Math Camp

Justin Grimmer

Associate Professor
Department of Political Science
Stanford University

September 4th, 2015
Where We’ve Been, Where We’re Going

Calculus: Analyze behavior of functions on real line
- Convergence
- Differentiation
- Integration

Linear Algebra
- Data stored in matrices
- Higher dimensional spaces
  - complex world, condition on many factors
  - flood of big data, store in many dimensions
- Linear Algebra:
  - Algebra of matrices
  - Geometry of high dimensional space
  - Calculus (multivariable) in many dimensions

Very important for regression(!!!!)
Points + Vectors

- A point in $\mathbb{R}^1$
  - 1
  - $\pi$
  - $e$

- An ordered pair in $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$
  - (1, 2)
  - (0, 0)
  - ($\pi$, $e$)

- An ordered triple in $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$
  - (3.1, 4.5, 6.11132)

...:

- An ordered n-tuple in $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}$
  - ($a_1, a_2, \ldots, a_n$)
Points and Vectors

Definition
A point $\mathbf{x} \in \mathbb{R}^n$ is an ordered $n$-tuple, $(x_1, x_2, \ldots, x_n)$. The vector $\mathbf{x} \in \mathbb{R}^n$ is the arrow pointing from the origin $(0, 0, \ldots, 0)$ to $\mathbf{x}$. 
One Dimensional Example
One Dimensional Example
One Dimensional Example
Two Dimensional Example
Two Dimensional Example
Two Dimensional Example
Three Dimensional Example

- \((\text{Latitude, Longitude, Elevation})\)
- \((1, 2, 3)\)
- \((0, 1, 0)\)
N-Dimensional Example

- Individual campaign donation records

\[ x = (1000, 0, 10, 50, 15, 4, 0, 0, 0, \ldots, 2400000000) \]

- Counties have proportion of vote for Obama

\[ y = (0.8, 0.5, 0.6, \ldots, 0.2) \]

- Run experiment, assess feeling thermometer of elected official

\[ t = (0, 100, 50, 70, 80, \ldots, 100) \]
Arithmetic with Vectors

Definition

Suppose \( u \) and \( v \) are vectors \( u, v \in \mathbb{R}^n \),

\[
\begin{align*}
    u &= (u_1, u_2, u_3, \ldots, u_n) \\
    v &= (v_1, v_2, v_3, \ldots, v_n)
\end{align*}
\]

We will say \( u = v \) if \( u_1 = v_1, u_2 = v_2, \ldots, u_n = v_n \)

Define the sum of \( u + v \) as

\[
    u + v = (u_1 + v_1, u_2 + v_2, u_3 + v_3, \ldots, u_n + v_n)
\]

Suppose \( k \in \mathbb{R} \). We will call \( k \) a scalar.

Define \( ku \) as the scalar product

\[
    ku = (ku_1, ku_2, \ldots, ku_n)
\]
Examples

Suppose:

\[
\begin{align*}
\mathbf{u} & = (1, 2, 3, 4, 5) \\
\mathbf{v} & = (1, 1, 1, 1, 1) \\
k & = 2
\end{align*}
\]

Then,

\[
\begin{align*}
\mathbf{u} + \mathbf{v} & = (1 + 1, 2 + 1, 3 + 1, 4 + 1, 5 + 1) = (2, 3, 4, 5, 6) \\
k \mathbf{u} & = (2 \times 1, 2 \times 2, 2 \times 3, 2 \times 4, 2 \times 5) = (2, 4, 6, 8, 10) \\
k \mathbf{v} & = (2 \times 1, 2 \times 1, 2 \times 1, 2 \times 1, 2 \times 1) = (2, 2, 2, 2, 2)
\end{align*}
\]
Properties of Arithmetic

Challenge Proofs—we can do these!

Theorem

Suppose that $u, v, w \in \mathbb{R}^n$ and $k$ and $l$ are scalars.

a) $u + v = v + u$
Properties of Arithmetic

Challenge Proofs—we can do these!

Theorem

Suppose that \( u, v, w \in \mathbb{R}^n \) and \( k \) and \( l \) are scalars.

a) \( u + v = v + u \)

Proof.

\[
\begin{align*}
  u + v & = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n) \\
 & = (v_1 + u_1, v_2 + u_2, \ldots, v_n + u_n) \\
 & = v + u
\end{align*}
\]
Properties of Arithmetic

Challenge Proofs—we can do these!

Theorem

Suppose that $u, v, w \in \mathbb{R}^n$ and $k$ and $l$ are scalars.

b) $u + 0 = 0 + u = u$
Properties of Arithmetic

Challenge Proofs—we can do these!

Theorem

Suppose that $u, v, w \in \mathbb{R}^n$ and $k$ and $l$ are scalars.

b) $u + 0 = 0 + u = u$

Proof.

\[
\begin{align*}
  u + 0 &= (u_1 + 0, u_2 + 0, \ldots, u_n + 0) \\
  &= (0 + u_1, 0 + u_2, \ldots, 0 + u_n) = 0 + u \\
  &= (u_1, u_2, \ldots, u_n) \\
  &= u
\end{align*}
\]
Properties of Arithmetic

Challenge Proofs—we can do these!

Theorem

Suppose that $u, v, w \in \mathbb{R}^n$ and $k$ and $l$ are scalars.

c) $(l + k)u = l(u) + k(u)$

Proof.

How can we show this?
Challenge Proofs

- Show that $1u = u$
- Show that $u + -1u = 0$
Inner Product

Definition

Suppose \( \mathbf{u} \in \mathbb{R}^n \) and \( \mathbf{v} \in \mathbb{R}^n \) then define \( \mathbf{u} \cdot \mathbf{v} \),

\[
\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n
\]

\[
= \sum_{i=1}^{N} u_i v_i
\]
Examples

Suppose \( \mathbf{u} = (1, 2, 3) \) and \( \mathbf{v} = (2, 3, 1) \). Then,

\[
\mathbf{u} \cdot \mathbf{v} = 1 \times 2 + 2 \times 3 + 3 \times 1 = 2 + 6 + 3 = 11
\]

Suppose \( \mathbf{y} = (y_1, y_2, \ldots, y_N) \) and \( \mathbf{1} = (1, 1, 1, \ldots, 1) \). Then,

\[
\mathbf{y} \cdot \mathbf{1} = y_1 + y_2 + \ldots + y_n = \sum_{i=1}^{n} y_i
\]
R Code

Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec <- c()
vec[1] <- 1
vec[3] <- 3
vec[4] <- 4
x1 <- c(2, 2, 3, 2)
x2 <- c(5, 3, 1, 3)
add <- x1 + x2
add

[1] 7 5 4 5

scalar <- 10 * x1
scalar

[1] 20 20 30 20

output <- x1 %*% x2
output

[,1]
[1,] 25
```
Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
```

```r
x1 <- c(2, 2, 3, 2)
x2 <- c(5, 3, 1, 3)
add <- x1 + x2
add
```

```r
scalar <- 10 * x1
scalar
```

```r
output <- x1 %*% x2
output
```

Justin Grimmer (Stanford University)
R Code

Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec <- c()
```

R Code

Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec<- c()
vec[1]<- 1
```

```r
x1<- c(2, 2, 3, 2)
x2<- c(5, 3, 1, 3)
add <- x1 + x2
add
```
```
[1] 7 5 4 5
```

```r
scalar<- 10 *x1
scalar
```
```
[1] 20 20 30 20
```

```r
output<- x1 %*% x2
output
```
```
[,1]
[1,] 25
```

R Code

Create a vector in R
vec <- c(1, 2, 3, 4, 5)
vec <- c()
vec[1] <- 1

add <- x1 + x2
add
[1] 7 5 4 5

scalar <- 10 * x1
scalar
[1] 20 20 30 20

output <- x1 %*% x2
output
[,1]
[1,] 25
Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec<- c()
vec[1]<- 1
vec[3]<- 3
```

```r
x1<- c(2, 2, 3, 2)
x2<- c(5, 3, 1, 3)
add <- x1 + x2
add
```

```r
[1]  7  5  4  5
```

```r
scalar<- 10 *x1
scalar
```

```r
[1] 20 20 30 20
```

```r
output<- x1 %*% x2
output
```

```r
[,1]
[1,] 25
```

Justin Grimmer (Stanford University)
Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec<- c()
vec[1]<- 1
vec[3]<- 3
vec[4]<- 4
```

```r
x1<- c(2, 2, 3, 2)
x2<- c(5, 3, 1, 3)
add <- x1 + x2
add
```

```r
[1] 7 5 4 5
```

```r
scalar<- 10 *x1
scalar
```

```r
[1] 20 20 30 20
```

```r
output<- x1 %*% x2
output
```

```r
[,1]
[1,] 25
```
R Code

Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec<- c()
vec[1]<- 1
vec[3]<- 3
vec[4]<- 4
```

```
x1<- c(2, 2, 3, 2)
x2<- c(5, 3, 1, 3)
add <- x1 + x2
add
```

```
[1]  7  5  4  5
```

```
scalar<- 10 *x1
scalar
```

```
[1] 20 20 30 20
```

```
output<- x1 %*% x2
output
```

```
[,1]
[1,,] 25
```
R Code

Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec <- c()
vec[1] <- 1
vec[3] <- 3
vec[4] <- 4
x1 <- c(2, 2, 3, 2)
```

```
x2 <- c(5, 3, 1, 3)
add <- x1 + x2
add
[1] 7 5 4 5
```

```
scalar <- 10 * x1
scalar
[1] 20 20 30 20
```

```
output <- x1 %*% x2
output
[,1]
[1,] 25
```

Justin Grimmer (Stanford University)
R Code

Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec<- c()
vec[1]<- 1
vec[3]<- 3
vec[4]<- 4
x1<- c(2, 2, 3, 2)
x2<- c(5, 3, 1, 3)
```

```r
add <- x1 + x2
add
[1] 7 5 4 5
```

```r
scalar<- 10 * x1
scalar
[1] 20 20 30 20
```

```r
output<- x1 %*% x2
output
[,1]
[1,] 25
```
R Code

Create a vector in R

```
vec <- c(1, 2, 3, 4, 5)
vec <- c()
vec[1] <- 1
vec[3] <- 3
vec[4] <- 4
x1 <- c(2, 2, 3, 2)
x2 <- c(5, 3, 1, 3)
add <- x1 + x2
```
Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec <- c()
vec[1] <- 1
vec[3] <- 3
vec[4] <- 4

x1 <- c(2, 2, 3, 2)
x2 <- c(5, 3, 1, 3)
add <- x1 + x2
add
```

```
[1] 7 5 4 5
```
R Code

Create a vector in R
vec <- c(1, 2, 3, 4, 5)
vec <- c()
vec[1] <- 1
vec[3] <- 3
vec[4] <- 4
x1 <- c(2, 2, 3, 2)
x2 <- c(5, 3, 1, 3)
add <- x1 + x2
add
[1] 7 5 4 5

scalar <- 10 * x1

output <- x1 %*% x2
output[,1]
[,1] 25
Create a vector in R
\begin{verbatim}
vec <- c(1, 2, 3, 4, 5)
vec<- c()
vec[1]<- 1
vec[3]<- 3
vec[4]<- 4
x1<- c(2, 2, 3, 2)
x2<- c(5, 3, 1, 3)
add <- x1 + x2
\end{verbatim}
\begin{verbatim}
add
[1] 7 5 4 5
\end{verbatim}
scalar<- 10 *x1
\begin{verbatim}
scalar
[1] 20 20 30 20
\end{verbatim}
R Code

Create a vector in R
vec <- c(1, 2, 3, 4, 5)
vec<- c()
vec[1]<- 1
vec[3]<- 3
vec[4]<- 4
x1<- c(2, 2, 3, 2)
x2<- c(5, 3, 1, 3)
add <- x1 + x2
add
[1] 7 5 4 5

scalar<- 10 *x1
scalar
[1] 20 20 30 20
output<- x1 %*% x2
output
[,1]
[1,] 25
R Code

Create a vector in R

```r
vec <- c(1, 2, 3, 4, 5)
vec <- c()
vec[1] <- 1
vec[3] <- 3
vec[4] <- 4
x1 <- c(2, 2, 3, 2)
x2 <- c(5, 3, 1, 3)
add <- x1 + x2
```

```
[1] 7 5 4 5
```

```r
scalar <- 10 * x1
```

```
[1] 20 20 30 20
```

```r
output <- x1 %*% x2
```

```
[,1]
[1,] 25
```
Challenge Problems

- Suppose $\mathbf{v} = (1, 4, 1, 4)$ and $\mathbf{w} = (4, 1, 4, 1)$. Calculate: $\mathbf{v} \cdot \mathbf{w}$
- Prove $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$
- Super hard: prove $\mathbf{v} \cdot \mathbf{v} \geq 0$ and $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$. 
Vector Length

Pythagorean Theorem:
- Side with length $a$
- Side with length $b$
- Right triangle $c = \sqrt{a^2 + b^2}$

This is generally true.
Vector Length

- Pythagorean Theorem:
  Side with length \( a \)
Vector Length

- Pythagorean Theorem:
  Side with length $a$
- Side with length $b$ and right triangle
Vector Length

- Pythagorean Theorem:
- Side with length $a$
- Side with length $b$ and right triangle
- $c = \sqrt{a^2 + b^2}$
Vector Length

- **Pythagorean Theorem:**
  - Side with length $a$
  - Side with length $b$ and right triangle
  - $c = \sqrt{a^2 + b^2}$
  - This is generally true
Vector Length

Definition

Suppose \( \mathbf{v} \in \mathbb{R}^n \). Then, we will define its length as

\[
\| \mathbf{v} \| = (\mathbf{v} \cdot \mathbf{v})^{1/2} = (v_1^2 + v_2^2 + v_3^2 + \ldots + v_n^2)^{1/2}
\]
Calculating a Length

Example 1: suppose \( \mathbf{x} = (1, 1, 1) \).

\[
\|\mathbf{x}\| = (\mathbf{x} \cdot \mathbf{x})^{1/2} = (1 + 1 + 1)^{1/2} = \sqrt{3}
\]

Example 2: R code for length function

```r
len.vec <- function(x) {
  p1 <- sqrt(x %*% t(x))
  return(p1)
}
x <- c(1, 1, 1)
len.vec(x)
```

\[,1\]

\[1,] 1.732051
Coding Problem

Let’s calculate the length of some vectors

- Write a function to assess the length of a vector.
- Use it to calculate the length of:
  - y <- c(10, 20, 30, 40)
  - x <- seq(1, 1000*pi, len=1000)
Texts in Space

\[ \text{Doc1} = (1, 1, 3, \ldots, 5) \]

\[ \text{Doc2} = (2, 0, 0, \ldots, 1) \]

\[ \text{Doc1}, \text{Doc2} \in \mathbb{R}^M \]

Provides many operations that will be useful.

Inner Product between documents:

\[ \text{Doc1} \cdot \text{Doc2} = (1 \times 2 + 1 \times 0 + 3 \times 0 + \ldots + 5 \times 1) = 7 \]
Texts in Space

$\text{Doc}_1 = (1, 1, 3, \ldots, 5)$
Texts in Space

\[ \begin{align*}
\text{Doc1} & = (1, 1, 3, \ldots, 5) \\
\text{Doc2} & = (2, 0, 0, \ldots, 1)
\end{align*} \]
Texts in Space

\[ \text{Doc1} = (1, 1, 3, \ldots, 5) \]
\[ \text{Doc2} = (2, 0, 0, \ldots, 1) \]
\[ \text{Doc1, Doc2} \in \mathbb{R}^M \]
Texts in Space

\[ \text{Doc1} = (1, 1, 3, \ldots, 5) \]
\[ \text{Doc2} = (2, 0, 0, \ldots, 1) \]

\[ \text{Doc1, Doc2} \in \mathbb{R}^M \]

Provides many operations that will be useful
Texts in Space

\[ \text{Doc1} = (1, 1, 3, \ldots, 5) \]
\[ \text{Doc2} = (2, 0, 0, \ldots, 1) \]
\[ \text{Doc1, Doc2} \in \mathbb{R}^M \]

Provides many operations that will be useful

**Inner Product** between documents:
Texts in Space

\[
\begin{align*}
\text{Doc}1 & = (1, 1, 3, \ldots, 5) \\
\text{Doc}2 & = (2, 0, 0, \ldots, 1)
\end{align*}
\]

\[
\text{Doc}1, \text{Doc}2 \in \mathbb{R}^M
\]

Provides many operations that will be useful

**Inner Product** between documents:

\[
\text{Doc}1 \cdot \text{Doc}2 = (1, 1, 3, \ldots, 5)'(2, 0, 0, \ldots, 1)
\]
Texts in Space

\[
\text{Doc1} = (1, 1, 3, \ldots, 5) \\
\text{Doc2} = (2, 0, 0, \ldots, 1)
\]

\[\text{Doc1, Doc2} \in \mathbb{R}^M\]

Provides many operations that will be useful

**Inner Product** between documents:

\[
\text{Doc1} \cdot \text{Doc2} = (1, 1, 3, \ldots, 5)'(2, 0, 0, \ldots, 1) \\
= 1 \times 2 + 1 \times 0 + 3 \times 0 + \ldots + 5 \times 1
\]
Texts in Space

\[ \text{Doc1} = (1, 1, 3, \ldots, 5) \]
\[ \text{Doc2} = (2, 0, 0, \ldots, 1) \]
\[ \text{Doc1, Doc2} \in \mathbb{R}^M \]

Provides many operations that will be useful

**Inner Product** between documents:

\[ \text{Doc1} \cdot \text{Doc2} = (1, 1, 3, \ldots, 5) (2, 0, 0, \ldots, 1) \]
\[ = 1 \times 2 + 1 \times 0 + 3 \times 0 + \ldots + 5 \times 1 \]
\[ = 7 \]
Length of document:
Length of document:

\[ \| \text{Doc1} \| = \sqrt{\text{Doc1} \cdot \text{Doc1}} \]
\[ = \sqrt{(1, 1, 3, \ldots, 5)'(1, 1, 3, \ldots, 5)} \]
\[ = \sqrt{1^2 + 1^2 + 3^2 + 5^2} \]
\[ = 6 \]
Length of document:

\[ \|\text{Doc1}\| \equiv \sqrt{\text{Doc1} \cdot \text{Doc1}} \]
\[ = \sqrt{(1, 1, 3, \ldots, 5)'(1, 1, 3, \ldots, 5)} \]
\[ = \sqrt{1^2 + 1^2 + 3^2 + 5^2} \]
\[ = 6 \]

Cosine of the angle between documents:
Length of document:

\[ \|\text{Doc1}\| \equiv \sqrt{\text{Doc1} \cdot \text{Doc1}} \]
\[ = \sqrt{(1, 1, 3, \ldots, 5)'(1, 1, 3, \ldots, 5)} \]
\[ = \sqrt{1^2 + 1^2 + 3^2 + 5^2} \]
\[ = 6 \]

Cosine of the angle between documents:

\[ \cos \theta \equiv \left( \frac{\text{Doc1}}{\|\text{Doc1}\|} \right) \cdot \left( \frac{\text{Doc2}}{\|\text{Doc2}\|} \right) \]
\[ = \frac{7}{6 \times 2.24} \]
\[ = 0.52 \]
Measuring Similarity

Documents in space → measure similarity/dissimilarity

- Maximum: document with itself
- Minimum: documents have no words in common (orthogonal)
- Increasing when more of same words used

\[ s(a, b) = s(b, a). \]
Measuring Similarity

Documents in space → measure similarity/dissimilarity

What properties should similarity measure have?

- Maximum: document with itself
- Minimum: documents have no words in common (orthogonal)
- Increasing when more of same words used
- \( s(a, b) = s(b, a) \)
Measuring Similarity

Documents in space → measure similarity/dissimilarity
What properties should similarity measure have?
  - Maximum: document with itself
Measuring Similarity

Documents in space $\rightarrow$ measure similarity/dissimilarity

What properties should similarity measure have?

- Maximum: document with itself
- Minimum: documents have no words in common (orthogonal)
Measuring Similarity

Documents in space → measure similarity/dissimilarity

What properties should similarity measure have?
- Maximum: document with itself
- Minimum: documents have no words in common (orthogonal)
- Increasing when more of same words used
Measuring Similarity

Documents in space \( \rightarrow \) measure similarity/dissimilarity

What properties should similarity measure have?

- Maximum: document with itself
- Minimum: documents have no words in common (orthogonal)
- Increasing when more of same words used
- \(? s(a, b) = s(b, a).\)
Measuring Similarity

Measure 1: Inner product

\[
\langle \mathbf{v}_1, \mathbf{v}_4 \rangle = 6
\]
Measuring Similarity

Measure 1: Inner product

\[(2, 1) \cdot (1, 4) = 6\]
Problem: length dependent

\[ (4, 2)' \cdot (1, 4) = 12 \]

\[ a \cdot b = ||a|| \times ||b|| \times \cos \theta \]
Problem(?): length dependent
Problem(?): length dependent

$$(4, 2)' (1, 4) = 12$$
Problem(?): length dependent

\[(4, 2)'(1, 4) = 12\]

\[a \cdot b = ||a|| \times ||b|| \times \cos \theta\]
Cosine Similarity

\[ \cos \theta : \text{removes document length from similarity measure} \]
Cosine Similarity

$\cos \theta$: removes document length from similarity measure

$$\cos \theta = \left( \frac{a}{\|a\|} \right) \cdot \left( \frac{b}{\|b\|} \right)$$
Cosine Similarity

cos $\theta$: removes document length from similarity measure

\[
\cos \theta = \left( \frac{a}{\|a\|} \right) \cdot \left( \frac{b}{\|b\|} \right)
\]

\[
\frac{(4, 2)}{\| (4, 2) \|} = (0.89, 0.45)
\]
Cosine Similarity

\[ \cos \theta : \text{removes document length from similarity measure} \]

\[
\cos \theta = \left( \frac{a}{\|a\|} \right) \cdot \left( \frac{b}{\|b\|} \right)
\]

\[
\frac{(4, 2)}{\| (4, 2) \|} = (0.89, 0.45)
\]

\[
\frac{(2, 1)}{\| (2, 1) \|} = (0.89, 0.45)
\]
Cosine Similarity

cos θ: removes document length from similarity measure

\[
\cos \theta = \left( \frac{a}{\|a\|} \right) \cdot \left( \frac{b}{\|b\|} \right)
\]

\[
\frac{(4, 2)}{\| (4, 2) \|} = (0.89, 0.45)
\]

\[
\frac{(2, 1)}{\| (2, 1) \|} = (0.89, 0.45)
\]

\[
\frac{(1, 4)}{\| (1, 4) \|} = (0.24, 0.97)
\]
Cosine Similarity

$\cos \theta$: removes document length from similarity measure

\[
\cos \theta = \frac{a}{\|a\|} \cdot \frac{b}{\|b\|}
\]

\[
\begin{align*}
\frac{(4, 2)}{\|(4, 2)\|} &= (0.89, 0.45) \\
\frac{(2, 1)}{\|(2, 1)\|} &= (0.89, 0.45) \\
\frac{(1, 4)}{\|(1, 4)\|} &= (0.24, 0.97) \\
(0.89, 0.45)'(0.24, 0.97) &= 0.65
\end{align*}
\]
Cosine Similarity

\[ \cos \theta \]: removes document length from similarity measure
Cosine Similarity

\[ \cos \theta : \text{removes document length from similarity measure} \]
Project onto Hypersphere
Cosine Similarity

\[ \cos \theta \]: removes document length from similarity measure

Project onto Hypersphere

\[ \cos \theta \rightarrow \text{Inverse distance on Hypersphere} \]
Cosine Similarity

\[ \cos \theta \text{: removes document length from similarity measure} \]

Project onto Hypersphere

\[ \cos \theta \rightarrow \text{Inverse distance on Hypersphere} \]

von Mises Fisher distribution : distribution on sphere surface
Matrices

Definition

A **Matrix** is a rectangular array of numbers

\[
A = \begin{pmatrix}
    a_{11} & a_{12} & \ldots & a_{1n} \\
    a_{21} & a_{22} & \ldots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1} & a_{m2} & \ldots & a_{mn}
\end{pmatrix}
\]

If \( A \) has \( m \) rows \( n \) columns we will say that \( A \) is an \( m \times n \) matrix.

Suppose \( X \) and \( Y \) are \( m \times n \) matrices. Then \( X = Y \) if \( x_{ij} = y_{ij} \) for all \( i \) and \( j \).
Simple Examples

If $I$ is an $n \times n$ matrix we will call an identity matrix.
Simple Examples

\[ X = \begin{pmatrix}
1 & 2 & 3 \\
2 & 1 & 4
\end{pmatrix} \]

\( X \) is an 2 \( \times \) 3 matrix
Matrix Algebra

Definition

Suppose \( X \) and \( Y \) are \( m \times n \) matrices. Then define

\[
X + Y = \begin{pmatrix}
    x_{11} & x_{12} & \cdots & x_{1n} \\
    x_{21} & x_{22} & \cdots & x_{2n} \\
        \vdots & \vdots & \ddots & \vdots \\
    x_{m1} & x_{m2} & \cdots & x_{mn}
\end{pmatrix} + \begin{pmatrix}
    y_{11} & y_{12} & \cdots & y_{1n} \\
    y_{21} & y_{22} & \cdots & y_{2n} \\
        \vdots & \vdots & \ddots & \vdots \\
    y_{m1} & y_{m2} & \cdots & y_{mn}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
    x_{11} + y_{11} & x_{12} + y_{12} & \cdots & x_{1n} + y_{1n} \\
    x_{21} + y_{21} & x_{22} + y_{22} & \cdots & x_{2n} + y_{2n} \\
        \vdots & \vdots & \ddots & \vdots \\
    x_{m1} + y_{m1} & x_{m2} + y_{m2} & \cdots & x_{mn} + y_{mn}
\end{pmatrix}
\]
Matrix Algebra

Definition
Suppose \( X \) is an \( m \times n \) matrix and \( k \in \mathbb{R} \). Then,

\[
kX = \begin{pmatrix}
kx_{11} & kx_{12} & \cdots & kx_{1n} \\
kx_{21} & kx_{22} & \cdots & kx_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
kx_{m1} & kx_{m2} & \cdots & kx_{mn}
\end{pmatrix}
\]

Prove theorems about this tonight
R Code

Using `matrix` command:

```r
mat1 <- matrix(NA, nrow=3, ncol=2) ##
```

Creating matrix:

```r
mat1[1,1] <- 1
mat1[1,2] <- 2
mat1[2,1] <- 1
mat1[2,2] <- 4
mat1[3,1] <- exp(1)
mat1[3,2] <- 4
```
R Code

Using `rbind`

```r
r1 <- c(1, 2)
r2 <- c(1, 4)
r3 <- c(exp(1), 4)
mat1 <- rbind(r1, r2, r3)
```
Using `cbind`

c1 <- c(1, 1, exp(1))
c2 <- c(2, 4, 4)
R Code

dim(mat1)
[1] 3 2
mat1 + mat1
[,1] [,2]
[1,] 2.000000 4
[2,] 2.000000 8
[3,] 5.436564 8
R Code

What if the matrices are of different dimension
mat1<- matrix(1, nrow=3, ncol=2)
mat2<- matrix(2, nrow=10, ncol=3)
mat1 + mat2
Error in mat1 + mat2 : non-conformable arrays
Matrix Transpose

We will use **matrix transpose** to flip the dimensionality of a matrix.
Matrix Transpose

We will use matrix transpose to flip the dimensionality of a matrix

\[ X = \begin{pmatrix}
  x_{11} & x_{12} & \ldots & x_{1n} \\
  x_{21} & x_{22} & \ldots & x_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{m1} & x_{m2} & \ldots & x_{mn}
\end{pmatrix} \]

If \( X \) is an \( m \times n \) then \( X' \) is \( n \times m \).

If \( X = X' \) then we say \( X \) is symmetric.
Matrix Transpose

We will use matrix transpose to flip the dimensionality of a matrix

\[ X = \begin{pmatrix}
  x_{11} & x_{12} & \ldots & x_{1n} \\
  x_{21} & x_{22} & \ldots & x_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{m1} & x_{m2} & \ldots & x_{mn}
\end{pmatrix} \]

\[ X' = \begin{pmatrix}
  x_{11} \\
  x_{12} \\
  \vdots \\
  x_{1n}
\end{pmatrix} \]
Matrix Transpose

We will use **matrix transpose** to flip the dimensionality of a matrix

\[
\mathbf{X} = \begin{pmatrix}
    x_{11} & x_{12} & \ldots & x_{1n} \\
    x_{21} & x_{22} & \ldots & x_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{m1} & x_{m2} & \ldots & x_{mn}
\end{pmatrix}
\]

\[
\mathbf{X}' = \begin{pmatrix}
    x_{11} & x_{21} \\
    x_{12} & x_{22} \\
    \vdots & \vdots \\
    x_{1n} & x_{2n}
\end{pmatrix}
\]

If \(\mathbf{X}\) is an \(m \times n\) matrix, then \(\mathbf{X}'\) is \(n \times m\). If \(\mathbf{X} = \mathbf{X}'\), then we say \(\mathbf{X}\) is symmetric.
Matrix Transpose

We will use \textit{matrix transpose} to flip the dimensionality of a matrix.

\[ X = \begin{pmatrix}
  x_{11} & x_{12} & \ldots & x_{1n} \\
  x_{21} & x_{22} & \ldots & x_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{m1} & x_{m2} & \ldots & x_{mn}
\end{pmatrix} \]

\[ X' = \begin{pmatrix}
  x_{11} & x_{21} & \ldots \\
  x_{12} & x_{22} & \ldots \\
  \vdots & \vdots & \ddots \\
  x_{1n} & x_{2n} & \ldots
\end{pmatrix} \]

If \( X \) is an \( m \times n \) then \( X' \) is \( n \times m \).

If \( X = X' \) then we say \( X \) is \textit{symmetric}. 
Matrix Transpose

We will use **matrix transpose** to flip the dimensionality of a matrix

\[
X = \begin{pmatrix}
    x_{11} & x_{12} & \ldots & x_{1n} \\
    x_{21} & x_{22} & \ldots & x_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{m1} & x_{m2} & \ldots & x_{mn}
\end{pmatrix}
\]

\[
X' = \begin{pmatrix}
    x_{11} & x_{21} & \ldots & x_{m1} \\
    x_{12} & x_{22} & \ldots & x_{m2} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{1n} & x_{2n} & \ldots & x_{mn}
\end{pmatrix}
\]

If \(X\) is an \(m \times n\) then \(X'\) is \(n \times m\). If \(X = X'\) then we say \(X\) is symmetric.
Matrix Transpose

We will use matrix transpose to flip the dimensionality of a matrix

\[
X = \begin{pmatrix}
x_{11} & x_{12} & \ldots & x_{1n} \\
x_{21} & x_{22} & \ldots & x_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
x_{m1} & x_{m2} & \ldots & x_{mn}
\end{pmatrix}
\]

\[
X' = \begin{pmatrix}
x_{11} & x_{21} & \ldots & x_{m1} \\
x_{12} & x_{22} & \ldots & x_{m2} \\
\vdots & \vdots & \ddots & \vdots \\
x_{1n} & x_{2n} & \ldots & x_{mn}
\end{pmatrix}
\]

If \( X \) is an \( m \times n \) then \( X' \) is \( n \times m \).
Matrix Transpose

We will use matrix transpose to flip the dimensionality of a matrix

\[ X = \begin{pmatrix}
  x_{11} & x_{12} & \ldots & x_{1n} \\
  x_{21} & x_{22} & \ldots & x_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{m1} & x_{m2} & \ldots & x_{mn}
\end{pmatrix} \]

\[ X' = \begin{pmatrix}
  x_{11} & x_{21} & \ldots & x_{m1} \\
  x_{12} & x_{22} & \ldots & x_{m2} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{1n} & x_{2n} & \ldots & x_{mn}
\end{pmatrix} \]

If \( X \) is an \( m \times n \) then \( X' \) is \( n \times m \).
If \( X = X' \) then we say \( X \) is symmetric.
Matrix Transpose

Example 1: \( \mathbf{X} = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} \) then \( \mathbf{X}' = \begin{pmatrix} 4 & 1 \\ 1 & 2 \\ 2 & 3 \end{pmatrix} \)

In R

```r
mat1 <- matrix(c(1, 2, 3), nrow=3, ncol=2)
mat2 <- t(mat1)
dim(mat1)
dim(mat2)
```

3 2
2 3
Matrix Multiplication

How do we multiply matrices?

Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition.

Suppose we have two matrices:

\[ X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]

We will create a new matrix \( A \) by matrix multiplication:

\[ A = XY = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 4 & 6 \end{pmatrix} \]
Matrix Multiplication

How do we multiply matrices?

Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition.
Matrix Multiplication

How do we multiply matrices?

Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition

Suppose we have two matrices

\[
X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}
\]

We will create a new matrix \( A \) by matrix multiplication:

\[
A = XY = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \\ 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 4 & 6 \end{pmatrix}
\]
Matrix Multiplication

How do we multiply matrices?
Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition

Suppose we have two matrices

\[ X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]
Matrix Multiplication

How do we multiply matrices?

Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition

Suppose we have two matrices

\[ X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]

We will create a new matrix \( A \) by matrix multiplication:
Matrix Multiplication

How do we multiply matrices?

Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition

Suppose we have two matrices

\[
X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}
\]

We will create a new matrix \( A \) by matrix multiplication:

\[
A = XY
\]
Matrix Multiplication

How do we multiply matrices?
Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition

Suppose we have two matrices

\[ \mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]

We will create a new matrix \( \mathbf{A} \) by matrix multiplication:

\[
\mathbf{A} = \mathbf{X} \mathbf{Y} \\
= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}
\]
Matrix Multiplication

How do we multiply matrices?

Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition.

Suppose we have two matrices,

\[
X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}
\]

We will create a new matrix \( A \) by matrix multiplication:

\[
A = XY = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}
\]

\[
= \begin{pmatrix} 1 \times 1 + 1 \times 3 \\ 1 \times 1 \end{pmatrix}
\]

\[
= \begin{pmatrix} 4 & 6 \\ 4 & 6 \end{pmatrix}
\]
Matrix Multiplication

How do we multiply matrices?
Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition.

Suppose we have two matrices
\[ X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]

We will create a new matrix \( A \) by matrix multiplication:

\[
A = XY \\
= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \\
= \begin{pmatrix} 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \end{pmatrix} \\
= \begin{pmatrix} 4 & 6 \\ 4 & 6 \end{pmatrix}
\]
Matrix Multiplication

How do we multiply matrices?

Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition

Suppose we have two matrices

\[ X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]

We will create a new matrix \( A \) by matrix multiplication:

\[
A = XY
\]

\[
= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}
\]

\[
= \begin{pmatrix} 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \\ 1 \times 1 + 1 \times 3 \end{pmatrix}
\]

\[
= \begin{pmatrix} 4 & 6 \\ 4 & 6 \end{pmatrix}
\]
Matrix Multiplication

How do we multiply matrices?

Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition

Suppose we have two matrices

\[ X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]

We will create a new matrix \( A \) by matrix multiplication:

\[
A = X Y = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]

\[
= \begin{pmatrix} 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \\ 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \end{pmatrix}
\]
Matrix Multiplication

How do we multiply matrices?
Because we want to use matrix multiplication to solve equations we won’t use an intuitive definition

Suppose we have two matrices

\[ \mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \]

We will create a new matrix \( \mathbf{A} \) by matrix multiplication:

\[
\mathbf{A} = \mathbf{XY} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}
= \begin{pmatrix} 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \\ 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \end{pmatrix}
= \begin{pmatrix} 4 & 6 \\ 4 & 6 \end{pmatrix}
\]
Matrix Multiplication

Definition

Suppose $X$ is an $m \times n$ matrix and $Y$ is an $n \times k$ matrix. Then define the matrix $A$ an $m \times k$ matrix that obtains from multiplying $X$ and $Y$ as,

$$A = XY$$

$$= \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{pmatrix} \begin{pmatrix} y_{11} & y_{12} & \ldots & y_{1k} \\ y_{21} & y_{22} & \ldots & y_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \ldots & y_{nk} \end{pmatrix}$$

$$= \begin{pmatrix} x_{11}y_{11} + x_{12}y_{21} + \ldots + x_{1n}y_{n1} & \ldots & x_{11}y_{1k} + x_{12}y_{2k} + \ldots + x_{1n}y_{nk} \\ \vdots & \ddots & \vdots \\ x_{m1}y_{11} + x_{m2}y_{21} + \ldots + x_{mn}y_{n1} & \ldots & x_{m1}y_{11} + x_{m2}y_{12} + \ldots + x_{mn}y_{nk} \end{pmatrix}$$
Definition

Suppose $X$ is an $m \times n$ matrix and $Y$ is an $n \times k$ matrix. Write the row vectors of $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$ and $Y$ as column vector $Y = (y_1 \ y_2 \ \cdots \ y_k)$. Then the $m \times k$ matrix $A = XY$ can be written as

$$A = \begin{pmatrix} x_1 \cdot y_1 & x_1 \cdot y_2 & \cdots & x_1 \cdot y_k \\ x_2 \cdot y_1 & x_2 \cdot y_2 & \cdots & x_2 \cdot y_k \\ \vdots & \vdots & \ddots & \vdots \\ x_m \cdot y_1 & x_m \cdot y_2 & \cdots & x_m \cdot y_k \end{pmatrix}.$$
Matrix Multiplication

Let’s work on an example together!

\[ X = \begin{pmatrix} 1 & 4 & 5 \\ 10 & 2 & 3 \end{pmatrix} \]

\[ Y = \begin{pmatrix} 2 & 3 \\ 1 & 5 \\ 3 & 5 \end{pmatrix} \]

What is \( XY \)?

Not all matrices can be multiplied. Matrix \( AB \) exists only if the number of columns in \( A \) equals the number of rows in \( B \). If \( AB \) exists, we will say the matrices are conformable.
Matrix Multiplication

Let’s work on an example together!

\[ X = \begin{pmatrix} 1 & 4 & 5 \\ 10 & 2 & 3 \\ 2 & 3 \end{pmatrix} \]

\[ Y = \begin{pmatrix} 2 & 3 \\ 1 & 5 \\ 3 & 5 \end{pmatrix} \]

What is \( X Y \)?

Not all matrices can be multiplied. Matrix \( AB \) exists only if the number of columns in \( A \) = number of rows in \( B \). If \( AB \) exists we will say the matrices are conformable.
Matrix Multiplication with a Vector

Suppose \( X = \begin{pmatrix} 2 & 3 & 4 & 5 \\ 1 & 5 & 1 & 2 \\ 3 & 5 & 3 & 4 \end{pmatrix} \) a 3 \( \times \) 4 matrix and that \( v = \begin{pmatrix} 3 \\ 3 \\ 4 \\ 10 \end{pmatrix} \) a 4 \( \times \) 1 matrix (or a column vector) what is \( Xv \)?
What is \( X'v \)?
Algebraic Properties

Suppose $X$ is an $m \times n$ matrix and $Y$ is an $n \times k$ matrix. Suppose that

$$I = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

as the identity matrix and that $k \in Re$.

- $XY \neq YX$ in general !!!! (but it could)
- $XI = X$ (let’s talk it out!)
- $(X')' = X$
- $(XY)' = Y'X'$
- $(kX)' = kX'$
- $(X + Y)' = X' + Y'$
Examples, Implementing in R

R and matrix multiplication
X <- matrix(NA, nrow=2, ncol=3)
Y <- matrix(NA, nrow=3, ncol=2)
X[1,] <- c(1, 4, 5)
X[2,] <- c(10, 2, 3)
Y[1,] <- c(2, 3)
Y[2,] <- c(1, 5)
Y[3,] <- c(3, 5)

A <- X %*% Y
> A
[,1] [,2]
[1,] 21 48
[2,] 31 55
Matrix Inversion

Big topic: suppose $X$ is an $n \times n$ matrix. We want to find the matrix $X^{-1}$ such that

$$X^{-1}X = XX^{-1} = I$$

where $I$ is the $n \times n$ identity matrix.

Why?

- Regression
- Solving systems of equations
- Will provide intuition about “colinearity”, “fixed effects”, “treatment designs” and what we can learn as social scientists

Calculate $\leadsto$ Properties of Inverses $\leadsto$ when do inverses exist $\leadsto$

Application to regression analysis
Some Motivating Examples

Consider the following equations:

\[
\begin{align*}
    x_1 + x_2 + x_3 &= 0 \\
    0x_1 + 0x_2 + x_3 &= 5
\end{align*}
\] (0.1)

\[
A = \begin{bmatrix}
    1 & 1 & 1 \\
    0 & 5 & 0 \\
    0 & 0 & 3
\end{bmatrix}
\]

\[
x = (x_1, x_2, x_3)
\]

\[
b = (0, 5, 6)
\]
Some Motivating Examples
Consider the following equations:

\[
\begin{align*}
\quad x_1 + x_2 + x_3 &= 0 \\
\quad x_1 + x_2 + 0x_3 &= 0 \\
\quad 0x_1 + x_2 + x_3 &= 0 \\
\quad x_1 + 0x_2 + x_3 &= 0 \\
\quad A &= \begin{pmatrix}
1 & 1 & 1 \\
0 & 5 & 0 \\
0 & 0 & 3 \\
\end{pmatrix} \\
\quad x &= (x_1, x_2, x_3) \\
\quad b &= (0, 5, 6)
\end{align*}
\]
Some Motivating Examples

Consider the following equations:

\[
\begin{align*}
    x_1 + x_2 + x_3 &= 0 \\
    0x_1 + 5x_2 + 0x_3 &= 5 \\
    0x_1 + 0x_2 + 3x_3 &= 6
\end{align*}
\]
Some Motivating Examples

Consider the following equations:

\[ x_1 + x_2 + x_3 = 0 \]
\[ 0x_1 + 5x_2 + 0x_3 = 5 \]
\[ 0x_1 + 0x_2 + 3x_3 = 6 \]

\[ A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix} \]
Some Motivating Examples

Consider the following equations:

\[
\begin{align*}
x_1 + x_2 + x_3 &= 0 \\
0x_1 + 5x_2 + 0x_3 &= 5 \\
0x_1 + 0x_2 + 3x_3 &= 6
\end{align*}
\]

\[
A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix}
\]

\[
x = (x_1, x_2, x_3)
\]
Some Motivating Examples

Consider the following equations:

\[ \begin{align*}
    x_1 + x_2 + x_3 &= 0 \\
    0x_1 + 5x_2 + 0x_3 &= 5 \\
    0x_1 + 0x_2 + 3x_3 &= 6
\end{align*} \]

\[ A = \begin{pmatrix}
    1 & 1 & 1 \\
    0 & 5 & 0 \\
    0 & 0 & 3
\end{pmatrix} \]

\[ x = (x_1, x_2, x_3) \]

\[ b = (0, 5, 6) \]
Some Motivating Examples

Consider the following equations:

\[ \begin{align*}
x_1 + x_2 + x_3 &= 0 \\
0x_1 + 5x_2 + 0x_3 &= 5 \\
0x_1 + 0x_2 + 3x_3 &= 6
\end{align*} \]

\[ \mathbf{A} = \begin{pmatrix}
1 & 1 & 1 \\
0 & 5 & 0 \\
0 & 0 & 3
\end{pmatrix} \]

\[ \mathbf{x} = (x_1, x_2, x_3) \]

\[ \mathbf{b} = (0, 5, 6) \]

The system of equations are now,
Some Motivating Examples

Consider the following equations:

\[ x_1 + x_2 + x_3 = 0 \]
\[ 0x_1 + 5x_2 + 0x_3 = 5 \]
\[ 0x_1 + 0x_2 + 3x_3 = 6 \]

\[ A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix} \]
\[ x = (x_1, x_2, x_3) \]
\[ b = (0, 5, 6) \]

The system of equations are now,

\[ Ax = b \]
Some Motivating Examples

Consider the following equations:

\[ x_1 + x_2 + x_3 = 0 \]
\[ 0x_1 + 5x_2 + 0x_3 = 5 \]
\[ 0x_1 + 0x_2 + 3x_3 = 6 \]

\[ A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix} \]
\[ x = (x_1, x_2, x_3) \]
\[ b = (0, 5, 6) \]

The system of equations are now,

\[ Ax = b \]

\( A^{-1} \) exists if and only if \( Ax = b \) has only one solution.
Matrix Inversion, Definition

Definition

Suppose $X$ is an $n \times n$ matrix. We will call $X^{-1}$ the inverse of $X$ if

$$X^{-1}X = XX^{-1} = I$$

If $X^{-1}$ exists then $X$ is invertible. If $X^{-1}$ does not exist, then we will say $X$ is singular.
Matrix Inversion

You’ll never invert a matrix by hand.

We’re going to use R

```r
X <- matrix(NA, nrow=3, ncol=3)
X[1,] <- c(2, 3, 4)
X[2,] <- c(3, 1, 3)
X[3,] <- c(2, 4, 2)
X.inv <- solve(X)
> X.inv
[,1] [,2] [,3]
[1,]  -0.5  0.5  0.25
[2,]   0.0 -0.2  0.30
[3,]   0.5 -0.1 -0.35
X.inv %*% X
[,1] [,2] [,3]
[1,]  1 0.000000e+00 -2.220446e-16
[2,]  0 1.000000e+00  0.000000e+00
[3,]  0 -2.220446e-16  1.000000e+00
```
Matrix Inversion

1) Calculate Inverses
2) Properties of Inverses
Matrix Inversion

1) Calculate Inverses
2) Properties of Inverses

Theorem

The inverse of matrix $\mathbf{X}$, $\mathbf{X}^{-1}$, is unique
Matrix Inversion

1) Calculate Inverses
2) Properties of Inverses

Theorem

The inverse of matrix $X$, $X^{-1}$, is unique

Proof.
By way of contradiction, suppose not. Then there are at least two matrices $A$ and $C$ such that $AX = I$ and $CX = I$
This implies that,

\[ AXC = (AX)C = IC = C \]
Matrix Inversion

But it also implies that

\[ AXC = A(XC) \]
\[ = A(I) \]
\[ = A \]

So \( C = AXC = A \) or \( C = A \) but this contradicts our assumption that there are two unique inverses.
Matrix Inversion

Theorem

Suppose $A$ has inverse $A^{-1}$ and $B$ has inverse $B^{-1}$. Then,

$$(AB)^{-1} = B^{-1}A^{-1}$$
Matrix Inversion

Theorem

Suppose \( A \) has inverse \( A^{-1} \) and \( B \) has inverse \( B^{-1} \). Then,

\[
(AB)^{-1} = B^{-1}A^{-1}
\]

Proof.

We need to show that \( (B^{-1}A^{-1})(AB) = (AB)(B^{-1}A^{-1}) = I \).

\[
(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I
\]
Matrix Inversion

\[(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I\]

So \(AB\) is invertible and \((AB)^{-1} = B^{-1}A^{-1}\).
Challenge Inversion Proofs

- Show that \((A^{-1})^{-1} = A\).
- Show that \((kA)^{-1} = \frac{1}{k} A^{-1}\)
Matrix Inversion

1) How to Calculate an Inverse
2) Inversion properties
3) When do inverses exist?

Linear Independence: not repeated information in matrix will be the key (for both inversion and regressions)
Matrix Inversion: Existence

Definition

Suppose we have a set of vectors $S = \{v_1, v_2, \ldots, v_r\}$
And consider the system of equations

$$k_1 v_1 + k_2 v_2 + \ldots + k_r v_r = 0$$

If the only solution is $k_1 = 0, k_2 = 0, k_3 = 0, \ldots, k_r = 0$ then we say that the set is **linearly independent**. If there are other solutions, then the set is **linearly dependent**.
Consider \( \mathbf{v}_1 = (1, 0, 0), \mathbf{v}_2 = (0, 1, 0), \mathbf{v}_3 = (0, 0, 1) \)
Can we write this as a combination of other vectors?
Matrix Inversion: Existence

Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, $\mathbf{v}_3 = (0, 0, 1)$
Can we write this as a combination of other vectors? no!
Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, $\mathbf{v}_3 = (0, 0, 1)$.
Can we write this as a combination of other vectors? \textit{no!}

Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, $\mathbf{v}_3 = (0, 0, 1)$, $\mathbf{v}_4 = (1, 2, 3)$.
Can we write this as a combination of other vectors?
Matrix Inversion: Existence

Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, $\mathbf{v}_3 = (0, 0, 1)$
Can we write this as a combination of other vectors? no!

Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, $\mathbf{v}_3 = (0, 0, 1)$, $\mathbf{v}_4 = (1, 2, 3)$.
Can we write this as a combination of other vectors?

$$\mathbf{v}_4 = \mathbf{v}_1 + 2\mathbf{v}_2 + 3\mathbf{v}_3$$
Matrix Inversion: Existence

Theorem

Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K \in \mathbb{R}^n$. If $K > n$ then the set is linearly dependent.
Matrix Inversion: Existence

Theorem

Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K \in \mathbb{R}^n$. If $K > n$ then the set is linearly dependent.

- $\mathbf{v}_1 = (v_{11}, v_{21}, \ldots, v_{n1})$
Matrix Inversion: Existence

Theorem

Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K \in \mathbb{R}^n$. If $K > n$ then the set is linearly dependent

- $\mathbf{v}_1 = (v_{11}, v_{21}, \ldots, v_{n1})$
- Says that if there are more vectors in the set than elements in each vector, one must be linearly dependent
We care because of the following theorem

Theorem

Suppose $\mathbf{X}$ is an $n \times n$ matrix. Recall we can write this matrix as

$$
\begin{pmatrix}
 x_1 \\
 x_2 \\
 \vdots \\
 x_n
\end{pmatrix}
$$

Then $\mathbf{X}$ has an inverse if and only if $S = \{x_1, x_2, \ldots, x_n\}$ is linearly independent.

If this is true, we say $\mathbf{X}$ has full rank.
Linear Regression

In 350a you learn about linear regression. For each \( i \) (individual) we observe covariates \( x_1, x_2, \ldots, x_k \) and independent variable \( Y \).

\[
Y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k}
\]

\[
Y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k}
\]

...    ...

\[
Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}
\]

\[
Y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk}
\]
Linear Regression

In 350a you learn about linear regression. For each $i$ (individual) we observe covariates $x_{i1}, x_{i2}, \ldots, x_{ik}$ and independent variable $Y_i$. Then,
In 350a you learn about linear regression. For each $i$ (individual) we observe covariates $x_{i1}, x_{i2}, \ldots, x_{ik}$ and independent variable $Y_i$. Then,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$
Linear Regression

In 350a you learn about linear regression. For each $i$ (individual) we observe covariates $x_{i1}, x_{i2}, \ldots, x_{ik}$ and independent variable $Y_i$. Then,

\[
Y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_k x_{1k} \\
Y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_k x_{2k} \\
\vdots \quad \vdots \quad \vdots \\
Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} \\
\vdots \quad \vdots \quad \vdots \\
Y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_k x_{nk}
\]
Linear Regression

- Define $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ik})$

- Define $\mathbf{X} =
\begin{pmatrix}
  \mathbf{x}_1 \\
  \mathbf{x}_2 \\
  \vdots \\
  \mathbf{x}_n
\end{pmatrix}$

- Define $\mathbf{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$

- Define $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$.

Then we can write

$$\mathbf{Y} = \mathbf{X}\mathbf{\beta}$$
Linear Regression

\[ Y = X\beta \]
\[ X'Y = X'X\beta \]
\[ (X'X)^{-1}X'Y = (X'X)^{-1}X'X\beta \]
\[ (X'X)^{-1}X'Y = \beta \]

Big question: is \((X'X)^{-1}\) invertible?
We’ll investigate in homework!