13 Apr 2026 SVD part 2.

Matrix $A \in \mathbb{R}^{m \times n}$. Rows $a_1, a_2, \ldots, a_m$.

Define vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ "right singular vectors"

Sequence of optimizations $\|\cdot\|$ denotes ℓ^2 norm $\sqrt{\sum x_i^2}$.

$$v_i = \arg\max_{\|x\|=1} \left\{ \|Ax\| : \langle x, v_j \rangle = 0 \ \forall j < i \right\}$$

not obvious how to solve efficiently.

Singular values $\sigma_i = \|Av_i\|$

Left singular vectors $u_i = \frac{1}{\sigma_i} \cdot Av_i$.

By construction, $\{v_1, \ldots, v_n\}$ are orthonormal basis of $\mathbb{R}^n$.

Lemma: $A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$.

Proof. Suffices to show $Av = \left[\sum_{i=1}^{n} \sigma_i u_i v_i^T\right] v$ for all $v \in \mathbb{R}^n$. In fact suffices to show this for $v \in \{v_1, \ldots, v_n\}$.

When $v = v_j$,

$$\left[\sum_{i=1}^{n} \sigma_i u_i v_i^T\right] v_j = \sum_{i=1}^{n} \sigma_i u_i \left(v_i^T v_j\right) = \sigma_j u_j$$
$$= Av_j$$

Letting $V_d = \text{span}\{v_1, \dots, v_d\}$, last lecture ended with the claim that V_d minimizes

$$\sum_{i=1}^{m} \text{dist}(a_i, V)^2$$

over all d-dimensional $V \subseteq \mathbb{R}^n$.

Proof. If $W \subseteq \mathbb{R}^n$ is dimensional with orthonormal basis $\{w_1, \dots, w_d\}$, then for all $a \in \mathbb{R}^n$ then projection of a on W is $\sum_{i=1}^{d} \langle a, w_i \rangle w_i$

$$\| \text{proj}_W(a) \|^2 = \sum_{i=1}^{d} \langle a, w_i \rangle^2$$

$$\text{dist}(a, W)^2 = \|a\|^2 - \sum_{i=1}^{d} (a \cdot w_i)^2$$

Minimizing $\sum_{j=1}^{m} \text{dist}(a_j, W)^2$ same as maximizing $\sum_{j=1}^{m} \sum_{i=1}^{d} (a_j \cdot w_i)^2$.

To prove V_d attains maximum of ↑, induct on d. Base case $d=1$: definition of v_1.

General case: suppose W is a d-dim subspace with orthonormal basis $\{w_1, \dots, w_d\}$.

Choose a basis with w_d orthogonal to each of $v_1, \ldots, v_{d-1}$.

(W has a 1-dimensional subspace of vectors orthogonal to $v_1, \ldots, v_{d-1}$. Choose w_d to be any unit vector in that subspace.)

$$\|w_d\| = 1, \quad \langle w_d, v_i \rangle = 0 \quad \forall i < d$$

$$\therefore \boxed{\|Aw_d\|^2 \leq \|Av_d\|^2}$$

b/c v_d was chosen to maximize $\|Av_d\|$ under some constraints.

Also, by ind hyp.

$$\sum_{j=1}^{m} \sum_{i=1}^{d-1} (a_j \cdot w_i)^2 \leq \sum_{j=1}^{m} \sum_{i=1}^{d-1} (a_j \cdot v_i)^2$$

$$\sum_{j=1}^{m} (a_j \cdot w_d)^2 \leq \sum_{j=1}^{m} (a_j \cdot v_d)^2$$

Add

$$\sum_{j=1}^{m} \sum_{i=1}^{d} (a_j \cdot w_i)^2 \leq \sum_{j=1}^{m} \sum_{i=1}^{d} (a_j \cdot w_i)^2$$

$\therefore$ induction step confirmed.

subspace W of dimension d

$$A = \begin{bmatrix} — a_1 — \\ \vdots \\ — a_m — \end{bmatrix} \Longrightarrow \begin{bmatrix} — \text{proj}_W(a_1) — \\ \vdots \\ — \text{proj}_W(a_m) — \end{bmatrix}$$

The following optimization problems are equiv.

① $\min \left\{ \sum_{i=1}^{m} \text{dist}(a_i, W)^2 : \dim W = d \right\}$

② $\max \left\{ \sum_{i=1}^{m} \text{proj}_W(a_i)^2 : \dim W = d \right\}$

③ $\min \left\{ \sum_{i=1}^{m} \| a_i - b_i \|^2 : \text{rank } B \leq d \right\}$

$$B = \begin{bmatrix} — b_1 — \\ \vdots \\ — b_m — \end{bmatrix}$$

We proved $\{v_1, \ldots, v_d\}$ span a space that optimizes ①, ②.

So the optimizer of ③ is the matrix B whose rows are projections of a_i onto $\text{span}(v_1, \ldots, v_d)$.

$$\text{proj}_{V_d}(a_i) = \sum_{j=1}^{d} (a_i \cdot v_j)\, v_j$$

$\text{proj}_{V_d}(a_i)$ as a row vector

$$= \sum_{j=1}^{d} (a_i \cdot v_j)\, v_j^T$$

$A v_j = \sigma_j u_j$

$$\begin{bmatrix} \text{---}\ \text{proj}_{V_d}(a_1)\ \text{---} \\ \vdots \\ \text{---}\ \text{proj}_{V_d}(a_n)\ \text{---} \end{bmatrix} = \sum_{j=1}^{d} (A v_j)\, v_j^T$$

$$= \sum_{j=1}^{d} \sigma_j u_j v_j^T .$$

$B = \sum_{j=1}^{d} \sigma_j u_j v_j^T$ minimizes

$\sum_i \sum_j (a_{ij} - b_{ij})^2$ over rank $\leq d$ matrices.

$$\| A - B \|_F^2$$

"squared Frobenius norm"

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

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

$$\underbrace{m\left\{\begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix}\right.}_{n} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}}_{n \times n} \underbrace{\begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_n^T & - \end{bmatrix}}_{n \times n}$$

If $m > n$ pick any $u_{n+1}, \dots, u_m$ unit-norm, orthogonal to $u_1, \dots, u_n$

$$A = \underbrace{\begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix}}_{m \times m} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_n^T & - \end{bmatrix}}_{n \times n}$$

If $m < n$ then $\sigma_{m+1}, \sigma_{m+2}, \dots, \sigma_n$ were already equal to zero, so the product is unchanged if we delete the columns $u_{m+1} \cdots u_n$.

In all cases,

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

- U: by construction columns are unit norm.
- Σ: diagonal with $\sigma_1, \ldots, \sigma_n$ on its diag.
- V^T: by construction, this matrix is orthogonal.

$$u_i = \frac{1}{\sigma_i}(Av_i) = \frac{Av_i}{\|Av_i\|}$$ ← unit vector.

Turns out U is also an orthogonal matrix. (Proof deferred.)

Calculating $\{u_1, \ldots, u_d\}$ and $\{v_1, \ldots, v_d\}$.

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T$$
$$= U\Sigma V^T V \Sigma^T U^T$$
$$= U(\Sigma\Sigma^T)U^T$$

$$= U \cdot \mathrm{Diag}(\sigma_1^2, \dots, \sigma_m^2) \cdot U^T$$

Claim $(AA^T) u_i = \sigma_i^2 u_i \quad \forall i \in [m]$.

$$U^T u_i = \begin{pmatrix} - u_1^T - \\ - u_2^T - \\ \vdots \\ - u_m^T - \end{pmatrix} u_i = e_i$$

$$\mathrm{Diag}(\sigma_1^2 \dots \sigma_m^2) \cdot e_i = \sigma_i^2 e_i$$

$$U(\sigma_i^2 e_i) = \sigma_i^2 (U e_i) = \sigma_i^2 u_i$$

$\therefore$ u_i is eigenvector of AA^T

eigenvalue σ_i^2.

Similar calculation shows

$$A^T A = V(\Sigma^T \Sigma) V^T$$

and v_i is eigenvect of $A^T A$

with eigenvalue σ_i^2.

TL;DR: right sing vect of A

= eigvect of $A^T A$

left sing vect of A

= eigvect of AA^T.