Econon 318 – Econometrics

Richard Schwinn

Spring 2015
MW 4:15-5:30 p.m.
Section 1

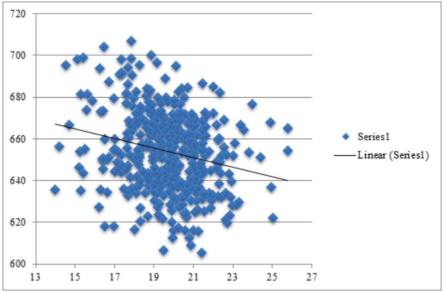
Text: *A Guide to Basic Econometric Techniques* by Elia Kacapyr

In God we trust, all others bring data.

–William Edwards Deming (1900-1993)

- This is data from the California public school system.
- The y-axis measures average test scores in classrooms for the range of student teacher ratios listed on the x-axis.
- What do you think explains this figure?

e



Simple Linear Regression

- We quantify the linear relationship between x and y by finding the equation of the line that "best" fits the data.
- That equation will be written in the form

$$\hat{y} = a + bx.$$

- The variable y represents the value that was actually observed.
- The variable $\hat{y}$ represents the value of y that is predicted by the model.

Simple Linear Regression

- Many possible lines will look *pretty good*.
- To choose the best one, we need to measure how well a line fits the data.
- How do we measure how well a line fits the data?

Linear Algebra

- Suppose I have data on longevity, education, income, and average temperature in the region where subjects live.
- How might I organize this information?
- How can I test whether the data agree with my intuition regarding these values?

- Suppose I have data on longevity, education, income, and average temperature in the region where subjects live.
- How might I organize this information?
- How can I test whether the data agree with my intuition regarding these values?

- Suppose I have data on longevity, education, income, and average temperature in the region where subjects live.
- How might I organize this information?
- How can I test whether the data agree with my intuition regarding these values?

- The following slides represent roughly 4 weeks of linear algebra compressed into one lecture. **To learn more see Hefferon's excellent and free text.**
- You don't need to memorize any definitions or operations. Just try to experience them in class.
- The important thing is to take away is the relationship between the observations in the dataframe:

$$\begin{pmatrix} y_{11} & x_{11} & x_{12} & \cdots & x_{1m} \\ y_{21} & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{n1} & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

and the data arranged into a into a linear model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

- The following slides represent roughly 4 weeks of linear algebra compressed into one lecture. **To learn more see Hefferon's excellent and free text.**
- You don't need to memorize any definitions or operations. Just try to experience them in class.
- The important thing is to take away is the relationship between the observations in the dataframe:

$$\begin{pmatrix} y_{11} & x_{11} & x_{12} & \dots & x_{1m} \\ y_{21} & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{n1} & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

and the data arranged into a into a linear model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

- The following slides represent roughly 4 weeks of linear algebra compressed into one lecture. **To learn more see Hefferon's excellent and free text.**
- You don't need to memorize any definitions or operations. Just try to experience them in class.
- The important thing is to take away is the relationship between the observations in the dataframe:

$$\begin{pmatrix} y_{11} & x_{11} & x_{12} & \cdots & x_{1m} \\ y_{21} & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{n1} & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

and the data arranged into a into a linear model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

- The following slides represent roughly 4 weeks of linear algebra compressed into one lecture. **To learn more see Hefferon's excellent and free text.**
- You don't need to memorize any definitions or operations. Just try to experience them in class.
- The important thing is to take away is the relationship between the observations in the dataframe:

$$\begin{pmatrix} y_{11} & x_{11} & x_{12} & \cdots & x_{1m} \\ y_{21} & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{n1} & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

and the data arranged into a into a linear model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

- The following slides represent roughly 4 weeks of linear algebra compressed into one lecture. **To learn more see Hefferon's excellent and free text.**
- You don't need to memorize any definitions or operations. Just try to experience them in class.
- The important thing is to take away is the relationship between the observations in the dataframe:

$$\begin{pmatrix} y_{11} & x_{11} & x_{12} & \cdots & x_{1m} \\ y_{21} & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{n1} & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

and the data arranged into a into a linear model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

Matrices

- Matrix: A rectangular array of numbers, e.g., $A \in \mathbb{R}^{n \times m}$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

Matrices

- Matrix: A rectangular array of numbers, e.g., $A \in \mathbb{R}^{n \times m}$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

Vectors

- Vector: A matrix consisting of only one column or one row, e.g., $x \in \mathbb{R}^n$

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

- Optionally, underset numbers tell us the number of rows in a matrix followed by its number of columns. e.g. $\underset{m,n}{A}$

Vectors

- Vector: A matrix consisting of only one column or one row, e.g., $x \in \mathbb{R}^n$

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

- Optionally, underset numbers tell us the number of rows in a matrix followed by its number of columns. e.g. $\underset{m,n}{A}$

Vectors

- Vector: A matrix consisting of only one column or one row, e.g., $x \in \mathbb{R}^n$

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

- Optionally, underset numbers tell us the number of rows in a matrix followed by its number of columns. e.g. $\underset{m,n}{A}$

Matrix and Vector Addition

- Matrix addition for a 2 by 2 matrices:

$$\underset{2,2}{A} + \underset{2,2}{B} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix} \quad (1)$$

$$= \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix} \quad (2)$$

- Now you try:

$$C = \begin{pmatrix} 9 & 1 \\ 2 & 3 \end{pmatrix}, \; D = \begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix} \rightarrow C + D = ?$$

Matrix and Vector Addition

- Matrix addition for a 2 by 2 matrices:

$$\underset{2,2}{A} + \underset{2,2}{B} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix} \quad (1)$$

$$= \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix} \quad (2)$$

- Now you try:

$$C = \begin{pmatrix} 9 & 1 \\ 2 & 3 \end{pmatrix}, \; D = \begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix} \rightarrow C + D = ?$$

Matrix and Vector Addition

- Matrix addition for a 2 by 2 matrices:

$$\underset{2,2}{A} + \underset{2,2}{B} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix} \quad (1)$$

$$= \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix} \quad (2)$$

- Now you try:

$$C = \begin{pmatrix} 9 & 1 \\ 2 & 3 \end{pmatrix}, \; D = \begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix} \rightarrow C + D = ?$$

Matrix and Vector Addition

- Vector addition for vectors of length 3:

$$\underset{3,1}{x} + \underset{3,1}{y} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} \quad (3)$$

$$= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix} \quad (4)$$

- Now you try:

$$v = \begin{pmatrix} 9 \\ 1 \\ 2 \end{pmatrix}, \ w = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \rightarrow v + w = ?$$

Econometrics

Notes 02

Matrix and Vector Addition

- Vector addition for vectors of length 3:

$$\underset{3,1}{x} + \underset{3,1}{y} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} \tag{3}$$

$$= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix} \tag{4}$$

- Now you try:

$$v = \begin{pmatrix} 9 \\ 1 \\ 2 \end{pmatrix}, \; w = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \rightarrow v + w = ?$$

Matrix and Vector Addition

- Vector addition for vectors of length 3:

$$\underset{3,1}{x} + \underset{3,1}{y} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} \tag{3}$$

$$= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix} \tag{4}$$

- Now you try:

$$v = \begin{pmatrix} 9 \\ 1 \\ 2 \end{pmatrix}, \; w = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \rightarrow v + w = ?$$

Scalar Multiplication

- A matrix with one row and one column is called a scalar. This is the same thing as the normal definition of a number that we're used to.
- When a matrix is multiplied by a scalar, every number in the array is multiplied by the scalar. Suppose c is a scalar$\rightarrow$

$$c * A = c * \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} c * a_{11} & c * a_{12} \\ c * a_{21} & c * a_{22} \end{pmatrix}$$

- For example:

$$5 * A = c * \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 10 \\ 15 & 20 \end{pmatrix}$$

Scalar Multiplication

- A matrix with one row and one column is called a scalar. This is the same thing as the normal definition of a number that we're used to.
- When a matrix is multiplied by a scalar, every number in the array is multiplied by the scalar. Suppose c is a scalar→

$$c * A = c * \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} c * a_{11} & c * a_{12} \\ c * a_{21} & c * a_{22} \end{pmatrix}$$

- For example:

$$5 * A = c * \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 10 \\ 15 & 20 \end{pmatrix}$$

Scalar Multiplication

- A matrix with one row and one column is called a scalar. This is the same thing as the normal definition of a number that we're used to.
- When a matrix is multiplied by a scalar, every number in the array is multiplied by the scalar. Suppose c is a scalar$\rightarrow$

$$c * A = c * \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} c * a_{11} & c * a_{12} \\ c * a_{21} & c * a_{22} \end{pmatrix}$$

- For example:

$$5 * A = c * \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 10 \\ 15 & 20 \end{pmatrix}$$

Econometrics

Notes 02

Matrix Multiplication

- If $\underset{m,n}{A} \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $C = AB$, then $C \in \mathbb{R}^{m \times p}$: $C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$.
- Special cases: Matrix-vector product, inner product of two vectors. e.g., with $x, y \in \mathbb{R}^n$:

$$x^T y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$$

- The product of two vectors is a scalar and equal to the length of the hypotenuse of the triangle formed by placing the vectors end-to-end.

$$v = \begin{pmatrix} 9 \\ 1 \\ 2 \end{pmatrix}, \; w = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \rightarrow v^T w = (9)(3) + (1)(4) + (2)(5) = 41$$

Matrix Multiplication

- If $\underset{m,n}{A} \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $C = AB$, then $C \in \mathbb{R}^{m \times p}$: $C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$.
- Special cases: Matrix-vector product, inner product of two vectors. e.g., with $x, y \in \mathbb{R}^n$:

$$x^T y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$$

- The product of two vectors is a scalar and equal to the length of the hypotenuse of the triangle formed by placing the vectors end-to-end.

$$v = \begin{pmatrix} 9 \\ 1 \\ 2 \end{pmatrix}, \ w = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \rightarrow v^T w = (9)(3) + (1)(4) + (2)(5) = 41$$

Matrix Multiplication

- If $\underset{m,n}{A} \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $C = AB$, then $C \in \mathbb{R}^{m \times p}$: $C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$.
- Special cases: Matrix-vector product, inner product of two vectors. e.g., with $x, y \in \mathbb{R}^n$:

$$x^T y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$$

- The product of two vectors is a scalar and equal to the length of the hypotenuse of the triangle formed by placing the vectors end-to-end.

$$v = \begin{pmatrix} 9 \\ 1 \\ 2 \end{pmatrix},\ w = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \rightarrow v^T w = (9)(3) + (1)(4) + (2)(5) = 41$$

Operators and Properties

Transposing a matrix swaps the row and column coordinates of each element of a matrix.

- Transpose: $A \in \mathbb{R}^{m \times n}$, then $A^T \in \mathbb{R}^{n \times m}$: $(A^T)_{ij} = A_{ji}$
- Properties:
  - $(A^T)^T = A$
  - $(AB)^T = B^T A^T$
  - $(A + B)^T = A^T + B^T$

The trace is just the sum of the diagonal of a matrix.

- Trace: $A \in \mathbb{R}^{n \times n}$, then: $tr(A) = \sum_{i=1}^{n} A_{ii}$
- Properties:
  - $tr(A) = tr(A^T)$
  - $tr(A + B) = tr(A) + tr(B)$
  - $tr(\lambda A) = \lambda tr(A)$
  - If AB is a square matrix, $tr(AB) = tr(BA)$

Operators and Properties

Transposing a matrix swaps the row and column coordinates of each element of a matrix.

- Transpose: $A \in \mathbb{R}^{m \times n}$, then $A^T \in \mathbb{R}^{n \times m}$: $(A^T)_{ij} = A_{ji}$
- Properties:
  - $(A^T)^T = A$
  - $(AB)^T = B^T A^T$
  - $(A + B)^T = A^T + B^T$

The trace is just the sum of the diagonal of a matrix.

- Trace: $A \in \mathbb{R}^{n \times n}$, then: $tr(A) = \sum_{i=1}^{n} A_{ii}$
- Properties:
  - $tr(A) = tr(A^T)$
  - $tr(A + B) = tr(A) + tr(B)$
  - $tr(\lambda A) = \lambda tr(A)$
  - If AB is a square matrix, $tr(AB) = tr(BA)$

Operators and Properties

Transposing a matrix swaps the row and column coordinates of each element of a matrix.

- Transpose: $A \in \mathbb{R}^{m \times n}$, then $A^T \in \mathbb{R}^{n \times m}$: $(A^T)_{ij} = A_{ji}$
- Properties:
  - $(A^T)^T = A$
  - $(AB)^T = B^T A^T$
  - $(A+B)^T = A^T + B^T$

The trace is just the sum of the diagonal of a matrix.

- Trace: $A \in \mathbb{R}^{n \times n}$, then: $tr(A) = \sum_{i=1}^{n} A_{ii}$
- Properties:
  - $tr(A) = tr(A^T)$
  - $tr(A+B) = tr(A) + tr(B)$
  - $tr(\lambda A) = \lambda tr(A)$
  - If AB is a square matrix, $tr(AB) = tr(BA)$

Operators and Properties

Transposing a matrix swaps the row and column coordinates of each element of a matrix.

- Transpose: $A \in \mathbb{R}^{m \times n}$, then $A^T \in \mathbb{R}^{n \times m}$: $(A^T)_{ij} = A_{ji}$
- Properties:
  - $(A^T)^T = A$
  - $(AB)^T = B^T A^T$
  - $(A + B)^T = A^T + B^T$

The trace is just the sum of the diagonal of a matrix.

- Trace: $A \in \mathbb{R}^{n \times n}$, then: $tr(A) = \sum_{i=1}^{n} A_{ii}$
- Properties:
  - $tr(A) = tr(A^T)$
  - $tr(A + B) = tr(A) + tr(B)$
  - $tr(\lambda A) = \lambda tr(A)$
  - If AB is a square matrix, $tr(AB) = tr(BA)$

Properties of Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B + C) = AB + AC$
- Non-commutative: $AB \neq BA$

Properties of Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B + C) = AB + AC$
- Non-commutative: $AB \neq BA$

Properties of Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B+C) = AB + AC$
- Non-commutative: $AB \neq BA$

Special types of matrices

- Identity matrix: $I = I_n \in \mathbb{R}^{n \times n}$:

$$I_{ij} = \begin{cases} 1 & \text{i=j,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\forall A \in \mathbb{R}^{m \times n}$: $AI_n = I_m A = A$
- Symmetric matrices: $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$.

Special types of matrices

- Identity matrix: $I = I_n \in \mathbb{R}^{n \times n}$:

$$I_{ij} = \begin{cases} 1 & \text{i=j,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\forall A \in \mathbb{R}^{m \times n}$: $AI_n = I_m A = A$
- Symmetric matrices: $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$.

- At this point, the arrangement of the data into a model should be clearer to you.
- Begin with the data: n observations, 1 response, and m variables.

$$\begin{pmatrix} y_{11} & x_{11} & x_{12} & \dots & x_{1m} \\ y_{21} & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{n1} & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

- Next it's arranged into a into a model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

- At this point, the arrangement of the data into a model should be clearer to you.
- Begin with the data: n observations, 1 response, and m variables.

$$\begin{pmatrix} y_{11} & x_{11} & x_{12} & \dots & x_{1m} \\ y_{21} & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{n1} & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

- Next it's arranged into a into a model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

- After the data is arranged into a into a model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \quad (5)$$

$$\underset{n,1}{Y} = \underset{n,m}{X} \underset{m,1}{B} \quad (6)$$

- Perform matrix multiplication and note that every entry in the first column of the X matrix is multiplied by β_0, every entry in the second column (i.e. those that correspond to the first x variable) by β_1, and so on:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} \beta_0 & \beta_1 x_{11} & \dots & \beta_m x_{1m} \\ \beta_0 & \beta_1 x_{21} & \dots & \beta_m x_{2m} \\ \vdots & \vdots & & \vdots \\ \beta_0 & \beta_1 x_{n1} & \dots & \beta_m x_{nm} \end{pmatrix}$$

- After the data is arranged into a into a model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \tag{5}$$

$$\underset{n,1}{Y} = \underset{n,m}{X} \underset{m,1}{B} \tag{6}$$

- Perform matrix multiplication and note that every entry in the first column of the X matrix is multiplied by β_0, every entry in the second column (i.e. those that correspond to the first x variable) by β_1, and so on:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} \beta_0 & \beta_1 x_{11} & \dots & \beta_m x_{1m} \\ \beta_0 & \beta_1 x_{21} & \dots & \beta_m x_{2m} \\ \vdots & \vdots & & \vdots \\ \beta_0 & \beta_1 x_{n1} & \dots & \beta_m x_{nm} \end{pmatrix}$$

- After the data is arranged into a into a model:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \tag{5}$$

$$\underset{n,1}{Y} = \underset{n,m}{X} \underset{m,1}{B} \tag{6}$$

- Perform matrix multiplication and note that every entry in the first column of the X matrix is multiplied by β_0, every entry in the second column (i.e. those that correspond to the first x variable) by β_1, and so on:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} \beta_0 & \beta_1 x_{11} & \dots & \beta_m x_{1m} \\ \beta_0 & \beta_1 x_{21} & \dots & \beta_m x_{2m} \\ \vdots & \vdots & & \vdots \\ \beta_0 & \beta_1 x_{n1} & \dots & \beta_m x_{nm} \end{pmatrix}$$

Matrix Inversion (Division)

Let $A = \begin{pmatrix} 4 & 3 \\ 3 & 2 \end{pmatrix}$ and $A^{-1} = \begin{pmatrix} -2 & 3 \\ 3 & -4 \end{pmatrix}$ and consider AA^{-1}

$$AA^{-1} = \begin{pmatrix} 4 & 3 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 3 & -4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Matrix Inversion (Division)

Let $A = \begin{pmatrix} 4 & 3 \\ 3 & 2 \end{pmatrix}$ and $A^{-1} = \begin{pmatrix} -2 & 3 \\ 3 & -4 \end{pmatrix}$ and consider AA^{-1}

$$AA^{-1} = \begin{pmatrix} 4 & 3 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 3 & -4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Matrix Inversion (Division)

- If $A \in \mathbb{R}^{n \times n}$, then the inverse of A, denoted A^{-1} is the matrix that: $AA^{-1} = A^{-1}A = I$. Recall that $IA = A$ for all conformable A.
- Properties:
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^T = (A^T)^{-1}$
- There is a problem in solving $Y = XB$. We can't simply multiply the inverse to both sides ($X^{-1}Y = X^{-1}XB$) to get B.
- Can anyone tell me why?*For bonus points? There is a hint on this page.*

Matrix Inversion (Division)

- If $A \in \mathbb{R}^{n \times n}$, then the inverse of A, denoted A^{-1} is the matrix that: $AA^{-1} = A^{-1}A = I$. Recall that $IA = A$ for all conformable A.
- Properties:
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^T = (A^T)^{-1}$
- There is a problem in solving $Y = XB$. We can't simply multiply the inverse to both sides ($X^{-1}Y = X^{-1}XB$) to get B.
- Can anyone tell me why? *For bonus points? There is a hint on this page.*

Matrix Inversion (Division)

- If $A \in \mathbb{R}^{n \times n}$, then the inverse of A, denoted A^{-1} is the matrix that: $AA^{-1} = A^{-1}A = I$. Recall that $IA = A$ for all conformable A.
- Properties:
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^T = (A^T)^{-1}$
- There is a problem in solving $Y = XB$. We can't simply multiply the inverse to both sides ($X^{-1}Y = X^{-1}XB$) to get B.
- Can anyone tell me why?*For bonus points? There is a hint on this page.*

Matrix Inversion (Division)

- If $A \in \mathbb{R}^{n \times n}$, then the inverse of A, denoted A^{-1} is the matrix that: $AA^{-1} = A^{-1}A = I$. Recall that $IA = A$ for all conformable A.
- Properties:
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^T = (A^T)^{-1}$
- There is a problem in solving $Y = XB$. We can't simply multiply the inverse to both sides ($X^{-1}Y = X^{-1}XB$) to get B.
- Can anyone tell me why?*For bonus points? There is a hint on this page.*

Matrix Inversion (Division)

- If $A \in \mathbb{R}^{n \times n}$, then the inverse of A, denoted A^{-1} is the matrix that: $AA^{-1} = A^{-1}A = I$. Recall that $IA = A$ for all conformable A.
- Properties:
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^{T} = (A^{T})^{-1}$
- There is a problem in solving $Y = XB$. We can't simply multiply the inverse to both sides $(X^{-1}Y = X^{-1}XB)$ to get B.
- Can anyone tell me why?*For bonus points? There is a hint on this page.*

Multiple Linear Regression

- The problem is that X is not a square matrix. This means that it does not have an equal number of rows and columns, so it cannot be inverted.
- The solution is a simple trick:
  - When you multiply a matrix by its transpose, the result is square.
  - So we multiply both sides of the equation by the transpose of X before inverting.

$$X^TY = X^TXB \tag{7}$$

$$(X^TX)^{-1}X^TY = (X^TX)^{-1}(X^TX)B \tag{8}$$

$$(X^TX)^{-1}X^TY = IB = B \tag{9}$$

- The best fitting fitting hyper-plane (or simply line, in the case of one x-variable, i.e. m=1, simple linear regression) is based on the parameters estimated using only X and Y of this equation: $\widehat{B} = (X^TX)^{-1}X^TY$.
- Now you've learned the most ubiquitous technique in academic research.

Multiple Linear Regression

- The problem is that X is not a square matrix. This means that it does not have an equal number of rows and columns, so it cannot be inverted.
- The solution is a simple trick:
  - When you multiply a matrix by its transpose, the result is square.
  - So we multiply both sides of the equation by the transpose of X before inverting.

$$X^TY = X^TXB \tag{7}$$

$$(X^TX)^{-1}X^TY = (X^TX)^{-1}(X^TX)B \tag{8}$$

$$(X^TX)^{-1}X^TY = IB = B \tag{9}$$

- The best fitting fitting hyper-plane (or simply line, in the case of one x-variable, i.e. m=1, simple linear regression) is based on the parameters estimated using only X and Y of this equation: $\widehat{B} = (X^TX)^{-1}X^TY$.
- Now you've learned the most ubiquitous technique in academic research.

Multiple Linear Regression

- The problem is that X is not a square matrix. This means that it does not have an equal number of rows and columns, so it cannot be inverted.
- The solution is a simple trick:
  - When you multiply a matrix by its transpose, the result is square.
  - So we multiply both sides of the equation by the transpose of X before inverting.

$$X^TY = X^TXB \tag{7}$$

$$(X^TX)^{-1}X^TY = (X^TX)^{-1}(X^TX)B \tag{8}$$

$$(X^TX)^{-1}X^TY = IB = B \tag{9}$$

- The best fitting fitting hyper-plane (or simply line, in the case of one x-variable, i.e. m=1, simple linear regression) is based on the parameters estimated using only X and Y of this equation: $\widehat{B} = (X^TX)^{-1}X^TY$.
- Now you've learned the most ubiquitous technique in academic research.

Multiple Linear Regression

- The problem is that X is not a square matrix. This means that it does not have an equal number of rows and columns, so it cannot be inverted.
- The solution is a simple trick:
  - When you multiply a matrix by its transpose, the result is square.
  - So we multiply both sides of the equation by the transpose of X before inverting.

$$X^T Y = X^T X B \tag{7}$$

$$(X^T X)^{-1} X^T Y = (X^T X)^{-1}(X^T X)B \tag{8}$$

$$(X^T X)^{-1} X^T Y = IB = B \tag{9}$$

- The best fitting fitting hyper-plane (or simply line, in the case of one x-variable, i.e. m=1, simple linear regression) is based on the parameters estimated using only X and Y of this equation: $\widehat{B} = (X^T X)^{-1} X^T Y$.
- Now you've learned the most ubiquitous technique in academic research.

Multiple Linear Regression

- The problem is that X is not a square matrix. This means that it does not have an equal number of rows and columns, so it cannot be inverted.
- The solution is a simple trick:
  - When you multiply a matrix by its transpose, the result is square.
  - So we multiply both sides of the equation by the transpose of X before inverting.

$$X^TY = X^TXB \tag{7}$$

$$(X^TX)^{-1}X^TY = (X^TX)^{-1}(X^TX)B \tag{8}$$

$$(X^TX)^{-1}X^TY = IB = B \tag{9}$$

- The best fitting fitting hyper-plane (or simply line, in the case of one x-variable, i.e. m=1, simple linear regression) is based on the parameters estimated using only X and Y of this equation: $\widehat{B} = (X^TX)^{-1}X^TY$.
- Now you've learned the most ubiquitous technique in academic research.

Multiple Linear Regression

- The problem is that X is not a square matrix. This means that it does not have an equal number of rows and columns, so it cannot be inverted.
- The solution is a simple trick:
  - When you multiply a matrix by its transpose, the result is square.
  - So we multiply both sides of the equation by the transpose of X before inverting.

$$X^T Y = X^T X B \tag{7}$$

$$(X^T X)^{-1} X^T Y = (X^T X)^{-1} (X^T X) B \tag{8}$$

$$(X^T X)^{-1} X^T Y = IB = B \tag{9}$$

- The best fitting fitting hyper-plane (or simply line, in the case of one x-variable, i.e. m=1, simple linear regression) is based on the parameters estimated using only X and Y of this equation: $\widehat{B} = (X^T X)^{-1} X^T Y$.
- Now you've learned the most ubiquitous technique in academic research.

Terminology 01

1. **Dependent variable (Y–variable)** – In an econometric model, this variable appears to the left of the equality sign. It is affected by the independent variable.
2. **Econometric model** (structural equation or regression equation) – A mathematical expression that captures the essence of the cause – and–effect relationship between two variables.
3. **Error term (residual or disturbance)** – This variable is attached to the end of an econometric model. It captures the difference between the observed value of the Y–variable and the value predicted by the econometric model.
4. **Independent variable (X–variable)** – In an econometric model, this variable appears to the right of the equality sign. It is affects by the dependent variable.
5. **Normal equation** – An equation that comes up in the derivation of the formulas for the ordinary least–squares estimators.

Terminology 01

1. **Dependent variable (Y–variable)** – In an econometric model, this variable appears to the left of the equality sign. It is affected by the independent variable.
2. **Econometric model** (structural equation or regression equation) – A mathematical expression that captures the essence of the cause – and–effect relationship between two variables.
3. **Error term (residual or disturbance)** – This variable is attached to the end of an econometric model. It captures the difference between the observed value of the Y–variable and the value predicted by the econometric model.
4. **Independent variable (X–variable)** – In an econometric model, this variable appears to the right of the equality sign. It is affects by the dependent variable.
5. **Normal equation** – An equation that comes up in the derivation of the formulas for the ordinary least–squares estimators.

Terminology 01

1. **Dependent variable (Y–variable)** – In an econometric model, this variable appears to the left of the equality sign. It is affected by the independent variable.
2. **Econometric model** (structural equation or regression equation) – A mathematical expression that captures the essence of the cause – and–effect relationship between two variables.
3. **Error term (residual or disturbance)** – This variable is attached to the end of an econometric model. It captures the difference between the observed value of the Y–variable and the value predicted by the econometric model.
4. **Independent variable (X–variable)** – In an econometric model, this variable appears to the right of the equality sign. It is affects by the dependent variable.
5. **Normal equation** – An equation that comes up in the derivation of the formulas for the ordinary least–squares estimators.

Terminology 01

1. **Dependent variable (Y–variable)** – In an econometric model, this variable appears to the left of the equality sign. It is affected by the independent variable.
2. **Econometric model** (structural equation or regression equation) – A mathematical expression that captures the essence of the cause – and–effect relationship between two variables.
3. **Error term (residual or disturbance)** – This variable is attached to the end of an econometric model. It captures the difference between the observed value of the Y–variable and the value predicted by the econometric model.
4. **Independent variable (X–variable)** – In an econometric model, this variable appears to the right of the equality sign. It is affects by the dependent variable.
5. **Normal equation** – An equation that comes up in the derivation of the formulas for the ordinary least–squares estimators.

Terminology 01

1. **Dependent variable (Y–variable)** – In an econometric model, this variable appears to the left of the equality sign. It is affected by the independent variable.
2. **Econometric model** (structural equation or regression equation) – A mathematical expression that captures the essence of the cause – and–effect relationship between two variables.
3. **Error term (residual or disturbance)** – This variable is attached to the end of an econometric model. It captures the difference between the observed value of the Y–variable and the value predicted by the econometric model.
4. **Independent variable (X–variable)** – In an econometric model, this variable appears to the right of the equality sign. It is affects by the dependent variable.
5. **Normal equation** – An equation that comes up in the derivation of the formulas for the ordinary least–squares estimators.

Terminology 02

1. **Ordinary least–squares** – A technique for estimating the structural parameters of an econometric model. This technique minimizes Σe_i^2 (
2. **Population regression function** – An econometric model estimated with error–free data that includes the entire population of interest.
3. **Sample regression function** – An econometric model estimated from sample data.
4. **Stochastic variable** – A variable that can take on different values depending on the sample data. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are stochastic variables, as are the e_i's.
5. **Structural parameter** – In an econometric model, $\widehat{beta_0}$ and $\widehat{\beta}_1$ are the structural parameters.

Terminology 02

1. **Ordinary least–squares** – A technique for estimating the structural parameters of an econometric model. This technique minimizes Σe_i^2 (
2. **Population regression function** – An econometric model estimated with error–free data that includes the entire population of interest.
3. **Sample regression function** – An econometric model estimated from sample data.
4. **Stochastic variable** – A variable that can take on different values depending on the sample data. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are stochastic variables, as are the e_i's.
5. **Structural parameter** – In an econometric model,$\widehat{beta_0}$ and $\widehat{\beta}_1$ are the structural parameters.

Terminology 02

1. **Ordinary least–squares** – A technique for estimating the structural parameters of an econometric model. This technique minimizes Σe_i^2 (
2. **Population regression function** – An econometric model estimated with error–free data that includes the entire population of interest.
3. **Sample regression function** – An econometric model estimated from sample data.
4. **Stochastic variable** – A variable that can take on different values depending on the sample data. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are stochastic variables, as are the e_i's.
5. **Structural parameter** – In an econometric model,$\widehat{beta_0}$ and $\widehat{\beta}_1$ are the structural parameters.

Terminology 02

1. **Ordinary least–squares** – A technique for estimating the structural parameters of an econometric model. This technique minimizes Σe_i^2 (
2. **Population regression function** – An econometric model estimated with error–free data that includes the entire population of interest.
3. **Sample regression function** – An econometric model estimated from sample data.
4. **Stochastic variable** – A variable that can take on different values depending on the sample data. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are stochastic variables, as are the e_i's.
5. **Structural parameter** – In an econometric model, $\widehat{beta_0}$ and $\widehat{\beta}_1$ are the structural parameters.

Terminology 02

1. **Ordinary least–squares** – A technique for estimating the structural parameters of an econometric model. This technique minimizes Σe_i^2 (
2. **Population regression function** – An econometric model estimated with error–free data that includes the entire population of interest.
3. **Sample regression function** – An econometric model estimated from sample data.
4. **Stochastic variable** – A variable that can take on different values depending on the sample data. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are stochastic variables, as are the e_i's.
5. **Structural parameter** – In an econometric model,$\widehat{beta_0}$ and $\widehat{\beta}_1$ are the structural parameters.

- *A Guide to Basic Econometric Techniques* by Elia Kacapyr
- **To learn more about linear algebra see Hefferon's excellent and free text.**
- Anonymous MIT notes on linear algebra (add link here).

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists \{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists \{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists\{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Linear Independence and Rank

- A set of vectors $\{x_1, \dots, x_n\}$ is linearly independent if $\nexists\{\alpha_1, \dots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A+B) \leq rank(A) + rank(B)$

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists \{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Properties of Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B + C) = AB + AC$
- Non-commutative: $AB \neq BA$
- Block multiplication: If $A = [A_{ik}]$, $B = [B_{kj}]$, where A_{ik}'s and B_{kj}'s are matrix blocks, and the number of columns in A_{ik} is equal to the number of rows in B_{kj}, then $C = AB = [C_{ij}]$ where $C_{ij} = \sum_k A_{ik} B_{kj}$
  Example: If $\overrightarrow{x} \in \mathbb{R}^n$ and $A = [\overrightarrow{a_1} | \overrightarrow{a_2} | \dots | \overrightarrow{a_n}] \in \mathbb{R}^{m \times n}$, $B = [\overrightarrow{b_1} | \overrightarrow{b_2} | \dots | \overrightarrow{b_p}] \in \mathbb{R}^{n \times p}$:

$$A\overrightarrow{x} = \sum_{i=1}^{n} x_i \overrightarrow{a_i}$$

$$AB = [A\overrightarrow{b_1} | A\overrightarrow{b_2} | \dots | A\overrightarrow{b_p}]$$

Properties of Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B+C) = AB + AC$
- Non-commutative: $AB \neq BA$
- Block multiplication: If $A = [A_{ik}]$, $B = [B_{kj}]$, where A_{ik}'s and B_{kj}'s are matrix blocks, and the number of columns in A_{ik} is equal to the number of rows in B_{kj}, then $C = AB = [C_{ij}]$ where $C_{ij} = \sum_k A_{ik}B_{kj}$
  Example: If $\overrightarrow{x} \in \mathbb{R}^n$ and $A = [\overrightarrow{a_1}|\overrightarrow{a_2}|\ldots|\overrightarrow{a_n}] \in \mathbb{R}^{m\times n}$, $B = [\overrightarrow{b_1}|\overrightarrow{b_2}|\ldots|\overrightarrow{b_p}] \in \mathbb{R}^{n\times p}$:

$$A\overrightarrow{x} = \sum_{i=1}^{n} x_i \overrightarrow{a_i}$$

$$AB = [A\overrightarrow{b_1}|A\overrightarrow{b_2}|\ldots|A\overrightarrow{b_p}]$$

Properties of Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B + C) = AB + AC$
- Non-commutative: $AB \neq BA$
- Block multiplication: If $A = [A_{ik}]$, $B = [B_{kj}]$, where A_{ik}'s and B_{kj}'s are matrix blocks, and the number of columns in A_{ik} is equal to the number of rows in B_{kj}, then $C = AB = [C_{ij}]$ where $C_{ij} = \sum_k A_{ik} B_{kj}$
  Example: If $\overrightarrow{x} \in \mathbb{R}^n$ and $A = [\overrightarrow{a_1} | \overrightarrow{a_2} | \ldots | \overrightarrow{a_n}] \in \mathbb{R}^{m \times n}$, $B = [\overrightarrow{b_1} | \overrightarrow{b_2} | \ldots | \overrightarrow{b_p}] \in \mathbb{R}^{n \times p}$:

$$A\overrightarrow{x} = \sum_{i=1}^{n} x_i \overrightarrow{a_i}$$

$$AB = [A\overrightarrow{b_1} | A\overrightarrow{b_2} | \ldots | A\overrightarrow{b_p}]$$

Properties of Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B+C) = AB + AC$
- Non-commutative: $AB \neq BA$
- Block multiplication: If $A = [A_{ik}]$, $B = [B_{kj}]$, where A_{ik}'s and B_{kj}'s are matrix blocks, and the number of columns in A_{ik} is equal to the number of rows in B_{kj}, then $C = AB = [C_{ij}]$ where $C_{ij} = \sum_k A_{ik}B_{kj}$

Example: If $\overrightarrow{x} \in \mathbb{R}^n$ and $A = [\overrightarrow{a_1}|\overrightarrow{a_2}|\ldots|\overrightarrow{a_n}] \in \mathbb{R}^{m\times n}$, $B = [\overrightarrow{b_1}|\overrightarrow{b_2}|\ldots|\overrightarrow{b_p}] \in \mathbb{R}^{n\times p}$:

$$A\overrightarrow{x} = \sum_{i=1}^{n} x_i \overrightarrow{a_i}$$

$$AB = [A\overrightarrow{b_1}|A\overrightarrow{b_2}|\ldots|A\overrightarrow{b_p}]$$

Properties of Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B + C) = AB + AC$
- Non-commutative: $AB \neq BA$
- Block multiplication: If $A = [A_{ik}]$, $B = [B_{kj}]$, where A_{ik}'s and B_{kj}'s are matrix blocks, and the number of columns in A_{ik} is equal to the number of rows in B_{kj}, then $C = AB = [C_{ij}]$ where $C_{ij} = \sum_k A_{ik}B_{kj}$
  Example: If $\overrightarrow{x} \in \mathbb{R}^n$ and $A = [\overrightarrow{a_1} | \overrightarrow{a_2} | \ldots | \overrightarrow{a_n}] \in \mathbb{R}^{m \times n}$, $B = [\overrightarrow{b_1} | \overrightarrow{b_2} | \ldots | \overrightarrow{b_p}] \in \mathbb{R}^{n \times p}$:

$$A\overrightarrow{x} = \sum_{i=1}^{n} x_i \overrightarrow{a_i}$$

$$AB = [A\overrightarrow{b_1} | A\overrightarrow{b_2} | \ldots | A\overrightarrow{b_p}]$$

Special types of matrices

- Identity matrix: $I = I_n \in \mathbb{R}^{n \times n}$:

$$I_{ij} = \begin{cases} 1 & \text{i=j,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\forall A \in \mathbb{R}^{m \times n}$: $AI_n = I_m A = A$
- Diagonal matrix: $D = diag(d_1, d_2, \ldots, d_n)$:

$$D_{ij} = \begin{cases} d_i & \text{j=i,} \\ 0 & \text{otherwise.} \end{cases}$$

- Symmetric matrices: $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$.
- Orthogonal matrices: $U \in \mathbb{R}^{n \times n}$ is orthogonal if $UU^T = I = U^T U$

Special types of matrices

- Identity matrix: $I = I_n \in \mathbb{R}^{n \times n}$:

$$I_{ij} = \begin{cases} 1 & \text{i=j,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\forall A \in \mathbb{R}^{m \times n}$: $AI_n = I_m A = A$
- Diagonal matrix: $D = diag(d_1, d_2, \ldots, d_n)$:

$$D_{ij} = \begin{cases} d_i & \text{j=i,} \\ 0 & \text{otherwise.} \end{cases}$$

- Symmetric matrices: $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$.
- Orthogonal matrices: $U \in \mathbb{R}^{n \times n}$ is orthogonal if $UU^T = I = U^T U$

Special types of matrices

- Identity matrix: $I = I_n \in \mathbb{R}^{n \times n}$:

$$I_{ij} = \begin{cases} 1 & \text{i=j,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\forall A \in \mathbb{R}^{m \times n}$: $AI_n = I_m A = A$
- Diagonal matrix: $D = diag(d_1, d_2, \ldots, d_n)$:

$$D_{ij} = \begin{cases} d_i & \text{j=i,} \\ 0 & \text{otherwise.} \end{cases}$$

- Symmetric matrices: $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$.
- Orthogonal matrices: $U \in \mathbb{R}^{n \times n}$ is orthogonal if $UU^T = I = U^T U$

Special types of matrices

- Identity matrix: $I = I_n \in \mathbb{R}^{n \times n}$:

$$I_{ij} = \begin{cases} 1 & \text{i=j,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\forall A \in \mathbb{R}^{m \times n}$: $AI_n = I_m A = A$
- Diagonal matrix: $D = diag(d_1, d_2, \dots, d_n)$:

$$D_{ij} = \begin{cases} d_i & \text{j=i,} \\ 0 & \text{otherwise.} \end{cases}$$

- Symmetric matrices: $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$.
- Orthogonal matrices: $U \in \mathbb{R}^{n \times n}$ is orthogonal if $UU^T = I = U^T U$

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists \{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists\{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists \{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists\{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Linear Independence and Rank

- A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists\{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$.
  - To understand this point think of each vector as a line segment pointing in a particular direction.
  - The note above says that if you place the vectors end-to-end, while maintaining their directions, that you can't arrange them in a way that they meet any of the other vectors, even if you're allowed to stretch them,
- Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)
- Properties:
  - $rank(A) \leq \min\{m, n\}$
  - $rank(A) = rank(A^T)$
  - $rank(AB) \leq \min\{rank(A), rank(B)\}$
  - $rank(A + B) \leq rank(A) + rank(B)$

Range and Nullspace of a Matrix

- Span: $span(\{x_1, \ldots, x_n\}) = \{\sum_{i=1}^{n} \alpha_i x_i | \alpha_i \in \mathbb{R}\}$
- Projection: $Proj(y; \{x_i\}_{1 \leq i \leq n}) = argmin_{v \in span(\{x_i\}_{1 \leq i \leq n})}\{||y - v||_2\}$
- Range: $A \in \mathbb{R}^{m \times n}$, then $\mathcal{R}(A) = \{Ax | x \in R^n\}$ is the span of the columns of A
- $Proj(y, A) = A(A^T A)^{-1} A^T y$
- Nullspace: $null(A) = \{x \in \mathbb{R}^n | Ax = 0\}$

Range and Nullspace of a Matrix

- Span: $span(\{x_1, \ldots, x_n\}) = \{\sum_{i=1}^{n} \alpha_i x_i | \alpha_i \in \mathbb{R}\}$
- Projection: $Proj(y; \{x_i\}_{1 \leq i \leq n}) = argmin_{v \in span(\{x_i\}_{1 \leq i \leq n})}\{||y - v||_2\}$
- Range: $A \in \mathbb{R}^{m \times n}$, then $\mathcal{R}(A) = \{Ax | x \in R^n\}$ is the span of the columns of A
- $Proj(y, A) = A(A^T A)^{-1} A^T y$
- Nullspace: $null(A) = \{x \in \mathbb{R}^n | Ax = 0\}$

Range and Nullspace of a Matrix

- Span: $span(\{x_1, \ldots, x_n\}) = \{\sum_{i=1}^{n} \alpha_i x_i | \alpha_i \in \mathbb{R}\}$
- Projection: $Proj(y; \{x_i\}_{1 \leq i \leq n}) = argmin_{v \in span(\{x_i\}_{1 \leq i \leq n})}\{||y - v||_2\}$
- Range: $A \in \mathbb{R}^{m \times n}$, then $\mathcal{R}(A) = \{Ax | \, x \in R^n\}$ is the span of the columns of A
- $Proj(y, A) = A(A^T A)^{-1} A^T y$
- Nullspace: $null(A) = \{x \in \mathbb{R}^n | \, Ax = 0\}$

Range and Nullspace of a Matrix

- Span: $span(\{x_1, \ldots, x_n\}) = \{\sum_{i=1}^{n} \alpha_i x_i | \alpha_i \in \mathbb{R}\}$
- Projection: $Proj(y; \{x_i\}_{1 \leq i \leq n}) = argmin_{v \in span(\{x_i\}_{1 \leq i \leq n})}\{||y - v||_2\}$
- Range: $A \in \mathbb{R}^{m \times n}$, then $\mathcal{R}(A) = \{Ax | x \in R^n\}$ is the span of the columns of A
- $Proj(y, A) = A(A^T A)^{-1} A^T y$
- Nullspace: $null(A) = \{x \in \mathbb{R}^n | Ax = 0\}$

Range and Nullspace of a Matrix

- Span: $span(\{x_1, \ldots, x_n\}) = \{\sum_{i=1}^{n} \alpha_i x_i | \alpha_i \in \mathbb{R}\}$
- Projection: $Proj(y; \{x_i\}_{1 \leq i \leq n}) = argmin_{v \in span(\{x_i\}_{1 \leq i \leq n})}\{||y - v||_2\}$
- Range: $A \in \mathbb{R}^{m \times n}$, then $\mathcal{R}(A) = \{Ax | x \in R^n\}$ is the span of the columns of A
- $Proj(y, A) = A(A^T A)^{-1} A^T y$
- Nullspace: $null(A) = \{x \in \mathbb{R}^n | Ax = 0\}$

Determinant

- $A \in \mathbb{R}^{n \times n}$, $a_1, \ldots, a_n$ the rows of A, then $det(A)$ is the volume of $S = \{\sum_{i=1}^{n} \alpha_i a_i | \, 0 \leq \alpha_i \leq 1\}$.
- Properties:
  - $det(I) = 1$
  - $det(\lambda A) = \lambda det(A)$
  - $det(A^T) = det(A)$
  - $det(AB) = det(A)det(B)$
  - $det(A) \neq 0$ if and only if A is invertible.
  - If A invertible, then $det(A^{-1}) = det(A)^{-1}$

Quadratic Forms and Positive Semidefinite Matrices

- $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$, $x^T Ax$ is called a quadratic form:

$$x^T Ax = \sum_{1 \le i,j \le n} A_{ij} x_i x_j$$

- A is positive definite if $\forall x \in \mathbb{R}^n : x^T Ax > 0$
- A is positive semidefinite if $\forall x \in \mathbb{R}^n : x^T Ax \ge 0$
- A is negative definite if $\forall x \in \mathbb{R}^n : x^T Ax < 0$
- A is negative semidefinite if $\forall x \in \mathbb{R}^n : x^T Ax \le 0$

Quadratic Forms and Positive Semidefinite Matrices

- $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$, $x^T A x$ is called a quadratic form:

$$x^T A x = \sum_{1 \leq i,j \leq n} A_{ij} x_i x_j$$

- A is positive definite if $\forall\, x \in \mathbb{R}^n : x^T A x > 0$
- A is positive semidefinite if $\forall\, x \in \mathbb{R}^n : x^T A x \geq 0$
- A is negative definite if $\forall\, x \in \mathbb{R}^n : x^T A x < 0$
- A is negative semidefinite if $\forall\, x \in \mathbb{R}^n : x^T A x \leq 0$

Eigenvalues and Eigenvectors

- $A \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of A with the corresponding eigenvector $x \in \mathbb{C}^n$ $(x \neq 0)$ if:
$$Ax = \lambda x$$
- eigenvalues: the n possibly complex roots of the polynomial equation $det(A - \lambda I) = 0$, and denoted as $\lambda_1, \ldots, \lambda_n$
- Properties:
  - $tr(A) = \sum_{i=1}^{n} \lambda_i$
  - $det(A) = \prod_{i=1}^{n} \lambda_i$
  - $rank(A) = |\{1 \leq i \leq n | \lambda_i \neq 0\}|$

Matrix Eigendecomposition

- $A \in \mathbb{R}^{n \times n}$, $\lambda_1, \dots, \lambda_n$ the eigenvalues, and $x_1, \dots, x_n$ the eigenvectors. $X = [x_1|x_2|\dots|x_n]$, $\Lambda = diag(\lambda_1, \dots, \lambda_n)$, then $AX = X\Lambda$.
- A called diagonalizable if X invertible: $A = X\Lambda X^{-1}$
- If A symmetric, then all eigenvalues real, and X orthogonal (hence denoted by $U = [u_1|u_2|\dots|u_n]$):

$$A = U\Lambda U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$$

- A special case of Singular Value Decomposition

Optimization

- A set of points S is convex if, for any $x, y \in S$ and for any $0 \leq \theta \leq 1$,

$$\theta x + (1 - \theta) y \in S$$

- A function $f : S \to \mathbb{R}$ is convex if its domain S is a convex set and

$$f(\theta x + (1 - \theta) y) \leq \theta f(x) + (1 - \theta) f(y)$$

  for all $x, y \in S$, $0 \leq \theta \leq 1$.
- A function $f : S \to \mathbb{R}$ is submodular if for any subset $A \subseteq B$,

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$$

- Convex functions can easily be minimized. Submodular functions allow approximate discrete optimization.

Proofs

- Induction:
  1. Show result on base case, associated with $n = k_0$
  2. Assume result true for $n \leq i$. Prove result for $n = i + 1$
  3. Conclude result true for all $n \geq k_0$

  Example: In a complete graph, $E = \frac{1}{2}N(N-1)$
- Contradiction (reductio ad absurdum):
  1. Assume result is false
  2. Follow implications in a deductive manner, until a contradiction is reached
  3. Conclude initial assumption was wrong, hence result true

  Example: Strongly connected components partition nodes

Graph theory

- Definitions: vertex/node, edge/link, loop/cycle, degree, path, neighbor, tree, clique,. . .
- Random graph (Erdos-Renyi): Each possible edge is present with some probability p
- (Strongly) connected component: subset of nodes that can all reach each other
- Diameter: longest minimum distance between two nodes
- Bridge: edge connecting two otherwise disjoint connected components

Basic algorithms

- BFS: explore by “layers”
- DFS: go as far as possible, then backtrack
- Greedy: maximize goal at each step
- Binary search: on ordered set, discard half of the elements at each step

Complexity

- Number of operations as a function of the problem parameters.
- Examples
  1. Find shortest path between two nodes:
     - DFS: very bad idea, could end up with the whole graph as a single path
     - BFS from origin: good idea
     - BFS from origin and destination: even better!
  2. Given a node, find its connected component
     - Loop over nodes: bad idea, needs N path searches
     - BFS or DFS: good idea