SVD

Orthogonality: Let $U = (u_1 \dots u_k) \in \mathbb{R}^{m \times k}$

$u_1 \dots u_m$ are orthogonal $\quad u_i^T u_j = 0$ for $i \neq j \iff U^T U$ is diagonal

$u_1 \dots u_k$ are orthonormal $\quad u_i^T u_j = 0 \quad i \neq j \iff U^T U = I$
$\quad u_i^T u_i = 1$

Historically:

square

Defn: $U = (u_1 \dots u_m) \in \mathbb{R}^{m \times m}$ is orthogonal if $u_1 \dots u_m$ are orthonormal

Notation: $U \in \mathcal{O}_m$

For this class:

Defn: $U = (u_1 \dots u_m) \in \mathbb{R}^{m \times k}$ is an "orthonormal matrix" if $u_1 \dots u_m$ are orthnorm.

$U \in \mathcal{V}_{k,m}$ (the stiefel manifold)

$\Rightarrow$ a square orthonormal matrix is an orthogonal matrix.

$\Rightarrow \mathcal{V}_{m,m} = \mathcal{O}_m$

Thm: If $U_1 \in \mathbb{R}^{m \times r}$ is orthonormal, there exists $U_2 \in \mathbb{R}^{m \times (m-r)}$ orthonormal

such that $U = (U_1\ U_2)$ is orthogonal, ie

$$U^T U = \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} U_1^T U_1 & U_1^T U_2 \\ U_2^T U_1 & U_2^T U_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

Interpretation

$U_1 \in \mathcal{V}_{r,m}$ defines an orthonormal basis for an r-dim subspace of $\mathbb{R}^m$

$$= \text{Col}(U_1) = \{x \in \mathbb{R}^m : x = a_1 \underset{\sim}{u}_1 + \dots + a_r \underset{\sim}{u}_r\}$$
$$= \{x \in \mathbb{R}^m : x = \underset{\sim}{U}\, \underset{\sim}{a}\}$$

U_2 def orth basis for null space of $\text{Col}(U_1)$

$tr(AB) = tr(BA) = \ldots$

$= tr(A^T B \ldots$

Norms:

(Frobenius) Norm: for $X \in \mathbb{R}^{m \times n}$, $\|X\|_F = \|X\| = \left(\sum_i \sum_j x_{ij}^2\right)^{1/2}$

$$\|X\|^2 = tr(X^T X) = tr(X X^T)$$

if $n=1$, $x \in \mathbb{R}^{m \times 1}$ is a vector,

$$\|x\|_F^2 = \sum x_i^2 = x^T x \;:\; \left(tr(xx^T) = tr(x^T x) = x^T x \right)$$

Vector space Norm: for $X \in \mathbb{R}^{m \times n}$, $\|X\|_v = \max_{u \in \mathbb{R}^n : u^T u = 1} \|Xu\|_F = \max_{u \in \mathbb{R}^n : u^T u = 1} (u^T X^T X u)^{1/2}$

$$= \max_{z \in \mathbb{R}^n} \frac{\|Xz\|_F}{\|z\|_F}$$

for $x^T \in \mathbb{R}^{1 \times n}$, $\max_{z \in \mathbb{R}^n} \dfrac{\|x^T z\|_F}{\|z\|_F} = \max_{z \in \mathbb{R}^n} \left[\dfrac{(x^T z)^2}{z^T z}\right]^{1/2} = (x^T x)^{1/2}$

$x \in Col(U_1) \Rightarrow x = U_1 a$

Ex: Let $y \in Col(U_2) \Rightarrow y = U_2 b$

then $x^T y = a^T U_1^T U_2 b = a^T \underset{r\times(m-r)}{0}\, b = 0$

$U = [U_1 \ U_2]$ is an called orthonormal basis for $\mathbb{R}^m$.

Norms & Rotations

Def: The (F) norm of a vector $x \in \mathbb{R}^n$ is $\|x\| = (x^T x)^{1/2} = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$

Def: The (F)-norm of a matrix $X \in \mathbb{R}^{n\times n}$ is $\|X\| = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}^2\right)^{1/2} = tr(X^T X)^{1/2} = tr(XX^T)^{1/2}$

(for diagonal A, $tr(A) = \sum_{i=1}^{m} a_{ii}$)

The (F)-norm of a matrix is sometimes called the "Frobenius" norm.

Rotations: A rotation (or rot + ref) of a vector $x \in \mathbb{R}^m$ is a multiplication of x by an orthogonal matrix

Ux is a rotation of x if $U \in \mathcal{O}(m)$

Rotations for matrices

Let $A \in \mathbb{R}^{m\times n}$, $U \in \mathcal{O}^m$, $V \in \mathcal{O}^n$

$UA = U(a_1 \ a_2 \cdots a_n) = (Ua_1 \cdots Ua_n)$

(rotation on the left rotates columns of A)

$AV^T = \begin{pmatrix} a_1 - \\ a_m - \end{pmatrix} V^T = \begin{pmatrix} a_1^T V^T \\ a_m^T V^T \end{pmatrix} = \begin{pmatrix} (Va_1)^T \\ (Va_m)^T \end{pmatrix}$

(rotation on right rotates the rows of A)

$\tilde{Y} = U \Upsilon V^T$

$$\|\tilde{Y}\|_V^2 = \sup_{w:\, w^Tw = I} \|U\Upsilon V^T w\|_F^2 = \sup_{w:\, w^Tw=I} \operatorname{tr}\left(U\Upsilon V^T w w^T V \Upsilon^T U^T\right)$$

$$= \sup_{\cdot\cdot} \operatorname{tr}\left(\Upsilon^T\Upsilon\, V^T w w^T V\right)$$

$$= \sup_{\cdot\cdot} \operatorname{tr}\left(\Upsilon^T\Upsilon\, \tilde{w}\tilde{w}^T\right)$$

$$= \sup \operatorname{tr}\left(\Upsilon^T\Upsilon\, w w^T\right)$$

$$= \sup \operatorname{tr}\left((\Upsilon w)^T(\Upsilon w)\right)$$

$$= \sup_{w:\, w^Tw=1} \|\Upsilon w\|_F^2 = \|\Upsilon\|_V^2$$

$$\left\|\Upsilon_{-1}\begin{pmatrix} d \\ w\end{pmatrix}\right\| \geq d_1^2 + w^Tw = \left(d_1^2 + w^Tw\right)^{1/2} \left\|\begin{pmatrix} d_1 \\ w\end{pmatrix}\right\|$$

$$\frac{\left\|\Upsilon_{-1}\begin{pmatrix} d \\ w\end{pmatrix}\right\|}{\left\|\begin{matrix} d_1 \\ w\end{matrix}\right\|} \geq \left(d_1^2 + w^Tw\right)^{1/2} \geq d_1$$

$\Upsilon_{-1} = \tilde{U}^T \Upsilon V$

as norm is invariant to rot.

$$\to \sup_{v^Tv=I} \|\Upsilon_{-1} v\| \geq \left(d_1^2 + w^Tw\right)^{1/2} \geq d_1 = \sup_{v^Tv=1} \|\Upsilon v\| = \sup_{v^Tv=1} \|\Upsilon_{-1} v\|$$

$\to$ $\underline{w = 0}$

Norm preserved under Rotation

Norms under rotations

Let $\tilde{x} = Ux \in \mathbb{R}^m$. Then $\|\tilde{x}\|^2 = \tilde{x}^T\tilde{x} = x^T U^T U x = x^T x = \|x\|^2$

Let $\tilde{Y} = UYV^T$. Then
$$\begin{aligned}\|\tilde{Y}\|^2 &= tr(\tilde{Y}\tilde{Y}^T) \\ &= tr(UYV^TVY^TU^T) \\ &= tr(UYY^TU^T) \\ &= tr(U^TUYY^T) = tr(YY^T) = \|Y\|^2\end{aligned}$$

Singular Value Decomposition

Thm: for $Y \in \mathbb{R}^{m\times n}$, $m \geq n$, $\exists\, U \in O(m)$, $V \in O(n)$, $D \in \mathcal{D}(m)$ s.t.

$$Y = UDV^T$$

Pf (Golub + van Loan, p. 70)

Let $v_1 = \underset{v\in\mathbb{R}^n : v^Tv = 1}{\arg\max} \|Yv\|$, let $d_1 = \|Yv\|$

Then $Yv_1 = d_1 u_1$ for some $u_1 \in \mathbb{R}^m : u_1^T u_1 = 1$

→ let $U = (u_1 \;\; U_2) \in O(m)$
$V = (v_1 \;\; V_2) \in O(n)$

then
$$U^TYV = \begin{pmatrix} u_1^TYv_1 & \overbrace{u_1^TYV_2}^{w^T} \\ U_2^TYv_1 & \underbrace{U_2^TYV_2}_{B} \end{pmatrix} \hat{=} Y_1$$

now $u_1^TYv_1 = u_1^T(d_1u_1) = d_1$
$U_2^TYv_1 = U_2^T(d_1u_1) = 0$

Claim: $\overset{w=}{u_1^TYV_2} = 0$: Pf: $\left\|Y_1\begin{pmatrix} d_1 \\ w\end{pmatrix}\right\| = \left\|\begin{pmatrix} d_1 & w^T \\ 0 & B\end{pmatrix}\begin{pmatrix} d_1 \\ w\end{pmatrix}\right\|$

norm wrt 1?

$$= \left\|\begin{pmatrix} d_1^2 + w^Tw \\ Bw\end{pmatrix}\right\| \geq d_1^2 + w^Tw$$

$$\left\| Y_1 \begin{pmatrix} d \\ w \end{pmatrix} \right\|^2 = d_1^2 + w^T w \geq d_1^2 = \sup_{v:\, v^Tv=1} \overbrace{\|Yv\|^2}^{\|Y\|_v^2} = \sup_{v:\, v^Tv=1} \overbrace{\|Y_{-1}\|^2}^{\|Y_{-1}\|_v^2}$$

$\to w = \underset{\sim}{0}$

$$\to \quad U^T Y V = \begin{pmatrix} d_1 & 0 \\ 0 & Y_{-1}' \end{pmatrix} \to Y = U \begin{bmatrix} d_1 & 0 \\ 0 & Y_{-1}' \end{bmatrix} V^T$$

now repeat argument on Y_{-1}' $\to$ $Y = U \begin{bmatrix} d_1 & 0 \\ 0 & \hat{U}\begin{pmatrix} d_2 & 0 \\ 0 & Y_{-2} \end{pmatrix} \hat{V}^T \end{bmatrix} V^T$

$$= \underbrace{U \begin{bmatrix} 1 & \underset{\sim}{0} \\ \underset{\sim}{0} & \hat{U} \end{bmatrix}}_{\in O(m)} \begin{bmatrix} d_1 & \underset{\sim}{0} \\ 0 & \begin{pmatrix} d_2 & \underset{\sim}{0} \\ \underset{\sim}{0} & Y_{-2}' \end{pmatrix} \end{bmatrix} \Bigg[\underbrace{V \begin{bmatrix} 1 & 0 \\ 0 & \tilde{V} \end{bmatrix}}_{\in O(n)} \Bigg]^T$$

<u>Uniqueness</u>: The representation is <u>not unique</u> <u>Why?</u>

$$Y = UDV^T = UP^TPDP^TPV^T = \tilde{U}\tilde{D}\tilde{V}^T, \text{ where } P = \text{permutation matrix}$$

$$= USSDSSV^T = \hat{U}\tilde{D}\hat{V}^T, \text{ where } S = \text{diag}(\pm 1)$$

for $Y \in \mathbb{R}^{m \times n}$, $m \geq n$: The (thin) SVD of Y is the triple (U, D, V) where

$Y = UDV^T$

$U \in \mathcal{V}_{n,m}$

$V \in \mathcal{O}_n$

$D = \text{diag}(d_1, \dots, d_n)$, $d_1 \geq d_2 \geq \dots \geq d_n \geq 0$

$U = (u_1 \dots u_n)$ are the <u>left</u> <u>singular vectors</u>

$V = (v_1 \dots v_n)$ are the right singular vectors

$d_1 \dots d_n$ are the singular values.