Lecture 12 September 1, 2015

Singular Value Decomposition:

Theorem: For any matrix $A \in \mathbb{R}^{(m \times n)}$ there exist orthogonal matrices $U \in \mathbb{R}^{(m \times m)}$ and $V \in \mathbb{R}^{(n \times n)}$, and $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_p \geq 0$ $[\sigma_i \in \mathbb{R}]$
$p = \min(m, n)$ s.t.

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

where $\Sigma \in \mathbb{R}^{m \times n}$, $\Sigma = \operatorname{diag}(\sigma_1, \ldots \sigma_p)$.

$m \leq n$

$$\underbrace{\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}}_{\text{Fat matrix}} = \underbrace{\begin{bmatrix} U \end{bmatrix}}_{(m \times m)} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_m & \end{bmatrix}}_{(m \times n)} \underbrace{\begin{bmatrix} & & \\ & V^T & \\ & & \end{bmatrix}}_{(n \times n)}$$

$m \geq n$

$$\underbrace{\begin{bmatrix} \\ A \\ \\ \end{bmatrix}}_{\text{(Tall matrix)}} = \underbrace{\begin{bmatrix} & & \\ & U & \\ & & \end{bmatrix}}_{(m \times m)} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix}}_{(m \times n)} \underbrace{\begin{bmatrix} V^T \end{bmatrix}}_{(n \times n)}$$

Note: SVD exist for <u>any</u> matrix

$\sigma_1, \sigma_2 \dots \sigma_p$ are called <u>singular values</u> of A

$U = [\vec{u}_1, \vec{u}_2 \dots \vec{u}_m] \Rightarrow$ Left Singular vectors

$V = [\vec{v}_1, \vec{v}_2, \dots \vec{v}_n] \Rightarrow$ Right " "

★ **Lemma 1**: If A is any arbitrary matrix $A \in \mathbb{R}^{(m \times n)}$ then AA^T and A^TA ~~are~~ Symmetric and PSD.

Proof: Symmetry:

$$(AA^T)^T = (A^T)^T A^T = \underline{\underline{AA^T}}$$

Similarly: $$(A^TA)^T = A^T (A^T)^T = \underline{\underline{A^TA}}$$

PSD: Let $x \in \mathbb{R}^n$ be any non-zero vector then

$$x^T (AA^T) x = (A^Tx)^T (A^Tx).$$
$$= \|A^Tx\|^2$$
$$\geq 0$$

Similarly you can prove for A^TA

★ **Lemma 2**: Let $A \in \mathbb{R}^{m \times n}$ then A^TA and AA^T ~~are~~ have the same ~~positive~~ non-zero eigenvalues

Proof: Let λ be a positive eigenvalue of AA^T and $\vec{v}$ be the corresponding eigenvector.

$$\Rightarrow AA^Tv = \lambda v.$$
$$\Rightarrow A^T(AA^Tv) = \lambda A^Tv.$$
$$\Rightarrow A^TA \underbrace{(A^Tv)}_{w} = \lambda (A^Tv) \Rightarrow A^TA\, w = \lambda w$$

Let us assume that $A = U \Sigma V^T$ — ①

From Lemma 1 AA^T is Symmetric & PSD

$A = (m \times n) \Rightarrow AA^T = (m \times m)$

we can write

$$AA^T = E \Lambda E^T \text{ — ② (Spectral Theorem)}$$

$E = [\vec{e}_1, \ldots \vec{e}_m]$ eigenvectors of AA^T $(m \times m)$

$\Lambda = \operatorname{diag}(\lambda_1, \lambda_2 \ldots \lambda_m)$ eigenvalues of AA^T

From ① & ②

$$AA^T = (U \Sigma V^T)(U \Sigma V^T)^T \qquad \{V^T V = I\}.$$

$$= U \Sigma \underline{\underline{V^T V}} \Sigma^T U^T$$

$$= U \Sigma \Sigma^T U^T$$

$$\Rightarrow \boxed{U \Sigma \Sigma^T U^T = E \Lambda E^T}$$

$\Rightarrow$ U = eigenvectors of AA^T

$\Sigma \Sigma^T = \Lambda$

$\because$ $\Sigma \Sigma^T$ is diagonal $(m \times m)$ matrix of σ_i^2

$\Rightarrow$ $\sigma_i = \sqrt{\lambda_i}$ (positive square root)

$\because$ AA^T is PSD $\Rightarrow$ $\lambda_i \geq 0$ $\forall i$

Similarly you can observe $A^T A$ and obtain that:

V = eigenvectors of $A^T A$

Observations :-

* We can decompose any matrix A into product of three matrices $U \Sigma V^T$ where

$U = [\vec{u}_1, \ldots \vec{u}_m]$

$\vec{u}_i$ = left singular vector = eigenvector of AA^T

$V = [\vec{v}_1, \ldots \vec{v}_n]$

$\vec{v}_i$ = right singular vector = eigenvectors of $A^T A$

*
$$A = U \Sigma V^T$$
$$\Rightarrow \quad AV = U\Sigma$$
$$\Rightarrow \quad \boxed{A\vec{v}_i = \sigma_i \vec{u}_i}$$

* Singular value decomposition of a matrix is not uniques. You are free to choose the eigenvectors of $\underline{AA^T}$ & $\underline{A^T A}$.

*

* Equivalence of SVD & EVD in case of Symmetric matrices

Let A be a symmetric square matrix

$\Rightarrow \quad A = E \Lambda E^T$ (Spectral theorem)

From SVD we know

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

Let us consider AA^T

$$AA^T = (E\Lambda E^T)(E\Lambda E^T)^T$$

$$= E\Lambda \underbrace{E^T E}_{= I} \Lambda^T E^T$$

$$= E \Lambda \Lambda^T E^T$$

$\Rightarrow$ Eigenvectors of AA^T are equal to eigenvectors of A whereas eigenvalues are squared $\Rightarrow U = E$

Similarly we can show that $V = E$

and $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2 \ldots \sigma_p)$

$\sigma_i = \lambda_i \quad ; \quad \lambda_i^2 =$ eigenvalue of AA^T

$\Rightarrow \quad \Sigma = \Lambda = \operatorname{diag}(\lambda_1, \lambda_2 \ldots \lambda_n)$.

* SVD and rank of a matrix.

<u>Claim</u>: $N(A) = N(A^T A)$

Let $\vec{x} \in N(A) \Rightarrow A\vec{x} = 0 \Rightarrow A^T(A\vec{x}) = 0$
$\Rightarrow \vec{x} \in N(A^T A)$

Let $\vec{x} \in N(A^T A) \Rightarrow A^T A\vec{x} = 0 \Rightarrow \vec{x}^T (A^T A)\vec{x} = 0$

$\Rightarrow \|A\vec{x}\|^2 = 0 \Rightarrow A\vec{x} = 0$

$\Rightarrow \vec{x} \in N(A)$.

$\Rightarrow N(A) = N(A^T A)$.

Now from rank & nullity theorem

$\text{rank}(A) = m - \dim(N(A))$

$= m - \dim(N(A^T A))$ $\quad \{A^T A = (m \times m)\}$

$= \text{rank}(A^T A)$

$= r$ (Number of non-zero eigenvalues)

= no. of non-zero singular values.

$\Rightarrow$ rank (A) = no. of non-zero singular values of A

Observation :-

For any $\vec{x} \in \mathbb{R}^m$ $\quad A\vec{x} = U\Sigma V^T\vec{x}$

$$U\Sigma V^T\vec{x} = U\Sigma \begin{bmatrix} \vec{v}_1^T\vec{x} \\ \vec{v}_2^T\vec{x} \\ \vdots \\ \vec{v}_m^T\vec{x} \end{bmatrix} = U \begin{bmatrix} \sigma_1 \vec{v}_1^T\vec{x} \\ \vdots \\ \sigma_r \vec{v}_r^T\vec{x} \\ 0 \end{bmatrix}$$

$$= \sum_{i=1}^{r} \left(\sigma_i \vec{v}_i^T \vec{x}\right) \cdot \vec{u}_i$$

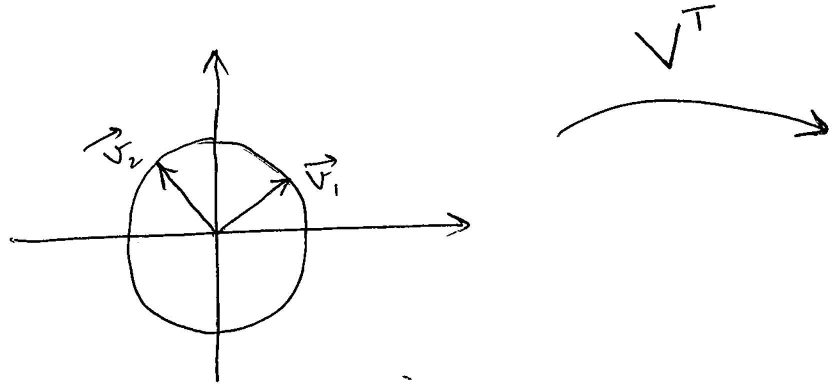
$$= \left[\sum_{i=1}^{r} \left(\sigma_i \vec{u}_i \vec{v}_i^T\right)\right] \vec{x}$$

$$\Rightarrow \boxed{A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T}$$

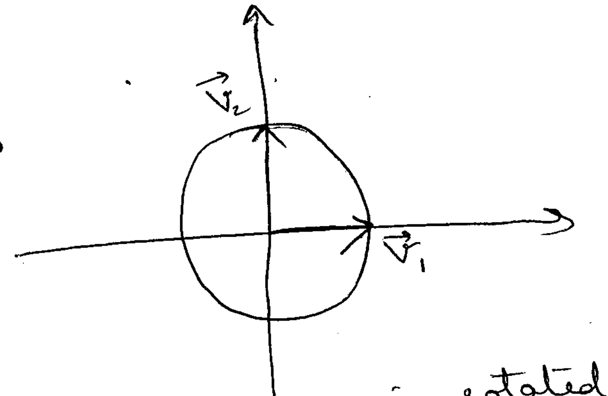
* Geometric Interpretation of SVD.

Let A be a (2×2) matrix $A = U\Sigma V^T$

$U = [\vec{u}_1 \;\; \vec{u}_2]$; $V = [\vec{v}_1 \;,\; \vec{v}_2]$; $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2)$

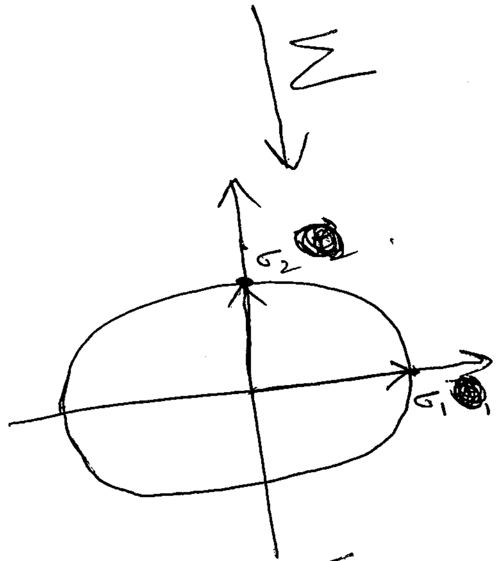


Unit circle with right singular vector.

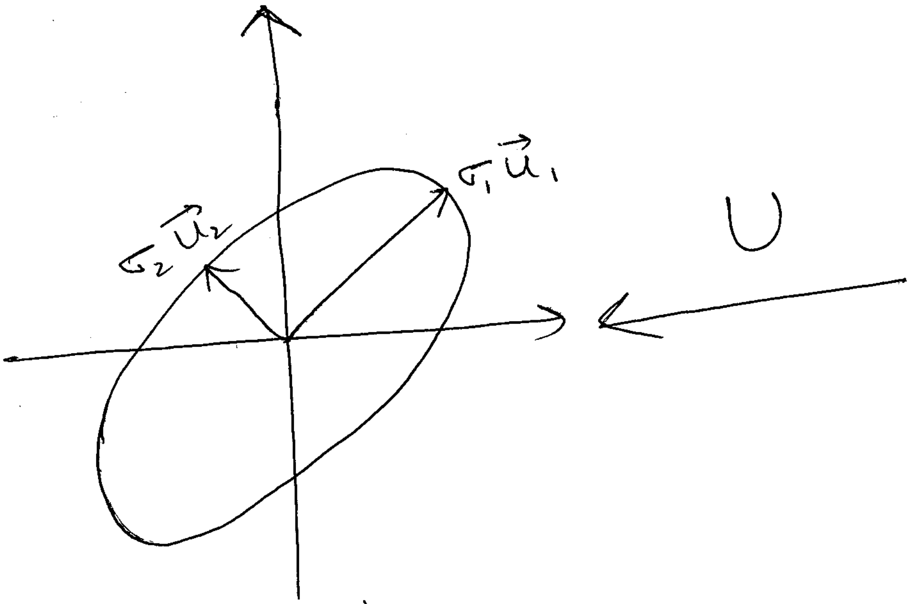


Unit circle is rotated by V^T.

$$\left(V^T \vec{v}_1 = \begin{bmatrix} \vec{v}_1^T \vec{v}_1 \\ \vec{v}_2 \vec{v}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)$$



Scaled by Σ result into ellipse



The ellipse is rotated by U

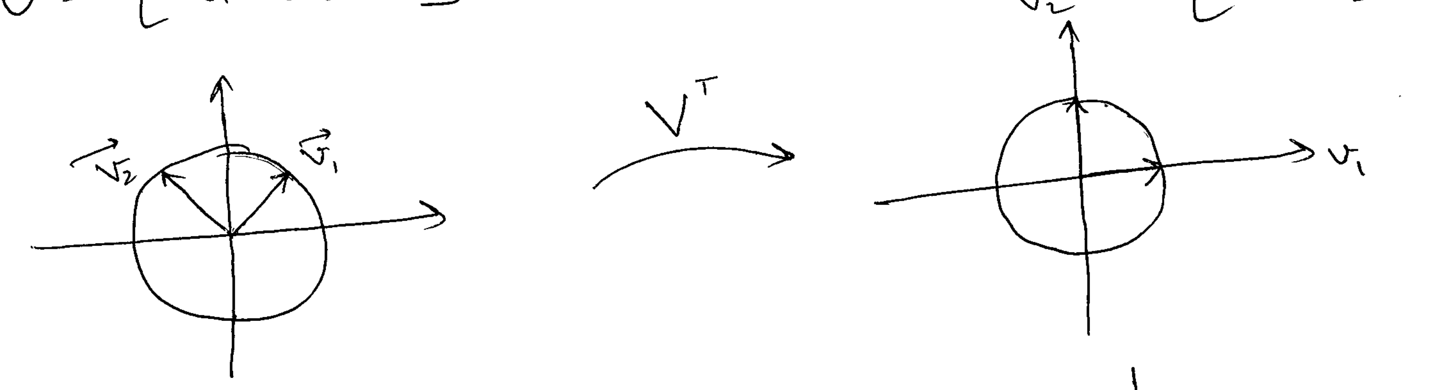
$$\Rightarrow \quad A = \underbrace{U}_{\text{Rotation}} \; \underbrace{\Sigma}_{\text{Scaling}} \; \underbrace{V^T}_{\text{Rotation}}$$

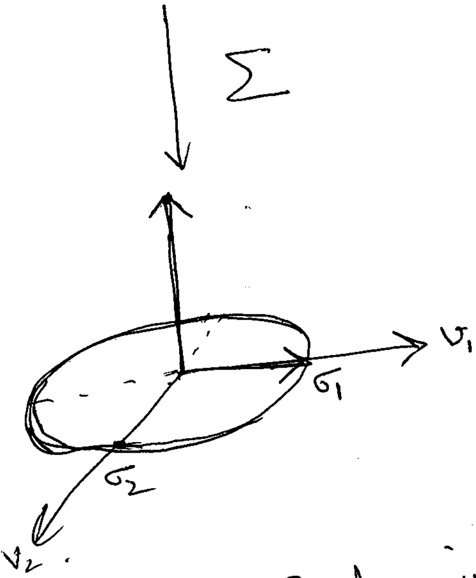
When matrix is rectangular then Σ does two things : scaling and changing dimension

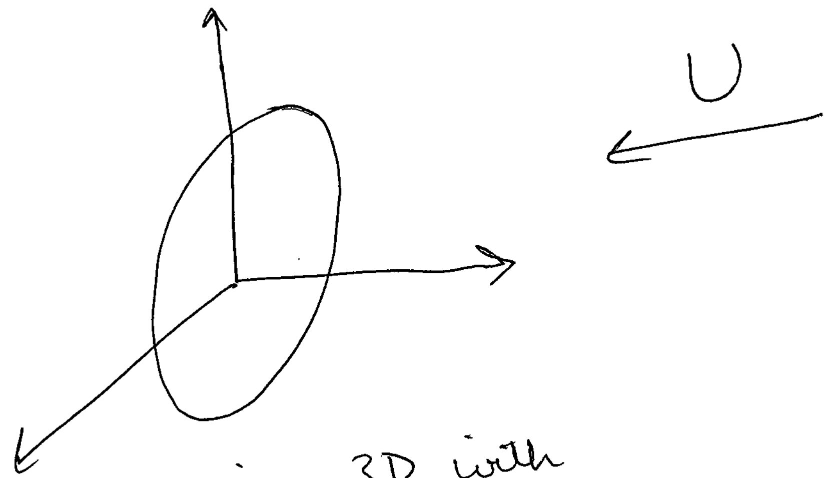
Example: If A is $(3\times 2) \Rightarrow A: \mathbb{R}^2 \to \mathbb{R}^3$

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

$$U = [\vec{u}_1 \ \vec{u}_2 \ \vec{u}_3] \ ; \quad V = [\vec{v}_1 \ , \vec{v}_2] \ ; \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix}$$







Ellipsoid in 3D with axes $\sigma_1 \vec{u}_1$ & $\sigma_2 \vec{u}_2$
note that it is a flat ellipse in 3D space.

ellipse in 3-dimensional space. It still lies in a 2D plane.

* Pseudoinverse

Recall least squares solution: $\boxed{Y = A\theta + e}$

$$\hat{\theta} = \underbrace{(A^TA)^{-1}A^T}\, Y = A^{+}Y$$

The quantity $(A^TA)^{-1}A^T$ is called the pseudoinverse of A

$$A^{+} = (A^TA)^{-1}A^T$$

Let us assume that $A = U\Sigma V^T$

$$\Rightarrow \quad A^{+} = \left[(U\Sigma V^T)^T(U\Sigma V^T)\right]^{-1}\left[U\Sigma V^T\right]^T$$

$$= \left[V\Sigma^T \underline{\underline{U^TU}}\, \Sigma V^T\right]^{-1}\left[V\Sigma^T U^T\right]$$

$$= \left[V\Sigma^T\Sigma V^T\right]^{-1} V\Sigma^T U^T$$

$\Sigma^T\Sigma = (m\times m)$ diagonal matrix of singular values.

and V is orthogonal $\Rightarrow V^T = V^{-1}$

$$\Sigma^T\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2 \ldots \sigma_m^2) = \Sigma^2$$

$$A^{+} = (V\Sigma^2V^T)^{-1}\, V\Sigma^T U^T$$

$$= (V^T)^{-1}(\Sigma^2)^{-1}\underline{V^{-1}V}\,\Sigma^T U^T$$

$$= V \underset{(m\times m)}{\begin{bmatrix} 1/\sigma_1^2 & & & \\ & 1/\sigma_2^2 & & \\ & & \ddots & \\ & & & 1/\sigma_m^2 \end{bmatrix}} \underset{\Sigma^T\,(m\times m)}{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & 0 \\ & & \ddots & \\ & & & \sigma_m \end{bmatrix}} U^T$$

$$= V \underbrace{\begin{bmatrix} 1/\sigma_1 & & & \\ \vdots & 1/\sigma_2 & & 0 \\ & & \ddots & \\ \vdots & & & 1/\sigma_m \end{bmatrix}}_{\Sigma^{+} = (m\times m)} U^T$$