# Singular Value Decomposition

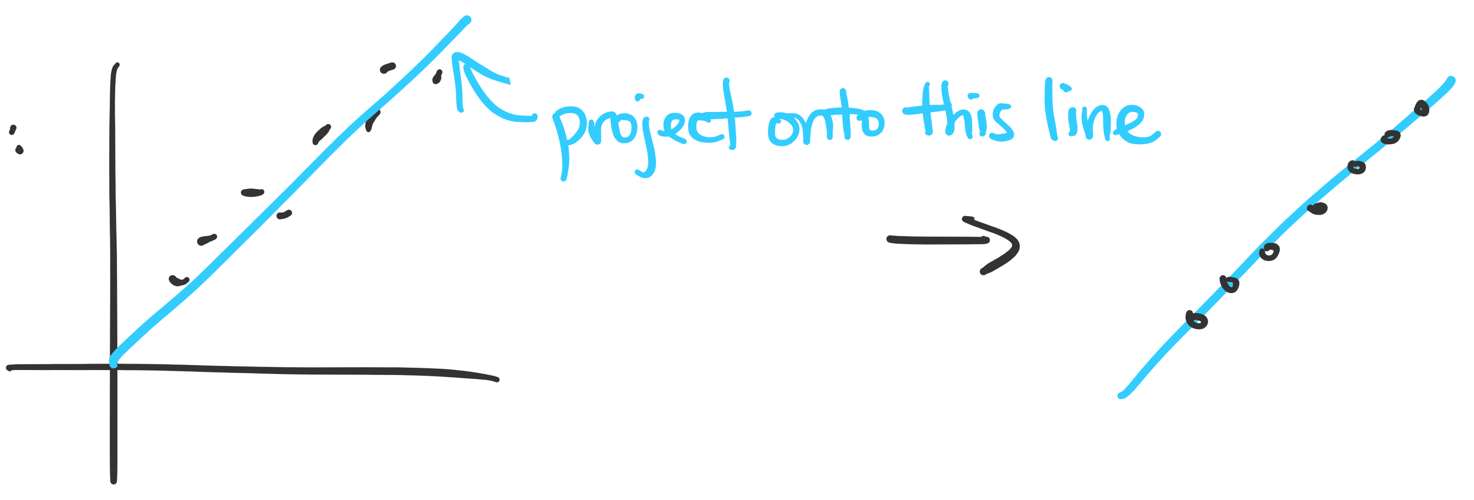
Tuesday, October 7, 2025 8:46 AM

Recall from Johnson-Lindenstrauss dimension reduction:

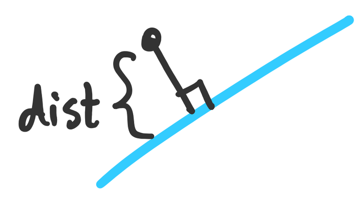
- Given data points $x_1, \dots, x_n \in \mathbb{R}^d$ in high dim, can compress data into $Ax_1, \dots, Ax_n \in \mathbb{R}^k$, $k = O\left(\frac{\log n}{\varepsilon^2}\right)$, s.t. distances are preserved up to $(1+\varepsilon)$.
- $A$ is just random Gaussian entries scaled by $\frac{1}{\sqrt{k}}$, oblivious to data points.
- Downside: $\frac{\log n}{\varepsilon^2}$ is still fairly large. What if want 1 dimension?

Singular Value Decomposition: compress to very low dimensions while maximizing "similarity"

Example (2D → 1D):



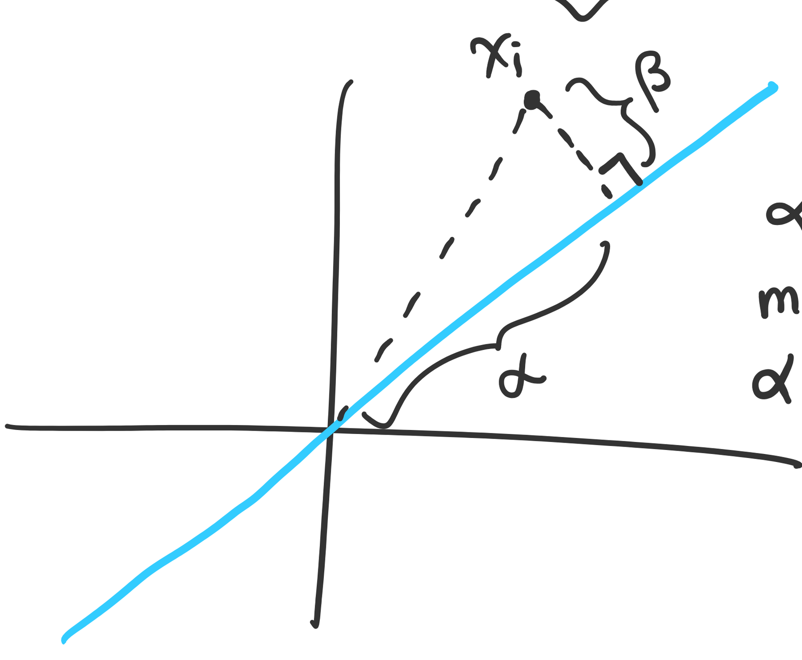

How to measure similarity? Minimize sum of squared errors:
Given $n$ points $x_1, \dots, x_n \in \mathbb{R}^d$ and integer $k \le d$, find subspace $V$ of dim $k$ minimizing $\sum \text{dist}(x_i, V)^2$.




Singular Value Decomposition computes this subspace.

Warm-up: $k=1$, $V = \text{span}(v)$ for some vector $v \in \mathbb{R}^d$.

Need to solve $\min_{\|v\|=1} \underbrace{\sum \text{dist}(x_i, \text{span}(v))^2}$.




$\alpha^2 + \beta^2 = \|x_i\|^2$
minimizing $\beta \iff$ maximizing $\alpha$.
$\alpha$ = length of projection of $x_i$ onto $v$
$= \langle x_i, v \rangle$.

So suffices to solve $\max_{\|v\|=1} \sum \langle x_i, v\rangle^2 = \max_{\|v\|=1} \left\| \begin{bmatrix} — x_1 — \\ — x_2 — \\ \vdots \\ — x_n — \end{bmatrix} \begin{bmatrix} v \end{bmatrix} \right\|_2^2$

Define the first (right) singular vector $v$ as the $v$ maximizing ↑.

For general $k$, define it iteratively:

$$v_1 = \underset{\|v\|=1}{\arg\max} \|Av\|$$

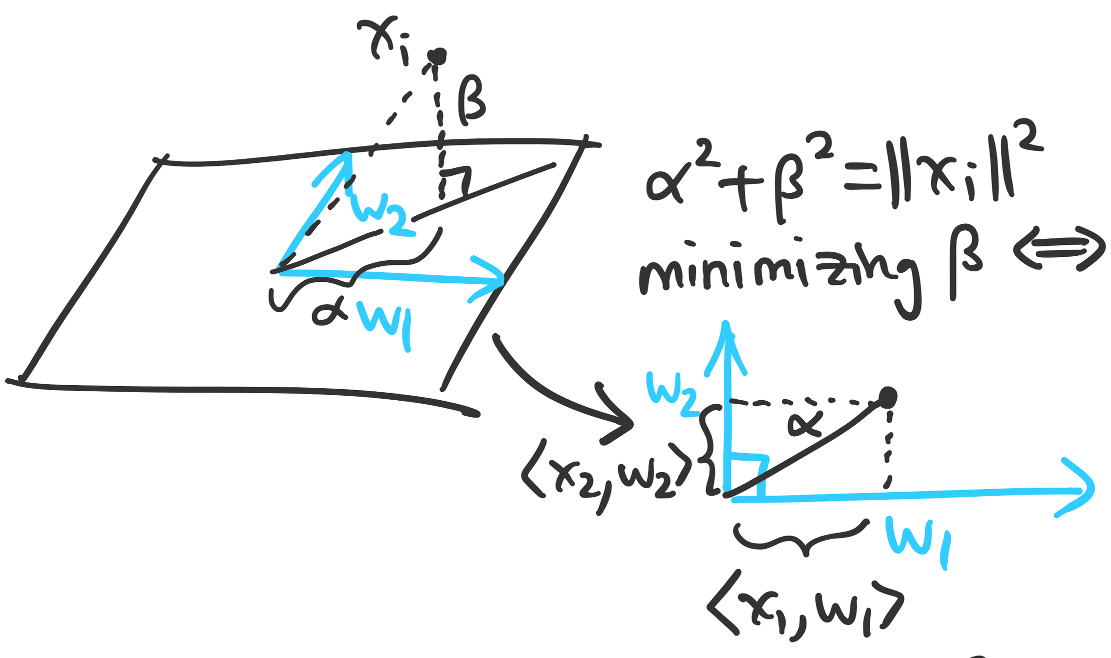
$$v_2 = \underset{\|v\|=1,\, v \perp v_1}{\arg\max} \|Av\|$$

$$v_3 = \underset{\|v\|=1,\, v \perp v_1, v_2}{\arg\max} \|Av\|$$

These are the right singular values of $A$.

Lemma: $V_k = \text{span}(v_1, \dots, v_k)$ minimizes $\sum \text{dist}(x_i, V)^2$ over all dim $k$ subspaces $V$.

Proof: consider $k=2$. Let $W_2$ be any other 2-dim subspace spanned by orthonormal $w_1, w_2$ where $w_2 \perp v_1$.




$\alpha^2 + \beta^2 = \|x_i\|^2$
minimizing $\beta \iff$ maximizing $\alpha$

$\alpha = \langle x_1, w_1\rangle^2 + \langle x_2, w_2 \rangle^2$.

So need to maximize $\|Aw_1\|_2^2 + \|Aw_2\|_2^2$.

Since $v_1 = \arg\max \|Av\|$, $\|Av_1\|_2^2 \ge \|Aw_1\|_2^2$

Since $v_2 = \underset{v \perp v_1}{\arg\max} \|Av\|$ and $w_2 \perp v_1$, $\|Av_2\|_2^2 \ge \|Aw_2\|_2^2$

So $(v_1, v_2)$ maximizes it.

---

Singular Value Decomposition: Define left singular vectors $u_i = \frac{Av_i}{\|Av_i\|} \in \mathbb{R}^n$ and singular values $\sigma_i = \|Av_i\|$ for $i = 1, \dots, d$.

Lemma: $A = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_d \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_d \end{bmatrix} \begin{bmatrix} — v_1 — \\ \vdots \\ — v_d — \end{bmatrix} = UDV^T$.

Proof: First, verify that $Av_i = UDV^T v_i$ for all $i = 1, \dots, d$.

$UDV^T v_i = UDe_i = U \cdot \sigma_i e_i = \sigma_i u_i = \|Av_i\| \cdot \frac{Av_i}{\|Av_i\|} = Av_i$.

If we can show $Av_i = UDV^T v_i$ for a basis $v_1, \dots, v_n$ then done.
Extend $v_1, \dots, v_d$ to a basis by adding $v_{d+1} \cdots v_n$ in kernel of $A$.
So for $i > d$, $Av_i = 0$ and since $v_i \perp v_1, \dots, v_d$, $V^T v_i = 0 \Rightarrow UDV^T v_i = 0$