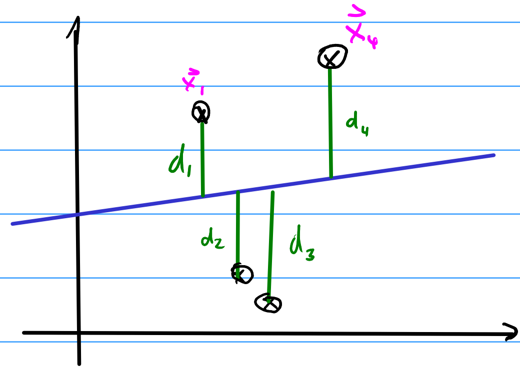
7 Singular Value Decomposition ("SVD")

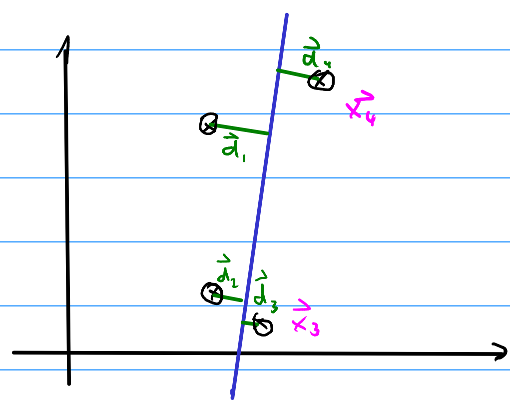
Given: data points "⊗" $\vec{x}_i$



Least squares: Can find line that minimizes $d_1^2+d_2^2+d_3^2+d_4^2$

But: Only vertical distance

Sometimes that's the right thing – but not always.



What if we'd like to find a line to minimize

$\|\vec{d}_1^2\| + \|\vec{d}_2\|^2 + \|d_3\|^2 + \|d_4\|^2$?

[Same answer as for least squares line fitting problem? No - see plc.

(Klein book: "trolley line problem")

Make problem easier (for now): Find line through origin that minimizes 2-norm distance to $(\vec{a}_i)$.

Mathematically: Find direction $\vec{v}$ (assume $\|\vec{v}\|_2=1$) s.t.

$$\sum_i \|\vec{d}_i\|_2^2 \quad \left(\text{where } \vec{d}_i \perp \vec{v},\ \vec{x}_i - \vec{d}_i \in \text{span}\{v\}\right)$$

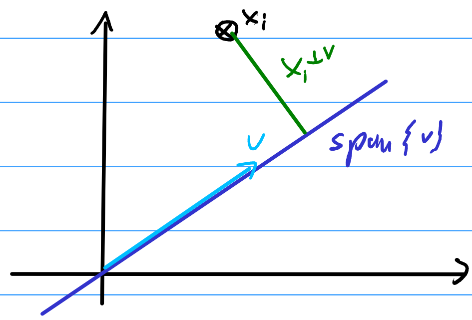
is minimized.

(Now drop vector arrows: $x_i = \vec{x}_i$, $d_i = \vec{d}_i$, $v = \vec{v}$)

[What is d_i? $x_i^{\perp v}$

$$x_i = x_i^{\perp v} + x_i^{\parallel v}$$

$$x_i = x_i^{\perp v} + (x_i, v)\, v$$



Pyth.
$$\Rightarrow \|x_i\|^2 = \|x_i^{\perp v}\|^2 + (x_i, v)^2 \underbrace{\|v\|^2}_{=1}$$

$$\rightsquigarrow \sum_i \|\vec{d}_i\|_2^2 = \sum_i \|x_i\|^2 - (x_i, v)^2$$

want to minimize

$$= \|X\|_F^2 - \|Xv\|_2^2$$

Frobenius

$\leftarrow$ Let $X = \begin{pmatrix} — x_1 — \\ \vdots \\ — x_n — \end{pmatrix}$.

[How do we minimize that? Maximize $\|Xv\|_2^2$.

[What's the maximal value of $\|Xv\|_2^2$? $\|X\|_2^2$

[How can we find v? Unclear (for now). Let's assume we can.

Terminology

v : first right singular vector of X $\rightsquigarrow$ rename to v_1

$\sigma_1 := \|Xv_1\|$ first singular value of X

lec 16

review right / left singular vectors

relate to trolley line example

review quiz

Q's:

Next, assume we look for a vector $v_2 \perp v_1$ such that

$$\|X v_2\|_2^2 \text{ is maximized.}$$

That's the second right singular vector of X.

And $\|X v_2\|$ is the second singular value.

Next, assume we look for a vector $v_3 \perp v_1, v_2 \ldots$

Then: $V^T = \begin{pmatrix} - v_1 - \\ \vdots \\ - v_n - \end{pmatrix}$ if X is an $m \times n$ matrix. (assume that for now)

same shape as X

$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix}$ where Σ is an $m \times n$ matrix in general.

not a sum – the Greek letter Σ.

$X = U \Sigma V^T$

$\rightsquigarrow X V = U \Sigma$ Demo: Finding the SVD

Result: The singular value decomposition factors any! $m \times n$ matrix into:

$$X = U \Sigma V^T$$

where

- U is $m \times m$ orthogonal ← columns: "left singular vectors"
- Σ is diagonal, $m \times n$ and has positive entries
- V is $n \times n$ orthogonal ← columns: "right singular vectors"

Numerical example

$$A = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\perp} \begin{pmatrix} 15 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \underbrace{\begin{pmatrix} -\frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix}}_{\perp}$$

Pseudo inverse

Define the pseudoinverse D^+ of an $m \times n$ diagonal matrix D as

$$D^+ = \underbrace{\overbrace{\begin{pmatrix} 1/\sigma_1 & & & \\ & 1/\sigma_2 & & \\ & & \ddots & \\ & & & 1/\sigma_{\min(m,n)} \\ & & 0 & \end{pmatrix}}^{m}}_{n} \quad \text{for } D = \begin{pmatrix} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_n & \end{pmatrix}$$

$$D^+ = \begin{pmatrix} 1/\sigma_1 & & & & \\ & 1/\sigma_2 & & & 0 \\ & & \ddots & & \\ & & & 1/\sigma_{\min(m,n)} & \end{pmatrix} \quad \text{for } D = \begin{pmatrix} \sigma_1 & & 0 \\ 0 & \ddots & \sigma_n \\ & 0 & \end{pmatrix}$$

[What to do about zero (or near-zero) σ_i? leave them as 0

Then define the pseudoinverse A^+ of a general matrix A by its SVD.
If $A = U\Sigma V^T$, then define $A^+ = V\Sigma^+ U^T$.

[Pseudoinverse of example?

Demo: Comparing the cost of LU, QR, SVD

lec 17

outline:

SVD apps

SVD in-class quiz

quiz review

eigval

practice exam submissions

review quiz

Q's: eigenvalues of A = eigenvalues of A^T?

Applications of the SVD

① Least squares problems

<u>Fact:</u> $A^{+}b$ solves the least squares problem $\min_{x} \|Ax-b\|_2$

This generalizes our QR-based method. $\rightarrow$ HW

[How? By also allowing underdetermined systems. (Fewer eqn's than unknowns, $A = \underset{\text{short}}{\overset{\text{fat}}{\square}}$)

② Principal component analysis $\rightarrow$ HW3

③ Computing $\|A\|_2$

$\|A\|_2 = \sigma_1$

④ Computing the 2-norm condition number

Assume A invertible.

<u>Recall:</u> $\kappa(A) = \|A\| \, \|A^{-1}\| = \sigma_1 / \sigma_n$

$$A = \begin{pmatrix} | & | & & | \\ & & U & \\ | & | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \begin{pmatrix} \text{---} \\ V^T \\ \text{---} \end{pmatrix} \qquad A^{-1} = \begin{pmatrix} | & | & & | \\ & & V & \\ | & | & & | \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \end{pmatrix} \begin{pmatrix} \text{---} \\ U^T \\ \text{---} \end{pmatrix}$$

$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n \geq 0$

Now, for <u>any</u> matrix (not just square/invertible), it turns out that

σ_1 / σ_n <u>is</u> the condition number.

(And that's also how it's computed. — Quite expensive!)

(5) Low-rank approximation

$$A = \begin{pmatrix} | & & | \\ & U & \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \begin{pmatrix} \text{—} & & \text{—} \\ & V^T & \\ \text{—} & & \text{—} \end{pmatrix}$$

$$= \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} \text{—} \ \sigma_1 v_1 \ \text{—} \\ \text{—} \ \sigma_2 v_2 \ \text{—} \\ \vdots \\ \text{—} \ \sigma_n v_n \ \text{—} \end{pmatrix} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_n u_n v_n^T$$

sum of outer products

= sum of rank-1 matrices

of decreasing norm!

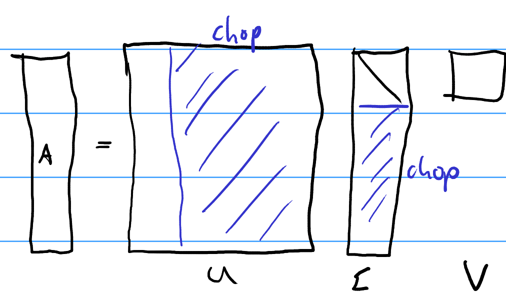
Idea: Could just use first few as an approximation.

In fact: $\sum_{i=1}^{k} \sigma_i u_i v_i^T$ is the closest rank-k matrix to A
(measured in the Frobenius-norm.)

[What happens if rank $A < k$? $A = \sum_i^k \sigma_i u_i v_i^T$!

[How does this relate to the trolley line problem? Closest rank-1!

Can define a variant of the SVD where U, V are not square.

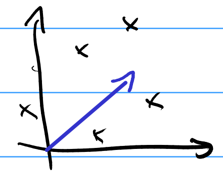


"thin" SVD

[Why is this important?
U consumes a large amount of memory.

Demo: image compression

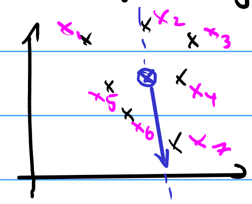
Original trolley line problem:



Line through the origin

<inclass-svd>

More generally - how about:



Line not through origin?

Just need one point on that line:
Can then subtract that point from all data.

Fact: $\frac{1}{n}\sum_{i=1}^{n} x_i$ is such a point.

"centroid"

[Relationship to PCA?