Course Overview

Welcome to EE445!

- Linear algebra, optimization, & machine learning models motivated by applications in areas including statistics, decision-making and control, communications, signal processing...
- See course website for all logistics, course information, pre-req's, and material
- Books:
  - "Vectors, Matrcies, and Least Squares (VMLS)" by Boyd & Vandenberghe (.pdf is online)
  - "Optimization Models" by Calafiore & El Ghaoui (check errata file too)
  - see other supplementary books Strang Axler
- Instructors: Maryam Fazel, Lillian Ratliff
- TA: Adhyyan Narang
- Important : HW0 – assessment HW is assigned (see announcement + website)

EE445 Mod1-Lec1: Linear Algebra I

References:

- [VMLS]: Chapters 1, 2, 3
- Topics: Vectors, Inner product, Linear functions, Regression model, Norm and distance

Vectors: notation

other notation:

$\boldsymbol{x}$ (boldface)

$\vec{x}$

- a vector is an ordered list of numbers: $(-1.2,\ 0,\ 3.6)$, or

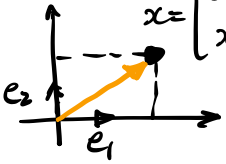
$$x = \begin{bmatrix} -1.2 \\ 0 \\ 3.6 \end{bmatrix}$$

- $x \in \mathbf{R}^n$ means n-vector with real entries $y \in \mathbb{C}^n$ (complex numbers), $z \in \mathbb{R}^n_+$ (nonnegative reals)
- x_i denotes the ith entry of x (warning: sometimes x_i refers to ith vector in a list of vectors)
- x^T denotes transpose
- $\mathbf{1}_n$ or $\mathbf{1}$ denotes vector of all ones $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$
- e_i denotes ith coordinate vector (1 in index i, zero elsewhere) $\begin{array}{cl} \left[\begin{matrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{matrix}\right] & \begin{matrix} \\ \\ \leftarrow i\text{th} \\ \\ \\ \end{matrix} \end{array}$
- suppose b, c, d are vectors of sizes m, n, p, can stack them into a larger block vector

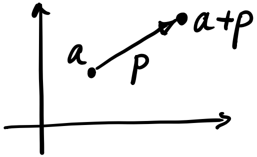
$$a = \begin{bmatrix} b \\ c \\ d \end{bmatrix} \begin{matrix} \updownarrow m \\ \updownarrow n \\ \updownarrow p \end{matrix} \qquad a \in \mathbb{R}^{m+n+p}$$

Examples

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 e_1 + x_2 e_2$



- (x_1, x_2) represents a location or a displacement in 2D



- examples:
- portfolio: entries give shares (or \$ value or fraction) held in each of n assets, with negative meaning short positions
- cash flow: x_i is payment in period i to us

  get a \$1 loan (in period 1) — repay with %10 interest (in period 2)

  e.g., $[1 \quad -1.1 \quad 0]$
- sampled audio signal: x_i is the acoustic pressure at sample time i
- feature vector: x_i is the value of ith feature or attribute of an entity
- word count vector

Word count example

- a short document:

"Word count vectors are used in computer based document analysis. Each entry of the word count vector is the number of times the associated dictionary word appears in the document."

- dictionary & word count vector:

$$\begin{array}{l}\text{word}\\ \text{in}\\ \text{number}\\ \text{flower}\\ \text{the}\\ \text{document}\end{array}\quad\begin{bmatrix}3\\2\\1\\0\\4\\2\end{bmatrix}$$

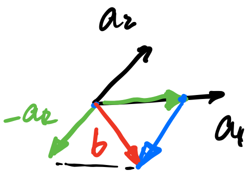
Linear combinations

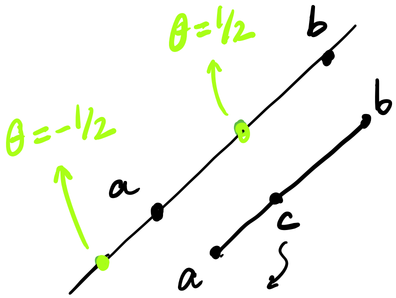
- for vectors $a_1, \dots, a_n$, scalars $\beta_1, \dots, \beta_n$,

$$y = \beta_1 a_1 + \dots + \beta_n a_n$$

  is a linear combination with coeff's $\beta_1, \dots \beta_n$

- 2D example: $b = 0.75a_1 - a_2$





- line and segment: a, b are n-vectors, $c = (1-\theta)a + \theta b$
  - when θ is any scalar: c on line passing through a, b
  - when $0 \leq \theta \leq 1$: c on line segment connecting a, b

c called mixture, or "convex combination" (more on this later!)

Inner product

- inner product (or dot product) of n-vectors a and b is

$$a^T b = \sum_{i=1}^{n} a_i b_i \qquad \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + \cdots + a_n b_n$$

  other notation used: $\langle a, b \rangle$, $a \cdot b$, $\langle a | b \rangle$

ex: $\begin{bmatrix} -1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = (-1)(1) + (2)(0) + (2)(-3) = -7$

Properties of inner product

$a, b \in \mathbb{R}^n, \ \gamma \in \mathbb{R}$

- $a^T b = b^T a$
- $(\gamma a)^T b = \gamma(a^T b)$
- $(a + b)^T c = a^T c + b^T c$

examples:

- $\underline{e_i^T a} = \underline{a_i}$ (e_i: ith coordinate vector)
- $\underline{\mathbf{1}^T a} = \sum_{i=1}^n a_i$ ($\mathbf{1}$: vector of all ones)
- $\underline{a^T a} = \sum_{i=1}^n a_i^2 = \|a\|^2$

Examples

- w is weight vector, f is feature vector; $w^T f$ is score e.g. credit score
- p is vector of asset prices, s gives portfolio holdings (in shares); $p^T s$ is total portfolio value
- c is cash flow, d is discount vector (with interest rate r):

$$d = (1, 1/(1+r), \ldots, 1/(1+r)^{n-1})$$

NPV

$d^T c$ is net present-value of cash flow

Linear functions

- $f : \mathbf{R}^n \to \mathbf{R}$ means f is a function mapping n-vectors to scalars (multivariate function of n variables)
- note: we'll also see $\mathbf{R}^n \to \mathbf{R}^m$ linear maps soon
- f satisfies superposition if $\boxed{f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)}$ holds for all scalars α, β and all $x, y \in \mathbf{R}^n$
- then f is called a *linear* function
- example: the 'inner product function' is linear:

f(lin. comb. of x,y) = lin. comb. of f(x) & f(y)

LHS: $a^T(\alpha x + \beta y) = \alpha a^T x + \beta a^T y$

RHS: $\alpha (a^T x) + \beta (a^T y)$

$\Big\}$ same

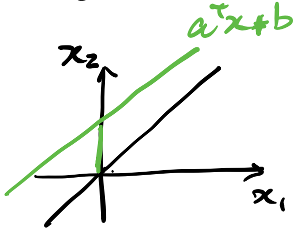
...and all linear functions are inner products

- suppose $f : \mathbf{R}^n \to \mathbf{R}$ is linear, then it can be expressed as $f(x) = a^T x$ for some a
- specifically: $\underline{a_i = f(e_i)}$
- follows from:

$$x = x_1 e_1 + \cdots + x_n e_n$$

$$f(x) = f(x_1 e_1 + \cdots + x_n e_n) = x_1 \overbrace{f(e_1)}^{a_1} + \cdots + x_n \overbrace{f(e_n)}^{a_n} = a^T x$$

- 'affine' function: a linear function plus a constant, general form:



$$f(x) = a^T x + \underline{b}$$

- sometimes (sloppily) we refer to this as linear also

Regression model

- *regression model* is (affine function of x):

$$\hat{y} = x^T \beta + v$$

- x is a feature vector, element x_i called regressors
- β is the *weight vector*, scalar v is the *offset*
- scalar $\hat{y}$ is the *prediction* (of some actual outcome y)

Example: house prices

- y is selling price of house (in \$1000)
- regressor is x=(house area in 1000 sqft, # bedrooms)
- weight vector and offset are: $\beta = (148.73, -18.85)$, $v = 54.4$ interpret as 'land value'
- will later see how to guess β and v from sales data
  in module 2
- data:

house	x_1 (area)	x_2 (bed)	y (price)	$\hat{y}$ (prediction)
1	0.846	1	115.00	161.37
2	1.324	2	234.50	213.61
3	1.150	3	198.00	168.88
4	3.0337	4	528.00	430.67
5	3.984	5	572.50	552.66

Example: house prices

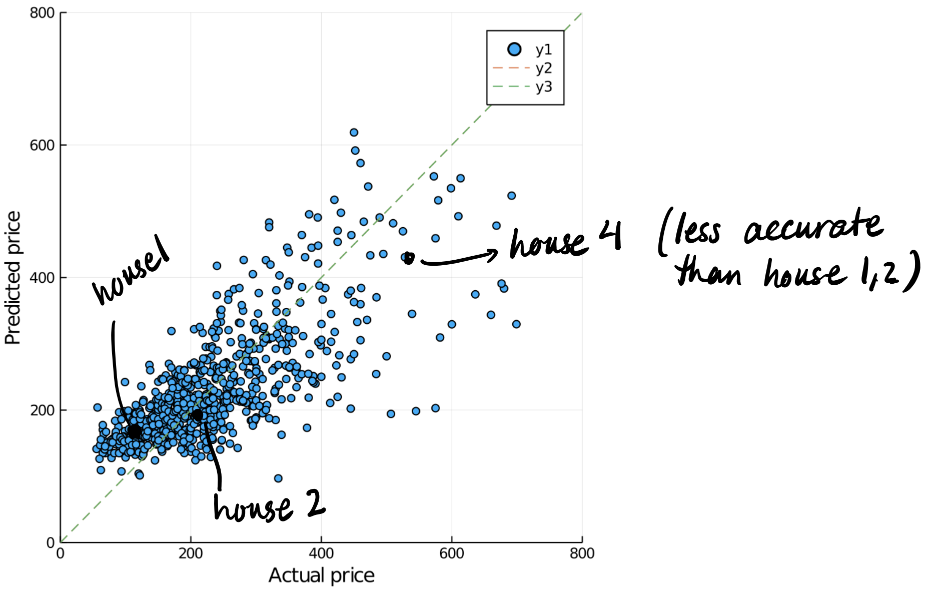
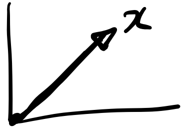


Figure: Scatter plot of actual and predicted sale prices for 774 houses sold in Sacramento in a 5-day period

Norm and distance

$x \in \mathbb{R}^n$



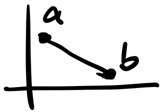
- Euclidean norm (2-norm) is $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$
- properties:
  - homogeneity: $\|\alpha x\| = |\alpha| \, \|x\|$
  - triangle inequality: $\|x+y\| \leq \|x\| + \|y\|$
  - nonnegativity: $\|x\| \geq 0$
  - definiteness: $\|x\| = 0$ only if $x = 0$

Norm and distance

- norm of block vectors: $\left\| \begin{bmatrix} a \\ b \\ c \end{bmatrix} \right\|^2 = [a \;\; b \;\; c] \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a^Ta + b^Tb + c^Tc$

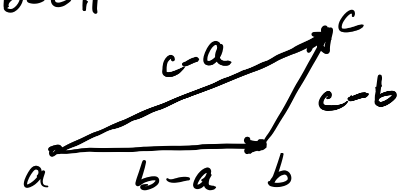
$$\|(a, b, c)\|^2 = a^Ta + b^Tb + c^Tc = \|a\|^2 + \|b\|^2 + \|c\|^2$$

- (Euclidean) *distance* between vectors a, b is: $\underline{\mathbf{dist}(a, b)} = \|a - b\|$



- triangle inequality: $\|a - c\| \le \|a - b\| + \|b - c\|$

$$\|a-c\| = \|\underline{a-b} + \underline{b-c}\| \le \|a-b\| + \|b-c\|$$



Chebyshev inequality

Simple but useful ineq. on relation between vector norm & vector entries :

$x \in \mathbb{R}^n$

- suppose that k of the numbers $|x_1|, \ldots, |x_n|$ are $\geq a$ some constant
- then k of the numbers $x_1^2, \ldots, x_n^2$ are $\geq a^2$
- so $\|x\|^2 = \sum_i x_i^2 \geq ka^2$, and we have $k \leq \|x\|^2/a^2$
- number of x_i with $|x_i| \geq a$ is no more than $\|x\|^2/a^2$
- this is the *Chebyshev inequality*

$k \leq \frac{\|x\|^2}{a^2}$

of x_i w/ $|x_i| \geq a$

define root-mean-square value: $\mathrm{rms}(x) = \sqrt{\frac{x_1^2 + \cdots + x_n^2}{n}} = \frac{\|x\|}{\sqrt{n}}$

- Chebyshev ineq. in terms of rms value:

  fraction of entries w/ $|x_i| \geq a$ (i.e., $\frac{k}{n}$) is no more than $\frac{\|x\|^2}{na^2} = \left(\frac{\mathrm{rms}(x)}{a}\right)^2$