EE2211 Introduction to Machine Learning

Lecture 4

Semester 2
2023/2024

Yueming Jin

ymjin@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, Vincent Tan, Robby Tan and Haizhou Li)

Course Contents

- Introduction and Preliminaries (Xinchao)
  - Introduction
  - Data Engineering
  - Introduction to Probability, Statistics, and Matrix
- Fundamental Machine Learning Algorithms I (Yueming)
  - Systems of linear equations
  - Least squares, Linear regression
  - Ridge regression, Polynomial regression
- Fundamental Machine Learning Algorithms II (Yueming)
  - 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 7 Wed)
Tutorial 4

Office hour via zoom:
Tuesday 9:30-10:30am

Course Contents

- Introduction and Preliminaries (Xinchao)
  - Introduction
  - Data Engineering
  - Introduction to Probability, Statistics, and Matrix
- Fundamental Machine Learning Algorithms I (Yueming)
  - Systems of linear equations
  - Least squares, Linear regression
  - Ridge regression, Polynomial regression
- Fundamental Machine Learning Algorithms II (Yueming)
  - 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

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:

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

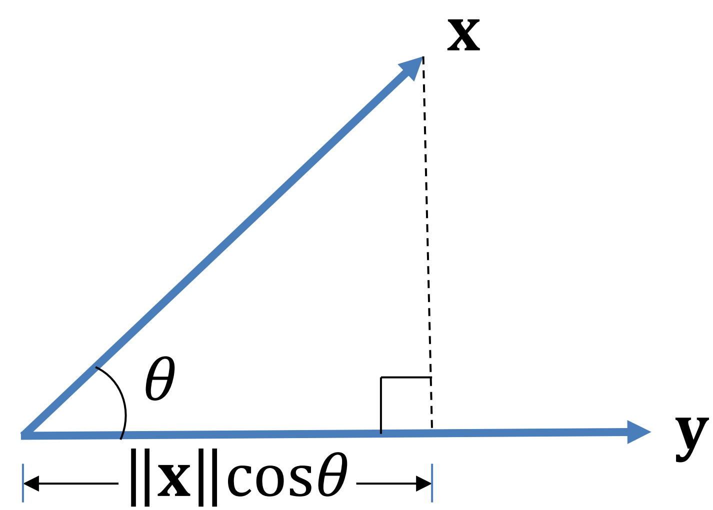
$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{bmatrix}, \quad \mathbf{X}^T = \begin{bmatrix} x_{1,1} & x_{2,1} & x_{3,1} \\ x_{1,2} & x_{2,2} & x_{3,2} \\ x_{1,3} & x_{2,3} & x_{3,3} \end{bmatrix}$$

Python demo 1

Operations on Vectors and Matrices

Dot Product or Inner Product of Vectors:

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



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

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

$$= [(x_1 w_{1,1} + x_2 w_{2,1}) \qquad (x_1 w_{1,2} + x_2 w_{2,2}) \qquad (x_1 w_{1,3} + x_2 w_{2,3})]$$

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)

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \dots 0 & 0 \\ 0 & 1 \quad 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \quad 1 & 0 \\ 0 & 0 \dots 0 & 1 \end{bmatrix}$$ 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

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

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

- 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 $a_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$

Its cofactor matrix is

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

$$\mathrm{adj}(\mathbf{A}) = \mathbf{C}^{\mathrm{T}}$$

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

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

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

Systems of Linear Equations

- Consider a system of m linear equations with d variables or unknowns $w_1, \dots, w_d$:

$$
\begin{gathered}
x_{1,1}w_1 + x_{1,2}w_2 + \cdots + x_{1,d}\, w_d = y_1 \\
x_{2,1}w_1 + x_{2,2}w_2 + \cdots + x_{2,d}\, w_d = y_2 \\
\vdots \\
x_{m,1}w_1 + x_{m,2}w_2 + \cdots + x_{m,d}\, w_d = y_m\,.
\end{gathered}
$$

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "**Introduction to Applied Linear Algebra**", Cambridge University Press, 2018 (Chp8.3)

Systems of Linear Equations

These equations can be written compactly in matrix-vector notation:

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

Where

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

Note:

- The data matrix $\mathbf{X} \in \mathcal{R}^{m \times d}$ and the target vector $\mathbf{y} \in \mathcal{R}^m$ are given
- The unknown vector of parameters $\mathbf{w} \in \mathcal{R}^d$ is to be learnt

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "**Introduction to Applied Linear Algebra**", Cambridge University Press, 2018 (Chp8.3)

Systems of Linear Equations

A set of linear equations can have no solution, one solution, or multiple solutions:

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

Where

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

$\mathbf{X}$ is Square	Even-determined	m = d	Equal number of equations and unknowns
$\mathbf{X}$ is Tall	Over-determined	m > d	More number of equations than unknowns
$\mathbf{X}$ is Wide	Under-determined	m < d	Fewer number of equations than unknowns

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "**Introduction to Applied Linear Algebra**", (Chp8.3 & 11) & [Ref 5] Tan's notes, (Chp 4)

Systems of Linear Equations

$\mathbf{X}\mathbf{w} = \mathbf{y}, \quad \mathbf{X} \in \mathcal{R}^{m\times d}, \mathbf{w} \in \mathcal{R}^{d\times 1}, \mathbf{y} \in \mathcal{R}^{m\times 1}$

1. Square or even-determined system: $m = d$

- Equal number of equations and unknowns, i.e., $\mathbf{X} \in \mathcal{R}^{d\times d}$
- **One unique solution** if $\mathbf{X}$ is invertible or all rows/columns of $\mathbf{X}$ are linearly independent
- If all rows or columns of $\mathbf{X}$ are linearly independent, then $\mathbf{X}$ is invertible.

Solution:

If $\mathbf{X}$ is invertible (or $\mathbf{X}^{-1}\mathbf{X} = \mathbf{I}$), then pre-multiply both sides by $\mathbf{X}^{-1}$

$$\mathbf{X}^{-1}\mathbf{X}\,\mathbf{w} = \mathbf{X}^{-1}\mathbf{y}$$

$$\Rightarrow \qquad \hat{\mathbf{w}} = \mathbf{X}^{-1}\mathbf{y}$$

(Note: we use a *hat* on top of $\mathbf{w}$ to indicate that it is a specific point in the space of $\mathbf{w}$)

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "**Introduction to Applied Linear Algebra**", Cambridge University Press, 2018 (Chp11)

Systems of Linear Equations

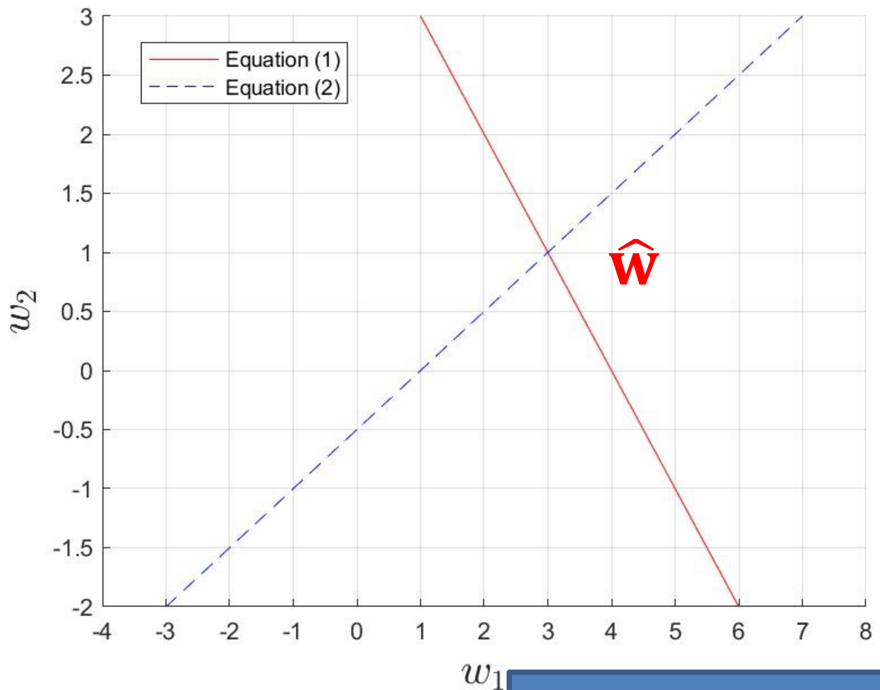
Example 1

$$w_1 + w_2 = 4 \qquad (1)$$
$$w_1 - 2w_2 = 1 \qquad (2)$$

Two unknowns
Two equations

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

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



Python demo 3

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

Systems of Linear Equations

$\mathbf{X}\mathbf{w} = \mathbf{y}, \quad \mathbf{X} \in \mathcal{R}^{m \times d}, \mathbf{w} \in \mathcal{R}^{d \times 1}, \mathbf{y} \in \mathcal{R}^{m \times 1}$

2. Over-determined system: $m > d$

- More equations than unknowns
- $\mathbf{X}$ is non-square (tall) and hence not invertible
- **Has no exact solution in general** *
- An **approximated solution** is available using the left inverse

If the left-inverse of $\mathbf{X}$ exists such that $\mathbf{X}^{\dagger}\mathbf{X} = \mathbf{I}$, then pre-multiply both sides by $\mathbf{X}^{\dagger}$ results in

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

Definition:

A matrix **B** that satisfies $\boldsymbol{B}_{d\,x\,m}\boldsymbol{A}_{m\,x\,d} = \mathbf{I}$ is called a **left-inverse** of $\mathbf{A}$.

The **left-inverse** of $\mathbf{X}$: $\mathbf{X}^{\dagger} = (\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}$ given $\mathbf{X}^{T}\mathbf{X}$ is invertible.

Note: * exception: when rank($\mathbf{X}$) = rank([$\mathbf{X}$,$\mathbf{y}$]), there is a solution.

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "**Introduction to Applied Linear Algebra**", Cambridge University Press, 2018 (Chp11.1-11.2, 11.5)

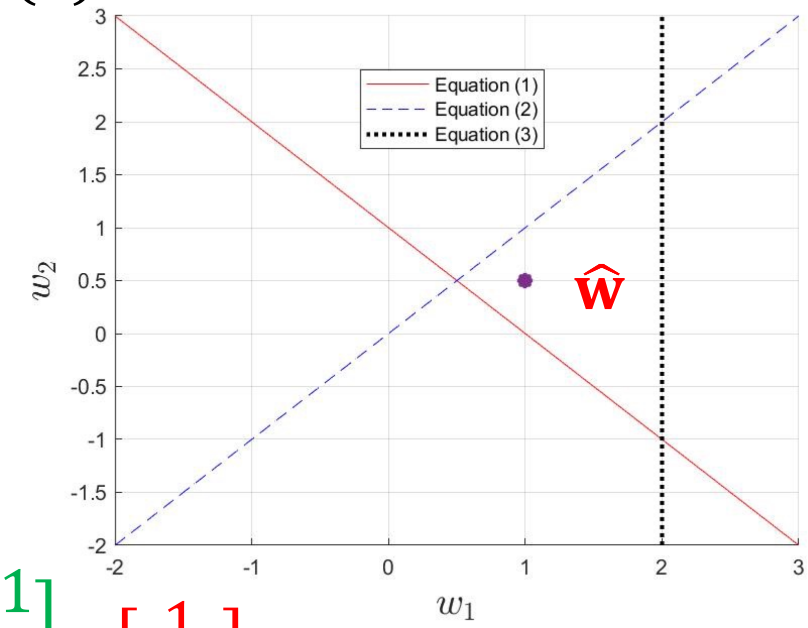
Systems of Linear Equations

Example 2

$$w_1 + w_2 = 1 \quad (1)$$
$$w_1 - w_2 = 0 \quad (2)$$
$$w_1 = 2 \quad (3)$$

Two unknowns
Three equations

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



No exact solution
Approximated solution

$$\hat{\mathbf{w}} = \mathbf{X}^{\dagger}\mathbf{y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$= \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

$\mathbf{X}^T\mathbf{X}$ is invertible

Python demo 4

Systems of Linear Equations

$\mathbf{X}\mathbf{w} = \mathbf{y}, \quad \mathbf{X} \in \mathcal{R}^{m\times d}, \mathbf{w} \in \mathcal{R}^{d\times 1}, \mathbf{y} \in \mathcal{R}^{m\times 1}$

3. Under-determined system: $m < d$

- More unknowns than equations
- Infinite number of solutions in general *

If the **right-inverse** of $\mathbf{X}$ exists such that $\mathbf{X}\mathbf{X}^{\dagger} = \mathbf{I}$, then the d-vector $\mathbf{w} = \mathbf{X}^{\dagger}\mathbf{y}$ (one of the infinite cases) satisfies the equation $\mathbf{X}\mathbf{w} = \mathbf{y}$, i.e.,

$$\mathbf{X}\mathbf{w} = \mathbf{y} \quad \Rightarrow \quad \mathbf{X}\mathbf{X}^{\dagger}\mathbf{y} = \mathbf{y}$$
$$\Rightarrow \quad \mathbf{I}\mathbf{y} = \mathbf{y}$$

Definition:

A matrix **B** that satisfies $\mathbf{A}_{m\,x\,d}\mathbf{B}_{d\,x\,m} = \mathbf{I}$ is called a **right-inverse** of $\mathbf{A}$.
The **right-inverse** of $\mathbf{X}$: $\mathbf{X}^{\dagger} = \mathbf{X}^{T}(\mathbf{X}\mathbf{X}^{T})^{-1}$ given $\mathbf{X}\mathbf{X}^{T}$ is invertible.

If $\mathbf{X}$ is right−invertible, we can find a unique constrained solution.

Note: * exception: no solution if the system is inconsistent rank($\mathbf{X}$) < rank([$\mathbf{X}$,$\mathbf{y}$])

Ref: [Book4] Stephen Boyd and Lieven Vandenberghe, "**Introduction to Applied Linear Algebra**", Cambridge University Press, 2018 (Chp11.1-11.2, 11.5)

Systems of Linear Equations

3. Under-determined system: $m < d$

Derivation:

$$\mathbf{X}\mathbf{w} = \mathbf{y}, \quad \mathbf{X} \in \mathcal{R}^{m\times d},\ \mathbf{w} \in \mathcal{R}^{d\times 1},\ \mathbf{y} \in \mathcal{R}^{m\times 1}$$

A unique solution is yet possible by constraining the search using $\mathbf{w} = \mathbf{X}^T\mathbf{a}$

If $\mathbf{X}\mathbf{X}^T$ is invertible, let $\mathbf{w} = \mathbf{X}^T\mathbf{a}$, then

$$\mathbf{X}\mathbf{X}^T\mathbf{a} = \mathbf{y}$$

$$\Rightarrow \quad \hat{\mathbf{a}} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{y}$$

$$\Rightarrow \hat{\mathbf{w}} = \mathbf{X}^T\hat{\mathbf{a}} = \underbrace{\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}}_{\mathbf{X}^\dagger}\mathbf{y}$$

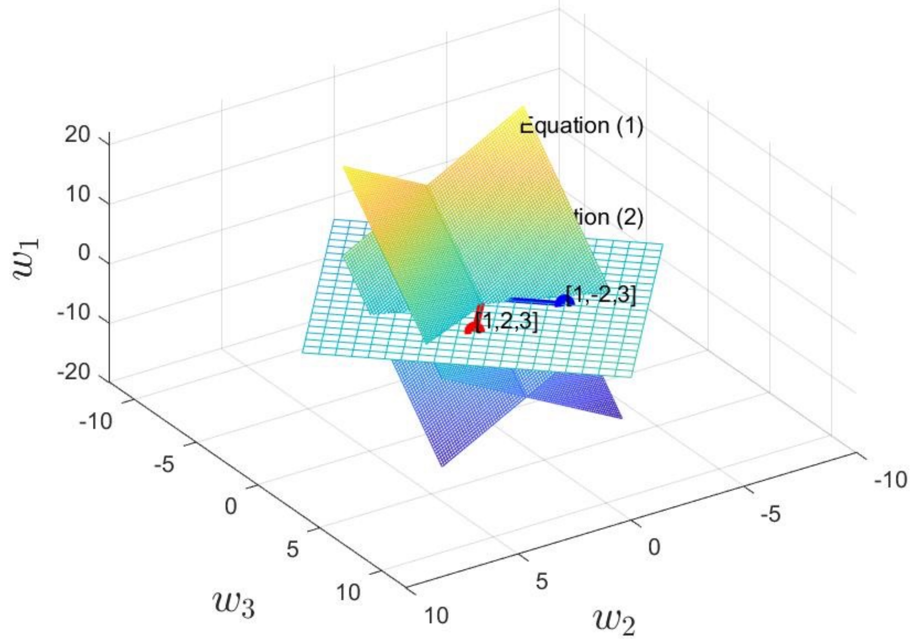
right-inverse

Systems of Linear Equations

Example 3 $w_1 + 2w_2 + 3w_3 = 2 \quad (1)$
$w_1 - 2w_2 + 3w_3 = 1 \quad (2)$

Three unknowns
Two equations

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



Infinitely many solutions along the intersection line

Here $\mathbf{X}\mathbf{X}^T$ is invertible

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

$$= \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 14 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.15 \\ 0.25 \\ 0.45 \end{bmatrix}$$

Constrained solution

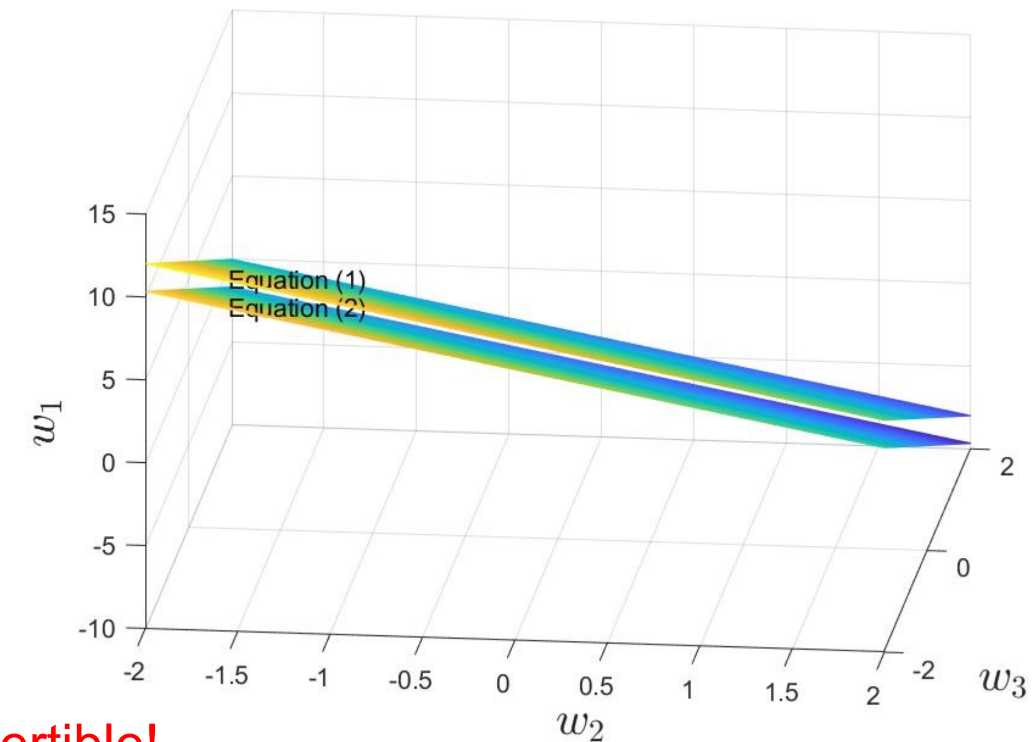
Systems of Linear Equations

Example 4 $w_1 + 2w_2 + 3w_3 = 2 \quad (1)$

$3w_1 + 6w_2 + 9w_3 = 1 \quad (2)$

Three unknowns
Two equations

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



Both $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$ are not invertible!

There is no solution for the system

Quick check 3*questions - Poll on PollEv.com/ymjin

Just “**skip**” if you are required to do registration

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

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:

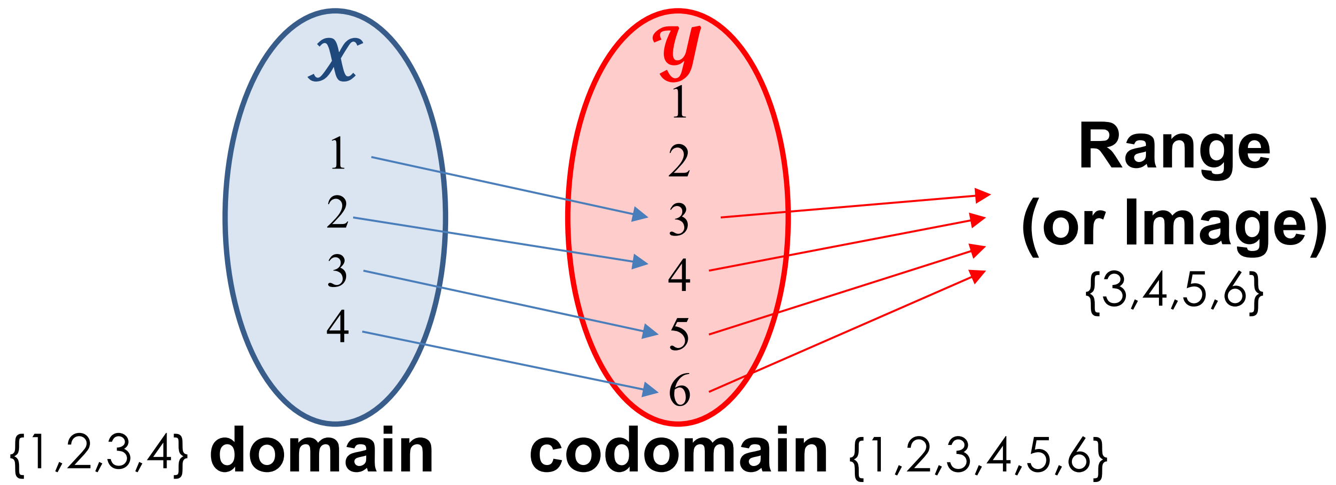
$$\mathcal{S}_3 \leftarrow \mathcal{S}_1 \cup \mathcal{S}_2$$

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 $f\colon \mathcal{R}^d \to \mathcal{R}$ 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., $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_d)$, $\mathbf{x} \in \mathcal{R}^d$, $f(\mathbf{x}) \in \mathcal{R}$
- Example: we can define a function $f\colon \mathcal{R}^4 \to \mathcal{R}$ 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)

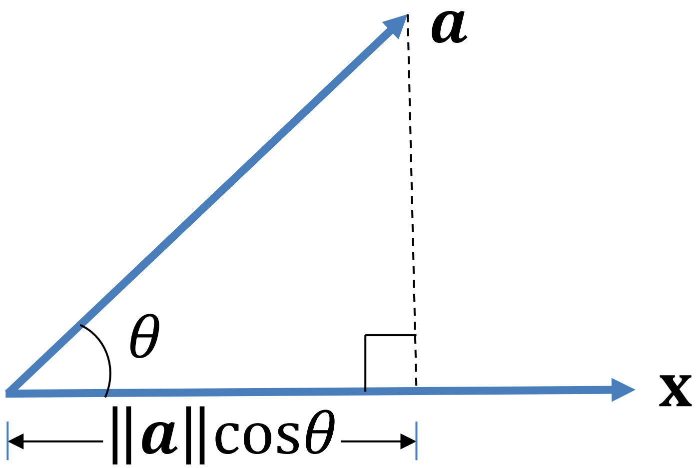
Functions

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 42) 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 $f: \mathcal{R}^d \to \mathcal{R}$ 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

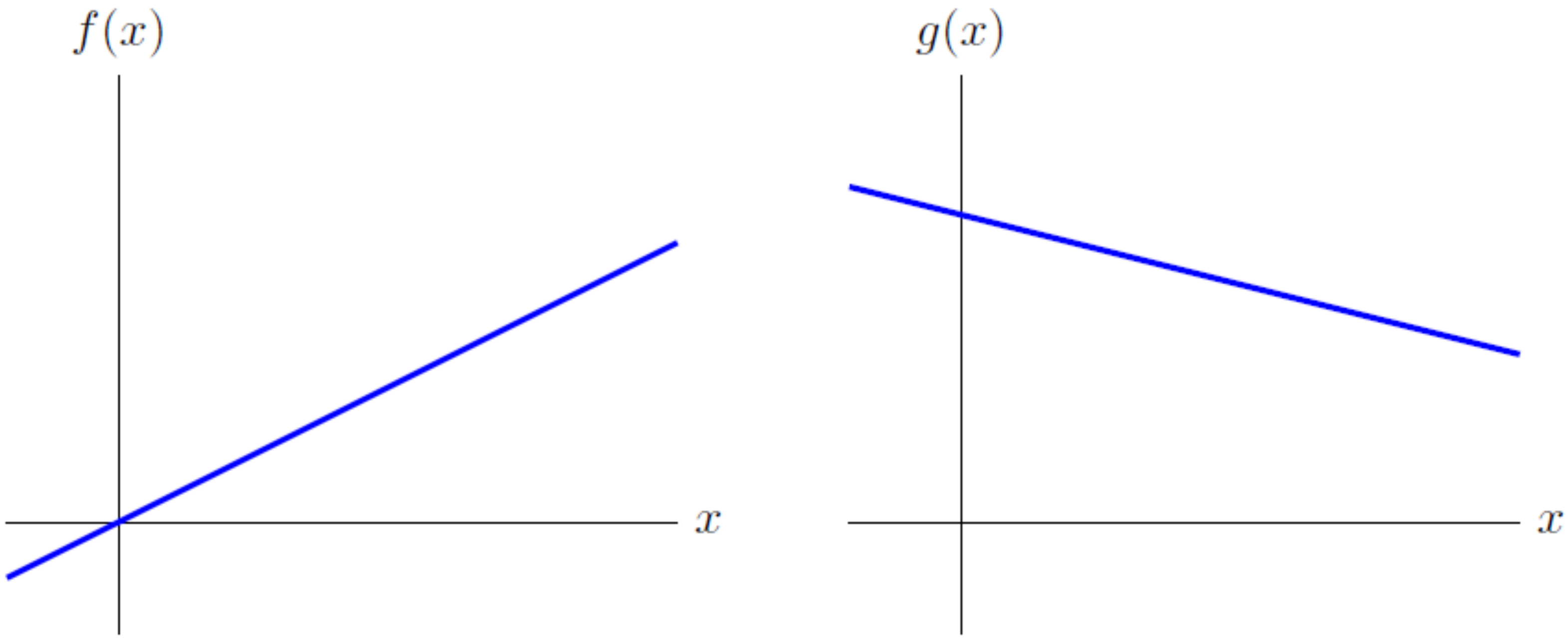


Figure 2.1 *Left.* The function f is linear. *Right.* The function g is affine, but not linear.

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 7 Wed)
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*