SVD & Regression

Singular Value Decomposition (SVD)

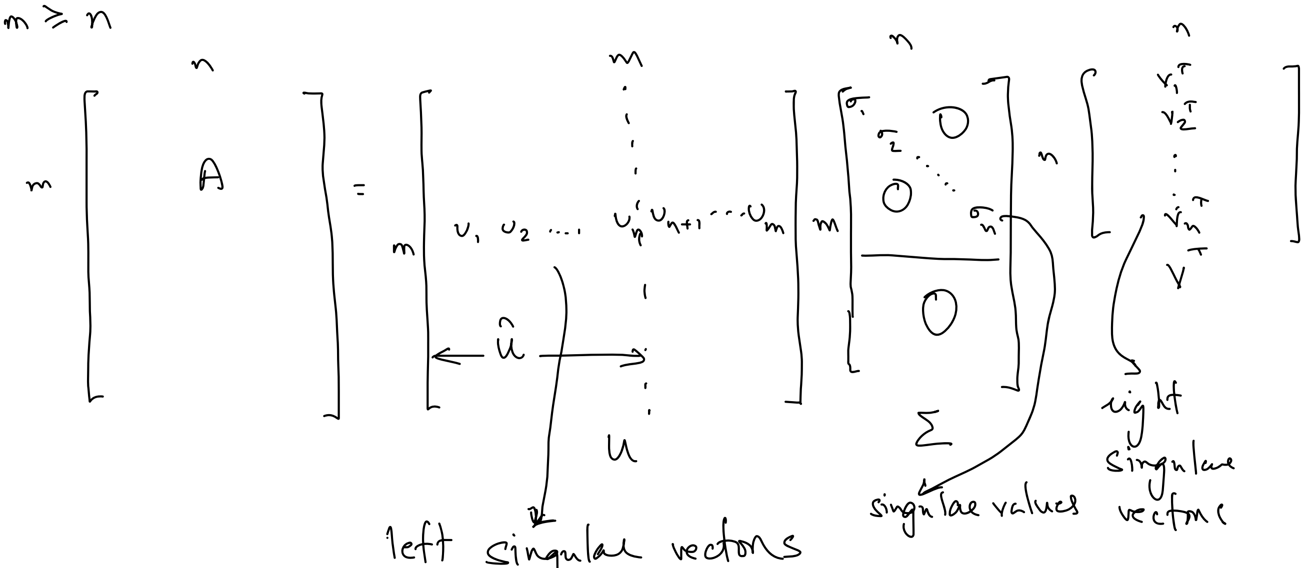
$A \in \mathbb{R}^{m \times n}$

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

$V^TV = I = VV^T$

(Full) SVD : $A = U \Sigma V^T$, where U & V are orthogonal matrices and Σ is ~~real~~ non-negative and diagonal matrix

$m \geq n$



$U^TU = I \Leftrightarrow u_i^Tu_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}, \quad 1 \leq i,j \leq m$

Since U is square and $U^TU = I \Rightarrow UU^T = I$

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

$V^TV = I \Leftrightarrow v_i^Tv_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}, \quad 1 \leq i,j \leq n$

Singular values : $\boxed{\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0}$, $\sigma_i \in \mathbb{R}_+$

$A = U\Sigma V^T \Rightarrow AV = U\Sigma \underbrace{V^TV}_{I} = U\Sigma$

$\Rightarrow AV = U\Sigma$

$Av_i = u_i\sigma_i, \quad 1 \leq i \leq n$

"Thin or Reduced SVD":

$A = \hat{U}\hat{\Sigma}\hat{V}^T, \quad \hat{U} \in \mathbb{R}^{m \times n}, \quad \hat{U}^T\hat{U} = I$

$\hat{\Sigma} \in \mathbb{R}^{n \times n} \rightarrow$ square, diagonal matrix

$A \in \mathbb{R}^{m\times n}$

$A^TA \in \mathbb{R}^{n\times n}$

$(xy)^T = y^Tx^T$

$\hat{V} \in \mathbb{R}^{n\times n}$, $\hat{V}^T\hat{V} = I = \hat{V}\hat{V}^T$

In general , $\hat{U}\hat{U}^T \neq I$

$\boxed{A = \hat{U}\hat{\Sigma}\hat{V}^T} \Rightarrow A^T = (\hat{U}\hat{\Sigma}\hat{V}^T)^T = \hat{V}\hat{\Sigma}^T\hat{U}^T = \boxed{\hat{V}\hat{\Sigma}\hat{U}^T}$

$A^TA = \hat{V}\hat{\Sigma}\underbrace{\hat{U}^T.\hat{U}}_{I}\hat{\Sigma}\hat{V}^T = \underline{\hat{V}\hat{\Sigma}^2\hat{V}^T}$ – Eigenvalue Decomposition of A^TA

$AA^T = \hat{U}\hat{\Sigma}\underbrace{\hat{V}^T\hat{V}}_{I}\hat{\Sigma}\hat{U}^T = \underline{\hat{U}\hat{\Sigma}^2\hat{U}^T}$ – Eigenvalue Decomposition of AA^T

A^TA & AA^T are positive semi-definite matrices (eigenvalues ≥ 0)

$m \geq n$

$A \in \mathbb{R}^{m\times n}$

Full SVD

$A = U\Sigma V^T$

$AV = U\Sigma$

$Av_i = u_i\sigma_i$, $1 \leq i \leq n$

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

$A^TU = V\Sigma^T$

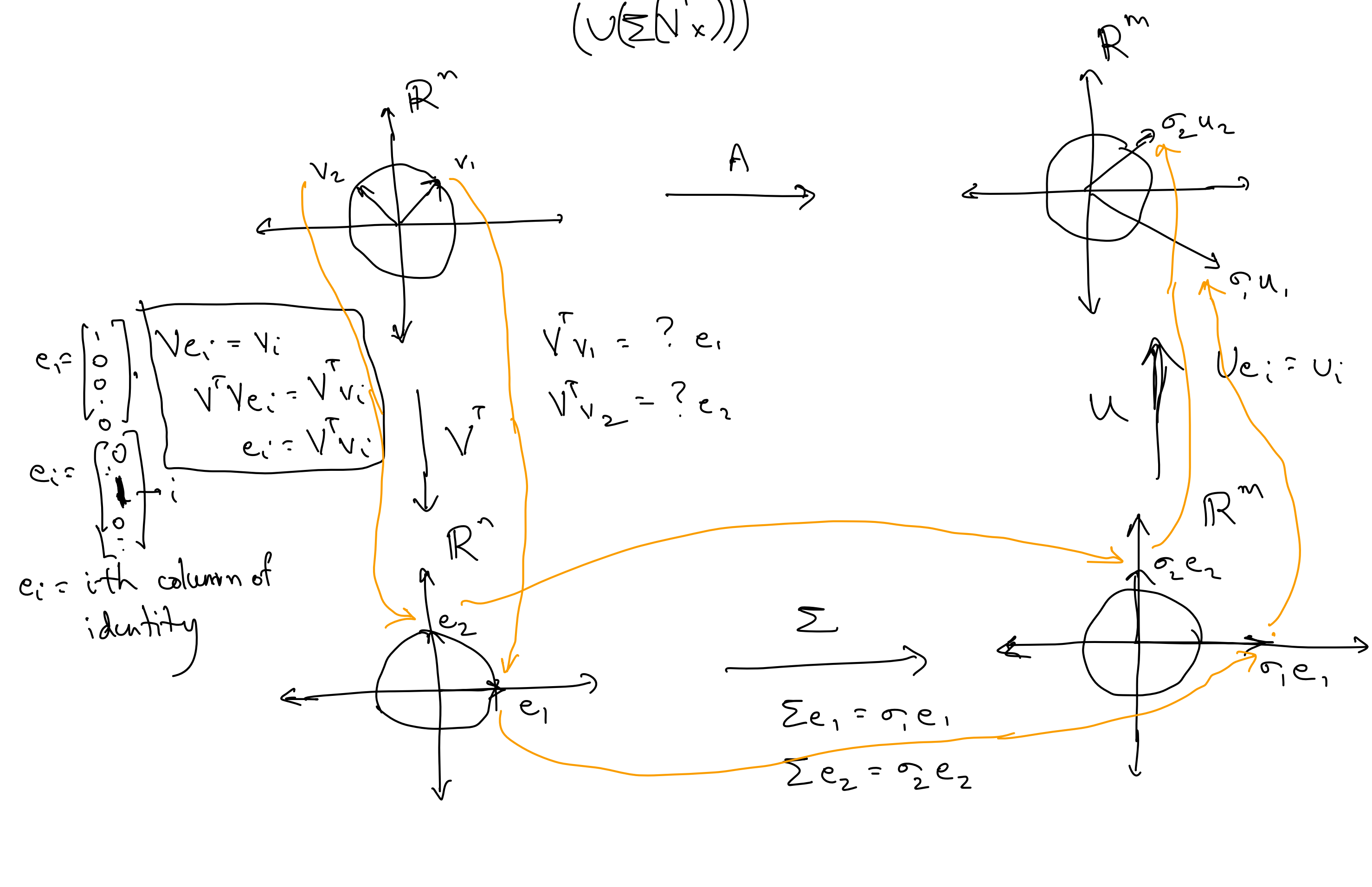
$A^Tu_i = v_i\sigma_i$, $1 \leq i \leq n$

$A^Tu_i = 0$, $n+1 \leq i \leq m$

$A : \mathbb{R}^n \to \mathbb{R}^m$

$x \mapsto Ax$

$(U(\Sigma(V^Tx)))$



If A has rank r $(r \le \min(m,n))$

$\sigma_1 \ge \sigma_2 \ge \dots \ge \sigma_r > 0$, $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_n = 0$

$A : \mathbb{R}^n \to \mathbb{R}^m$

$v_1 \mapsto u_1\sigma_1$
$v_2 \mapsto u_2\sigma_2$
$\vdots$
$v_r \to u_r\sigma_r$
$v_{r+1} \mapsto 0$
$\vdots$
$v_n \mapsto 0$

$A^T : \mathbb{R}^m \to \mathbb{R}^n$

$u_1 \mapsto v_1\sigma_1$
$u_2 \mapsto v_2\sigma_2$
$\vdots$
$u_r \mapsto v_r\sigma_r$
$u_{r+1} \mapsto 0$
$\vdots$
$u_m \mapsto 0$

SVD provides orthogonal basis for the four fundamental subspaces of A:

Column Space $= \mathcal{R}(A) = \langle u_1, u_2, \dots u_r \rangle$
Row Space $= \mathcal{R}(A^T) = \langle v_1, v_2, \dots, v_r \rangle$
Null Space$(A) = \mathcal{N}(A) = \langle v_{r+1}, v_{r+2} \dots v_n \rangle$
Null Space $(A^T) = \mathcal{N}(A^T) = \langle u_{r+1}, u_{r+2}, \dots\dots u_m \rangle$

SVD is the "Rolls Royce"™ as well as the "Swiss Army Knife" of Matrix Decompositions

$A \in \mathbb{R}^{m \times n}$

Truncated SVD, $\rightarrow$ $\boxed{A_k = U_k \Sigma_k V_k^T}$, $U_k \in \mathbb{R}^{m \times k}$, $U_k^T U_k = I$,

for any k, $1 \le k \le \min(m,n)$

$\Sigma_k \in \mathbb{R}^{k \times k}$

$V_k \in \mathbb{R}^{k \times n}$, $V_k^T V_k = I$

Fix k. rank of A_k is k

Among all rank-k approximations of A,

A_k is the "best"

$$A_k = \underset{B \text{ is of rank } k}{\arg\min} \|A - B\|_2 , \quad A_k = \underset{B \text{ is of rank } k}{\arg\min} \|A - B\|_F^2$$

True for all k. $\left(A_k = U_k \Sigma_k V_k^T\right)$

Regression $(x_i, y_i),\ 1 \le i \le N,\ x \in \mathbb{R}^d,\ y_i \in \mathbb{R}$

$$X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_N \end{bmatrix}, \quad X \in \mathbb{R}^{(d+1)\times N}$$

$$\min_w \|y - X^T w\|_2^2$$

Least Squares Solution : $XX^T w^* = Xy$

$$\boxed{w^* = (XX^T)^{-1} X y}$$

(x_i, y_i)

$$\boxed{\hat{y} = X^T w^* = X^T (XX^T)^{-1} X y} \quad \text{- Prediction on training data set}$$

Let $X^T = U \Sigma V^T$ be the reduced SVD of X^T

$$XX^T = V \Sigma \underbrace{U^T \cdot U}_{I} \Sigma V^T = V \Sigma^2 V^T$$

$$(XX^T)^{-1} = (V\Sigma^2 V^T)^{-1} = (V^T)^{-1} (\Sigma^2)^{-1} V^{-1} = V\Sigma^{-2} V^T \quad (V^{-1} = V^T)$$

$$X^T (XX^T)^{-1} X = U\Sigma \underbrace{V^T (V}_{I} \Sigma^{-2} \underbrace{V^T) V}_{I} \Sigma U^T$$

$$= U \underbrace{\Sigma \Sigma^{-2} \Sigma}_{I} U^T = UU^T$$

$$U = [u_1\ u_2 \ldots u_{d+1}]$$

$$U^T U = I, \quad UU^T \ne I$$

$$UU^T = [u_1\ u_2 \ldots u_{d+1}] \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_{d+1}^T \end{bmatrix} = u_1 u_1^T + u_2 u_2^T + \ldots + u_{d+1} u_{d+1}^T$$

UU^T — orthogonal projection onto the range space of X^T

$$UU^T y = (u_1 u_1^T + u_2 u_2^T + \ldots + u_{d+1} u_{d+1}^T)\, y$$

$$= u_1 (u_1^T y) + u_2 (u_2^T y) + \ldots + u_{d+1} (u_{d+1}^T y)$$