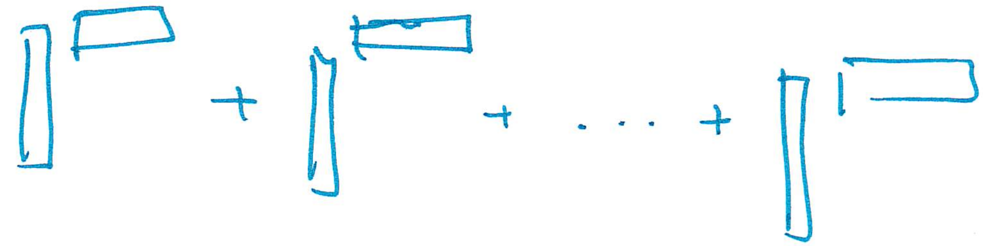
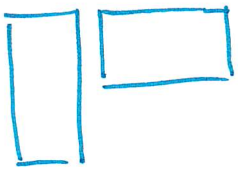
last time: matrix multiplication

$$TW = \sum_{k=1}^{r} t_k w_k =$$





$n \times r \quad r \times p$

t_k: k^{th} col of T

w_k: k^{th} row of W

= sum of rank-1 matrices

$\Rightarrow$ TW has rank r <u>iff</u> cols of T and row of W are linearly independent

Imagine cols of T are <u>not</u> linearly independent

e.g. $t_1 = \alpha t_2 + \beta t_3$.

$$TW = t_1 w_1 + t_2 w_2 + t_3 w_3$$

$$= (\alpha t_2 + \beta t_3) w_1 + t_2 w_2 + t_3 w_3$$

$$= t_2 (\alpha w_1 + w_2) + t_3 (\beta w_1 + w_3)$$

$\Rightarrow$ rank is atmost $\underline{2}$ not $\underline{3}$

$$A = \begin{bmatrix} 1 & 4 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 0 \end{bmatrix}$$

rank $= 2$

TW

$$\begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix}}_{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}$$

Exercise: consider the following linear systems.

For each, how many solutions are there? (zero, one, or many)

If one (or more) solutions exist, then find it (or a couple).

Why do different cases have different numbers of soln's?

a) $3x_1 + 2x_2 = 1$
$3x_1 + x_2 = 0$

one soln
$x_1 = -1/3$
$x_2 = 1$

$x_2 = -3x_1$
$-x_2 + 2x_2 = 1$
$x_2 = 1$

b) $3x_1 + x_2 = 0$

underdetermined

∞ solns.

c) $\begin{bmatrix} 3 & 2 \\ 3 & 1 \\ 3 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ $\quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

overdetermined

one soln
$x = \begin{bmatrix} -1/3 \\ 1 \end{bmatrix}$

d) $3x_1 + 2x_2 = 1$
$3x_1 + x_2 = 0$
$2x_1 + 2x_2 = 2$

zero solutions

e) $3x_1 + x_2 = 1$
$6x_1 + 2x_2 = 2$

∞ solns.

$\begin{bmatrix} 3 & 1 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

Linear systems of equations

$$\underbrace{\underline{y}}_{n \times 1} = \underbrace{X}_{n \times p} \underbrace{\underline{w}}_{p \times 1}$$

given y, X, solve for $\underline{w}$

$$y = w_1 \underline{x}_1 + w_2 \underline{x}_2 + \cdots + w_p \underline{x}_p$$

= weighted sum of cols of X.

$$\boxed{\text{what are weights?}}$$

ex.

$$y = \begin{array}{cl} \text{calories} & \\ \begin{bmatrix} 110 \\ 110 \\ 210 \end{bmatrix} & \begin{array}{l} \text{frosted flakes} \\ \cancel{\text{FMNT}}\ \text{GN} \\ \cancel{\text{Grape Nuts}} \\ \text{FMNT} \end{array} \end{array}
\qquad
X = \begin{array}{ccc} \text{carbs} & \text{fat} & \text{protein} \end{array}
\begin{bmatrix} 25 & 0 & 1 \\ 20 & 1 & 2 \\ 40 & 1 & 6 \end{bmatrix}$$

Gaussian Elimination

$$\begin{bmatrix} 25 & 0 & 1 \\ 20 & 1 & 2 \\ 40 & 1 & 6 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 110 \\ 110 \\ 210 \end{bmatrix}$$

row 3 − 2(row 2)

$$40w_1 + 1w_2 + 6w_3 = 210$$

$$\cancel{20}\,w_1 + \cancel{1}\,w_2 + \cancel{2}\,w_3 = \cancel{110}$$

$$+ \quad -40 \qquad -2 \qquad -4 \qquad -220$$

$$0w_1 - 1w_2 + 2w_3 = -10$$

$$\begin{bmatrix} 25 & 0 & 1 \\ 20 & 1 & 2 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 110 \\ 110 \\ -10 \end{bmatrix}$$

$5r_2 - 4r_1$

$$\begin{array}{cccc} 100 & 5 & 10 & \\ -100 & 0 & -4 & \\ \hline 0 & 5 & 6 & 110 \end{array}$$

$$\Downarrow$$

$$\begin{bmatrix} 25 & 0 & 1 \\ 0 & 5 & 6 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 110 \\ 110 \\ -10 \end{bmatrix}$$

$5r_3 + r_2$:

$\cancel{0 \quad 15}$

$$\begin{array}{ccc|c} 0 & -5 & 10 & -50 \\ 0 & 5 & 6 & 110 \\ \hline 0 & 0 & 16 & 60 \end{array}$$

$$\Rightarrow \begin{bmatrix} 25 & 0 & 1 \\ 0 & 5 & 6 \\ 0 & 0 & 16 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 110 \\ 110 \\ 60 \end{bmatrix} \leftarrow$$

$w_3 = 60/16$

$5w_2 = 100 - 6w_3$

$w_2 = 17.5$

$\vdots$

$w_1 = 4.25$

General Linear Systems:

$$X \in \mathbb{R}^{n \times p}$$

$$\underline{y} \in \mathbb{R}^{n}$$

$$\underline{w} \in \mathbb{R}^{p}$$

$n = p$ (X is square) $\Rightarrow$ # eqns = # unknowns

$n > p$ $\Rightarrow$ more equations than unknowns

"overdetermined"

$n < p$

~~$p < n$~~

$\Rightarrow$ fewer equations than unknowns

"underdetermined" $\rightarrow$ many solns.

Today: how to find $\underline{w}$?

- search of space of candidate weights $\underline{w}$.
- choose the one so that $\hat{\underline{y}} = X\underline{w}$ is "as close as possible" to $\underline{y}$.

how do we measure "close"?

Vector norms:

e.g. have n training samples, $(\underline{x}_i, y_i)$ for $i=1,\dots,n$

we learn a predictor $\underline{\hat{y}} = X\underline{w}$ $(\hat{y}_i = \underline{x}_i^T \underline{w})$

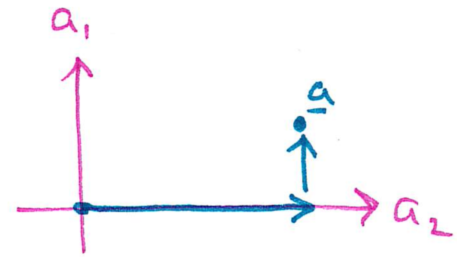
prediction error $y_i - \hat{y}_i$. how big are error totals? ~~to~~ How close is $\hat{\underline{y}}$ to $\underline{y}$?

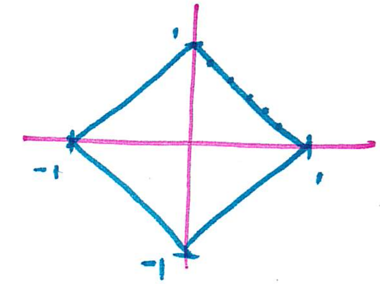
Let $\underline{a} \in \mathbb{R}^n$ be a vector (e.g. $a_i = y_i - \hat{y}_i$)

$$\|a\|_1 = \sum_{i=1}^{n} |a_i|^1$$

taxicab / Manhattan norm

l_1



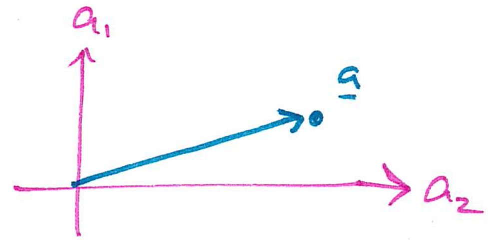


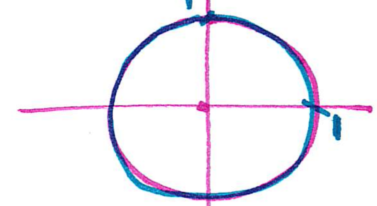
$$\|a\|_2 = \left(\sum_{i=1}^{n} a_i^2 \right)^{1/2}$$

$$= \sqrt{\underline{a}^T \underline{a}}$$

Euclidean norm

l_2





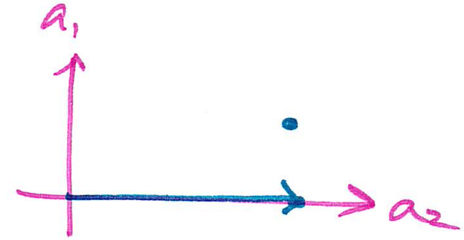
unit cicle

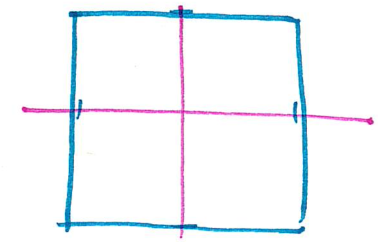
$$\|a\|_\infty = \max_i |a_i|$$

sup norm

max norm

l_∞





$$\|a\|_q = \left(\sum_{i=1}^{n} |a_i|^q \right)^{1/q},$$

l_q norm

$$q \geq 1$$

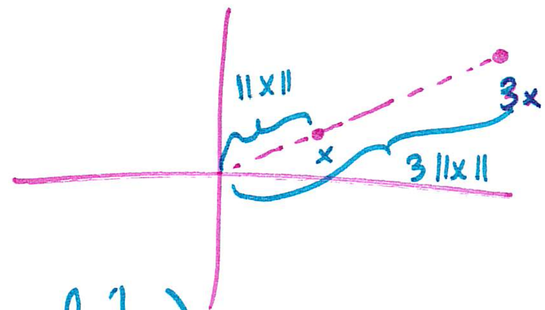
These are all examples of norms

Vector norms = function $\|\cdot\|$ mapping from $\mathbb{R}^n$ to $\mathbb{R}$ with the following properties:

(i) $\|x\| \geq 0 \quad \forall$ (for all) x

(ii) $\|x\| = 0$ iff (if and only if) $x = 0$

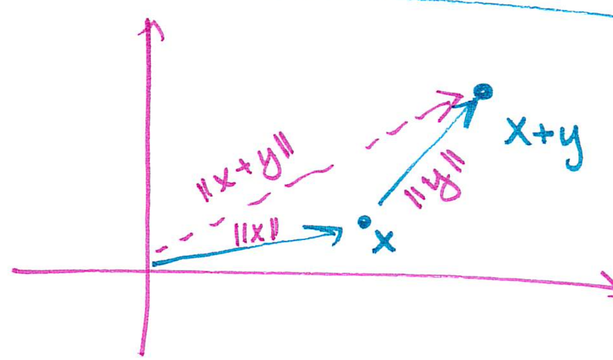
(iii) $\|\alpha x\| = |\alpha| \, \|x\| \quad \forall \quad \alpha \in \mathbb{R}, \; x \in \mathbb{R}^n$



(iv) $\|x+y\| \leq \|x\| + \|y\|$ (triangle inequality)

Helpful fact:

$$\|x\|_{q'} \leq \|x\|_q \quad \text{if} \quad 1 \leq q < q' \leq \infty$$



e.g. $\|x\|_\infty \leq \|x\|_1$ "inclusion of l_p norms"

i.e. $\max\limits_i |x_i| \leq \sum\limits_i |x_i|$

back to solving linear systems:

$$\underset{\underline{w}}{\text{minimize}} \; \|\underline{y} - \underline{\hat{y}}\| \quad \text{where } \hat{y} = X\underline{w}$$

$$\Rightarrow \underset{\underline{w}}{\text{minimize}} \; \|y - X\underline{w}\|$$

Least squares: $n \geq p$, $\underset{\underline{w}}{\text{minimize}} \; \|y - X\underline{w}\|_2$

$$\equiv \underset{\underline{w}}{\text{minimize}} \; \|y - X\underline{w}\|_2^2$$