VECTOR NORMS

ℓ^p $\quad \|\vec{v}\|_p = \left(\sum |v_i|^p\right)^{1/p}$

$p = 1, 2$ most important

ℓ^∞ $\quad \|\vec{v}\|_\infty = \max\{|v_1|, \ldots, |v_n|\}$

$\|\vec{v}\|$ says how big $\vec{v}$ is

$d(\vec{v}, \vec{w}) = \|\vec{v} - \vec{w}\|$ distance, metric

between two vector (points)

Examples

$V^* =$ TRUE SOLN

$V =$ computed soln.

error $= \|V - V^*\|$

$\vec{w} =$ training data

$F_a(\vec{w})$ output of training data with parameters a

$\vec{v} =$ actual true output or label.

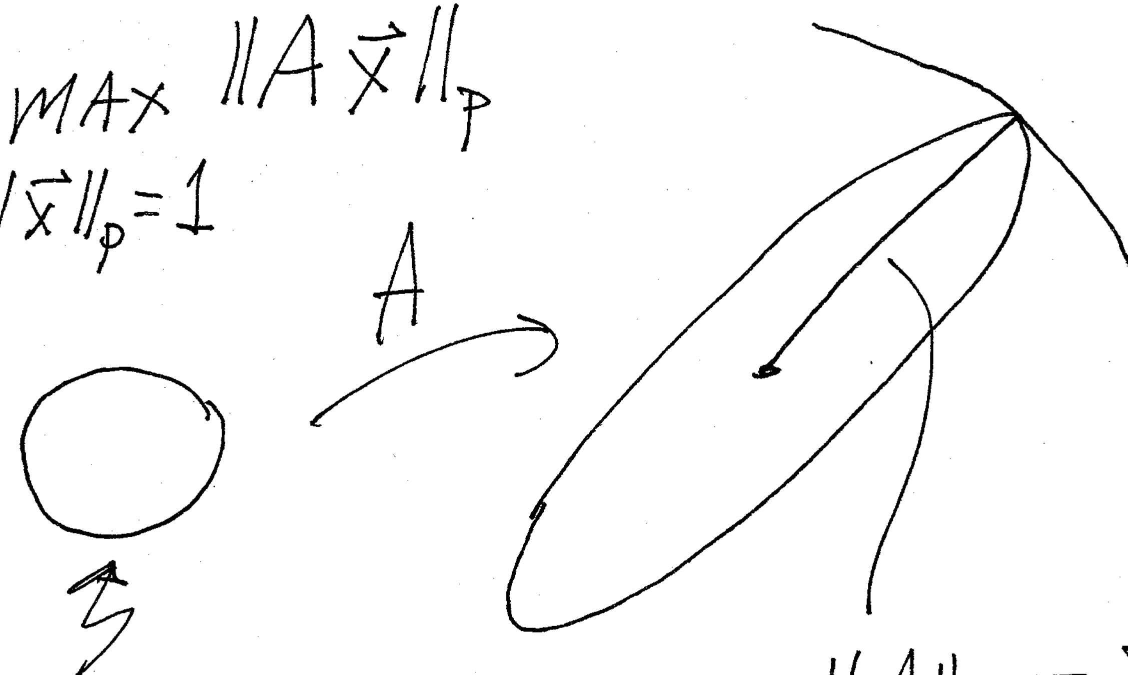
loss function: $\|F_{\underline{\underline{a}}}(\vec{w}) - \vec{v}\|$

Matrix norms

Induced norms from vector norms

$$\|A\|_p = \max_{\|\vec{x}\|_p = 1} \|A\vec{x}\|_p$$

$p = 2$



$\|\vec{x}\|_2 = 1$

$\|A\|_2 = \sigma_1$

or

$$\|A\|_p = \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p} = \max_{\vec{x} \neq \vec{0}} \|A \underbrace{\frac{\vec{x}}{\|\vec{x}\|_p}}_{\text{length } 1}\|_p$$

<u>Fröbenius norm</u> – Treats A as a single vector

$$\|A\|_F = \left(\sum_{j=1}^{n} \sum_{i=1}^{m} |A_{ij}|^2 \right)^{1/2}$$

This implies that

$$\|A\|_F = \sqrt{\text{trace}(A^T A)} = \sqrt{\text{trace}(A A^T)}$$

where B is $k \times k$, $\text{trace}(B) = \sum_{i=1}^{k} B_{ii}$.

Properties of all these norms

(1) $\|A\| \geq 0$ and $\|A\| = 0$ only when $A = 0$

(2) $\|\alpha A\| = |\alpha| \, \|A\|$

(3) $\|A+B\| \leq \|A\| + \|B\|$ ($\triangle \neq$)

(4) $\|AB\| \leq \|A\| \, \|B\|$ Submultiplicative

One important property for SVD

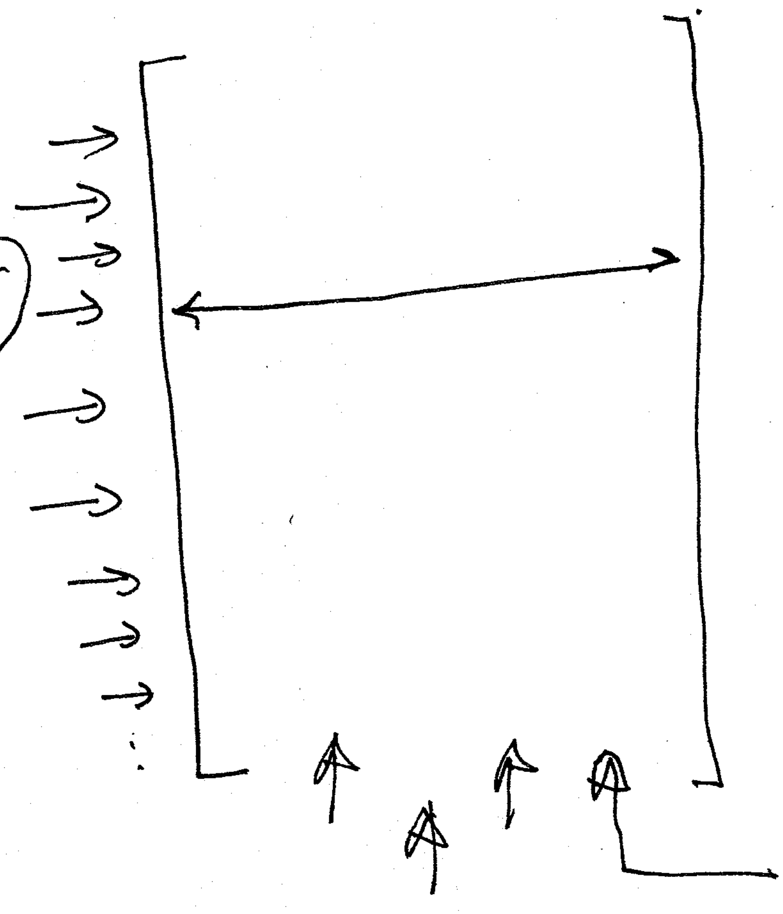
If Q is orthogonal then

$$\|QA\|_2 = \|AQ\|_2 = \|A\|_2.$$

because $\|Q\vec{x}\|_2 = \|\vec{x}\|_2 = \sqrt{\vec{x} \cdot \vec{x}}$

BACK TO SVD

each customer



$m > n$

each Row is
a data VECTOR

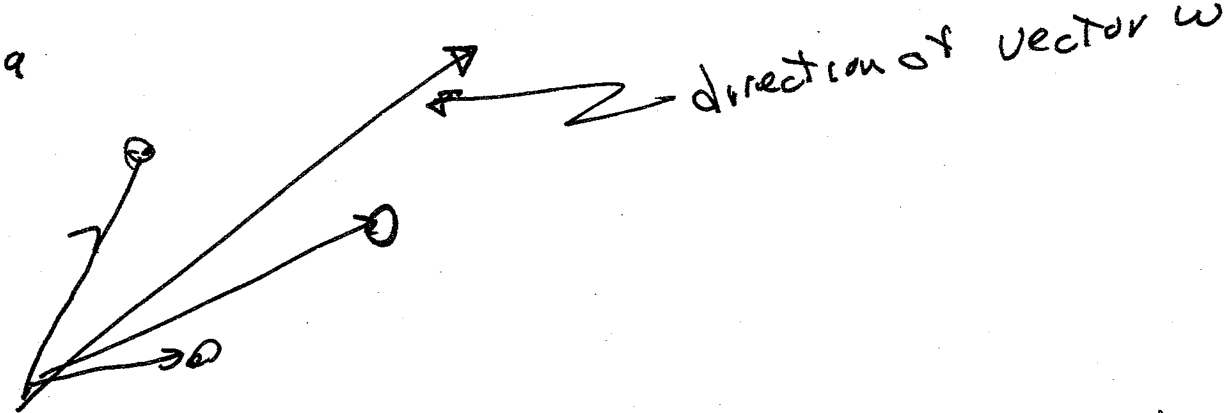
Features

money spent on
chocalate/year

→ is there some combination of features
which best characterize a given customer?

One interpretation – 1st a single quantity

Find the one direction that best captures the data



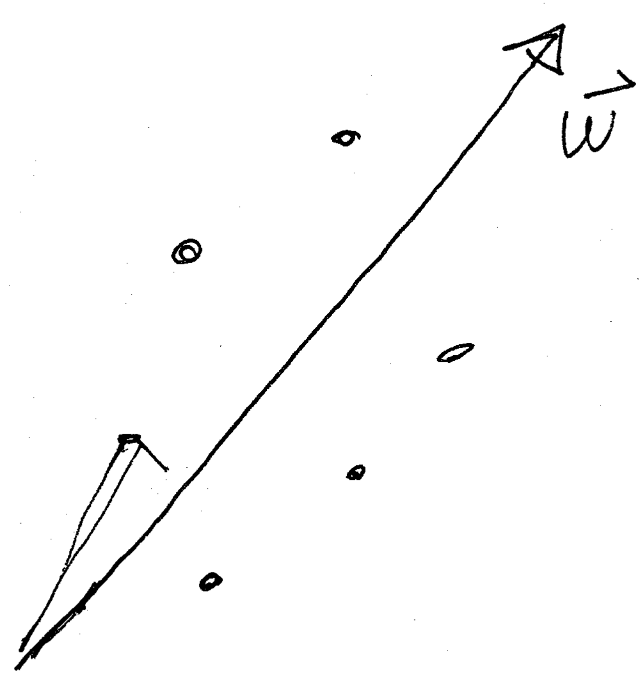
in the sense of maximizes the sum of the squared projections onto w

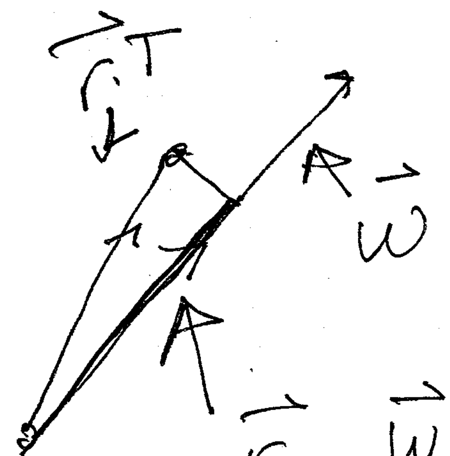
Mathematical formulation

$$A = \begin{array}{c} \\ m \\ \\ \end{array}\begin{bmatrix} \vec{r}_1^{\,T} \\ \vdots \\ \vec{r}_m^{\,T} \end{bmatrix}$$
$$\qquad\qquad n$$

data matrix

Then $\vec{r}_i$ give a collection of m different pts in $\mathbb{R}^n$.





$\vec{r}_j \cdot \vec{w}$

$= \vec{r}_j^{\,T} \vec{w}$

$\boxed{\vec{w} = \text{unit vector}}$

each $\vec{r}_i$ contributes a projection

$$\vec{r}_i \circ \vec{w}$$

TOTAL sum of Squared projections

$$\Phi(\vec{w}) = \sum_{i=1}^{m} (\vec{r}_i \circ \vec{w})^2$$

Find unit vector $\vec{w}$ that **MAXIMIZES**.

NOTE: $\Phi(\vec{w}) = \sum_{i=1}^{m} (r_i^T \vec{w})^2$

But

$$A\vec{w} = \begin{bmatrix} \vec{r}_1^{\,T} \\ \vdots \\ \vec{r}_m^{\,T} \end{bmatrix} \vec{w} = \begin{bmatrix} r_1^T \vec{w} \\ r_2^T \vec{w} \\ \vdots \\ \vec{r}_m^{\,T} \vec{w} \end{bmatrix}$$

__So__ $\Phi(w) = \|A\vec{w}\|_2^2$

Reformulated problem:

Find unit $\vec{w}$ which MAXIMIZES $\|A\vec{w}\|_2^2$

__OR__ $\boxed{\underset{\|\vec{w}\|=1}{\operatorname{argmax}} \|A\vec{w}\|_2^2}$

Theorem: If $A = U\Sigma V^T$ with $V = \left[\vec{v}_1 \cdots \vec{v}_n\right]$

$$\Rightarrow \vec{v}_1 = \underset{\|\vec{w}\|=1}{\operatorname{argmax}} \left(\|A\vec{w}\|_2^2\right)$$

and $\sigma_1^2 = \underset{\|\vec{w}\|=1}{\max} \left(\|A\vec{w}\|_2^2\right)$.

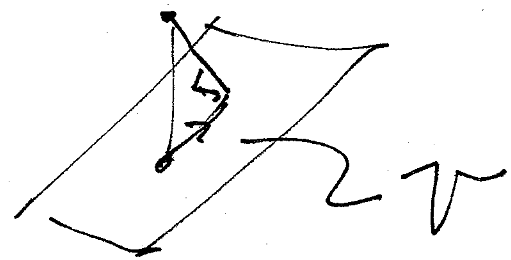
NOTE! Some books, articles put the samples in the columns and features in **Rows** (like Strang). $\Rightarrow$ Maximal direction is $\vec{u}_1$, the first left singular

We also have.

Theorem The k-dimensional subspace V "

that maximizes

$$\sum_{i=1}^{m} \left(\text{magnitude of projection of } \vec{r}_i \text{ orthogonally to } V \right)^2$$



is

$$\mathrm{Span}\{\underbrace{\vec{v}_1, \ldots, \vec{v}_k}_{\text{right sing vectors}}\}.$$