TENSOR DECOMPOSITION MODELS

Why do we want to factorize a matrix/tensor?

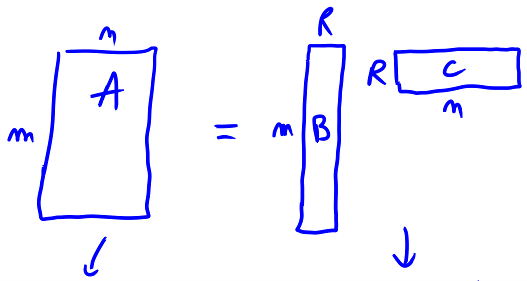
- ↳ $\boxed{\text{Compression / reduce model size}}$
- ↳ Discover structure in data (PCA)
- ↳ Infer missing entries (completion)

(I) MATRIX LOW RANK FACTORIZATION

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

$\text{rank}(A) \leq R \iff$ There exists $B \in \mathbb{R}^{m \times R}$ and $C \in \mathbb{R}^{R \times n}$ such that $\underline{A = BC}$

↳ "rank R factorization of A"



mn parameters vs. $R(m+n)$ parameters

Sometimes, an approximate factorization is enough:

$$\underset{m\times n}{A} \approx \underset{m\times R}{B}\,\underset{R\times n}{C}$$

Low rank approximation Problem: Given $A \in \mathbb{R}^{m \times n}$, and a target rank R

$$\min_{\substack{B \in \mathbb{R}^{m\times R} \\ C \in \mathbb{R}^{R \times n}}} \|A - BC\|_F$$

Def (Frobenius norm): $\|A\|_F = \sqrt{\sum_{i,j} (A_{ij})^2}$

Def (Singular Value Decomposition, SVD)

Any matrix $A \in \mathbb{R}^{m\times n}$ $(m \leq n)$ can be decomposed into

(1) $$\underset{m\times n}{A} = \underset{m\times m}{U}\,\underset{m\times m}{D}\,\underset{m\times n}{V^T}$$

where U and V are orthonormal (i.e. $U^TU = I$ and $V^TV = I$) and D is a diagonal matrix with non-negative entries.

The columns of U are the left singular vectors: $u_1, u_2, \dots, u_m \in \mathbb{R}^m$

The columns of V are the right singular vectors: $v_1, v_2, \dots, v_m \in \mathbb{R}^n$

The diagonal entries of D are the singular values $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_m \geq 0$

Remark:
- We can rewrite the SVD (1): $\quad A = \sum_{i=1}^{m} \sigma_i u_i v_i^T$
- If $\mathrm{rank}(A) = R$, then $\sigma_{R+1} = \sigma_{R+2} = \dots = \sigma_m = 0$ and $A = \sum_{i=1}^{R} \sigma_i u_i v_i^T$

Theorem (Eckart-Young)

Let $A \in \mathbb{R}^{m\times n}$ and let $A = UDV^T = \sum_{i=1}^{m} \sigma_i u_i v_i^T$ be its SVD.

Then the solution of

$$\min_{X \in \mathbb{R}^{m\times n}} \|A - X\|_F^2 \quad \text{subject to } \mathrm{rank}(X) \leq R$$

is given by $X^* = \sum_{i=1}^{R} \sigma_i u_i v_i^T$. (truncated SVD)

Low rank approximation Problem: Given $A \in \mathbb{R}^{m\times n}$, and a target rank R

$$\min_{\substack{B \in \mathbb{R}^{m\times R}\\ C \in \mathbb{R}^{R\times n}}} \|A - BC\|_F$$

$\boxed{\text{SOLUTION}}$ Compute SVD of $A = UDV^T$

keep the first R columns of U and V and the first R diagonal elements of D:

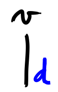
$$A = \Big[\, \overbrace{u_1 \; \cdots \; u_R}^{U_R} \; \cdots \; u_m \,\Big] \begin{bmatrix} \overbrace{\sigma_1 \; \sigma_2 \; \cdots \; \sigma_R}^{D_R} & \\ & \ddots \;\; \sigma_m \end{bmatrix} \begin{bmatrix} \overbrace{\begin{matrix} - v_1^T - \\ \vdots \\ - v_R^T - \end{matrix}}^{V_R^T} \\ \vdots \\ - v_m^T - \end{bmatrix}$$

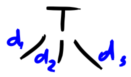
$$A \approx \underset{m\times R}{U_R}\,\underset{R\times R}{D_R}\,\underset{R\times n}{V_R^T} = \underbrace{(U_R D_R)}_{\substack{B\\ m\times R}} \underbrace{(V_R^T)}_{\substack{C\\ R\times n}}$$

II TENSOR NETWORKS (TN)

TN are graphs representing operations between tensors:

- Nodes represent tensors
- The arity (# of incoming edges / # of legs) of a node correspond to the order of the tensor:

$v \in \mathbb{R}^d$:  $\quad A \in \mathbb{R}^{m \times n}$: $\underset{m}{-}A\underset{n}{-}$

$T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$: 

- Edges represent contractions (summations):

+ Two matrices: $\underset{m}{-}A\underset{n}{-}$ $\qquad$ $\underset{n}{-}B\underset{k}{-}$

$$\left(\underset{m}{-}A\underset{n}{-}B\underset{k}{-}\right)_{i,j} = \sum_{k=1}^{n} A_{ik} B_{kj} = (AB)_{ij}$$

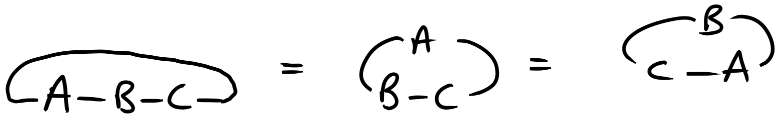
$\hookrightarrow$ $-A-B- = AB$

+ $u \underset{d}{-} v = \sum_{i=1}^{d} u_i v_i = \langle u, v \rangle$

+ $A \in \mathbb{R}^{m \times n}$, $v \in \mathbb{R}^n$ $\qquad \left(\underset{m}{-}A\underset{n}{-}v\right)_i = \sum_{j=1}^{m} A_{ij} v_j = (Av)_i$

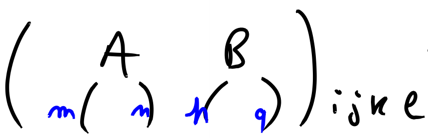
+ $A \in \mathbb{R}^{m \times m}$ $\qquad$  $= \sum_{i=1}^{m} A_{ii} = \mathrm{Tr}(A)$

+ Proof of $\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB)$ $\quad (\neq \mathrm{Tr}(BAC))$

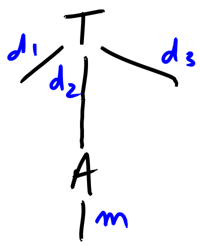


+ $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$ $\qquad$  $= u_i v_j = (u \circ v)_{ij}$

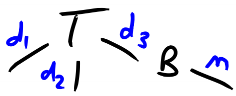
$\hookrightarrow$  $= u \circ v$

$A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{k \times q}$  $= A_{ij} B_{kl}$

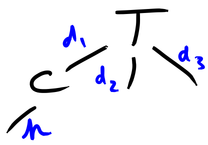
+ $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, $A \in \mathbb{R}^{m \times d_2}$, $B \in \mathbb{R}^{n \times d_3}$, $C \in \mathbb{R}^{h \times d_1}$

$\mathbb{R}^{d_1 \times m \times d_2} \ni$ 

$= T \times_2 A$

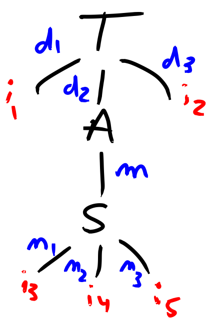
$\mathbb{R}^{d_1 \times d_2 \times n} \ni$ 

$= T \times_3 B$

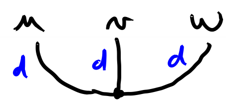
$\mathbb{R}^{h \times d_2 \times d_3} \ni$ 

$= T \times_1 C$

+ $T \in R^{d_1 \times d_2 \times d_3}$, $A \in R^{d_2 \times m}$, $S \in R^{m_1 \times m_2 \times m_3 \times m}$



$= \sum_{j=1}^{d_2} \sum_{k=1}^{m} T_{i_1 j i_2} A_{jk} S_{i_3 i_4 i_5 k}$

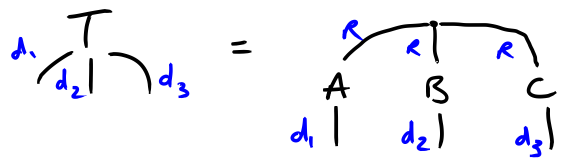
+ $u, v, w \in \mathbb{R}^d$: $\mathbb{R} \ni$ 

$= \sum_{i=1}^{d} u_i v_i w_i$

III 3 Tensor Decomposition Models

1) CP decomposition

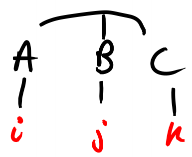
↳ (CANDECOMP | PARAFAC)

☆ 

← rank R CP decomposition of T

$A \in \mathbb{R}^{d_1 \times R}$, $B \in \mathbb{R}^{d_2 \times R}$, $C \in \mathbb{R}^{d_3 \times R}$

Def: The <u>CP rank</u> of a tensor T is the smallest R such that a rank R CP decomposition of T exists.

☆ : T_{ijk} = 

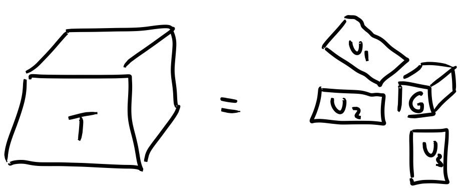
$= \sum_{r=1}^{R} A_{ir} B_{jr} C_{kr}$

If $A = \begin{pmatrix} | & & | \\ a_1 & \cdots & a_R \\ | & & | \end{pmatrix}$, $B = \begin{pmatrix} | & & | \\ b_1 & \cdots & b_R \\ | & & | \end{pmatrix}$, $C = \begin{pmatrix} | & & | \\ c_1 & \cdots & c_R \\ | & & | \end{pmatrix}$

then $T_{ijk} = \sum_{r=1}^{R} (a_r)_i (b_r)_j (c_r)_k = \sum_{r=1}^{R} (a_r \circ b_r \circ c_r)_{ijk}$

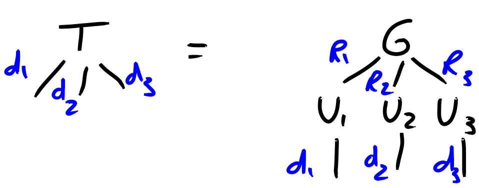
↳ $T = \sum_{r=1}^{R} a_r \circ b_r \circ c_r$

2) Tucker decomposition



$T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, $G \in \mathbb{R}^{R_1 \times R_2 \times R_3}$, $U_i = \mathbb{R}^{d_i \times R_i}$ for $i = 1,2,3$

↳ "core tensor" ↳ "factor matrices"

(△) 

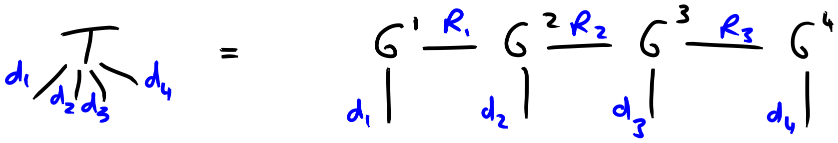
← rank (R_1, R_2, R_3) Tucker decomposition of T.

def : The <u>multilinear rank</u> (Tucker rank) of $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is the smallest tuple (R_1, R_2, R_3) s.t. a rank (R_1, R_2, R_3) Tucker decomposition of T exists.

↳ The Tucker decomposition (△) can be written as $T = G \times_1 U_1 \times_2 U_2 \times_3 U_3$

3) Tensor train decomposition (TT)

$T \in \mathbb{R}^{d_1 \times d_2 \times d_3 \times d_4}$, $\underbrace{G^1 \in \mathbb{R}^{d_1 \times R_1},\ G^2 \in \mathbb{R}^{R_1 \times d_2 \times R_2},\ G^3 \in \mathbb{R}^{R_2 \times d_3 \times R_3},\ G^4 \in \mathbb{R}^{R_3 \times d_4}}_{\text{"core tensors"}}$



↳ rank (R_1, R_2, R_3) TT decomposition of T.

Def: The TT rank of a tensor $T \in \mathbb{R}^{d_1 \times d_2 \times d_3 \times d_4}$ is the smallest (R_1, R_2, R_3) such that rank (R_1, R_2, R_3) TT decomposition of T exists.

(IV) TENSOR LOW RANK APPROXIMATION

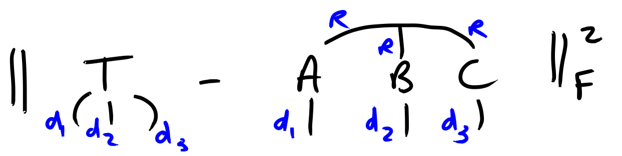
For $T \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_N}$ $\quad$ ($d_1 d_2 \cdots d_N$ parameters)

	rank	# parameters	EXACT DECOMPOSITION	LOW RANK APPROXIMATION
CP	R	$R(d_1 + d_2 + \cdots + d_N)$	NP-hard	NP
TUCKER	$(R_1, R_2, \ldots, R_N)$	$R_1 R_2 \cdots R_N + \sum_{i=1}^{N} d_i R_i$	P (easy)	NP
TT	$(R_1, \ldots, R_{N-1})$	$d_1 R_1 + R_1 d_2 R_2 + R_2 d_3 R_3 + \cdots + R_{N-2} d_{N-1} R_{N-1} + R_{N-1} d_N$	P	NP

Low rank approximation problem :

1) CP $\qquad$ Given $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and a target rank R :

$\min_{A \in \mathbb{R}^{d_1 \times R},\ B \in \mathbb{R}^{d_2 \times R},\ C \in \mathbb{R}^{d_3 \times R}}$



Def : The Frobenius norm of $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$: $\quad \|T\|_F = \sqrt{\sum_{ijk} (T_{ijk})^2}$

2) TUCKER

Given $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and a target rank (R_1, R_2, R_3)

$$\min_{\substack{G \in \mathbb{R}^{R_1 \times R_2 \times R_3} \\ U_1 \in \mathbb{R}^{d_1 \times R_1} \\ U_2 \in \mathbb{R}^{d_2 \times R_2} \\ U_3 \in \mathbb{R}^{d_3 \times R_3}}} \| T - G \times_1 U_1 \times_2 U_2 \times_3 U_3 \|_F^2$$

3) TENSOR TRAIN

Given $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and a target rank (R_1, R_2)

$\min_{G_1 \in \mathbb{R}^{d_1 \times R_1},\ G_2 \in \mathbb{R}^{R_1 \times d_2 \times R_2},\ G_3 \in \mathbb{R}^{R_2 \times d_3}}$