Tensor SVD

Matrix SVD + rank

$Y \in \mathbb{R}^{m_1 \times m_2}$ $\quad m_1 \geq m_2 \quad . \quad Y = UDV^T \quad D = \operatorname{diag}(d_1 \ldots d_{m_2})$ (thin SVD)

where $d_1 \geq d_2 \geq \cdots \geq d_R \geq 0, \; d_s = 0, \; s > R$

Matrix rank (notion 1)

"Def": M is a rank-1 $m_1 \times m_2$ matrix if $M = \underset{\sim}{a}\underset{\sim}{b}^T = \underset{\sim}{a} \circ \underset{\sim}{b}$, some $a \in \mathbb{R}^{m_1}, b \in \mathbb{R}^{m_2}$
$(M_{ij} = a_i b_j)$ $\quad ab^T \neq 0$.

"Def": the rank of a matrix Y is the smallest # R for which

$$Y = \sum_{r=1}^{R} a_r b_r^T \quad \text{for } \underset{\sim}{a}_r \in \mathbb{R}^{m_1}, \underset{\sim}{b}_r \in \mathbb{R}^{m_2}$$

SVD: the SVD shows that $\quad Y = \sum_{r=1}^{R} d_r u_r v_r^T$

$\Rightarrow \quad \operatorname{rank}(Y) = $ # of non-zero sing. values

Extension to arrays

Def: M is a rank-1 $m_1 \times m_2 \times m_3$ array if $M = a \circ b \circ c$
$(m_{ijk} = a_i b_j c_k)$

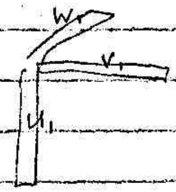
for some $\underset{\sim}{a} \in \mathbb{R}^{m_1}$
$b \in \mathbb{R}^{m_2}$
$c \in \mathbb{R}^{m_3}$, $\quad a \circ b \circ c \neq 0$.

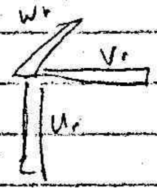
Define: the rank of an $m_1 \times m_2 \times m_3$-array is the smallest # R for which

$$Y = \sum_{r=1}^{R} a_r \circ b_r \circ c_r \quad \left(y_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} \right)$$

One notion of an "array SVD" is a representation of Y in terms of this rank

$$Y = \sum_{r=1}^{R} d_r \, \underset{\sim}{u}_r \circ \underset{\sim}{v}_r \circ \underset{\sim}{w}_r \qquad \left(u_r^T u_r = v_r^T v_r = w_r^T w_r = I \right)$$

$= d_1$ 

$+ \cdots + d_r$ 

This is representation is called the "CP" decomposition

(PARAFAC Harshman (1970)
CANDECOMP Carroll & Chang (1970))

Matrix Rank (Notion 2)

$Y \in \mathbb{R}^{m_1 \times m_2}$, $\quad Y = UDV^T$ (thin svd)

U = orthogonal basis for space spanned by $\begin{pmatrix} y_1 & \cdots & y_{m_2} \\ | & & | \end{pmatrix}$
(although describes "row variation")

$U = \begin{pmatrix} u_1 & \cdots & u_{m_2} \\ | & & | \end{pmatrix}$. Let $\underset{\sim}{x} = b_1 y_1 + \cdots + b_{m_2} y_{m_2} \in \text{span}(y_1 \ldots y_{m_2})$

$$\begin{aligned} &= Y \underset{\sim}{b} \\ &= UDV^T \underset{\sim}{b} \\ &= U \tilde{\underset{\sim}{b}} \\ &= \tilde{b}_1 \underset{\sim}{u}_1 + \tilde{b}_2 \underset{\sim}{u}_2 + \cdots + \tilde{b}_m \underset{\sim}{u}_m \in \text{span}(u_1 \ldots u_{m_2}) \end{aligned}$$

similarly, V is an orthogonal basis of $\begin{pmatrix} y_1^T & \text{---} \\ y_2^T & \text{---} \\ \vdots & \\ y_{m_1}^T & \text{---} \end{pmatrix}$

Reduced rank case

Suppose $D = \begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix}$

then $Y\underset{\sim}{b} = UDV^T b$

$= UD\underset{\sim}{\check{b}} = U \begin{pmatrix} d_1 \check{b}_1 \\ d_2 \check{b}_2 \\ \vdots \\ d_r \check{b}_r \\ 0 \\ 0 \\ 0 \end{pmatrix} = U \begin{bmatrix} \tilde{b}_r \\ 0 \end{bmatrix}$

$= \tilde{b}_1 \underset{\sim}{u}_1 + \cdots + \tilde{b}_r \underset{\sim}{u}_r + 0 + \cdots + 0$

$\Rightarrow$ every linear combination of columns of Y can be expressed as a point in a r-dim subspace of $\mathbb{R}^{m_1}$

$\dim(\text{span}(\underset{\sim}{Y}_1 \cdots \underset{\sim}{Y}_{m_2})) = r$

"Def": the "column rank" of a matrix is the dimension of the linear span of its columns.

"Def": the "row rank" " span of its rows.

Results: (1) row rank $(Y) = \dim(\text{span}(\ (\underset{\sim}{Y}_1) \cdots (\underset{\sim}{Y}_{m_1})\ ))$

$= \dim(\text{span}(\ (\underset{\sim}{Y}_1) \cdots (\underset{\sim}{Y}_{m_2})\ )) = \text{col rank}(Y)$

$= \text{rank}(Y)$

(2) $\text{rank}(Y) = \#$ non zero singular values

Extension to Tensors

For a matrix Y, a "column vector" $\begin{bmatrix} y_{1j} \\ \vdots \\ y_{m_1 j} \end{bmatrix}$ are entries of Y with fixed second index, varying first index.

The left singular vectors $U = (u_1 \cdots u_{m_2})$

- all have length $= m_1$
- describe variation along the rows (1st mode)
- provide an orth basis for l.c.s of $\left\{ \begin{bmatrix} y_{1j} \\ \vdots \\ y_{m_1 j} \end{bmatrix}, j=1,\dots,m_2 \right\}$

What is a column vector for a tensor?

Mode fibers

For an array Y, a "mode 1 fiber" is the vector $\begin{bmatrix} y_{1jk} \\ y_{2jk} \\ \vdots \\ y_{m_1 jk} \end{bmatrix}$, entries of Y with fixed second + 3rd index, varying 1st index.

Goal: Describe variation along 1st mode

$\iff$ describe row variation of $\left\{ \begin{bmatrix} y_{1jk} \\ \vdots \\ y_{m_1 jk} \end{bmatrix}, j=1,\dots,m_2, k=1,\dots,m_3 \right\}$

$\iff$ obtained from an orthonormal basis for $\left\{ \begin{bmatrix} y_{1jk} \\ \vdots \\ y_{m_1 jk} \end{bmatrix}, \dots \right\}$

IDEA: $Y_{(1)} = \begin{bmatrix} y_{111} & y_{121} & \cdots \\ y_{211} & y_{221} & \\ \vdots & \vdots & \\ y_{m_1 11} & & \end{bmatrix} = \begin{bmatrix} y_{(1)11} & y_{(1)21} & \cdots & y_{(1)m_2 m_3} \\ | & | & & | \end{bmatrix}$

Let $Y_{(1)} = U_1 D_1 V_1^T$

$\Rightarrow$ U_1 is an orthonormal basis for the column space of $Y_{(1)}$

describes variation along the rows (1st mode) of $Y_{(1)}$

Let R_1 = # nonzero values of D_1

$\Rightarrow$ $R_1 = \dim(\text{span}(\text{col}(Y_{(1)})))$
= dim of variation along the 1st mode

def: R_1 is the "1-rank" of the tensor Y.

Def: (K-rank) Let $Y \in \mathbb{R}^{m_1 \times \cdots \times m_K}$. The K-rank of Y is the vector $(R_1 \ldots R_K)$, where

$R_k = \dim(\text{span}(\text{col}(Y_{(k)})))$ = # nonzero SV's of $Y_{(k)}$.

Tucker Decomposition / HOSVD

Thm: Every K-mode tensor $Y \in \mathbb{R}^{m_1 \times \cdots \times m_K}$ has a representation of the form

$Y = S \times \{U_1, \ldots U_K\}$, where

- $U_k \in \mathcal{O}_{m_k}$ are the k-mode singular vectors
- $S \in \mathbb{R}^{m_1 \times \cdots \times m_K}$ is the "core array" which has the properties of
  - all-orthogonality
  - ordering
  - k-mode singular value recovery

Pf:

1) Let $U_1 \dots U_n$ be left s.v.'s of $Y_{(1)} \dots Y_{(n)}$, $\left[U_k \in O(n_k) \right]$

define $S = Y \times \{U_1^T \dots U_K^T\}$

then $\underline{S \times \{U_1 \dots U_n\}} = Y \times \{U_1^T \dots U_n^T\} \times \{U \dots U_n\}$
$= Y \times \{I \dots I\}$
$= \underline{Y}$

Properties of S

S is "sort of" an analogue to D in the matrix SVD. It has the following properties

"All-orthogonality"

$\langle S_{[i_1,,]}, S_{[i_2,,]} \rangle = 0$ if $i_1 \neq i_2$

subtensor subtensor

$$\left[D = \begin{pmatrix} d_1 & 0 & 0 & \cdot \\ 0 & d_2 & \cdot & - \\ | & & \ddots & \\ & & & d_{m_1} \end{pmatrix} \Rightarrow \begin{array}{l} \langle D_{[i_1,]}, D_{[i_2,]} \rangle = 0 \\ \langle D_{[,j_1]}, D_{[,j_2]} \rangle = 0 \end{array} \right]$$

2) ordering

$\| S_{[1,,]} \| \geq \| S_{[2,,]} \| \geq \dots \geq \| S_{[m_1,,]} \|$

$$\left[\begin{array}{l} d_j = \| D_{[j,]} \| \\ d_1 \geq d_2 \geq \dots \geq d_{m_1} \end{array} \right.$$

Proof: these follow by equating $SVD(Y_{(1)})$ with $(hosvd(Y))_{(1)}$

$Y_{(1)} = U_1 D_1 V_1^T = (S \times \{U_1 \dots U_n\})_{(1)}$
$= U_1 S_{(1)} (U_K^T \otimes \dots \otimes U_2^T)$

$D_1 V_1^T = S_{(1)} (U_n^T \otimes \dots \otimes U_2^T)$

$\dots V^T (U_n \otimes \dots \otimes U_2) = D_1 \hat{V}^T$

if $m_1 \le m_{-1}$

$$\underset{m_1 \times m_1}{D_1}\ \underset{m_1 \times m_{-1}}{\tilde{V}_1^T}$$

$$S_{(1)} = D_1 \tilde{V}_1^T$$

$$= \begin{pmatrix} d_{11} & & \\ & \ddots & \\ & & d_{m_1} \end{pmatrix} \begin{pmatrix} \tilde{v}_1^T — \\ \vdots \\ \tilde{v}_{m_1}^T — \end{pmatrix} = \begin{pmatrix} d_{11} \tilde{v}_1^T — \\ \vdots \\ d_{m_1} \tilde{v}_{m_1}^T — \end{pmatrix}$$

$\Rightarrow$ all orthogonality : $\langle S_{[i_1,\cdot]}, S_{[i_2,\cdot]} \rangle = d_{i_1} d_{i_2} \tilde{v}_{i_1}^T \tilde{v}_{i_2} = 0$

$\Rightarrow$ ordering : $\| S_{[i_1,\cdot]} \| = (d_{i_1}^2 \tilde{v}_{i_1}^T \tilde{v}_{i_1})^{1/2} = d_{i_1}$

also recovers mode 1 sv's

Truncated HOSVD / TSVD

Model: $Y = M + E, \quad Y, M, E \in \mathbb{R}^{m_1 \times \cdots \times m_K}$

$\text{rank}(M) = (r_1 \ldots r_K)$

$E = \{e_{ij}\} \sim \text{iid } N(0, \sigma^2) \Rightarrow \text{rank}(Y) = m_1 \times \cdots \times m_K$

ML/OLS Estimator of M

$$\hat{M} = \underset{M : \text{rank}(M) = \underset{\sim}{r}}{\arg\min} \|Y - M\|^2$$

What does it mean for $\text{rank}(M) = \underset{\sim}{r} < \underset{\sim}{m} = \dim(M)$??

$M = S \times \{U_1 \ldots U_K\}$

$M_{(1)} = U_1 S_{(1)} (U_K^T \otimes \cdots \otimes U_2^T) \qquad \text{rank}_1(M) = r_1 \Rightarrow \text{rank } M_{(1)} = r_1$

$= \underset{m_1}{U_1} D_1 V_1^T$

$$= \underset{m_1 \times r_1}{\tilde{U}_1}\ \underset{r_1 \times r_1}{\tilde{D}_1}\ \underset{r_1 \times (m_2 \cdots m_K)}{\tilde{V}_1^T}$$

(thin SVD)

Implications for $S_{(1)}$

$$D_1 V_1^T = \begin{bmatrix} \overset{r_1\times r_1}{\tilde{D}_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \\ v_{r+1}^T \\ \vdots \\ v_{m_1}^T \end{bmatrix} = \begin{bmatrix} \tilde{D}_1 \tilde{V}_1^T \\ 0 \end{bmatrix}_{m_1 \times m_2 \cdots m_n}$$

$$= S_{(1)} \mathcal{U}_{-1}^T \qquad \left(\mathcal{U}_{-1} = (\mathcal{U}_n \otimes \cdots \otimes \mathcal{U}_2) \right)$$

$$S_{(1)} = \begin{bmatrix} \overbrace{\tilde{D}_1 \tilde{V}_1^T}^{r_1 \times m_2 \cdots m_n} \\ \underbrace{0}_{(m_1 - r_1) \times m_2 \cdots m_n} \end{bmatrix} \mathcal{U}_{-1}^T = \begin{bmatrix} \tilde{D}_1 \tilde{V}_1^T \mathcal{U}_{-1}^T \\ 0 \end{bmatrix} = \underline{\underline{\begin{bmatrix} \tilde{S}_{(1)} \\ 0 \end{bmatrix}}}$$

$$\rightarrow \quad M_{(1)} = \tilde{\mathcal{U}}_1 \underbrace{\tilde{S}_{(1)}}_{r_1 \times m_2 \cdots m_n} \left(\mathcal{U}_n^T \otimes \cdots \otimes \mathcal{U}_1 \right)^T$$

Repeating for each mode:

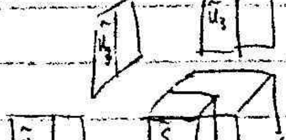
$\operatorname{rank}(M) = \underset{\sim}{r} \iff M = S \times \{\mathcal{U}_1 \ldots \mathcal{U}_K\}$

$S \in \mathbb{R}^{r_1 \cdots r_n}, \quad \mathcal{U}_n \in \mathcal{V}_{r_n, m_n}$

Returning to our problem

minimize $\|Y - S \times \{\mathcal{U}_1 \ldots \mathcal{U}_n\}\|^2$ over $\begin{cases} \mathcal{U}_n \in \mathcal{V}_{r_n, m_n} \\ S \in \mathbb{R}^{r_1 \times \cdots \times r_n} \end{cases}$

Attempt 1: By analogy w/ Eckart-Young Thm

$Y =$  , maybe take $\hat{M} =$ 

$= \tilde{S} \times \{\tilde{\mathcal{U}}_1, \hat{\mathcal{U}}_2, \hat{\mathcal{U}}_3\}$

i.e. as $y_{ijk} = \sum_{i'=1}^{m_1} \sum_{j'=1}^{m_2} \sum_{k'=1}^{m_3} s_{i'j'k'} \, u^{(1)}_{ii'} u^{(2)}_{jj'} u^{(3)}_{kk'}$

Get $\hat{M}_{ijk} = \sum_{1}^{r_1} \sum_{1}^{r_2} \sum_{1}^{r_3}$ "

this is "original" Tucker Decomposition (Tucker I)
also called the Truncated HOSVD

Algorithm: given Y, $\underline{r}$

for k=1...K {
  let $U_k = \text{svd}(Y_{(k)})\$u[,1:r_k]$
  }

let $S = \underset{m_1 \times \cdots \times m_K}{Y} \times \{ \underset{r_1 \times m_1}{U_1^T}, \ldots, \underset{r_K \times m_K}{U_K^T} \} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ $\quad (S_{(1)} = U_1^T Y_{(1)} [U_K \otimes \cdots \otimes U_2])$

Check that this gives the "truncated" S

let $\text{svd}(Y_{(k)})\$u \to [U_k \; W_k]$ $\quad U_k^T W_k = 0_{r_k \times (m_k - r_k)}$

$(\text{core}(Y))_{(1)} = \begin{bmatrix} \tilde{S}_Y \\ \check{S}_Y \end{bmatrix}$

$S_{(1)} = U_1^T Y_{(1)} [U_K \otimes \cdots \otimes U_2]$

$= U_1^T \Big[\underbrace{[U_1 W_1] \begin{bmatrix} \tilde{S}_Y \\ \check{S}_Y \end{bmatrix} \left([U_K W_K]^T \otimes \cdots [U_2 W_2]^T \right)}_{Y_{(1)}} \Big] [U_K \otimes \cdots \otimes U_2]$

$= [I \;\; 0] \begin{bmatrix} \tilde{S}_Y \\ \check{S}_Y \end{bmatrix} \left[\begin{pmatrix} I \\ 0 \end{pmatrix} \otimes \begin{pmatrix} I \\ 0 \end{pmatrix} \cdots \otimes \begin{pmatrix} I \\ 0 \end{pmatrix} \right]$

$= \tilde{S}_Y \left[\begin{pmatrix} I \\ 0 \end{pmatrix} \otimes \cdots \begin{pmatrix} I \\ 0 \end{pmatrix} \right]$

repeat along each mode

Back to OLS estimation

Let $\check{Y}_{\underset{\sim}{r}}$ be the $\underset{\sim}{r}$ - truncated HOSVD of Y

unfortunately, $\check{Y}_{\underset{\sim}{r}}$ is not generally the OLS estimate of M
(unless Y is already of rank r)

OLS / ML estimate can be found w/ BCD (ALS)

Procedure: iteratively obtain $U_1 \dots U_n$, S to minimize $\|Y - S \times \{U\}\|$

task k: obtain U_k to minimize $\|Y - S \times \{U_1 \dots U_n\}\|^2$

WLOG, let $k=1$: $\|Y - S \times \{U_1 \dots U_n\}\|^2 = \|Y_{(1)} - U_1 S_{(1)} U_{-1}^T\|^2$

$$= tr(Y_{(1)}^T Y_{(1)}) - 2\, tr(Y_{(1)}^T U_1 S_{(1)} U_{-1}^T) + tr(U_1 S_{(1)} U_{-1}^T U_{-1} S_{(1)}^T U_1^T)$$

doesn't dep on U_1 $\qquad Y_1^T U_{-1} S_1^T U_1 \qquad = |S|^2$

$$\text{MIN SSR} \iff \max \quad tr(U_1 [S_{(1)} U_{-1}^T Y_{(1)}^T])$$

$$= tr(U_1^T [Y_{(1)} U_{-1} S_1^T])$$

$$= tr(U_1^T Z) \qquad Z \in \mathbb{R}^{m_1 \times r_1}$$

Lemma: $tr(U^T Z)$ is maximized in $U \in \mathcal{V}_{r,m}$ by

$$\hat{U} = [svd(Z)\$u][svd(Z)\$v]^T$$

(i.e. $\hat{U} = Z$ but without any singular values)

min is unique if Z has distinct singular values.

task S: obtain S to minimize $\|Y - S\times\{U_1 \dots U_K\}\|^2$

Solution: (OLS): $\begin{cases} y = \text{vec}(Y) \\ s = \text{vec}(S) \end{cases}$, $\text{vec}(S\times\{U_1 \dots U_K\}) = (U_K \otimes \dots \otimes U_1)s$

$\|Y - S\times\{U_1 \dots U_K\}\|^2 = (y - Us)^T(y - Us)$ $\quad y \in \mathbb{R}^{m_1 \cdots m_K}$ $\quad s \in \mathbb{R}^{r_1 \cdots r_K}$ $\quad U \in \mathbb{R}^{(m_1 \dots m_K)\times(r_1 \dots r_K)}$

This is a regression problem: $(y - X\beta)^T(y - X\beta)$

$$\hat{s} = (U^T U)^{-1} U^T y = U^T y$$

converting back to array form:

$$S = Y \times \{U_1^T, \dots, U_K^T\}$$

note $S\times\{U_1 \dots U_K\} \neq Y$:

$S\times\{U_1 \dots U_K\} = Y\times\{U_1^T \dots U_K^T\}\times\{U_1 \dots U_K\}$
$= Y\times\{U_1U_1^T \dots U_KU_K^T\}$

ALS for best rank r approximation

Given starting values $U_1 \dots U_K$, iteratate the following until convergence

① let $S = Y\times\{U_1^T \dots U_K^T\}$

② for $k = 1 \dots K$

a) compute $Z_k = Y_{(k)} U_{-k} S_{(k)}^T$

b) let $U_k = [\text{svd}(Z_k)\$u][\text{svd}(Z_k)\$v]^T$