EE2211 Introduction to Machine Learning

Lecture 4

Semester 2
2022/2023

Vincent Tan
vtan@nus.edu.sg

Electrical and Computer Engineering Department
National University of Singapore

Acknowledgement:
EE2211 development team
(Kar-Ann Toh, Xinchao Wang, Thomas Yeo, Helen Zhou, Chen Khong, Helen Zhou, Robby Tan and Haizhou Li)

About Me

- Associate Professor of Mathematics and ECE
- Joined NUS in 2014
- Undergraduate in Information Engineering from Cambridge University 2005
- Ph.D. in Electrical Engineering and Computer Science (EECS) from MIT in 2011
- Research interests in Information Theory, Signal Processing, and Machine Learning
- I teach probability, stochastic processes, information theory, machine learning and mathematical analysis at NUS.

Course Contents

- Introduction and Preliminaries (Xinchao)
  - Introduction
  - Data Engineering
  - Introduction to Probability, Statistics, and Matrix
- Fundamental Machine Learning Algorithms I (Vincent)
  - Systems of linear equations
  - Least squares, Linear regression
  - Ridge regression, Polynomial regression
- Fundamental Machine Learning Algorithms II (Vincent)
  - Over-fitting, bias/variance trade-off
  - Optimization, Gradient descent
  - Decision Trees, Random Forest
- Performance and More Algorithms (Xinchao)
  - Performance Issues
  - K-means Clustering
  - Neural Networks

Assignment 1 (week 6)
Tutorial 4

Systems of Linear Equations

Module II Contents

- Operations on Vectors and Matrices
- Systems of Linear Equations
- Set and Functions
- Derivative and Gradient
- Least Squares, Linear Regression
- Linear Regression with Multiple Outputs
- Linear Regression for Classification
- Ridge Regression
- Polynomial Regression

Fundamental ML Algorithms: Linear Regression

References for Lectures 4-6:

Main

- [Book1] Andriy Burkov, "**The Hundred-Page Machine Learning Book**", 2019. (**read first, buy later**: http://themlbook.com/wiki/doku.php)
- [Book2] Andreas C. Muller and Sarah Guido, "**Introduction to Machine Learning with Python**: A Guide for Data Scientists", O'Reilly Media, Inc., 2017

Supplementary

- [Book3] Jeff Leek, "**The Elements of Data Analytic Style**: A guide for people who want to analyze data", Lean Publishing, 2015.
- [Book4] Stephen Boyd and Lieven Vandenberghe, "**Introduction to Applied Linear Algebra**", Cambridge University Press, 2018 (available online) http://vmls-book.stanford.edu/
- [Ref 5] **Professor Vincent Tan's notes (chapters 4-6): (**useful**)** **https://vyftan.github.io/papers/ee2211book.pdf**

Recap on Notations, Vectors, Matrices

Scalar	Numerical value	15, -3.5
Variable	Take scalar values	*x* or *a*
Vector	An ordered list of scalar values	$\mathbf{x}$ or $\mathbf{a}$
	Attributes of a vector	$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$
Matrix	A rectangular array of numbers arranged in rows and columns	$\mathbf{X} = \begin{bmatrix} 2 & 4 \\ 21 & -6 \end{bmatrix}$

Capital Sigma $\sum_{i=1}^{m} x_i = x_1 + x_2 + \ldots + x_{m-1} + x_m$

Capital Pi $\prod_{i=1}^{m} x_i = x_1 \cdot x_2 \cdot \ldots \cdot x_{m-1} \cdot x_m$

Operations on Vectors and Matrices

Operations on Vectors: summation and subtraction

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

$$\mathbf{x} - \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \end{bmatrix}$$

Operations on Vectors and Matrices

Operations on Vectors: scalar

$$a\,\mathbf{x} = a\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \end{bmatrix}$$

$$\frac{1}{a}\mathbf{x} = \frac{1}{a}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{a}x_1 \\ \frac{1}{a}x_2 \end{bmatrix}$$

Operations on Vectors and Matrices

Matrix or Vector Transpose:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \mathbf{x}^T = \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$

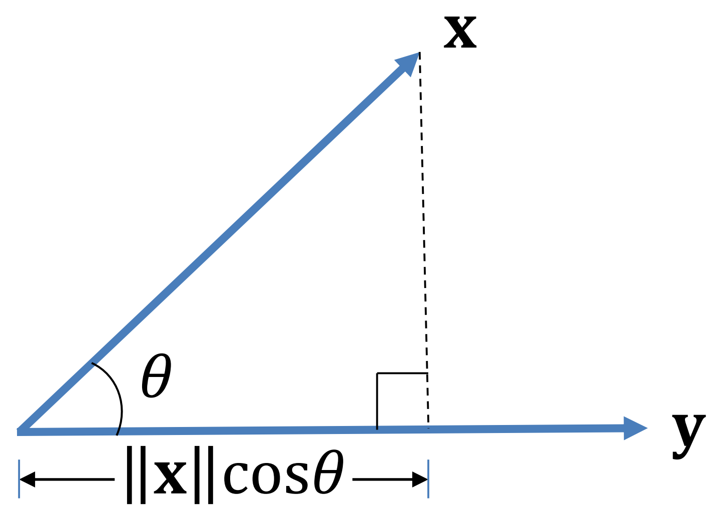
$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{bmatrix}, \quad \mathbf{X}^T = \begin{bmatrix} x_{1,1} & x_{2,1} & x_{3,1} \\ x_{1,2} & x_{2,2} & x_{3,2} \\ x_{1,3} & x_{2,3} & x_{3,3} \end{bmatrix}$$

Python demo 1

Operations on Vectors and Matrices

Dot Product or Inner Product of Vectors:

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$$
$$= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$
$$= x_1 \, y_1 + x_2 y_2$$



Geometric definition:

$$\mathbf{x} \cdot \mathbf{y} = ||\mathbf{x}|| \; ||\mathbf{y}|| \cos\theta$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$,
and $||\mathbf{x}|| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ is the Euclidean length of vector $\mathbf{x}$

$\mathbf{E.g.}\ \mathbf{a} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ➔ $\mathbf{a} \cdot \mathbf{c}$ = 2*1 + 3 *0 = 2

Operations on Vectors and Matrices

Matrix-Vector Product

$$\mathbf{Wx} = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ w_{2,1} & w_{2,2} & w_{2,3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \begin{bmatrix} w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 \\ w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 \end{bmatrix}$$

Operations on Vectors and Matrices

Vector-Matrix Product

$$\mathbf{x}^T\mathbf{W} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ w_{2,1} & w_{2,2} & w_{2,3} \end{bmatrix}$$

$$= \begin{bmatrix} (x_1 w_{1,1} + x_2 w_{2,1}) & (x_1 w_{1,2} + x_2 w_{2,2}) & (x_1 w_{1,3} + x_2 w_{2,3}) \end{bmatrix}$$

Operations on Vectors and Matrices

Matrix-Matrix Product

$$\mathbf{XW} = \begin{bmatrix} x_{1,1} & \dots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,d} \end{bmatrix} \begin{bmatrix} w_{1,1} & \dots & w_{1,h} \\ \vdots & \ddots & \vdots \\ w_{d,1} & \dots & w_{d,h} \end{bmatrix}$$

$$= \begin{bmatrix} (x_{1,1}w_{1,1} + \cdots + x_{1,d}w_{d,1}) & \dots & (x_{1,1}w_{1,h} + \cdots + x_{1,d}w_{d,h}) \\ \vdots & \ddots & \vdots \\ (x_{m,1}w_{1,1} + \cdots + x_{m,d}w_{d,1}) & \dots & (x_{m,1}w_{1,h} + \cdots + x_{m,d}w_{d,h}) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{d} x_{1,i}w_{i,1} & \dots & \sum_{i=1}^{d} x_{1,i}w_{i,h} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{d} x_{m,i}w_{i,1} & \dots & \sum_{i=1}^{d} x_{m,i}w_{i,h} \end{bmatrix}$$

If **X** is *m* x *d* and **W** is *d* x *h, then* the outcome is a *m* x *h* matrix

Operations on Vectors and Matrices

Matrix inverse

Definition:

A *d-by-d* square matrix **A** is **invertible** (also **nonsingular**)

if there exists a *d-by-d* square matrix **B** such that

$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$ (identity matrix)

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \ldots 0 & 0 \\ 0 & 1 \;\; 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \;\; 1 & 0 \\ 0 & 0 \ldots 0 & 1 \end{bmatrix}$$ *d-by-d* dimension

Ref: https://en.wikipedia.org/wiki/Invertible_matrix

Operations on Vectors and Matrices

Matrix inverse computation

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\operatorname{adj}(\mathbf{A})$$

- $\det(\mathbf{A})$ is the determinant of $\mathbf{A}$
- $\operatorname{adj}(\mathbf{A})$ is the adjugate or adjoint of $\mathbf{A}$

Determinant computation

Example: 2x2 matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\det(\mathbf{A}) = |\mathbf{A}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Ref: https://en.wikipedia.org/wiki/Invertible_matrix

Operations on Vectors and Matrices

- $\mathrm{adj}(\mathbf{A})$ is the adjugate **or** adjoint of $\mathbf{A}$
- $\mathrm{adj}(\mathbf{A})$ is the transpose of the **cofactor matrix** $\mathbf{C}$ of $\mathbf{A}$ → $\mathrm{adj}(\mathbf{A}) = \mathbf{C}^{\mathrm{T}}$
- **Minor** of an element in a matrix $\mathbf{A}$ is defined as the determinant obtained by deleting the row and column in which that element lies

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\text{Minor of } \mathrm{a}_{12} \text{ is } M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$$

- The (i, j) entry of the **cofactor matrix** $\mathbf{C}$ is the minor of (i, j) element times a sign factor

$$\text{Cofactor } C_{ij} = (-1)^{i+j} M_{ij}$$

- The **determinant** of $\mathbf{A}$ can also be defined by minors as

$$\det(\mathbf{A}) = \sum_{j=1}^{k} = a_{ij} C_{ij} = (-1)^{i+j} a_{ij} M_{ij}$$

Ref: https://en.wikipedia.org/wiki/Invertible_matrix

Operations on Vectors and Matrices

Minor of a_{12} is $M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$ $\quad$ $\mathrm{adj}(\mathbf{A}) = \mathbf{C}^{\mathrm{T}}$

Cofactor $C_{ij} = (-1)^{i+j} M_{ij}$ $\quad$ $\det(\mathbf{A}) = \sum_{j=1}^{k} (-1)^{i+j} a_{ij} M_{ij}$

- E.g. $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$\mathbf{C} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

- $\mathrm{adj}(\mathbf{A}) = \mathbf{C}^T = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ $\quad$ $\det(\mathbf{A}) = |\mathbf{A}| = ad - bc$

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathrm{adj}(\mathbf{A}) = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Ref: https://en.wikipedia.org/wiki/Invertible_matrix

Operations on Vectors and Matrices

Determinant computation $\det(\mathbf{A}) = \sum_{j=1}^{k} (-1)^{i+j} a_{ij} M_{ij}$

Example: 3x3 matrix, use the first row (i = 1)

$$|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} \square & \square & \square \\ \square & e & f \\ \square & h & i \end{vmatrix} - b \begin{vmatrix} \square & \square & \square \\ d & \square & f \\ g & \square & i \end{vmatrix} + c \begin{vmatrix} \square & \square & \square \\ d & e & \square \\ g & h & \square \end{vmatrix}$$

$$= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

$$= a(ei - fh) - b(di - fg) + c(dh - eg)$$

Python demo 2

Ref: https://en.wikipedia.org/wiki/Determinant

Operations on Vectors and Matrices

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The minor of $a_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix}.$$

Ref: https://en.wikipedia.org/wiki/Determinant

Operations on Vectors and Matrices

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The minor of $a_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ & & \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ & & \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix}.$$

Ref: https://en.wikipedia.org/wiki/Determinant

Operations on Vectors and Matrices

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The minor of $a_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix}.$$

Ref: https://en.wikipedia.org/wiki/Determinant

Operations on Vectors and Matrices

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The minor of $a_{21} = \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix}.$$

Ref: https://en.wikipedia.org/wiki/Determinant

Operations on Vectors and Matrices

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The minor of $a_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix}.$$

Ref: https://en.wikipedia.org/wiki/Determinant

Operations on Vectors and Matrices

Example

Find the cofactor matrix of $\mathbf{A}$ given that $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 0 & 6 \end{bmatrix}$.

Solution:

$$a_{11} \Rightarrow \begin{vmatrix} 4 & 5 \\ 0 & 6 \end{vmatrix} = 24, \quad a_{12} \Rightarrow -\begin{vmatrix} 0 & 5 \\ 1 & 6 \end{vmatrix} = 5, \quad a_{13} \Rightarrow \begin{vmatrix} 0 & 4 \\ 1 & 0 \end{vmatrix} = -4,$$

$$a_{21} \Rightarrow -\begin{vmatrix} 2 & 3 \\ 0 & 6 \end{vmatrix} = -12, \quad a_{22} \Rightarrow \begin{vmatrix} 1 & 3 \\ 1 & 6 \end{vmatrix} = 3, \quad a_{23} \Rightarrow -\begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix} = 2,$$

$$a_{31} \Rightarrow \begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} = -2, \quad a_{32} \Rightarrow -\begin{vmatrix} 1 & 3 \\ 0 & 5 \end{vmatrix} = -5, \quad a_{33} \Rightarrow \begin{vmatrix} 1 & 2 \\ 0 & 4 \end{vmatrix} = 4,$$

The cofactor matrix C is thus $\begin{bmatrix} 24 & 5 & -4 \\ -12 & 3 & 2 \\ -2 & -5 & 4 \end{bmatrix}$.

Ref: https://www.mathwords.com/c/cofactor_matrix.htm

Systems of Linear Equations

Module II Contents

- Operations on Vectors and Matrices
- **Systems of Linear Equations**
- Set and Functions
- Derivative and Gradient
- Least Squares, Linear Regression
- Linear Regression with Multiple Outputs
- Linear Regression for Classification
- Ridge Regression
- Polynomial Regression

EE2211 Lecture 4:
Nature of Solutions of Linear Systems

Vincent Y. F. Tan

Department of Electrical and Computer Engineering,
National University of Singapore

EE2211 Spring 2023

Nature of Solutions of Linear Systems

- A set of linear equations can have no solution, one solution, or multiple solutions.
- The system is characterized by a matrix (called the design matrix) $\mathbf{X} \in \mathbb{R}^{m \times d}$ and a vector (called the target vector) $\mathbf{y} \in \mathbb{R}^m$ and we seek to find (if possible) a solution (the weights) $\mathbf{w} \in \mathbb{R}^d$ satisfying

$$\mathbf{X}\mathbf{w} = \mathbf{y}.$$

- Note that

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,d} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad \text{and} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}.$$

Rank of a Matrix

The rank of a matrix is the number of pivots in its reduced row-echelon form (RREF) (cf. MA1508E). Let's try to find the RREF of the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & -2 & 4 & -2 \\ 2 & 1 & 10 & 7 \\ -4 & 4 & -8 & 4 \\ 4 & -1 & 14 & 6 \end{bmatrix}$$

Perform $R_2 - R_1$ and $2R_1 + R_3$ and $R_4 - 2R_1$. We get

$$\begin{bmatrix} 2 & -2 & 4 & -2 \\ 0 & 3 & 6 & 9 \\ 0 & 0 & 0 & 0 \\ 0 & 3 & 6 & 10 \end{bmatrix}$$

Now do $R_4 - R_2$ and swap R_3 and R_4. We get

$$\mathbf{A}_{\mathrm{REF}} = \begin{bmatrix} 2 & -2 & 4 & -2 \\ 0 & 3 & 6 & 9 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The matrix is in row-echelon form.

Rank of a Matrix

To get it into its RREF, we can divide R_1 by 2, divide R_2 by 3, do $R_3 + R_1$, $R_2 - 3R_3$, $R_2 + R_1$ to get

$$\mathbf{A}_{\text{RREF}} = \begin{bmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

There are 3 pivots, indicated in red above. Hence, the rank of $\mathbf{A}$ is 3.

Alternatively (simpler method), the rank of a matrix is the number of linearly independent rows or columns or $\mathbf{A}$. Note that in $\mathbf{A}$,

$$R_3 = -2R_1$$

so there are no more than three linearly independent rows (e.g., R_1, R_2, and R_4). We need to confirm that R_1, R_2 and R_4 are indeed linearly independent. You can show that R_4 cannot be expressed as a linear combination of R_1 and R_2. Hence, the rank of $\mathbf{A}$ is 3.

Form the augmented matrix

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \mathbf{y} \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d} & y_1 \\ x_{2,1} & x_{2,2} & \dots & x_{2,d} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m,1} & x_{m,2} & \dots & x_{m,d} & y_m \end{bmatrix} \in \mathbb{R}^{m\times(d+1)}.$$

(i) The system $\mathbf{X}\mathbf{w} = \mathbf{y}$ admits a unique solution if and only if

$$\mathrm{rank}(\mathbf{X}) = \mathrm{rank}(\tilde{\mathbf{X}}) = d;$$

(ii) The system $\mathbf{X}\mathbf{w} = \mathbf{y}$ has no solution if and only if

$$\mathrm{rank}(\mathbf{X}) < \mathrm{rank}(\tilde{\mathbf{X}});$$

(iii) The system $\mathbf{X}\mathbf{w} = \mathbf{y}$ has infinitely many solutions if and only if

$$\mathrm{rank}(\mathbf{X}) = \mathrm{rank}(\tilde{\mathbf{X}}) < d.$$

Example of Case (i)

Consider the following even determined system in which $m = 2$ and $d = 2$:

$$\mathbf{X} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} 4 \\ 10 \end{bmatrix}.$$

The augmented matrix is

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \mathbf{y} \end{bmatrix} = \begin{bmatrix} 2 & 1 & 4 \\ 4 & 3 & 10 \end{bmatrix}$$

Then $\mathrm{rank}(\mathbf{X}) = \mathrm{rank}(\tilde{\mathbf{X}}) = 2$ and there is a unique solution.

Example of Case (ii): Usual Situation

Consider the following over-determined system in which $m = 3$ and $d = 2$:

$$\mathbf{X} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \\ 5 & 6 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

The augmented matrix is

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \mathbf{y} \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 2 \\ 5 & 6 & 3 \end{bmatrix}.$$

- In this case, $\mathrm{rank}(\mathbf{X}) = 2$ and $\mathrm{rank}(\tilde{\mathbf{X}}) = 3$.
- This is case (ii) of the RC Theorem and there is no solution.
- This is the usual case for over-determined systems.
- In Python, you can find the rank of a matrix (2D array) $\mathbf{A}$ using `np.linalg.matrix_rank(A)`.

Example of Case (i): Unusual Situation

Consider the following over-determined system in which $m = 3$ and $d = 2$:

$$\mathbf{X} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \\ 5 & 6 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} 4 \\ 10 \\ 17 \end{bmatrix}.$$

- In this case $\mathrm{rank}(\mathbf{X}) = 2$ and $\mathrm{rank}(\tilde{\mathbf{X}}) = 2$.
- This is case (i) of the Rouché-Capelli Theorem and there is a unique solution even though the system is over-determined.
- Note that $\mathbf{y}$ is one times the first column of $\mathbf{X}$ plus two times the second column of $\mathbf{X}$, so it is in the linear span of the columns of $\mathbf{X}$.

Example of Case (iii): Usual Situation

Consider the following under-determined system in which $m = 2$ and $d = 3$:

$$\mathbf{X} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 5 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}.$$

- In this case, $\mathrm{rank}(\mathbf{X}) = 2$ and $\mathrm{rank}(\tilde{\mathbf{X}}) = 2$ but $d = 3$.
- This is case (iii) of the Rouché-Capelli Theorem and there are infinitely many solutions.
- This is the usual case for under-determined systems.

Example of Case (iii): Unusual Situation

Consider the following over-determined system in which $m = 3$ and $d = 2$:

$$\mathbf{X} = \begin{bmatrix} 2 & 1 \\ 4 & 2 \\ 6 & 3 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} 8 \\ 16 \\ 24 \end{bmatrix}.$$

- In this case $\mathrm{rank}(\mathbf{X}) = 1$ and $\mathrm{rank}(\tilde{\mathbf{X}}) = 1$ and both these ranks are less than $d = 2$.
- This is case (iii) of the Rouché-Capelli Theorem and there are infinitely many solutions even though the system is over-determined.
- Note that the three columns of $\tilde{\mathbf{X}}$ are collinear.

Example of Case (iii): Unusual Situation

Consider the following under-determined system in which $m = 2$ and $d = 3$:

$$\mathbf{X} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 6 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

- In this case, $\mathrm{rank}(\mathbf{X}) = 1$ and $\mathrm{rank}(\tilde{\mathbf{X}}) = 2$ because $\mathbf{y}$ is not in the range of $\mathbf{X}$.
- This is case (ii) of the Rouché-Capelli Theorem and there is no solution.
- Note that $\mathbf{y}$ boosts the rank of $\mathbf{X}$ by 1 in the augmented matrix $\tilde{\mathbf{X}}$, i.e., $\mathbf{y}$ is not in the column space of $\mathbf{X}$, which is the ray $\{[t, 2t]^\top : t \in \mathbb{R}\}$.

Questions

- For under-determined systems ($m < d$), briefly explain why we cannot have case (i) of the Rouché-Capelli Theorem?
- We have seen that for over-determined systems ($m > d$) all three cases of the Rouché-Capelli Theorem are possible.
- Why is there an asymmetry here?

Solving A Linear System

- Consider the square or even-determined system

$$\mathbf{X}\mathbf{w} = \mathbf{y} \quad \Longleftrightarrow \quad \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,d} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

  where $\mathbf{X} \in \mathbb{R}^{m \times d}$ and $m = d$.
- Same number of equations m and unknowns d.
- If $\mathrm{rank}(\mathbf{X}) = d = m$, there is a unique solution (why in the context of Rouché–Capelli theorem?).
- In this case, the inverse $\mathbf{X}^{-1}$ exists ($\mathbf{X}$ is invertible or non-singular) and the solution can be found by

$$\mathbf{X}^{-1}\mathbf{X}\hat{\mathbf{w}} = \mathbf{X}^{-1}\mathbf{y} \quad \Longleftrightarrow \quad \hat{\mathbf{w}} = \mathbf{X}^{-1}\mathbf{y}.$$

Example of Square or Even-Determined System

Consider the example with two equations and two unknowns

$$w_1 + w_2 = 4 \quad \text{and} \quad w_1 - 2w_2 = 1$$

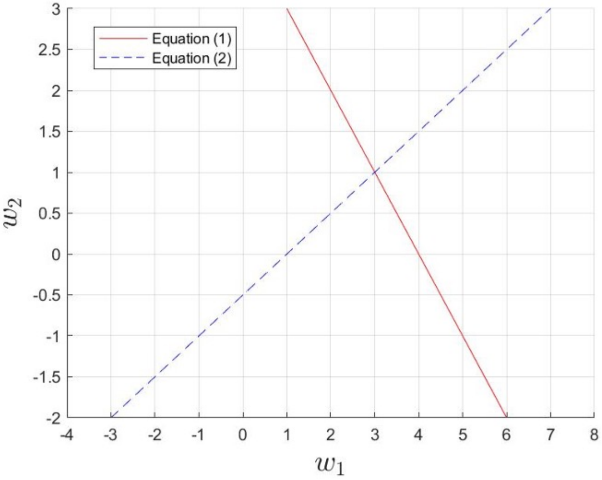
We can rewrite this as

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}}_{=\mathbf{X}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{=\mathbf{w}} = \underbrace{\begin{bmatrix} 4 \\ 1 \end{bmatrix}}_{=\mathbf{y}}$$

Since $\mathbf{X}$ is invertible (why?),

$$\hat{\mathbf{w}} = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \frac{-1}{3} \begin{bmatrix} -2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Example of Square or Even-Determined System



Over-Determined System ($m > d$)

- Now, we have more equations m than unknowns d, i.e., the system of equations

$$\mathbf{X}\mathbf{w} = \mathbf{y} \quad \Longleftrightarrow \quad \begin{bmatrix} x_{1,1} & \cdots & x_{1,d} \\ x_{2,1} & \cdots & x_{2,d} \\ x_{3,1} & \cdots & x_{3,d} \\ \vdots & \ddots & \vdots \\ x_{m-1,1} & \cdots & x_{m-1,d} \\ x_{m,1} & \cdots & x_{m,d} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{m-1} \\ y_m \end{bmatrix}$$

- $\mathbf{X}$ is tall (non-square). Not invertible.
- No unique solution in the usual case (see slide “Example of Case (ii): Usual Situation”).
- Can find an approximate solution – called the least squares solution

Over-Determined Systems ($m > d$): Least Squares

- The least squares solution of an over-determined system $\mathbf{X}\mathbf{w} = \mathbf{y}$ is

$$\hat{\mathbf{w}} = \mathbf{X}^{\dagger}\mathbf{y},$$

where $\mathbf{X}^{\dagger} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}$ is the left-inverse of $\mathbf{X}$.
- The left inverse $\mathbf{X}^{\dagger}$ satisfies $\mathbf{X}^{\dagger}\mathbf{X} = \mathbf{I}$, the identity matrix.
- The left inverse exists if $\mathbf{X}$ has full column rank, i.e., number of linearly independent columns equals number of columns.
- What are the column ranks of

$$\mathbf{X}_1 = \begin{bmatrix} 2 & 1 \\ 4 & 3 \\ 5 & 6 \end{bmatrix} \quad \mathbf{X}_2 = \begin{bmatrix} 2 & 4 \\ 4 & 8 \\ 5 & 10 \end{bmatrix} \quad \mathbf{X}_3 = \begin{bmatrix} 1 & 4 \\ 2 & 8 \\ 3 & 10 \end{bmatrix}?$$

- Important: The least squares solution is

$$\hat{\mathbf{w}} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{y}.$$

Example of Over-Determined System

Consider an example with three equations and two unknowns

$$w_1 + w_2 = 1, \qquad w_1 - w_2 = 0, \qquad w_1 = 2.$$

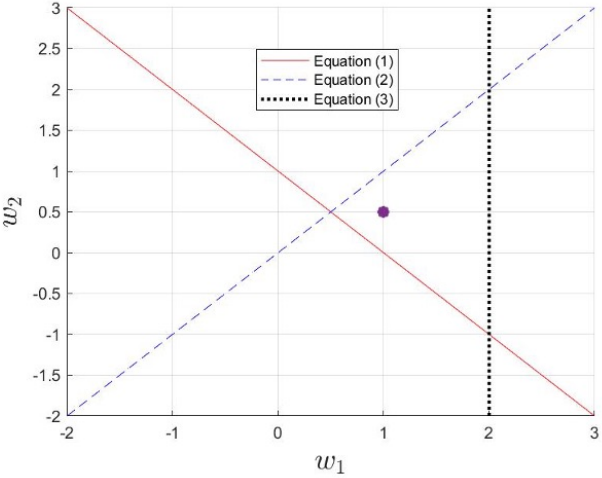
We can rewrite this as

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}}_{=\mathbf{X}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{=\mathbf{w}} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}}_{=\mathbf{y}}$$

Since $\mathbf{X}$ has full column rank ($\mathbf{X}^\top\mathbf{X}$ is invertible)

$$\hat{\mathbf{w}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

Example of Over-Determined System



Under-Determined System ($m < d$)

- Now, we have fewer equations m than unknowns d ($m < d$), i.e., the system of equations

$$\mathbf{X}\mathbf{w} = \mathbf{y} \quad \Longleftrightarrow \quad \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{1,d-1} & x_{1,d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,1} & \cdots & x_{m,d-1} & x_{m,d} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{d-1} \\ w_d \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

- $\mathbf{X}$ is fat (non-square). Not invertible.
- Multiple solutions in the usual case (see slide "Example of Case (iii): Usual Situation").
- Among the many, find one – called the least norm solution.

Under-Determined Systems ($m < d$): Least Norm

- The least norm solution of an under-determined system $\mathbf{X}\mathbf{w} = \mathbf{y}$ is

$$\hat{\mathbf{w}} = \mathbf{X}^{\dagger}\mathbf{y},$$

  where $\mathbf{X}^{\dagger} = \mathbf{X}^{\top}(\mathbf{X}\mathbf{X}^{\top})^{-1}$ is the right inverse of $\mathbf{X}$.
- The right inverse $\mathbf{X}^{\dagger}$ satisfies $\mathbf{X}\mathbf{X}^{\dagger} = \mathbf{I}$, the identity matrix.
- The right inverse exists if $\mathbf{X}$ has full row rank, i.e., number of linearly independent rows equals number of rows.
- What are the row ranks of

$$\mathbf{X}_1 = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 3 & 2 \end{bmatrix} \quad \mathbf{X}_2 = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 6 \end{bmatrix}?$$

- Important: The least norm solution of an under-determined system is

$$\hat{\mathbf{w}} = \mathbf{X}^{\top}(\mathbf{X}\mathbf{X}^{\top})^{-1}\mathbf{y}.$$

Example of Under-Determined System

Consider an example with $m = 2$ equations and $d = 3$ unknowns

$$w_1 + 2w_2 + 3w_3 = 2, \qquad w_1 - 2w_2 + 3w_3 = 1.$$

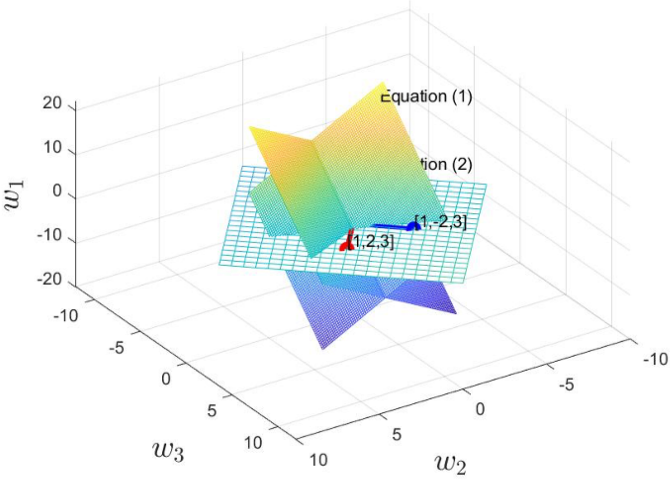
We can rewrite this as

$$\underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 1 & -2 & 3 \end{bmatrix}}_{=\mathbf{X}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}}_{=\mathbf{w}} = \underbrace{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}_{=\mathbf{y}}$$

Since $\mathbf{X}$ has full row rank ($\mathbf{X}\mathbf{X}^\top$ is invertible)

$$\hat{\mathbf{w}} = \mathbf{X}^\top (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{y} = \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 14 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.15 \\ 0.25 \\ 0.45 \end{bmatrix}$$

Example of an Under-Determined System



Final Example

Consider an example with $m = 2$ equations and $d = 3$ unknowns

$$w_1 + 2w_2 + 3w_3 = 2, \qquad 3w_1 + 6w_2 + 9w_3 = 1.$$

i.e.,

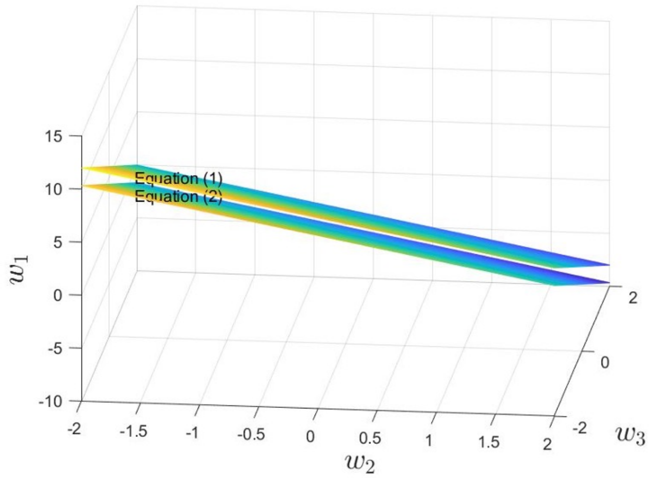
$$\mathbf{X}\mathbf{w} = \mathbf{y} \iff \begin{bmatrix} 1 & 2 & 3 \\ 3 & 6 & 9 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

- Does $\mathbf{X}$ have full column rank/full row rank? No in both cases.
- Neither $(\mathbf{X}^\top \mathbf{X})^{-1}$ nor $(\mathbf{X}\mathbf{X}^\top)^{-1}$ exist. Note that

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} 10 & 20 & 30 \\ 20 & 40 & 60 \\ 30 & 60 & 90 \end{bmatrix} \quad \text{and} \quad \mathbf{X}\mathbf{X}^\top = \begin{bmatrix} 14 & 42 \\ 42 & 126 \end{bmatrix}$$

  No least squares or least norm solution because $\mathrm{rank}(\mathbf{X}^\top \mathbf{X}) = \mathrm{rank}(\mathbf{X}\mathbf{X}^\top) = 1$.
- $1 = \mathrm{rank}(\mathbf{X}) < \mathrm{rank}(\tilde{\mathbf{X}}) = 2$. Case (ii) of RC Theorem.

Final Example



Systems of Linear Equations

<u>Module II Contents</u>

- Operations on Vectors and Matrices
- Systems of Linear Equations
- **Set and Functions**
- Derivative and Gradient
- Least Squares, Linear Regression
- Linear Regression with Multiple Outputs
- Linear Regression for Classification
- Ridge Regression
- Polynomial Regression

Notations: Set

- A **set** is an unordered collection of unique elements
  - Denoted as a calligraphic capital character e.g., $\mathcal{S}, \mathcal{R}, \mathcal{N}$ etc
  - When an element x belongs to a set $\mathcal{S}$, we write $x \in \mathcal{S}$
- A set of numbers can be finite - include a fixed amount of values
  - Denoted using accolades, e.g. {1, 3, 18, 23, 235} or $\{x_1, x_2, x_3, x_4, \ldots, x_d\}$
- A set can be infinite and include all values in some interval
  - If a set of real numbers includes all values between *a* and *b*, including *a* and *b*, it is denoted using square brackets as [*a*, *b*]
  - If the set does not include the values *a* and *b*, it is denoted using parentheses as (*a*, *b*)
- Examples:
  - The special set denoted by $\mathcal{R}$ includes all real numbers from minus infinity to plus infinity
  - The set [0, 1] includes values like 0, 0.0001, 0.25, 0.9995, and 1.0

Ref: [Book1] Andriy Burkov, "The Hundred-Page Machine Learning Book", 2019 (p4 of chp2).

Notations: Set operations

- **Intersection** of two sets:

$$\mathcal{S}_3 \leftarrow \mathcal{S}_1 \cap \mathcal{S}_2$$

Example: {1,3,5,8} ∩ {1,8,4} = {1,8}

- **Union** of two sets:

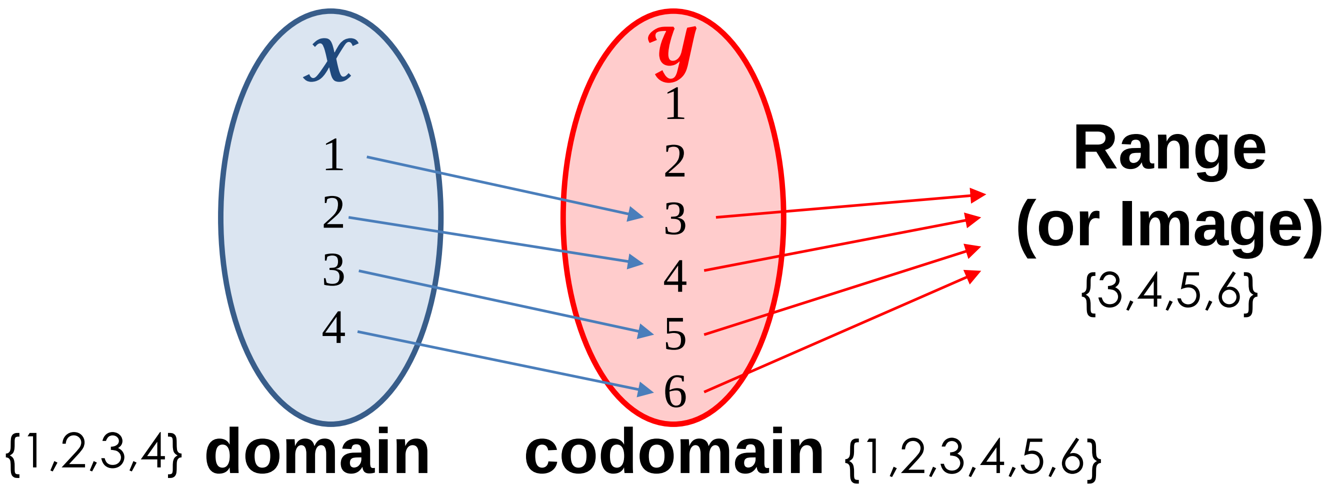
$$\mathcal{S}_3 \leftarrow \mathcal{S}_1 \cup \mathcal{S}_2$$

Example: {1,3,5,8} ∪ {1,8,4} = {1,3,4,5,8}

Ref: [Book1] Andriy Burkov, “The Hundred-Page Machine Learning Book”, 2019 (p4 of chp2).

Functions

- A **function** is a relation that associates each element x of a set $\mathcal{X}$, the **domain** of the function, to a single element y of another set $\mathcal{Y}$, the **codomain** of the function
- If the function is called f, this relation is denoted $y = f(x)$
  - The element x is the argument or input of the function
  - y is the value of the function or the output
- The symbol used for representing the input is the variable of the function
  - $f(x)$ f is a function of the variable x; $f(x, w)$ f is a function of the variable x and w



Ref: [Book1] Andriy Burkov, "The Hundred-Page Machine Learning Book", 2019 (p6 of chp2).

Functions

- A scalar function can have vector argument
  - E.g. $y = f(\mathbf{x}) = x_1 + x_2 + 2x_3$
- A vector function, denoted as $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is a function that returns a vector $\mathbf{y}$
  - Input argument can be a **vector** $\mathbf{y} = \mathbf{f}(\mathbf{x})$ or a **scalar** $\mathbf{y} = \mathbf{f}(x)$
  - E.g. $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ x_2 \end{bmatrix}$
  - E.g. $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -2x_1 \\ 3x_1 \end{bmatrix}$

Ref: [Book1] Andriy Burkov, "The Hundred-Page Machine Learning Book", 2019 (p7 of chp2).

Functions

- The notation $f: \mathcal{R}^d \to \mathcal{R}$ means that f is a function that maps real d-vectors to real numbers
  - i.e., f is a scalar-valued function of d-vectors
- If $\mathbf{x}$ is a d-vector argument, then $f(\mathbf{x})$ denotes the value of the function f at $\mathbf{x}$
  - i.e., $f(\mathbf{x}) = f(x_1, x_2, \dots, x_d)$, $\mathbf{x} \in \mathcal{R}^d$, $f(\mathbf{x}) \in \mathcal{R}$

- Example: we can define a function $f: \mathcal{R}^4 \to \mathcal{R}$ by

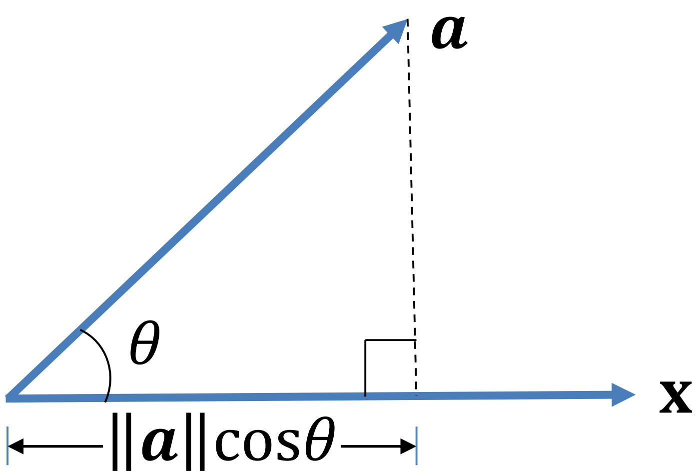
$$f(\mathbf{x}) = x_1 + x_2 - x_4^2$$

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "Introduction to Applied Linear Algebra", 2018 (Ch 2, p29)

Functions

The inner product function

- Suppose $\boldsymbol{a}$ is a *d*-vector. We can define a scalar valued function f of *d*-vectors, given by
$$f(\mathbf{x}) = \boldsymbol{a}^T\mathbf{x} = a_1x_1 + a_2x_2 + \cdots a_dx_d \qquad (1)$$
  for any *d*-vector $\mathbf{x}$
- The inner product of its *d*-vector argument $\mathbf{x}$ with some (fixed) *d*-vector $\boldsymbol{a}$
- We can also think of f as forming a **weighted sum** of the elements of $\mathbf{x}$; the elements of $\boldsymbol{a}$ give the weights



Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "Introduction to Applied Linear Algebra", Cambridge University Press, 2018 (p30)

Functions

Linear Functions

A function $f: \mathcal{R}^d \rightarrow \mathcal{R}$ is **linear** if it satisfies the following two properties:

- **Homogeneity**
  - For any *d*-vector $\mathbf{x}$ and any scalar α, $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$
  - Scaling the (vector) argument is the same as scaling the function value

- **Additivity**
  - For any *d*-vectors $\mathbf{x}$ and $\mathbf{y}$, $f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$
  - Adding (vector) arguments is the same as adding the function values

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "Introduction to Applied Linear Algebra", Cambridge University Press, 2018 (p31)

Functions

Linear Functions

Superposition and linearity

- The inner product function $f(\mathbf{x}) = \boldsymbol{a}^T\mathbf{x}$ defined in equation (1) (slide 9) satisfies the property

$$\begin{aligned} f(\alpha\mathbf{x} + \beta\mathbf{y}) &= \boldsymbol{a}^T(\alpha\mathbf{x} + \beta\mathbf{y}) \\ &= \boldsymbol{a}^T(\alpha\mathbf{x}) + \boldsymbol{a}^T(\beta\mathbf{y}) \\ &= \alpha(\boldsymbol{a}^T\mathbf{x}) + \beta(\boldsymbol{a}^T\mathbf{y}) \\ &= \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) \end{aligned}$$

  for all d-vectors $\mathbf{x}$, $\mathbf{y}$, and all scalars α, β.

- This property is called superposition, which consists of homogeneity and additivity
- A function that satisfies the superposition property is called linear

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "Introduction to Applied Linear Algebra", Cambridge University Press, 2018 (p30)

Functions

Linear Functions

- If a function f is linear, superposition extends to linear combinations of any number of vectors:
$$f(\alpha_1 \mathbf{x}_1 + \cdots + \alpha_k \mathbf{x}_k) = \alpha_1 f(\mathbf{x}_1) + \cdots + \alpha_k f(\mathbf{x}_k)$$
for any d vectors $\mathbf{x}_1 + \cdots + \mathbf{x}_k$, and any scalars $\alpha_1 + \cdots + \alpha_k$.

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "Introduction to Applied Linear Algebra", Cambridge University Press, 2018 (p30)

Functions

Linear and Affine Functions

A linear function plus a constant is called an affine function

A linear function $f: \mathcal{R}^d \rightarrow \mathcal{R}$ is **affine** if and only if it can be expressed as $f(\mathbf{x}) = \boldsymbol{a}^T\mathbf{x} + b$ for some d-vector $\boldsymbol{a}$ and scalar b, which is called the offset (or bias)

Example:

$$f(\mathbf{x}) = 2.3 - 2x_1 + 1.3x_2 - x_3$$

This function is affine, with $b = 2.3$, $\boldsymbol{a}^T = [-2,\ 1.3,\ -1]$.

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "Introduction to Applied Linear Algebra", Cambridge University Press, 2018 (p32)

Functions

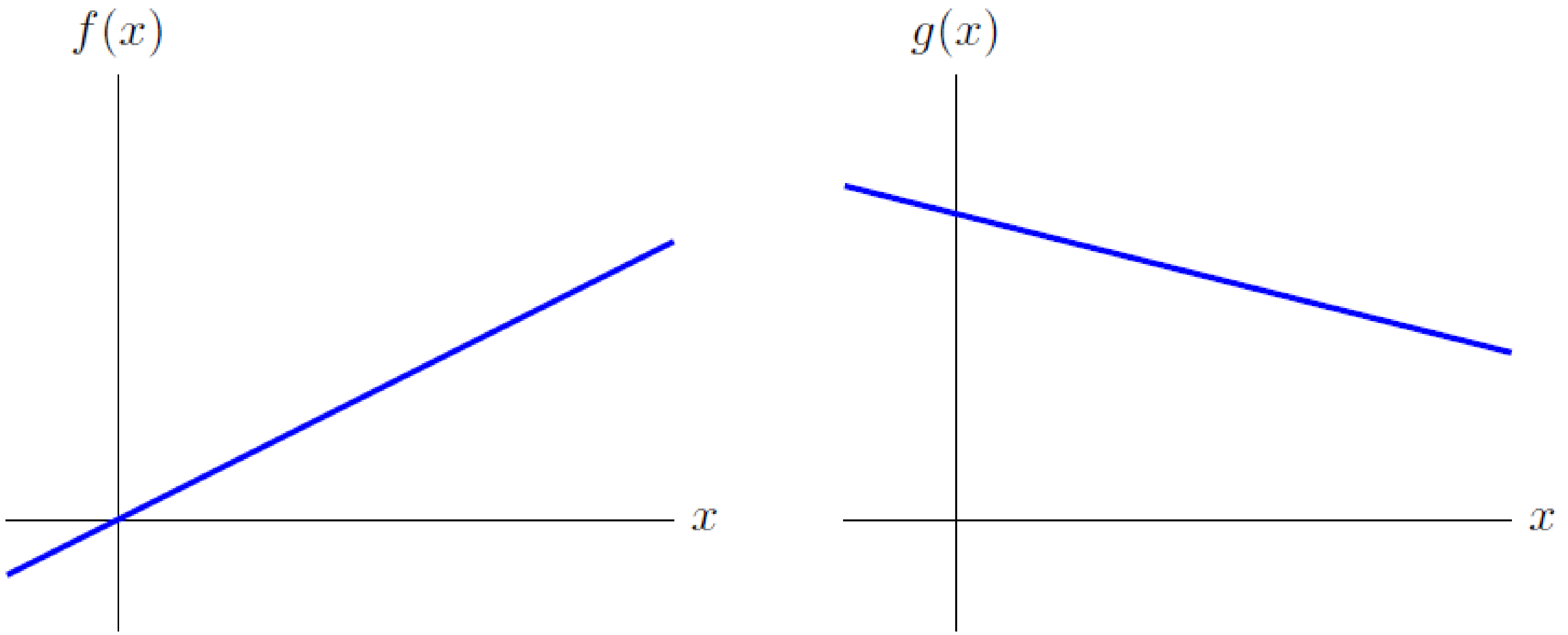


Figure 2.1 *Left.* The function f is linear. *Right.* The function g is affine, but not linear.

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "Introduction to Applied Linear Algebra", Cambridge University Press, 2018 (p33)

Summary

- Operations on Vectors and Matrices
  - Dot-product, matrix inverse
- Systems of Linear Equations $\mathbf{X}\mathbf{w} = \mathbf{y}$
  - Matrix-vector notation, linear dependency, invertible
  - Even-, over-, under-determined linear systems
- Set and Functions

Assignment 1 (week 6)
Tutorial 4

$\mathbf{X}$ is Square	Even-determined	m = d	One unique solution in general	$\hat{\mathbf{w}} = \mathbf{X}^{-1}\mathbf{y}$
$\mathbf{X}$ is Tall	Over-determined	m > d	No exact solution in general; An approximated solution	$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ Left-inverse
$\mathbf{X}$ is Wide	Under-determined	m < d	Infinite number of solutions in general; Unique constrained solution	$\hat{\mathbf{w}} = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{y}$ Right-inverse

  - Scalar and vector functions
  - Inner product function
  - Linear and affine functions

python package *numpy*
Inverse: *numpy.linalg.inv(X)*
Transpose: *X.T*