

L15: Principal Component Analysis via SVD

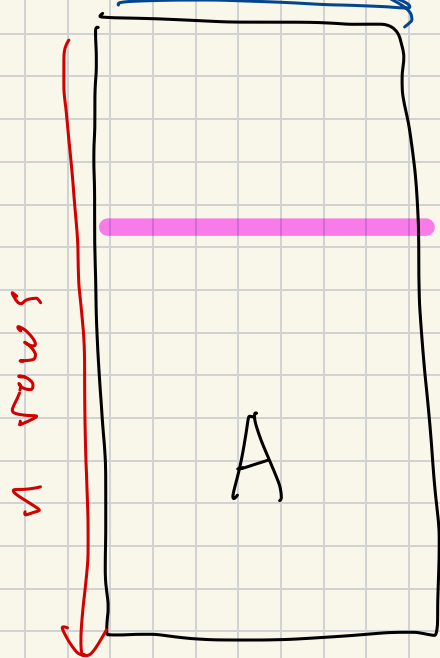
Data Mining
10.22.2026

Jeff M. Phillips



Input $A \in \mathbb{R}^{n \times d}$

d columns



$a_i \in \mathbb{R}^d$

$$A = \{a_1, a_2, \dots, a_n\} \in \mathbb{R}^d$$

dimension d too big

Dimensionality Reduction

- each column is same units compatible

$$\|a_i - p\|_2 = \sqrt{\sum_{j=1}^d (a_{ij} - p_j)^2}$$

Output

compare
draw

Mapping $\pi : \mathbb{R}^d \rightarrow \mathbb{R}^k$
 $\pi(a_i) \in \mathbb{R}^d$
 $\pi(a_i) \in \mathbb{R}^k$

$$k \ll d$$

NEW CUYAMA

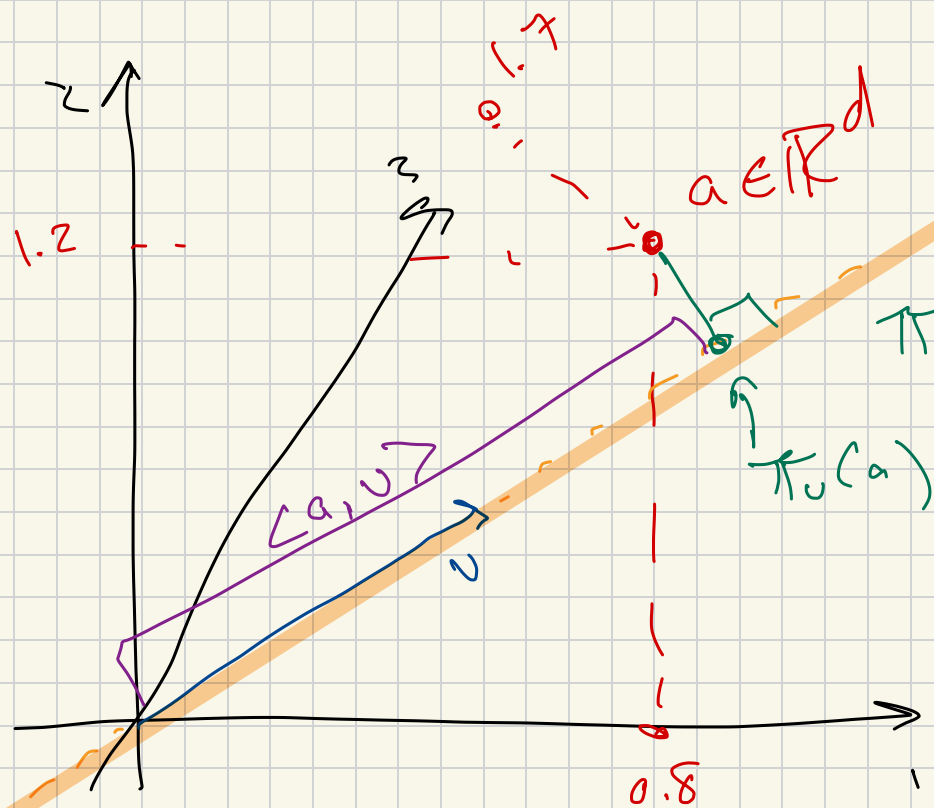
Population	562
Ft. above sea level	2150
Established	1951

TOTAL	4663
-------	------

Don't do this!

Dot Product = Inner Product $\Rightarrow$ Projection

$$\langle a, u \rangle = \sum_{j=1}^d a_j u_j$$

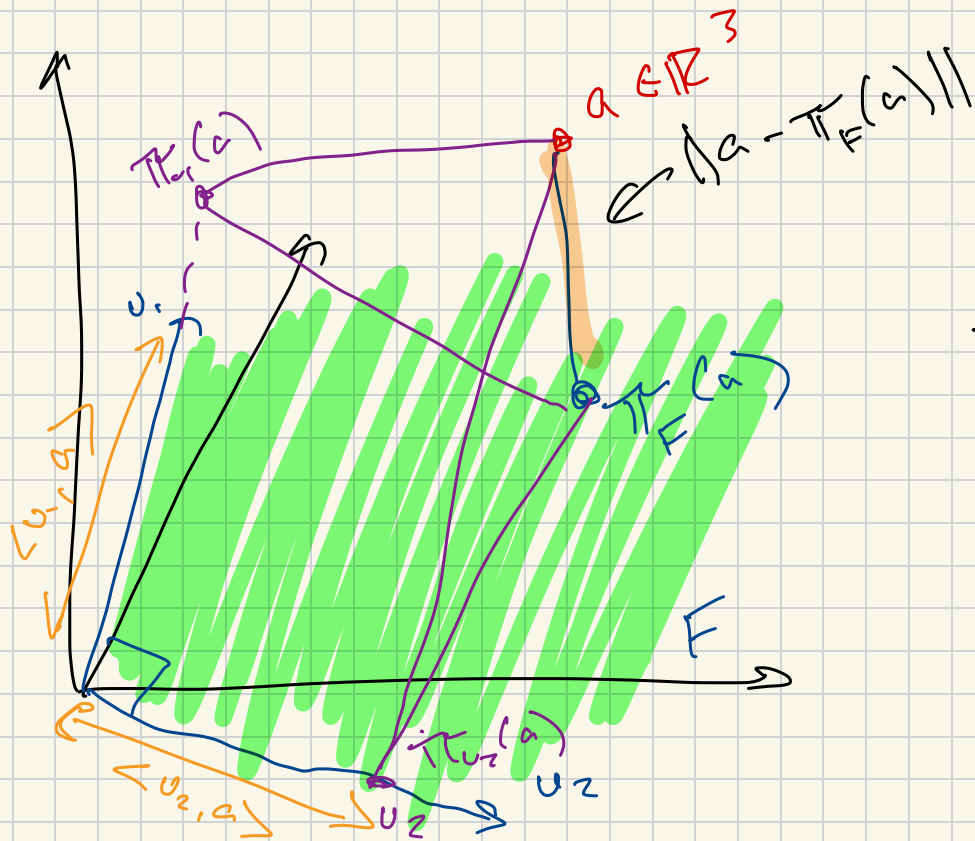


$$\pi_u(a) = \underbrace{\langle a, u \rangle}_{\text{how far}} \underbrace{u}_{\text{what direction}} \in \mathbb{R}^d$$

u unit vector

$$u \in \mathbb{R}^d$$

$$\|u\| = 1$$



$F = \{u_1, u_2\}$
 $\underbrace{\hspace{1cm}}_{\text{unit vectors in } \mathbb{R}^d}$

$\langle u_1, u_2 \rangle = 0$
 or orthogonal

$$\Pi_F(a) = \langle a, u_1 \rangle u_1 + \langle a, u_2 \rangle u_2$$

$F = \{u_1, u_2, \dots, u_n\}$ unit, orthogonal

$$\Pi_F(a) = \sum_{j=1}^n \langle a, u_j \rangle u_j$$

Project to k -dimensional basis $\bar{\Pi}_F$

$$\bar{\Pi}_F : \mathbb{R}^d \rightarrow \mathbb{R}^k$$

$$\bar{\Pi}_F(a) = (\langle a, v_1 \rangle, \langle a, v_2 \rangle, \dots, \langle a, v_k \rangle)$$

Sum of Squared Errors

$$SSE(A, F) = \sum_{a_i \in A} \|a_i - \pi_F(a_i)\|^2$$

Goal: Find k -dim subspace $F = \{v_1, v_2, \dots, v_k\}$

$$F^* = \underset{F}{\text{argmin}} SSE(A, F)$$

$$\|v_j\| = 1$$

$$\langle v_i, v_j \rangle = 0$$

→ solved by Singular Value Decomposition
SVD

Not yet! solve Principal Component Analysis
PCA

Singular Value Decomposition

SVD

$$\begin{matrix} n \\ \downarrow \\ A \in \mathbb{R}^{n \times d} \end{matrix}$$

=

$$\begin{matrix} \begin{matrix} u_j \\ \downarrow \\ U \in \mathbb{R}^{n \times n} \end{matrix} & \begin{matrix} S \in \mathbb{R}^{d \times d} \\ \text{diagonal} \end{matrix} \end{matrix}$$

$$V^T \in \mathbb{R}^{d \times d}$$

jth right sine vector
 $\|v_j\|=1$
 $\langle v_j, v_{j'} \rangle = 0$

$S_{jj} = \sigma_j$
 jth singular value

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_d \geq 0$
 decreasing order

$$A =$$

$$U S V^T$$

orthogonal
 u_j - jth left sine vector
 $\|u_j\|=1$
 $\langle u_j, u_{j'} \rangle = 0$
 $u_j \in \mathbb{R}^n$

$$A = U S V^T = \sum_{j=1}^d u_j \sigma_j v_j^T = \sum_{j=1}^d \sigma_j \underbrace{u_j v_j^T}_{\in \mathbb{R}^{n \times d}}$$

Low-Rank Approx of A

$$A_k = \sum_{j=1}^k \sigma_j u_j v_j^T \in \mathbb{R}^{n \times d}$$

minimize

$$\|A - A_k\|_{2,F}$$

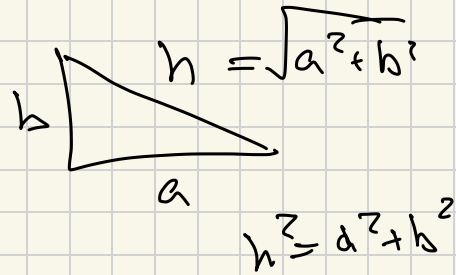
$$\hookrightarrow F_k^* = \{u_1, u_2, \dots, u_k\}$$

$$\Pi_{F_k^*}(A) = A_k$$

$$\sum_i \| a_i - \pi_{F^*}(a_i) \|^2$$

$$F^* = \{v_1, v_2 \dots v_k\} \subset \text{top } k \text{ } RSV_s$$

$$\Rightarrow \sum_i \left\| \sum_{j=1}^d v_j \langle v_j, a_i \rangle - \sum_{j=1}^k v_j \langle v_j, a_i \rangle \right\|^2$$



$$h = \sqrt{a^2 + b^2}$$

$$h^2 = a^2 + b^2$$

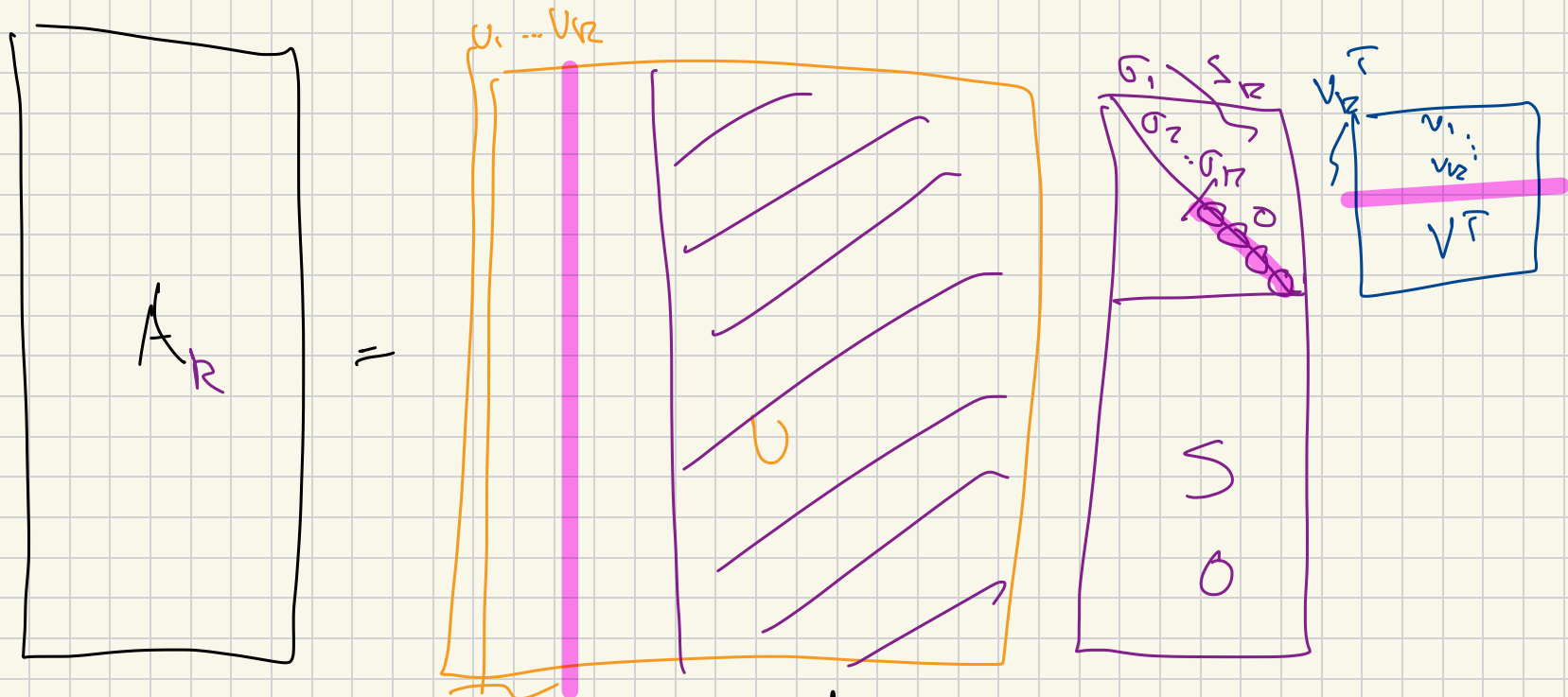
$$= \sum_i \left\| \sum_{j=k+1}^d v_j \langle v_j, a_i \rangle \right\|^2$$

Pythagorean

$$= \sum_i \sum_{j=k+1}^d \| v_j \langle v_j, a_i \rangle \|^2$$

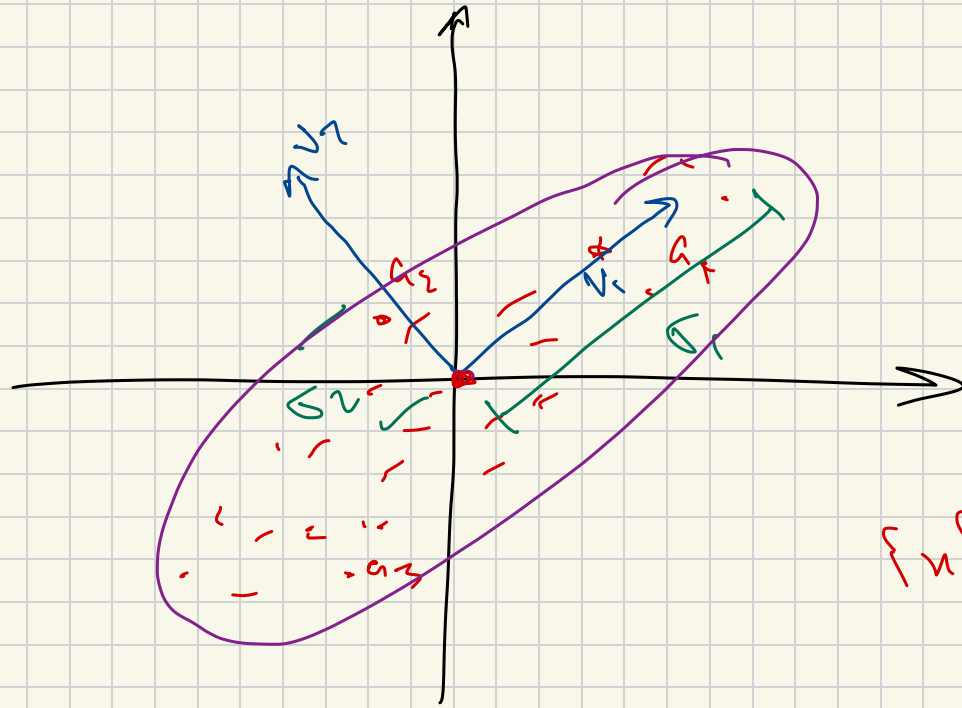
$$= \sum_{j=k+1}^d \left[\sum_i \langle v_j, a_i \rangle^2 \right] = \sum_{j=k+1}^d \| A v_j \|^2 = \sum_{j=k+1}^d \sigma_j^2$$

best rank k approx



$$A_R = U_k S_k V_k^T = \sum_{j=1}^k u_j \sigma_j v_j^T$$

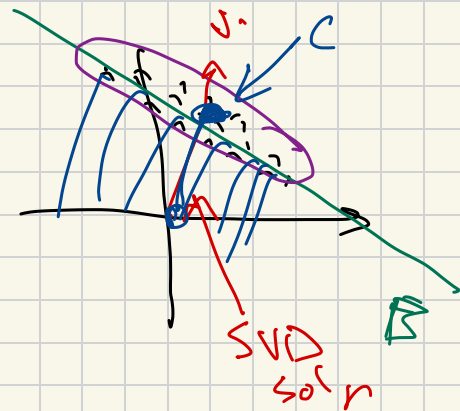
$q_i \in \mathbb{A}$



inflation
requires
A centred
origin

Principal Component Analysis (PCA)

Input $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^d$
 $X \in \mathbb{R}^{n \times d}$



Goal Find k -dim subspace B

$$\text{minimize } \sum_{i=1}^n \|x_i - \pi_B(x_i)\|^2$$

PCA

1. $A = \text{center}(X)$
2. $\text{SVD}(A)$
3. $\text{Project}(A, V_k)$

center

$$a_i = x_i - C \quad \text{for all } x_i \in X$$

$$C = \frac{1}{n} \sum_{i=1}^n x_i = \text{mean}(X) \xrightarrow{A}$$