Xc\neq 0_{n\times 1} \;\text{ for all }\; c\neq 0_{k\times 1}$$
Then $c^{\mathrm{T}}X^{\mathrm{T}}Xc = (Xc)^{\mathrm{T}}Xc > 0$ for all $c\neq 0_{k\times 1}$, i.e., $X^{\mathrm{T}}X$ is positive definite

Session 4.6

Session 4.6 Principal Component Analysis

- Eigendecomposition of symmetric matrices, without proofs
- Application to Principal Component Analysis

No discussion of eigenvalues or eigenvectors, see readings if interested

Eigendecomposition of Symmetric Matrices

Another very important fact about symmetric matrices (Eigendecomposition)

Every $k \times k$ symmetric matrix A can be decomposed in the following way

$$A = Q\Lambda Q^{\mathrm{T}} = \begin{bmatrix} q_1 & q_2 & \cdots & q_k \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}} \\ q_2^{\mathrm{T}} \\ \vdots \\ q_k^{\mathrm{T}} \end{bmatrix} = \sum_{i=1}^{k} \lambda_i q_i q_i^{\mathrm{T}}$$

- λ_i, $i = 1, \dots, k$ are real numbers called **eigenvalues** (usually ranked $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$)
- The $k \times 1$ vectors q_i, $i = 1, \dots, k$ are the corresponding **eigenvectors**
- Q satisfies the property $Q^{\mathrm{T}}Q = I_k$

(see BPT Chapter 10)

Sample Variance-Covariance Matrix

Let X be a $n \times k$ data matrix, each column contains n observations of some variable, mean removed. Then

$$X = \begin{bmatrix} x_{11} - \overline{x_1} & x_{12} - \overline{x_2} & \cdots & x_{1k} - \overline{x_k} \\ x_{21} - \overline{x_1} & x_{22} - \overline{x_2} & \cdots & x_{2k} - \overline{x_k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \overline{x_1} & x_{n2} - \overline{x_2} & \cdots & x_{nk} - \overline{x_k} \end{bmatrix} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1k} \\ \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{n1} & \tilde{x}_{n2} & \cdots & \tilde{x}_{nk} \end{bmatrix}$$

and $\dfrac{1}{n-1}X^{\mathrm{T}}X$ is the symmetric sample variance-covariance matrix

$$\frac{1}{n-1}X^{\mathrm{T}}X = \begin{bmatrix} \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i1}^2 & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i1}\tilde{x}_{i2} & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i1}\tilde{x}_{ik} \\ \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i2}\tilde{x}_{i1} & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i2}^2 & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i2}\tilde{x}_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{ik}\tilde{x}_{i1} & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{ik}\tilde{x}_{i2} & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{ik}^2 \end{bmatrix}$$

Principal Components

Apply eigendecomposition to the sample variance-covariance matrix

$$\frac{1}{n-1}X^{\mathrm{T}}X = Q\Lambda Q^{\mathrm{T}}$$

Construct the following data matrix

$$Y = XQ = X\begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = \begin{bmatrix} Xq_1 & Xq_2 & \cdots & Xq_n \end{bmatrix}$$

Each $n \times 1$ vector Xq_i, $i = 1, \dots, k$ is an "index variable" formed by a linear combination of the k X variables

$$Xq_i = \begin{bmatrix} X_{*1} & X_{*2} & \cdots & X_{*k} \end{bmatrix} \begin{bmatrix} q_{1i} \\ q_{2i} \\ \vdots \\ q_{ki} \end{bmatrix} = q_{1i}X_{*1} + q_{2i}X_{*2} + \cdots + q_{ki}X_{*k}$$

Principal Component Analysis

The sample variance-covariance matrix of Y is

$$\frac{1}{n-1}Y^{\mathrm{T}}Y = \frac{1}{n-1}Q^{\mathrm{T}}X^{\mathrm{T}}XQ = Q^{\mathrm{T}}\left(\frac{1}{n-1}X^{\mathrm{T}}X\right)Q = Q^{\mathrm{T}}Q\Lambda Q^{\mathrm{T}}Q = \Lambda$$

That is, the $n \times k$ matrix Y contain observations of k **uncorrelated** variables

- first column has obs. of the first index variable, which has the greatest variance
- second column has obs. of the second index variable, which has the second highest variance

These index variables are called **principal components**

Often first two or three indexes account for most of the variance in the data – dimension reduction

Principal Component Analysis

`causes-of-death-by-state.csv`

- 51×14 data matrix
- age-adjusted number of deaths per 100,000, all races, both sexes, all ages, over the period 2016-2020
- across the 51 US states plus District of Columbia (rows)
- 14 causes of death (columns): accidents & adverse effects (accident), Alzheimer's disease (Alzheimers), cancer, cerebrovascular diseases (cerebrovascular), chronic lower respiratory disease (respiratory), chronic liver disease & cirrhosis (liver), diabetes mellitus (diabetes), heart disease (heart), homicide & legal intervention (homicide), influenza, kidney disease - nephritis & nephrosis (kidney), pneumonia, septicemia, suicide & self-inflicted injury (suicide).

Qn: How do states differ by cause of death?

Principal Component Analysis

```r
library(tidyverse)
library(ggrepel)
df <- read.csv("data/causes-of-death-by-state.csv")
row.names(df) <- df[,1]
df <- df[,-1]
head(df, 4) # Show data for first four states
```

```
         Accident Alzheimers Cancer Cerebrovascular Respiratory Liver Diabetes
Alabama      55.2       46.1  166.9            51.8        56.0  13.6     20.5
Alaska       62.3       26.0  146.8            37.1        35.0  16.7     19.7
Arizona      58.6       33.8  132.3            30.9        40.2  14.9     23.9
Arkansas     51.6       40.9  169.7            42.9        61.6  12.7     30.8
         Heart Homicide Influenza Kidney Pneumonia Septicemia Suicide
Alabama  225.1     12.9       1.5   16.9      17.0       17.0    16.2
Alaska   136.1      9.4       2.1   10.0       8.2        8.3    26.5
Arizona  139.1      6.7       1.6    7.1       9.4        4.8    18.2
Arkansas 222.2     10.4       2.2   18.6      15.7       12.9    18.9
```


Principal Component Analysis

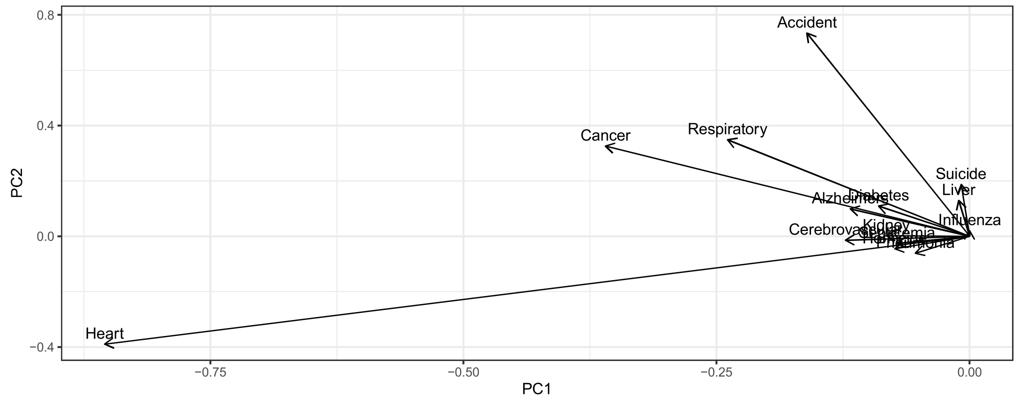
```r
# Importance of each PC
dfs <- scale(df, scale=FALSE)  # remove the mean but don't standardize
dfpca1 <- prcomp(dfs)
summary(dfpca1)
```

```
Importance of components:
                           PC1     PC2     PC3     PC4     PC5     PC6     PC7
Standard deviation     32.8587 12.5414 9.57876 7.44281 6.38255 4.53854 3.63309
Proportion of Variance  0.7258  0.1057 0.06168 0.03724 0.02738 0.01385 0.00887
Cumulative Proportion   0.7258  0.8315 0.89319 0.93043 0.95781 0.97166 0.98053
                           PC8     PC9    PC10    PC11    PC12    PC13    PC14
Standard deviation     2.89666 2.50515 2.28355 2.17946 1.54667 1.37289 0.24006
Proportion of Variance 0.00564 0.00422 0.00351 0.00319 0.00161 0.00127 0.00004
Cumulative Proportion  0.98617 0.99039 0.99389 0.99709 0.99869 0.99996 1.00000
```

Principal Component Analysis

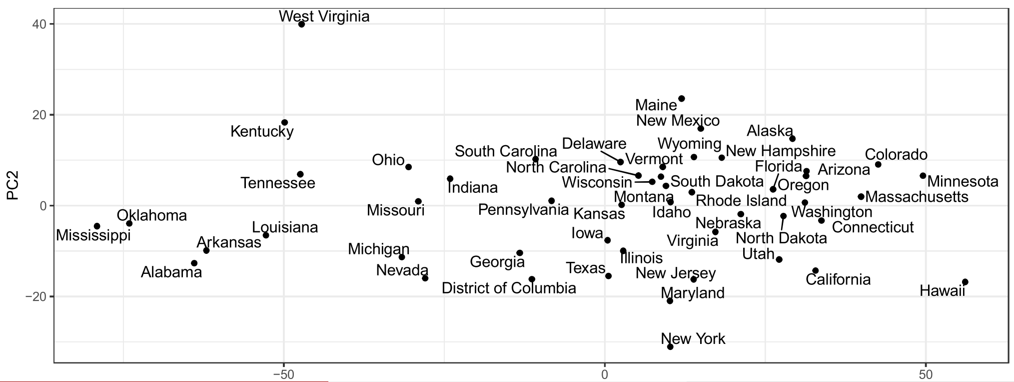
```
# Cause of Death Loading for First Two PCs

pc1and2 <- as.data.frame(dfpca1$x[,1:2])  # Collect first two PCs into a data fram

loadings1and2 <- data.frame(xstart = 0,                # A data frame containing the
                            ystart = 0,                # loadings (weights) placed on each
                            PC1 = dfpca1$rotation[,1], # cause of death in the first two
                            PC2 = dfpca1$rotation[,2]) # principal components

p1 <- ggplot(loadings1and2,
         aes(x = xstart, y = ystart, xend = PC1, yend = PC2)) +
      geom_segment(arrow = arrow(length=unit(0.1, "inches"))) + ylab("PC2") + xlab("PC1") +
      annotate("text", x=loadings1and2$PC1, y=loadings1and2$PC2+0.04,
            label=rownames(loadings1and2), size=4) +
      theme_bw()
```


Principal Component Analysis



Principal Component Analysis

```
# Each States PC1 and PC2 scores
ggplot(as.data.frame(pc1and2), aes(PC1, PC2, label = rownames(pc1and2))) +
  geom_point() + geom_text_repel(size=4, box.padding = 0.1) + theme_bw()
```



Roadmap

- (Previous) Session 1: Statistics Review
- (Previous) Session 2: Simple Linear Regression
- (Previous) Session 3: Estimator Standard Errors; Multiple Linear Regression
- **This Session 4: Matrix Algebra**
- *Next Session 5: OLS using Matrix Algebra*
- Session 6: Hypothesis Testing
- Session 7: Prediction
- Session 8: Instrumental Variable Regression
- Session 9: Logistic and Other Regressions
- Session 10: Panel Data Regressions
- Session 11: Introduction to Time Series
- Session 12: Time Series Regressions