Todays lecture is all about applications of SVD

FYI only, not required

Review of SVD

- Singular Value Decomposition (SVD)

For any $A \in \mathbb{R}^{n\times n}$, there are $U, \Sigma, V \in \mathbb{R}^{n\times n}$ s.t.

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

① $\Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}$, $\sigma_i \geq 0$ are singular values

$$\sigma(A) = \sqrt{\lambda(A^TA)} = \sqrt{\lambda(AA^T)}$$

↳ eigenvalue

② Cols of U are orthonormal : left singular vectors

$$U^TU = UU^T = I$$

③ Cols of V are orthonormal : right singular vectors

$$V^TV = VV^T = I \qquad A = U\Sigma V^T \Rightarrow A^TA = V\Sigma^2V^T$$

④ $A = U\Sigma V^T$

$\Leftrightarrow A = U\Sigma V^{-1}$

$\Leftrightarrow AV = U\Sigma$

$\Leftrightarrow A[\vec{v}_1\ \vec{v}_2\ \vec{v}_3] = [\vec{u}_1\ \vec{u}_2\ \vec{u}_3]\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$

$\Leftrightarrow [A\vec{v}_1\ A\vec{v}_2\ A\vec{v}_3] = [\sigma_1\vec{u}_1\ \ \sigma_2\vec{u}_2\ \ \sigma_3\vec{u}_3]$

$\Leftrightarrow A\vec{v}_i = \sigma_i\vec{u}_i$ The order must match

⑤ $A = U\Sigma V^T \Rightarrow AA^T = U\Sigma V^T V \Sigma U^T = U\Sigma^2 U^T$

Review: A is real positive definite if ① $A = A^T$ ② $\vec{x}^T A \vec{x} > 0$, $\forall \vec{x} \in \mathbb{R}^n$, $\vec{x} \neq \vec{0}$

$= \Box \|$

- Theorem: A real symmetric $A \in \mathbb{R}^{n\times n}$ is positive (semi-)definite if and only if all eigenvalues of A are positive. (non-negative)

$\Rightarrow \begin{cases} \vec{u}_i \text{ are orthonormal eigenvectors of } AA^T. \\ \sigma_i^2 \text{ are eigenvalues of } AA^T. \end{cases}$

$A = U\Sigma V^T \Rightarrow A^TA = V\Sigma U^T U \Sigma V^T = V\Sigma^2 V^T$

$\Rightarrow \begin{cases} \vec{v}_i \text{ are orthonormal eigenvectors of } A^TA. \\ \sigma_i^2 \text{ are eigenvalues of } A^TA. \end{cases}$

And we have pick right order and signs of $\vec{u}_i, \vec{v}_i$ s.t.

$$A\vec{v}_i = \sigma_i \vec{u}_i$$

Example: $A = \begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & -2 & 0 \end{bmatrix}$. Find its SVD.

① $A^TA = \begin{bmatrix} 2 & 2 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & -2 & 0 \end{bmatrix} = \begin{bmatrix} 8 & 2 & 0 \\ 2 & 5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$\lambda(A^TA) = 9, 4, 0$

$\Rightarrow \sigma(A) = 3, 2, 0$

② Corresponding eigenvectors of A^TA are

$\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Orthonormal eigenvectors $\left(\dfrac{\vec{v}_i}{\|\vec{v}_i\|} \right)$:

$$\frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \quad \frac{2}{\sqrt{5}}\begin{pmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$V = \begin{pmatrix} 2/\sqrt{5} & -1/\sqrt{5} & 0 \\ 1/\sqrt{5} & 2/\sqrt{5} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$A\vec{v}_i = \sigma_i \vec{u}_i$

$\iff \vec{u}_i = \dfrac{A\vec{v}_i}{\sigma_i}, \ \sigma_i \neq 0$

③ $$AA^T = \begin{pmatrix} 4 & 4 & 0 \\ 4 & 5 & -2 \\ 0 & -2 & 4 \end{pmatrix}$$

$$\lambda(AA^T) = 9, \ 4, \ 0$$

eigenvectors: $\begin{pmatrix} -4 \\ -5 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$

orthonormal eigen-vectors $\pm\dfrac{1}{3\sqrt{5}}\begin{pmatrix} 4 \\ 5 \\ -2 \end{pmatrix}, \pm\dfrac{1}{\sqrt{5}}\begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix}, \pm\dfrac{1}{3}\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$

The sign is to be determined by $A\vec{v}_i = \sigma_i \vec{u}_i$

$$A\vec{v}_2 = \begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & -2 & 0 \end{bmatrix} \frac{2}{\sqrt{5}}\begin{bmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix} = \underset{\substack{\| \\ 2}}{\sigma_2} \cdot \frac{1}{\sqrt{5}}\begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix}$$

$$A\vec{v}_3 = \begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & -2 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \underset{\substack{\| \\ 0}}{\sigma_3} \cdot (\pm)\frac{1}{3}\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$$

$$\Rightarrow \quad U = \begin{pmatrix} \frac{4}{3\sqrt5} & \frac{-1}{\sqrt5} & -\frac{2}{3} \\ \frac{5}{3\sqrt5} & 0 & \frac{2}{3} \\ \frac{-2}{3\sqrt5} & \frac{-2}{\sqrt5} & \frac{1}{3} \end{pmatrix}$$

$$\Rightarrow A = U \begin{pmatrix} 3 & & \\ & 2 & \\ & & 0 \end{pmatrix} V^T \quad \Leftrightarrow A\vec{v}_i = \sigma_i \vec{u}_i$$

- $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, the Frobenius norm of A is defined as

$$\|A\|_F = \sqrt{\sum_{j=1}^{3}\sum_{i=1}^{3} a_{ij}^2}$$

- Useful formula: $\|A\|_F = \sqrt{tr(A^TA)} = \sqrt{tr(AA^T)}$

$$tr\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = 1+5+9$$

- $A = U\Sigma V^T \Rightarrow A^TA = (U\Sigma V^T)^T U\Sigma V^T$

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

$$= V\Sigma^2 V^T$$

$$= V \begin{pmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \sigma_3^2 \end{pmatrix} V^T$$

$$tr(A^TA) = tr(V\Sigma^2 V^T) = tr(\Sigma^2 V^T V)$$

$$tr(ABC) = tr(BCA)$$

$$= tr(\Sigma^2) = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$$

$$\Rightarrow \|A\|_F = \sqrt{\sigma_1^2+\sigma_2^2+\sigma_3^2}$$

- Theorem: The number of nonzero singular values is the rank of A

Proof: $A = U\Sigma V^T = U\begin{bmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\\ 0 & 0 & 0\end{bmatrix}V^T$, $\sigma_1, \sigma_2 > 0$.

$$A = [\vec{u}_1\ \vec{u}_2\ \vec{u}_3]\begin{bmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\\ 0 & 0 & 0\end{bmatrix}V^T$$

$$= [\sigma_1\vec{u}_1\ \ \sigma_2\vec{u}_2\ \ \vec{0}]\,V^T$$

$$= [\sigma_1\vec{u}_1\ \ \sigma_2\vec{u}_2\ \ \vec{0}]\begin{bmatrix}* & * & *\\ * & * & *\\ * & * & *\end{bmatrix}$$

$\Rightarrow$ cols of A are linear combinations of $\sigma_1\vec{u}_1$ and $\sigma_2\vec{u}_2$

$\Rightarrow$ Rank$(A) = 2$.

- Low-Rank Approximation Theorem

For $A \in \mathbb{R}^{n\times n}$, let $A = U\Sigma V^T$ be its SVD with singular values $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_n \geq 0$

For any $r = 1, 2, \ldots, n$, let $\tilde{\Sigma} = \begin{pmatrix}\sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0\end{pmatrix}$.

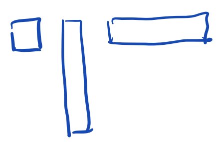
Then $\| A - U\tilde{\Sigma}V^T \|_F \le \| A - B \|_F$ for any rank-r matrix B.

This means $U\tilde{\Sigma}V$ is the best rank-r approximation to A.

- Example: $r=1$, $U\tilde{\Sigma}V^T = [\vec{u}_1\ \vec{u}_2\ \vec{u}_3]\begin{bmatrix}\sigma_1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix} V^T$

$$= [\sigma_1\vec{u}_1\ \vec{0}\ \vec{0}]\begin{bmatrix}\vec{v}_1^T\\ v_2^T\\ v_3^T\end{bmatrix}$$

$$= \sigma_1 \vec{u}_1 \vec{v}_1^T$$



- Principal Component Analysis (PCA):

  If $A = U\Sigma V^T$, then $\sigma_i \vec{u}_i$ are principal components.

- Low-rank compression:

  if we replace A by $\tilde{A} = U\tilde{\Sigma}V^T$

  then the error is $\| A - \tilde{A} \|_F = \| U(\Sigma - \tilde{\Sigma})V^T \|_F$

$$= \| U \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & \sigma_{r+1} & \\ & & & & \ddots & \\ & & & & & \sigma_n \end{pmatrix} V^T \|_F$$

A

$$= \sqrt{\sigma_{r+1}^2 + \cdots + \sigma_n^2}$$

For $\tilde{A}$, we only need to store $\approx$

① r scalars: $\sigma_1, \dots, \sigma_r$

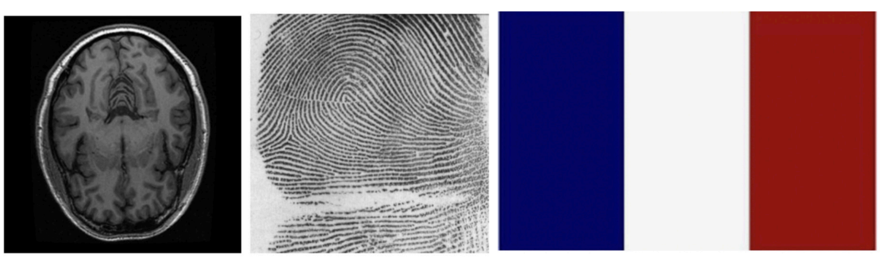
② $2r$ vectors: $\vec{u}_1, \dots, \vec{u}_r$

$\vec{v}_1, \dots, \vec{v}_r$

$r<n$: $U \tilde{\Sigma} V^T = \quad \tilde{\Sigma}$

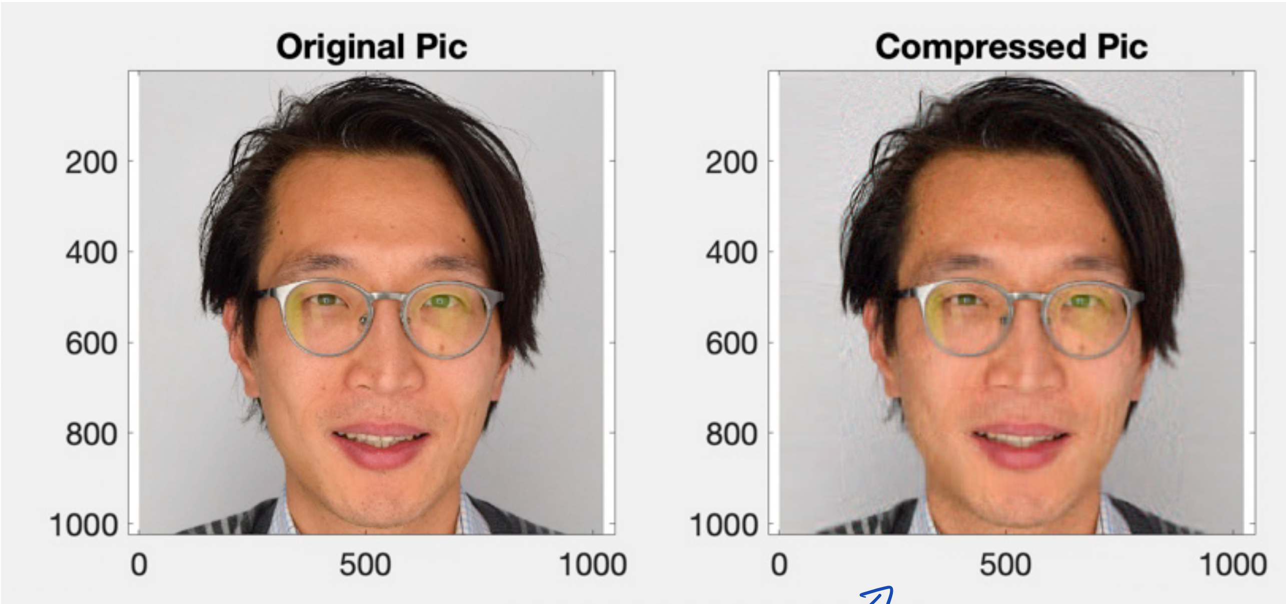
Example: $r=1$, $U \tilde{\Sigma} V^T = \sigma \vec{u}_1 \vec{v}_1^T =$

$$\begin{bmatrix} 2 & 2 & 3 & 3 & 4 & 4 \\ 2 & 2 & 3 & 3 & 4 & 4 \\ 2 & 2 & 3 & 3 & 4 & 4 \\ 2 & 2 & 3 & 3 & 4 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 2 & 3 & 3 & 4 & 4 \end{bmatrix}$$



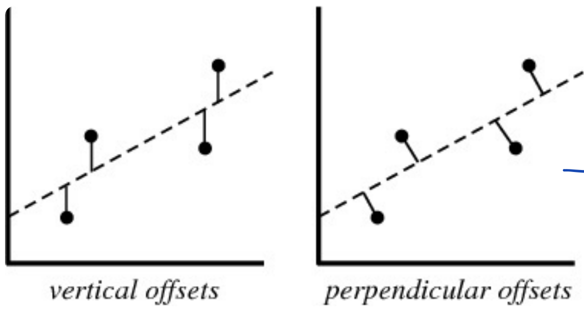
(a) MRI Image (b) Finger Print (c) France Flag

- Pictures are naturally stored as matrices: a 1024×1024 black and white picture is stored as a 1024×1024 matrix.
- For many pictures, if we store it as a matrix, it is a waste: MRI image has information only inside the curve boundary; finger prints have repeated/similar structure.
- Compression of data: we do not need to store the whole $n \times n$ pixels for a $n \times n$ picture.
- There are many ways to compress data: low-rank approximation is one of them.
- Example: France Flag is a rank-1 matrix. SVD is $A = U\Sigma V^* = \mathbf{u}_1 \sigma_1 \mathbf{v}_1^*$. So we only need to store two vectors $\mathbf{u}_1$ and $\mathbf{v}_1$ and σ_1.



Guess what the rank is ?

Line Fitting



→ computed by PCA/SVD of a nonsquare matrix

↓ Computed by least square

• SVD for $A \in \mathbb{R}^{m\times n}$, $m > n$.

$$\underset{m\times n}{\boxed{A}} = \underset{m\times n}{\boxed{U}}\;\underset{n\times n}{\boxed{\Sigma}}\;\underset{n\times n}{\boxed{V^T}}$$

$$\Sigma = \begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n\end{pmatrix}$$

$$\sigma_i(A) = \sqrt{\lambda_i(A^TA)}$$

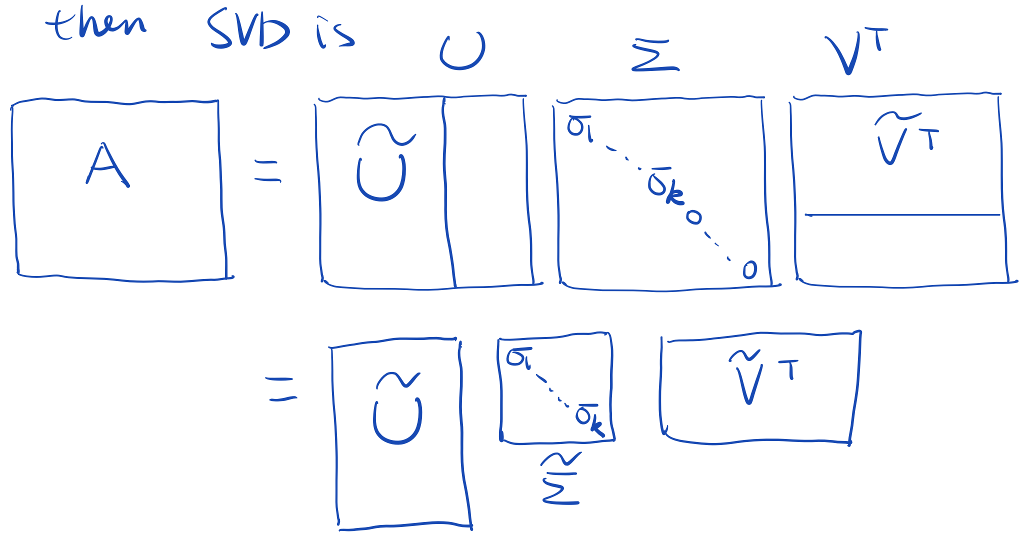
The eigen-decomposition: $\boxed{A^TA} = \boxed{V}\begin{pmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2\end{pmatrix}\boxed{V^T}$

$$\boxed{AA^T} = \boxed{U \;\;\big|\;\;}\begin{pmatrix}\sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_n^2 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0\end{pmatrix}\boxed{\begin{matrix}U^T \\ \hline \\ \end{matrix}}$$

$$= \boxed{U}\begin{pmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2\end{pmatrix}\boxed{U^T}$$

- Compact SVD: for $A \in \mathbb{R}^{n\times n}$, if $\operatorname{rank}(A) = k < n$,

  then SVD is



$= \tilde{U}\tilde{\Sigma}\tilde{V}^T$ is called compact SVD.

- Compact SVD, low-rank approximation and PCA are used for almost all matrix applications.

- σ_i contain lots of information:

  even for a nonsquare $A \in \mathbb{R}^{m\times n}$, $m > n$

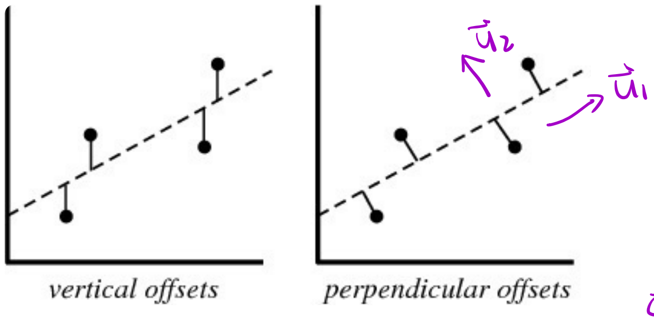
  ① $\sqrt{\sigma_1^2 + \cdots + \sigma_n^2} = \|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$

  ② $\operatorname{rank}(A) = \#$ of nonzero σ_i

To find the best line fit using perpendicular offsets for (x_i, y_i)

Step I: $\bar{x} = \frac{1}{4}\sum_{i=1}^{4} x_i \quad \bar{y} = \frac{1}{4}\sum_{i=1}^{4} y_i$

$\hat{x}_i = x_i - \bar{x}, \quad \hat{y}_i = y_i - \bar{y}$ So that line passes the origin



vertical offsets
perpendicular offsets

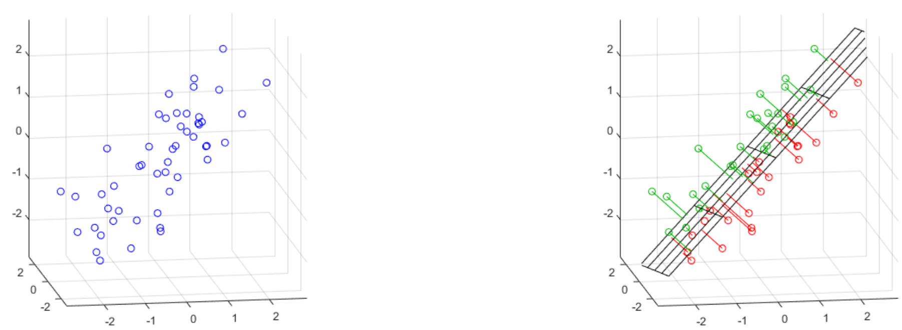
$$A = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \bar{x}_4 \\ \bar{y}_1 & \bar{y}_2 & \bar{y}_3 & \bar{y}_4 \end{bmatrix}$$

AA^T : 2×2

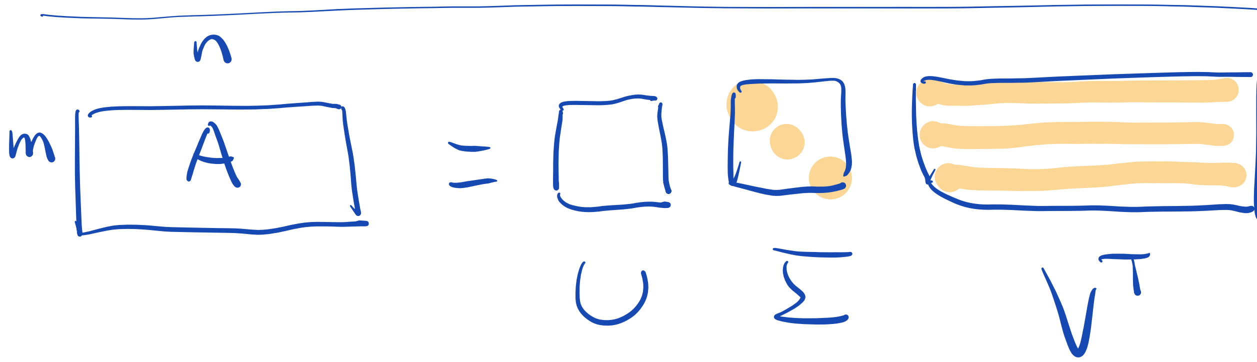
σ_1 $\vec{u}_1$ $\vec{u}_2$ σ_2

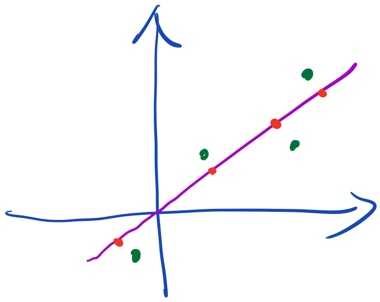
Why PCA minizes perpendicular offsets?

More general setup: 3D data and a plane.



$$A = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \\ \bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_n \\ \bar{z}_1 & \bar{z}_2 & \cdots & \bar{z}_n \end{bmatrix}$$





$$A = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \end{bmatrix}$$

$$\boxed{A} = \boxed{U}\,\boxed{\Sigma}\,\boxed{V^T}$$

$$\approx \boxed{\vec{u}_1}\,\boxed{\sigma_1}\,\boxed{\vec{v}_1^T}$$