Last example

Suppose we want to predict annual income of a person based on the following

finish HS ? yes / no ? → yes 1, no = 0
finish bachelor ? yes / no ?
finish grad ? yes / no ?
age over 20 (number

$(1, 1, 0, \underbrace{17}_{\text{37 years old}})$

↑
x

$\hat{y}$ predicted income

model: v number

$b = (b_1, b_2, b_3, b_4)$

parameters of the model

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

$= b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + v$

linear $\underbrace{\text{regression}}$

prediction

x_i regressors

$[\hat{y}] = \$$

$b_1 \rightarrow$ additional income for having completed HS.

$V \longrightarrow$ expected move ~~of~~ 20 %
with no HS diploma

Norms + distance

B

$(3,7)$

A

$(0,0)$

How far is B from A?

$$\sqrt{3^2 + 7^2}$$ Euclidean distance

Given <u>any</u> vector $x \in \mathbb{R}^n$

$$\|x\| = \left(x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2\right)^{1/2}$$

$$x = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 4 \end{bmatrix} \qquad \|x\| = \sqrt{1^2 + 2^2 + 1^2 + 4^2}$$

$$= (1 + 4 + 1 + 16)^{1/2}$$

$$= \sqrt{22}$$

Properties $\quad \|x\| \geq 0$ and $\|x\| = 0 \Rightarrow x = 0$

$$x = 0 \Rightarrow \|x\| = 0$$

$$\|x\| = 0 \Leftrightarrow x = 0$$

$$\|x\| = \left(x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2\right)^{1/2}$$

$$\|7x\| = \left((7x_1)^2 + (7x_2)^2 + \cdots + (7x_n)^2 \right]^{1/2}$$

$$= \left(7^2 x_1^2 + 7^2 x_2^2 + \cdots + 7^2 x_n^2 \right)^{1/2}$$

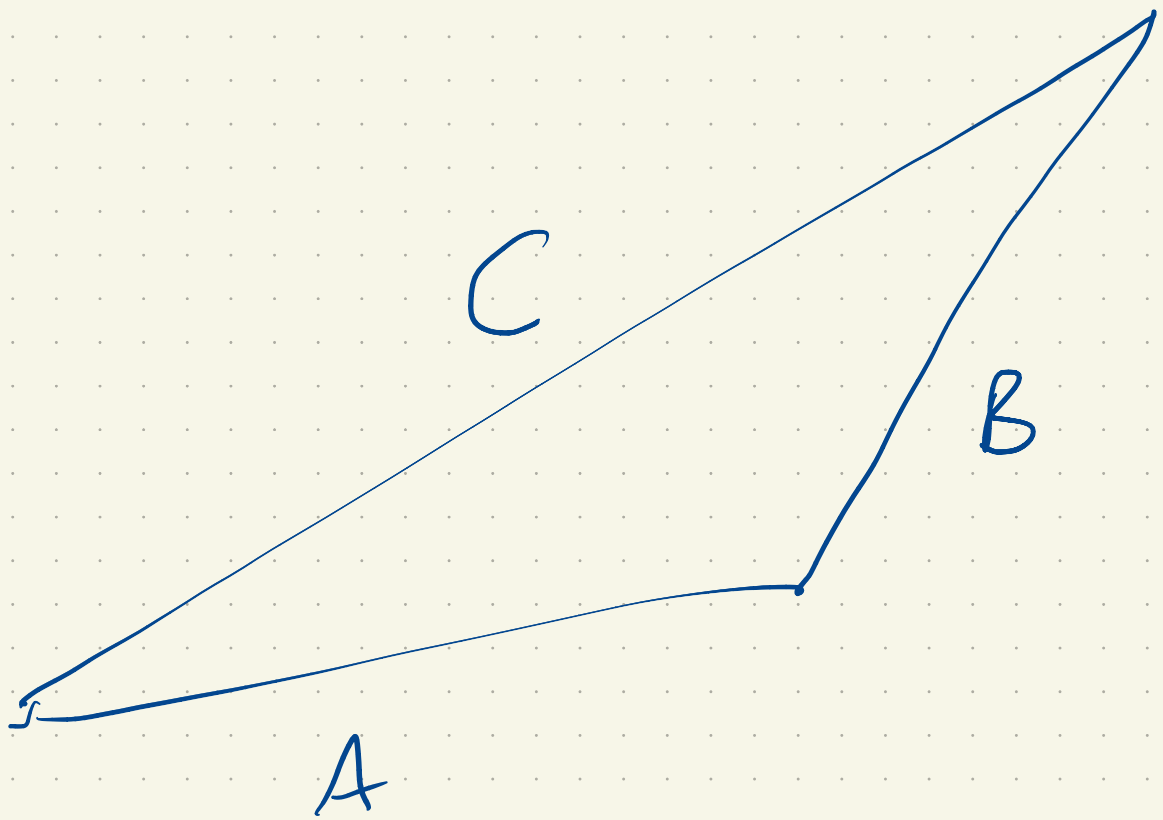
$$= \left(7^2 \left(x_1^2 + \cdots + x_n^2 \right) \right)^{1/2}$$

$$= (7^2)^{1/2} \|x\|$$

$$(\alpha^2)^{1/2} = |\alpha|$$

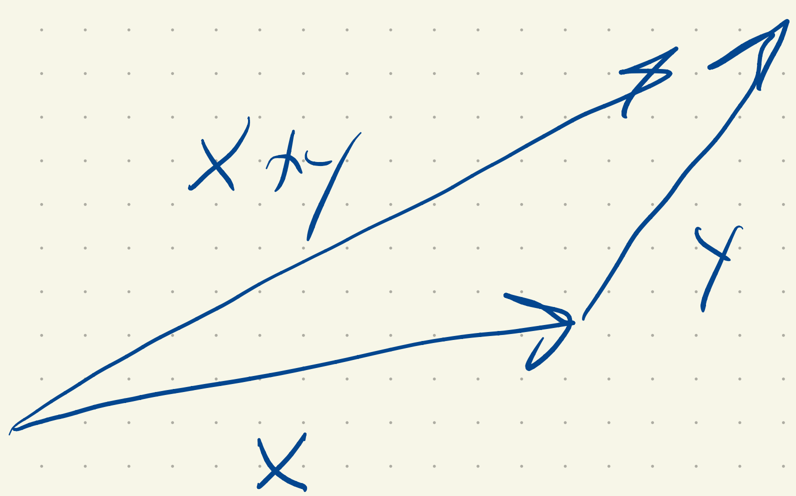
$$= 7 \|x\|$$

$$\|\alpha x\| = |\alpha| \, \|x\|$$



C vs A and B

$C \leq A + B$



$$\|x+y\| \leq \|x\| + \|y\|$$

I'll prove it later

Triangle inequality

$\|x\|$ is called the Euclidean length or norm of x.

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

$$x = (1, 1, 0, 17) \longrightarrow \left(1, 1, 0, \frac{17}{20}\right)$$

$$\|x\| = \left(1^2 + 1^2 + 0^2 + 17^2\right)^{1/2}$$

$$\left(1^2 + 1^2 + 0^2 + \left(\frac{17}{20}\right)^2\right)$$

See text 3.28

$$\|1_n\| = (1^2 + 1^2 + \cdots + 1^2)^{1/2} = \sqrt{n}$$

This tells you about the size of the elements of the vector <u>and</u> the number of entries (the dimension)

$$(26, 21, 25, 19, 30, \ldots, 42)$$

$$\mathrm{rms}(x) = \frac{\|x\|}{\sqrt{n}} \qquad \mathrm{rms}(1_n) = \frac{\sqrt{n}}{\sqrt{n}} > 1$$

rms root mean square

typical size (in absolute value) of the

entries of x.