$\boxed{\text{3rd optimization property of SVD}}$

$$A = U\Sigma V^T = \sum_{j=1}^{r} \sigma_j \vec{u}_j \vec{v}_j^T \qquad \sigma_1 \ge \sigma_2 \ge \dots$$

r = # of nonzero sing. val.

Define $A_l = \sum_{j=1}^{l} \sigma_j \vec{u}_j \vec{v}_j^T$ is the best rank l approx to A.

Theorem (Eckart–Young)

$$\min_{\operatorname{rank}(B)=l} \|B-A\|_2 = \|A_l - A\|_2 = \sigma_{l+1}$$

$$\min_{\operatorname{rank}(B)=l} \|B-A\|_F = \|A_l - A\|_F = \sqrt{\sigma_{l+1}^2 + \dots + \sigma_r^2}$$

$\|A_L - A\|$ = error in using A_L to approximate A in the given norm.

$\boxed{\text{Proof in Books}}$ – But two observation

1. $\text{rank}(A_L) = L$

$$A_L = \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_L \vec{u}_L \vec{v}_L^T$$

$$= \begin{bmatrix} \vec{u}_1 \cdots \vec{u}_L \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_L \end{bmatrix} \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_L^T \end{bmatrix}$$

$$= \begin{bmatrix} \vec{u}_1 \cdots \vec{u}_L \cdots \vec{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_L & \\ 0 & & & 0 \; 0 \; 0 \end{bmatrix} \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix}$$

$$= \quad U \quad \Sigma_L \quad V^T$$

③

$A_L = U \Sigma_L V^T$ is the SVD of A_L.

rank A_L = # of nonzero sing val = L

namely $\sigma_1, \ldots, \sigma_L$ since $\Sigma_L = \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_L & & \\ & & & 0 & \\ 0 & & & & 0 \end{bmatrix}$

---

② $\|A_L - A\|_2 = \sigma_{L+1}$

$$\begin{aligned} A - A_L &= U \Sigma V^T - U \Sigma_L V^T \\ &= U(\Sigma - \Sigma_L) V^T \\ &= U \begin{bmatrix} 0 & & & & & \\ & 0 & & & & \\ & & \ddots & & & \\ & & & 0 & & \\ & & & & \sigma_{L+1} & \\ & & & & & \ddots \\ & & & & & & \sigma_r \end{bmatrix} V^T \end{aligned}$$

FACT: 2 norm of matrix is its largest singular value.

$$\|A - A_L\|_2 = \left\| U \begin{bmatrix} 0 & \cdots & & & 0 \\ & 0 & & & \\ & & \sigma_{L+1} & & \\ & & & \ddots & \\ 0 & & & & \sigma_r \; 0 \end{bmatrix} V^T \right\|_2$$

$$= \left\| \begin{bmatrix} 0 & 0 & \cdots & & 0 \\ & & \sigma_{L+1} & & \\ & & & \ddots & \\ 0 & & & & \sigma_r \; 0 \end{bmatrix} \right\|_2 = \sigma_{L+1}$$

Hint for Frobenius norm result

$$\|UA\|_F = \|AU\|_F = \|A\|_F$$

proof <u>via</u> $\|A\|_F^2 = \text{trace}(A^TA)$

$$= \text{trace}(AA^T).$$

2^{nd} optimization property of the SVD

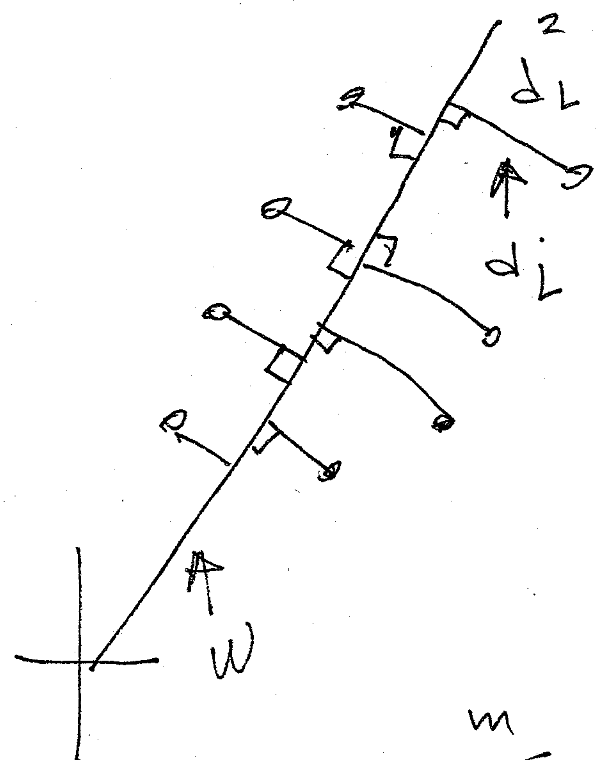
$$\begin{array}{r c} & n \\ \text{Samples} \rightarrow & \\ \text{cust} \quad m & \begin{bmatrix} \longleftarrow \vec{r}_i^{\,T} \longrightarrow \\ \\ \\ \\ \end{bmatrix} \\ & \uparrow \end{array}$$

categories ← products

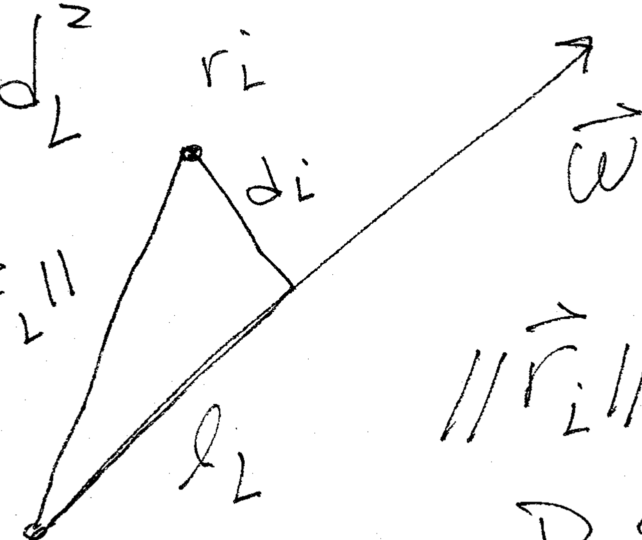
Treat $\{\vec{r}_i\}$ as m points in $\mathbb{R}^n$

(6)

Find unit vector $\vec{w}$ which <u>minimizes</u> the sum of the squares of the perpendicular distances to $\vec{w}$.



$$\underline{\underline{\min}} \quad \Lambda(\vec{w}) = \sum_{i=1}^{m} d_i^2$$



$$\|\vec{r}_i\|^2 = d_i^2 + l_i^2$$

<u>Pythag.</u>

$$d_i^2 = \|\vec{r}_i\|^2 - l_i^2$$

2

$$\operatorname{Argmin} \sum_{i=1}^{m} d_i^2 = \operatorname{Argmin} \sum_{i=1}^{m} \|\vec{r}_i\|_2^2 - \underbrace{(\vec{r}_i \cdot \vec{w})}_{l_i}{}^2$$

$$= \operatorname{Argmin} \|A\|_F^2 - \sum_{i=1}^{m} (\vec{r}_i \cdot \vec{w})^2$$

$$= \operatorname{Argmin} (\underbrace{\|A\|_F^2}_{const} - \Phi(\vec{w}))$$

1st opt. problem

$$= \operatorname{ArgMAX} \Phi(\vec{w})$$

$$= \vec{V}_1,$$ 1st right sing vect from 1st opt problem.

Theorem: If $A = U \Sigma V^T$ then $\vec{w} = \vec{v}_1$ is the direction that minimizes the sum of the squares of the perperpendicular distances.

And for each $k \leq r$ then the subspace $W_k = \text{span}\{\vec{v}_1, .., \vec{v}_k\}$ is the k-dimensional subspaces which minimizes the sum of the squares of the perpendicular distaces to W_k. — "Best fit k-dim subspace"

More on norms

If A has SVD $A = U\Sigma V^T$ and $\operatorname{rank}(A) = r$.

changed the names to make end result look nice

(1) $\|A\|_2 = \sigma_1$

(2) $\|A\|_F = (\sigma_1^2 + \dots + \sigma_r^2)^{1/2}$.

(3) A is square, invertible, $n \times n$

$\Rightarrow \|A^{-1}\| = 1/\sigma_n$ Proof $A = V\Sigma U^T$

$$\Rightarrow A^{-1} = (U^T)^{-1}\Sigma^{-1}V^{-1} = U \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}^{-1} V^T$$

$$= U \begin{bmatrix} 1/\sigma_1 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_n \end{bmatrix} V^T$$

Numerics or large scale computations

Errors, magnification of errors, conditioning, stability, round off error.

conditioning = perturbation behaviour of the math problem. — error in the input gives rise to what error in the output or solution

stability := perturbation behaviour when implemented.

Abstract structure

$$\mathbb{X} \xrightarrow{F} \mathbb{Y}$$

Inputs or data.

Outputs or solutions

$$x \longrightarrow F(x)$$

$$x+\delta x \longrightarrow F(x+\delta).$$

Error

derivative

$$\frac{F(x+\delta x)-F(x)}{\delta x} \cong \underline{\underline{DF}}$$