PS200C Section: Basic Linear Algebra Review

Soonhong Cho

2022-04-14

Section 1

Definitions

Definition (Matrix)

A **matrix** is a rectangular array of numbers. An $(n \times k)$ matrix has n rows and k columns.

$$\underset{(n \times k)}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix}$$

The shorthand $\mathbf{A} = [a_{ij}]$ can be used to define matrix operations.

- We'll often organize our data such that each row is an observation, and each column is a covariate (properties of that observation)–which is called "model matrix," or **X** in regression context.
- Matrices (and vectors) are typically written in bold, so I'm following in this slides to be more explicit (but will not later in the course, as typing them is really annoying–you always have to type `mathbf{X}` instead of `X`)

Definition (Row Vector)

A $(1 \times k)$ matrix is a **row vector**.

$$\underset{(1\times k)}{\mathbf{y}} = \begin{bmatrix} y_{11} & \cdots & y_{1k} \end{bmatrix}$$

Definition (Column Vector)

A $(n \times 1)$ matrix is a **column vector**.

$$\underset{(n\times 1)}{\mathbf{y}} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{n1} \end{bmatrix}$$

Definition (Scalar)

A scalar is a (1×1) matrix.

- Remember that vectors are typically written out as column vectors rather than row vectors: it's just a convention. For example, when you see a 2-dim vector v alone in any text, then it usually means a column vector $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.
- A corresponding row vector is written as $v' = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}' = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$ which reads the *transpose* of the column vector v (we'll see what is transpose in a minute).
- A scalar is just an ordinary number; it can be viewed as a (1×1) matrix, or a (1×1) vector.

Definition (Square Matrix)

A **square matrix** is one where $n = k$: the same number of rows and columns.

$$\underset{(n\times n)}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Definition (Diagonal Matrix)

A **diagonal matrix** is a square matrix where $a_{ij} = 0$ if $i \neq j$.

$$\underset{(n\times n)}{\mathbf{A}} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

Definition (Identity Matrix)

An **identity matrix** is a diagonal matrix whose diagonal elements (a_{ij} where $i = j$) are one, and off-diagonal elements (a_{ij} where $i \neq j$) are zero.

$$\underset{(n\times n)}{\mathbf{I}} \equiv \mathbf{I}_n \equiv \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

The identity matrix is the matrix version of "1."

Definition (Zero Matrix)

A **zero matrix** is an $n \times k$ matrix whose elements are all zero.

$$\underset{(n\times k)}{\mathbf{0}} \equiv \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

The zero matrix is the matrix version of "0."

Section 2

Matrix Operations

Addition

We can add two matrices **A** and **B** if they have the same dimensions. It is conducted element by element, and represented by $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$.

$$\underset{n\times k}{\mathbf{A}} + \underset{n\times k}{\mathbf{B}} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1k}+b_{1k} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2k}+b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \cdots & a_{nk}+b_{nk} \end{bmatrix}$$

For example: if $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$,

Addition

We can add two matrices **A** and **B** if they have the same dimensions. It is conducted element by element, and represented by $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$.

$$\underset{n \times k}{\mathbf{A}} + \underset{n \times k}{\mathbf{B}} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1k} + b_{1k} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2k} + b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nk} + b_{nk} \end{bmatrix}$$

For example: if $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$, then $\mathbf{A} + \mathbf{B} = \begin{bmatrix} 3 & 6 \\ 9 & 12 \end{bmatrix}$

Scalar multiplication

For any scalar $c \in \mathbb{R}^1$, $c\mathbf{A} = [c \cdot a_{ij}]$:

$$c\mathbf{A} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1k} \\ ca_{21} & ca_{22} & \cdots & ca_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nk} \end{bmatrix}$$

Matrix multiplication

Remember that matrix multiplication is different from scalar multiplication to matrix! It's confusing at first look to most people. But it's fundamental to grab how estimators (and many other things) you'll work with are constructed—e.g., the OLS estimator; so let's see what matrix multiplication is actually doing (and the resulting dimensions).

Matrix multiplication

Remember that matrix multiplication is different from scalar multiplication to matrix! It's confusing at first look to most people. But it's fundamental to grab how estimators (and many other things) you'll work with are constructed—e.g., the OLS estimator; so let's see what matrix multiplication is actually doing (and the resulting dimensions).

1. The matrix multiplication $\underset{(n\times m)}{\mathbf{A}}\underset{(m\times k)}{\mathbf{B}}$ is only possible if they are **conformable**: the number of columns of the left one **A** must be equal to the number of rows of the right one **B**. The dimension of resulting matrix **AB** is $(n \times m)(m \times k) = (n \times k)$. However, the result of $\underset{(m\times k)}{\mathbf{B}}\underset{(n\times m)}{\mathbf{A}}$ is undefined, because the matrices are not conformable. So, **ORDER MATTERS**. You should get comfortable with the dimensions of matrices multiplied together.

Matrix multiplication

2. The definition is not that helpful:

$$\mathbf{AB} = \left[\sum_{d=1}^{m} a_{id} b_{dj} \right],$$

where each element of **AB** is ab_{ij}, for $i = 1, \cdots, n$ and $j = 1, \cdots, k$.

One trick for easier life is thinking column-wise way: the rows of **B** give a "recipe" to combine the columns of **A**. Here is one example: say we want to compute **AB** for

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{n1} & a_{n2} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} {\color{blue}b_{11}} & {\color{red}b_{12}} \\ {\color{blue}b_{21}} & {\color{red}b_{22}} \end{bmatrix}.$$

They are conformable and the dimension of the resulting matrix must be $(3 \times 2) \times (2 \times 2) = (3 \times 2)$. We have

One trick for easier life is thinking column-wise way: the rows of **B** give a "recipe" to combine the columns of **A**. Here is one example: say we want to compute **AB** for

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{n1} & a_{n2} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \color{blue}{b_{11}} & \color{red}{b_{12}} \\ \color{blue}{b_{21}} & \color{red}{b_{22}} \end{bmatrix}.$$

They are conformable and the dimension of the resulting matrix must be $(3 \times 2) \times (2 \times 2) = (3 \times 2)$. We have

$$\underset{(3\times\color{blue}{2})}{\mathbf{A}}\underset{(\color{blue}{2}\times 2)}{\mathbf{B}} = \begin{bmatrix} \color{blue}{b_{11}}\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + \color{blue}{b_{21}}\begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} & \color{red}{b_{12}}\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + \color{red}{b_{22}}\begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} \color{blue}{b_{11}}a_{11} + \color{blue}{b_{21}}a_{12} & \color{red}{b_{12}}a_{11} + \color{red}{b_{22}}a_{12} \\ \color{blue}{b_{11}}a_{21} + \color{blue}{b_{21}}a_{22} & \color{red}{b_{12}}a_{21} + \color{red}{b_{22}}a_{22} \\ \color{blue}{b_{11}}a_{31} + \color{blue}{b_{21}}a_{32} & \color{red}{b_{12}}a_{31} + \color{red}{b_{22}}a_{32} \end{bmatrix},$$

where the first equality is our "trick," and the second one is the definition.

Matrix multiplication

A more specific example: say $\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then, we have:

Matrix multiplication

A more specific example: say $\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then, we have:

$$\mathbf{AB} = \begin{bmatrix} 1\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0\begin{pmatrix} 3 \\ 4 \end{pmatrix} & 1\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 1\begin{pmatrix} 3 \\ 4 \end{pmatrix} \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 6 \end{bmatrix}.$$

3. In R, do not use the usual multiplication mark $*$ as it's for element-wise multiplication (in the above example, it will return $\begin{bmatrix} 1 & 3 \\ 0 & 4 \end{bmatrix}$). You have to use %*% for matrix multiplication!

```r
A <- matrix(c(1, 2, 3, 4), byrow=FALSE, ncol=2)
A
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

```r
B <- cbind(c(1, 0), c(1, 1)) #"column vector" way
B
```

```
##      [,1] [,2]
## [1,]    1    1
## [2,]    0    1
```

```
A*B #element-wise multiplication ("Hadamard product")
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    0    4
```

```
A %*% B #matrix multiplication
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    6
```

Properties of matrix multiplication

1. The commutative property does not hold! Notice that **BA** is NOT conformable:

$$\underset{(n\times m)}{\mathbf{A}}\underset{(m\times k)}{\mathbf{B}} \neq \underset{(m\times k)}{\mathbf{B}}\underset{(n\times m)}{\mathbf{A}}$$

2. Associative property:

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

3. Distributive properties (both left and right hold):

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC} \text{ and } (\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA}+\mathbf{CA}$$

4. Multiplicative identity property:

$$\mathbf{IA} = \mathbf{A} \text{ and } \mathbf{AI} = \mathbf{A}$$

5. Dimension property:

The product of an $(n\times m)$ matrix and an $(m\times k)$ matrix is an $(n\times k)$ matrix.

Transpose

Definition (Transpose)

Let $\mathbf{A} = [a_{ij}]$ be an $(n \times k)$ matrix. The **transpose** of $\mathbf{A}$, denoted $\mathbf{A}'$ or $\mathbf{A}^T$ is the $k \times n$ matrix whose elements are $\mathbf{A}' \equiv [a_{ji}]$

When you transpose, you essentially "mirror" the matrix about its diagonal. This is the same as saying that column i of a matrix is row i of its transpose.

For example, if $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$, then

Transpose

Definition (Transpose)

Let $\mathbf{A} = [a_{ij}]$ be an $(n \times k)$ matrix. The **transpose** of $\mathbf{A}$, denoted $\mathbf{A}'$ or $\mathbf{A}^T$ is the $k \times n$ matrix whose elements are $\mathbf{A}' \equiv [a_{ji}]$

When you transpose, you essentially "mirror" the matrix about its diagonal. This is the same as saying that column i of a matrix is row i of its transpose.

For example, if $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$, then $\mathbf{A}' = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

Properties of Transpose

1. $(\mathbf{A}')' = \mathbf{A}$
2. $\alpha' = \alpha$, where α is a scalar
3. $(\alpha\mathbf{A})' = \alpha\mathbf{A}'$
4. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$
5. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, where $\mathbf{A}$ is $(n \times k)$ and $\mathbf{B}$ is $(k \times m)$. Note that they become $(m \times n)$ matrices.
6. Inner product: let $\mathbf{x}$ be a $(n \times 1)$ vector. Then $\mathbf{x}'\mathbf{x} = \sum_i^n x_i^2$. For example, if $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, then

$$\mathbf{x}'\mathbf{x} = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1^2 + x_2^2 + x_3^2 = \sum_{i=1}^{3} x_i^2.$$

Expanding this result, you will likely run into the construction $\mathbf{X}'\mathbf{X}$ (hint: in OLS estimator); this is the matrix world equivalent of "squaring and summing".

```
A
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

```
t(A) #transpose: function `t`
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    3    4
```

Symmetric Matrix

Definition (Symmetric Matrix)

A square matrix $\mathbf{A}$ is **symmetric** if and only if $\mathbf{A}' = \mathbf{A}$. If $\mathbf{A}$ is a $(n \times k)$ non-square matrix, then $\mathbf{A}'\mathbf{A}$ is always defined and is symmetric.

Let's see if $\mathbf{A}'\mathbf{A}$ is symmetric using the definition and some properties:

$$(\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A},$$

so $\mathbf{A}'\mathbf{A}$ is symmetric.

Trace

Definition (Trace)

For any $(n \times n)$ symmetric matrix $\mathbf{A}$, the trace of $\mathbf{A}$, denoted $tr(\mathbf{A})$, is the sum of its diagonal elements: $tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$.

- You'll encounter the trace for the proof that the OLS estimator's variance is unbiased for the true variance (and many other things), but I'm not sure if we'll use it in this course.
- Some useful properties include:
  1. $tr(\mathbf{I}_n) = n$
  2. $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$
  3. $tr(\alpha\mathbf{A}) = \alpha tr(\mathbf{A})$
  4. $tr(\mathbf{AB}) = tr(\mathbf{BA})$, where $\mathbf{A}$ is $(n \times k)$ and $\mathbf{B}$ is $(k \times n)$. This one is really useful when you learn more advanced matrix algebra in the context of statistics.

```
A
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

```
sum(diag(A)) #sum of diagonal terms
```

```
## [1] 5
```

Inverse

Definition (Inverse)

An $(n \times n)$ square matrix $\mathbf{A}$ has an **inverse** $\mathbf{A}^{-1}$ if and only iff $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I_n}$.

In this case $\mathbf{A}$ is said to be *invertible* or *non-singular*. Otherwise, $\mathbf{A}$ is called *non-invertible* or *singular*.

Inverse

- The mechanics behind inverse are not important, so I'll skip it (so annoying and you'll never do it by hand, maybe). The thing is: it's an operation that is similar in function to the inverse of a number (like division).
- An inverse is **only defined for square matrices** (think about comformability), in addition to a few other criteria. One that you should know: the matrix must be **full-rank**. A matrix that is not full-rank is said to be rank deficient.
- So: "full-rank = non-singular → invertible," and "rank deficient = singular → non-invertible." I'm still confusing when I talk about them.

Inverse

- I'll not define *rank* here (since we have to spend at least a few hours to get there). This is linear algebra word to mean that no column is an exact multiple of any other column, or a (linear) combination of other columns (i.e., a square matrix is invertible when its columns are "linearly independent"). Intuitively, if a matrix is full rank, it means that every row/column provides unique and new information that we can't get from any of the other rows/columns or their some combinations.
- This is why OLS does not work under "perfect multicollinearity" (it was in the first assignment, True or False quiz): because the OLS estimator $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ needs to invert the quantity $(\mathbf{X}'\mathbf{X})$, it requires that $\mathbf{X}$ to be full-rank.

Inverse

- Example: are they invertible?

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 3 & 4 & 1 \\ 1 & 3 & 2 \\ 4 & 6 & 2 \end{bmatrix}$$

Inverse

- Example: are they invertible?

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 3 & 4 & 1 \\ 1 & 3 & 2 \\ 4 & 6 & 2 \end{bmatrix}$$

Y, Y, N (sum of column 1 and column 3 is what?).

```
solve(A) #R function for inverse is `solve`
```

```
##      [,1] [,2]
## [1,]   -2  1.5
## [2,]    1 -0.5
```

```
C <- matrix(c(3,1,4,4,3,6,1,2,2), nrow=3)
solve(C) #not invertible ("singular")
```

```
## Error in solve.default(C): Lapack routine dgesv: system is exactly s
```

Properties of the Inverse

1. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, if they are invertible.
2. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$
3. The inverse of a matrix is *unique*, if exists.

Quadratic Form

Definition (Quadratic Form)

Let **A** be an $(n \times n)$ symmetric matrix. The quadratic form associated with **A** is the real-valued function defined for all $(n \times 1)$ vectors **x**:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j>1} a_{ij}x_ix_j$$

What is the dimension of $\mathbf{x}'\mathbf{A}\mathbf{x}$?:

Quadratic Form

Definition (Quadratic Form)

Let **A** be an $(n \times n)$ symmetric matrix. The quadratic form associated with **A** is the real-valued function defined for all $(n \times 1)$ vectors **x**:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j>1} a_{ij}x_i x_j$$

What is the dimension of $\mathbf{x}'\mathbf{A}\mathbf{x}$?: $(1 \times n)(n \times n)(n \times 1)$

Idempotent

Definition (Idempotent)

Let $\mathbf{A}$ be an $(n \times n)$ symmetric matrix. Then $\mathbf{A}$ is idempotent if and only if $\mathbf{AA} = \mathbf{A}^2 = \mathbf{A}$.

So, an idempotent matrix is a matrix which, when multiplied by itself, produces itself. It's useful when you want to understand the linear regression *geometrically*.

Idempotent example: projection matrix

Definition (Projection)

Let $\mathbf{X}$ be an $(n \times k)$ full-rank matrix. Then, the **projection** matrix of $\mathbf{X}$ is defined as

$$P \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

- It is "projection" because the product $\mathbf{Py}$ for an $(n \times 1)$ vector $\mathbf{y}$ is the projection of $\mathbf{y}$ onto the space "spanned" by the columns of $\mathbf{X}$ (roughly, a span describes the whole space reachable by linear combinations of some given vectors. $\mathbf{X}$).
- Let's check if it's idempotent (i.e., prove $\mathbf{PP} = \mathbf{P}$):

$$\mathbf{PP} = \mathbf{X}\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{=\mathbf{I}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{P}.$$

- In the context of linear regression, we can see the connection:

$$\hat{\beta}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$
$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}_{\text{OLS}} = \underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{=\mathbf{P}_\mathbf{X}}\mathbf{y},$$

so the regression coefficient vector is the results from projecting **y** onto the space spanned by **X**. The projection matrix here is sometimes called "hat" matrix **H**, since when you (left-)multiply it to the data vector **y** it puts a hat on it (Hat typically means some quantity *predicted* by some model, or "fitted values").

Matrix Differentiation

- When we derive a regression model in matrix form, we have to use some partial derivative of a matrix. A **partial derivative** of a function of some variables is its derivative with respect to one of those variables, with the others held constant.
- Let $\mathbf{A}$ be a $(n \times n)$ symmetric matrix, $\mathbf{a}$ be a $(n \times 1)$ vector of constants (i.e., not random variables), and $\mathbf{y}$ be a $(n \times 1)$ vector of **random variables**.

Matrix Differentiation

- Consider a linear function $f(\mathbf{y}) = \mathbf{a}'\mathbf{y} = (a_1 y_1 + \cdots + a_n y_n)$.
- Then the partial derivative of if with regard to the vector $\mathbf{a}$ is:

Matrix Differentiation

- Consider a linear function $f(\mathbf{y}) = \mathbf{a}'\mathbf{y} = (a_1 y_1 + \cdots + a_n y_n)$.
- Then the partial derivative of if with regard to the vector **a** is:

$$\frac{\partial f(\mathbf{y})}{\partial \mathbf{y}} = \frac{\partial \mathbf{a}'\mathbf{y}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial \mathbf{a}'\mathbf{y}}{\partial y_1} \\ \vdots \\ \frac{\partial \mathbf{a}'\mathbf{y}}{\partial y_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial (a_1 y_1 + \cdots + a_n y_n)}{\partial y_1} \\ \vdots \\ \frac{\partial (a_1 y_1 + \cdots + a_n y_n)}{\partial y_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \mathbf{a}.$$

- So matrix (inner) product can loosely be thought of as analogous to constants in univariate calculus, except that a transpose pops out! (if **a** were $(1 \times n)$ row vector: $\frac{\partial \mathbf{a}\mathbf{y}}{\partial \mathbf{y}} = \mathbf{a}'$)

Matrix Differentiation

- Similarly:

$$\frac{\partial \mathbf{y}'\mathbf{y}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial \mathbf{y}'\mathbf{y}}{\partial y_1} \\ \vdots \\ \frac{\partial \mathbf{y}'\mathbf{y}}{\partial y_n} \end{pmatrix} = \begin{pmatrix} 2y_1 \\ \vdots \\ 2y_n \end{pmatrix} = 2\mathbf{y}.$$

Matrix Differentiation

- Note that $\mathbf{A}$ is assumed symmetric. What about $\mathbf{y}'\mathbf{A}\mathbf{y}$ (it's called "quadratic form." If you want to know the derivation, try google. Now you might be able to understand it):

$$\frac{\partial \mathbf{y}'\mathbf{A}\mathbf{y}}{\partial \mathbf{y}} = (\mathbf{A} + \mathbf{A}')\mathbf{y} = 2\mathbf{A}\mathbf{y}.$$

Section 3

Random Vectors

Expectations of random vector/matrix

Definition (Expectation of random vector)

If $\mathbf{X}$ is an $(n \times k)$ random matrix, the expectation of $\mathbf{X}$, denoted by $\mathbb{E}[\mathbf{X}]$, is the matrix of expected values: $\mathbb{E}[\mathbf{X}] = (\mathbb{E}[x_{ij}])$

Expectations of random vector/matrix

Definition (Expectation of random vector)

If $\mathbf{X}$ is an $(n \times k)$ random matrix, the expectation of $\mathbf{X}$, denoted by $\mathbb{E}[\mathbf{X}]$, is the matrix of expected values: $\mathbb{E}[\mathbf{X}] = (\mathbb{E}[x_{ij}])$

- If $\mathbf{y}$ is a $(k \times 1)$ random vector, $\mathbf{A}$ is an $(n \times k)$ non-random matrix, and $\mathbf{b}$ is a non-random $(n \times 1)$ vector, then

$$\mathbb{E}[\mathbf{Ay} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{y}] + \mathbf{b}$$

- If $\mathbf{X}$ is a $(n \times k)$ random matrix, $\mathbf{A}$ is an $(m \times n)$ non-random matrix, and $\mathbf{B}$ is a non-random $(k \times p)$ non-random matrix, then

$$\mathbb{E}[\mathbf{AXB}] = \mathbf{A}\mathbb{E}[\mathbf{X}]\mathbf{B}$$

- That is, we can take out constants out of expectation (same as random variable case)
- You'll need it to prove the unbiasedness of the OLS estimator

Variance of random vector/matrix

Definition (Variance of random vector)

If **y** is an $(n \times 1)$ random vector, its variance-covariance matrix is defined as

$$\mathbb{V}(\mathbf{y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix},$$

where $\sigma_i^2 = \mathbb{V}(y_j)$ and $\sigma_{ij} = \mathcal{C}\wr\sqsubseteq(y_i, y_j)$.

- Can you notice some obvious properties as a matrix?

Variance of random vector/matrix

Definition (Variance of random vector)

If $\mathbf{y}$ is an $(n \times 1)$ random vector, its variance-covariance matrix is defined as

$$\mathbb{V}(\mathbf{y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix},$$

where $\sigma_i^2 = \mathbb{V}(y_j)$ and $\sigma_{ij} = \mathcal{C}\wr\sqsubseteq(y_i, y_j)$.

- Can you notice some obvious properties as a matrix? It's square and symmetric!

Some properties

- Let $\mathbf{a}$ be an $(n \times 1)$ non-random vector and $\mathbf{y}$ an $(n \times 1)$ random vector. Then it follows that

1. $\mathbb{V}(\mathbf{y}) = \mathbb{E}[(\mathbf{y} - \mu)(\mathbf{y} - \mu)']$, where $\mu = \mathbb{E}[\mathbf{y}]$.
2. If the elements of $\mathbf{y}$ are not correlated with each other, then $\mathbb{V}(\mathbf{y})$ is diagonal:
$$\mathbb{V}(\mathbf{y}) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$
3. If the diagonal elements are equal, say $\sigma_{ii} = \sigma$, then $\mathbb{V}(\mathbf{y}) = \sigma^2 \mathbf{I}_n$.
4. If $\mathbf{A}$ is a $(k \times n)$ non-random matrix and $\mathbf{b}$ is an $(n \times 1)$ non-random vector, then
$$\mathbb{V}(\mathbf{A}\mathbf{y} + \mathbf{b}) = \mathbf{A}[\mathbb{V}(\mathbf{y})]\mathbf{A}'$$

- You'll need some of them to derive the variance of the OLS estimator.

Section 4

Regression in matrix form

Setting

- The (multiple) regression model in matrix form is

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

where

$$\underset{(n\times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \underset{(n\times(k+1))}{\mathbf{X}} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix},$$

$$\underset{((k+1)\times 1)}{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}, \underset{(n\times 1)}{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Prepare for derivation

- Then $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the solution to the least square problem:

$$\hat{\beta}_{OLS} = \underset{\beta}{\text{argmin}}(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

- We call $\mathbf{Y} - \mathbf{X}\beta = \epsilon$ as error terms, which is defined by the model ($\mathbf{y} = \mathbf{X}\beta + \epsilon$).
- Recall that inner product of a matrix is like "squaring" the matrix: so we call

$$\epsilon'\epsilon = \begin{pmatrix} \epsilon_1 & \cdots & \epsilon_n \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} = \epsilon_1^2 + \cdots + \epsilon_n^2 = \sum_{i=1}^{n} \epsilon_i$$

as the "sum of squared errors." We wish to find $\hat{\beta}$ that minimize it $\epsilon'\epsilon$

Derivation and others

- Recall that a minimum is obtained by the so-called first order condition (i.e., take first derivative w.r.t $\hat{\beta}$, the parameter vector we wish to estimate, and set it to zero). Solving for $\hat{\beta}$ gives the OLS estimator $\hat{\beta}_{\text{OLS}}$.
- Please review the Pset 1 OLS problems, with what we've gone over today, for practice. FYI: Link here

Derivation and others

- Recall that a minimum is obtained by the so-called first order condition (i.e., take first derivative w.r.t $\hat{\beta}$, the parameter vector we wish to estimate, and set it to zero). Solving for $\hat{\beta}$ gives the OLS estimator $\hat{\beta}_{\text{OLS}}$.
- Please review the Pset 1 OLS problems, with what we've gone over today, for practice. FYI: Link here
- Do we still need this? For example, here is another least squares estimator:

$$\hat{\beta}_{\text{IV}} = (\mathbf{Z}'\mathbf{D})^{-1}\mathbf{Z}'\mathbf{Y},$$

where Z is instrumental variable.

1. Regress **D** on **Z**. Grab the projection, $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D} \quad \cdots (1)$
2. Regress **y** on **Z**. Grab the projection, $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \quad \cdots (2)$
3. Divide (2) by (1)

$$\hat{\beta}_{\text{IV}} = \frac{\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}}{\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}} = (\mathbf{Z}'\mathbf{D})^{-1}\mathbf{Z}'\mathbf{Y}$$