# 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
  - see other supplementary books
- Instructors: Maryam Fazel, Lillian Ratliff
- TA: Adhyyan Narang

# 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

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

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

- $x \in \mathbf{R}^n$ means $n$-vector with real entries
- $x_i$ denotes the $i$th entry of $x$ (warning: sometimes $x_i$ refers to $i$th vector in a list of vectors)
- $x^T$ denotes transpose
- $\mathbf{1}_n$ or $\mathbf{1}$ denotes vector of all ones
- $e_i$ denotes $i$th coordinate vector (1 in index $i$, zero elsewhere)
- suppose $b, c, d$ are vectors of sizes $m, n, p$, can stack them into a larger block vector

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

# Examples

- $(x_1, x_2)$ 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
  sampled audio signal: $x_i$ is the acoustic pressure at sample time $i$
  feature vector: $x_i$ is the value of $i$th 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:

  word
  in
  number
  flower
  the
  document

# Linear combinations

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

$$\beta_1 a_1 + \ldots + \beta_n a_n$$

  is a linear combination with coeff's $\beta_1, \ldots \beta_n$
- 2D example: $b = 0.75a_1 - a_2$
- line and segment: $a, b$ are $n$-vectors, $c = (1 - \theta)a + \theta b$
  - when $\theta$ is any scalar: $c$ on line passing through $a, b$
  - when $0 \leq \theta \leq 1$: $c$ on line segment connecting $a, b$

# Inner product

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

$$a^T b = \sum_{i=1}^{n} a_i b_i$$

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

# Properties of inner product

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

examples:

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

# Examples

- $w$ is weight vector, $f$ is feature vector; $w^T f$ is 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 $r$):

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

  $d^T c$ 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 $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ holds for all scalars $\alpha, \beta$ and all $x, y \in \mathbf{R}^n$
- then $f$ is called a *linear* function
- example: the 'inner product function' is linear:

# . . . 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: $a_i = f(e_i)$
- follows from:

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

$$f(x) = a^T x + b$$

- 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
- $\beta$ 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: $\beta = (148.73, -18.85)$, $\quad v = 54.4$
- will later see how to guess $\beta$ and $v$ from sales data
- 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

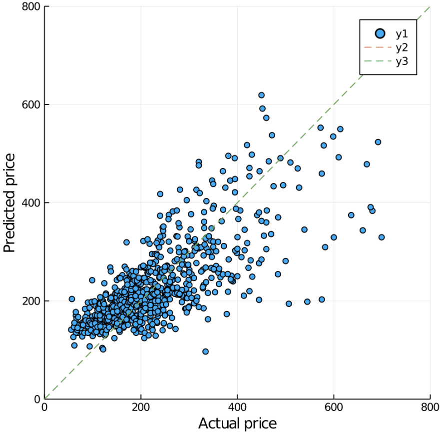


**Figure:** Scatter plot of actual and predicted sale prices for 774 houses sold in Sacramento in a 5-day period

# Norm and distance

- Euclidean norm (2-norm) is $\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}$
- properties:
  - homogeneity:
  - triangle inequality:
  - nonnegativity:
  - definiteness:
- norm of block vectors:

$$\|(a, b, c)\|^2 = a^T a + b^T b + c^T = \|a\|^2 + \|b\|^2 + \|c\|^2$$

- (Euclidean) *distance* between vectors $a, b$ is: $\mathbf{dist}(a, b) = \|a - b\|$
- triangle inequality: $\|a - c\| \leq \|a - b\| + \|b - c\|$

# Chebyshev inequality

- suppose that $k$ of the numbers $|x_1|, \ldots, |x_n|$ are $\geq a$
- 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*