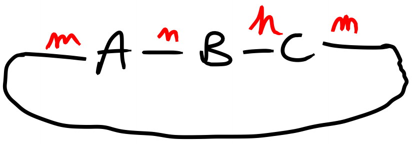
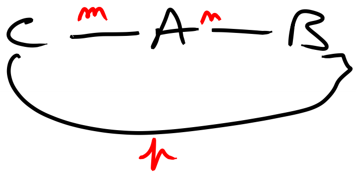
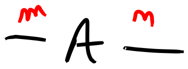
# TENSOR NETWORKS
# FOR
# ANALYSIS OF BENEFITS of DEPTH in RNNs
# &
# EXPLAINABILITY.

$$\mathrm{Tr}(\overset{\curvearrowleft}{ABC}) = \mathrm{Tr}(CAB) \neq \mathrm{Tr}(CBA)$$

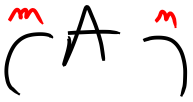


=

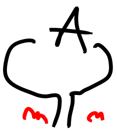




$$A = \begin{pmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{m \times n}$$

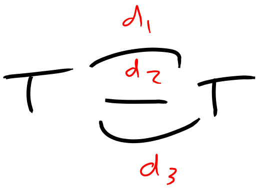


$$\mathrm{vec}(A) = \begin{pmatrix} | \\ a_1 \\ | \\ | \\ a_2 \\ | \\ \vdots \\ | \\ a_m \\ | \end{pmatrix} \in \mathbb{R}^{mn}$$

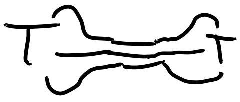


$$T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$$

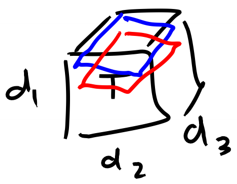
$$\| T \|_F^2 =$$



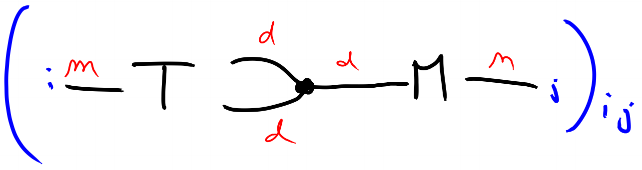
=



$$= \| \mathrm{vec}(T) \|_F^2$$

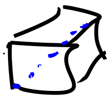


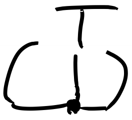
$$\rightsquigarrow \quad T_{(1)} \in \mathbb{R}^{d_1 \times d_2 d_3}$$



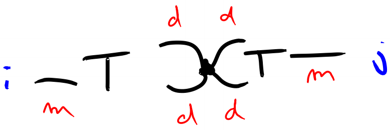
$$= \sum_k T_{ikk} M_{kj}$$

$$T \in \mathbb{R}^{d \times d \times d}$$





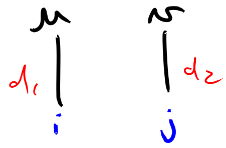
$$= \sum_{i=1}^{d} T_{iii}$$



$$= \sum_{k=1}^{d} T_{ikk} T_{jkk}$$

---

Outer product $u \in \mathbb{R}^{d_1}$ $v \in \mathbb{R}^{d_2}$



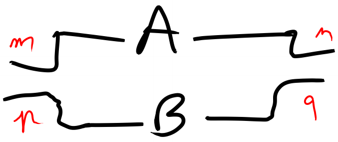
$= u_i v_j$

$= uv^T$

$= A_{ij} B_{kl}$

$\mathbb{R}^{m \times n \times p \times q}$

$A \circ B$



$= A \otimes B$

↑

$\mathbb{R}^{mp \times nq}$

Kronecker Product.

# SEQUENTIAL MODELS

$h_0,$

$$h_t = \phi(h_{t-1}, x_t)$$

$$y_t = \psi(h_t)$$

Vanilla 1st order RNN :

$$h_t = \sigma(U h_{t-1} + V x_t)$$

$$y_t = z(W h_t)$$

$\cap|$

2nd order RNN

$$h_t = \sigma\left( h_{t-1} \overset{m}{—} \underset{\underset{x_t}{|_d}}{A} \overset{m}{—} \right) \in \mathbb{R}^m$$

( linear $2^{nd}$ order RNN $\equiv$ WFA )

$\cup|$

BENEFITS OF DEPTH FOR LONG-TERM MEMORY OF RECURRENT NETWORKS

Yoav Levine, Or Sharir & Amnon Shashua
The Hebrew University of Jerusalem
{yoavlevine,or.sharir,shashua}@cs.huji.ac.il

RAC (recurrent arithmetic circuit)

$$h_t = (U h_{t-1}) \odot (V x_t)$$

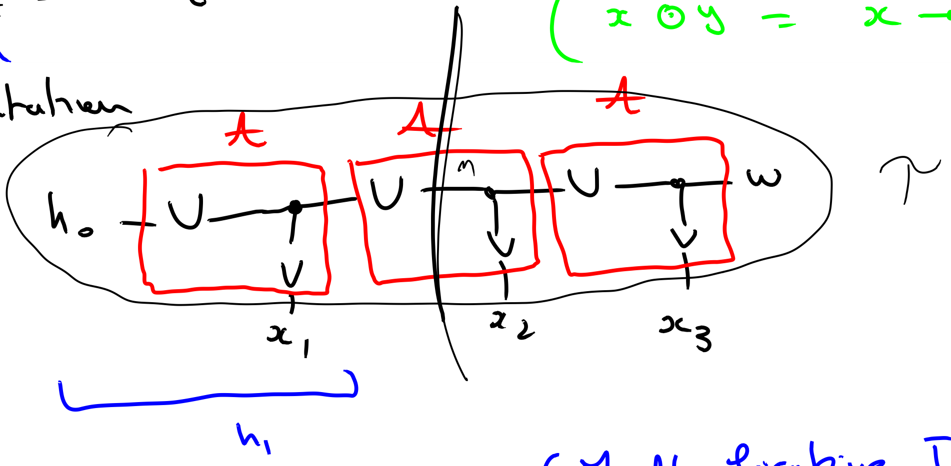
Hadamard / Component wise product.

$$y_t = W h_t$$

( $x \odot y = x \overset{}{—\!\!\!\bullet\!\!\!—} y$ )

with 1 layer RAC computation

$y_t =$



(Multiplicative Interaction RNN)

RAC with L layers:

$$h_t^{(1)} = (U^{(1)} h_{t-1}^{(1)}) \odot (V^{(1)} x_t)$$

$$h_t^{(2)} = (U^{(2)} h_{t-1}^{(2)}) \odot (V^{(1)} h_t^{(1)})$$

$$\vdots$$

$$h_t^{(L)} = (U^{(L)} h_{t-1}^{(L)}) \odot (V^{(L)} h_t^{(L)})$$

$$y_t = W h_t^{(L)}$$

**Def**: Separation rank.

Let $f : \mathcal{X}^T \longrightarrow \mathbb{R}$. For any $t = 1 \dots T$

$\uparrow$ $(x_1, x_2, \dots, x_T)$ $\uparrow$ $\in \mathcal{X}$

$$\operatorname{sep}_t(f) = \min \Big\{ R \in \mathbb{N} \;\Big|\; \exists\, g_1, \dots, g_R \cdot h_1, \dots, h_R$$

$$f(x_1, \dots, x_T) = \sum_{r=1}^{R} g_r(x_1, \dots, x_t)\, h_r(x_{t+1}, \dots, x_T)$$

$$\forall\, x_1, x_2, \dots, x_T \in \mathcal{X} \Big\}$$

Intuition: If $\operatorname{sep}_t(f) = 1$, then

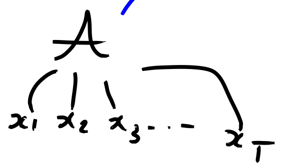
$$f(x_1, \dots, x_T) = g(x_1, \dots, x_t)\, h(x_{t+1}, \dots, x_T)$$

$$f : x_1, x_2, \dots, x_{T/2} \;\vdots\; x_{T/2+1}, \dots, x_T$$

$$\operatorname{sep}_{T/2}(f)$$

Lemma 1: If $\mathcal{X} = \mathbb{R}^d$ and

$$f : x_1, x_2, \dots, x_T \longmapsto$$



$\langle \mathcal{A}, x_1 \circ x_2 \circ \dots \circ x_T \rangle$

$\in \mathbb{R}$

where $\mathcal{A} \in \mathbb{R}^{\underbrace{d \times \dots \times d}_{T \text{ times}}}$

$$\operatorname{sep}_t(f) = \operatorname{rank}\left( [\![ \mathcal{A} ]\!]_{(1, \dots, t), (t+1, \dots, T)} \right)$$

**Def**: Let $f : \mathcal{X}^T \rightarrow \mathbb{R}$ be an arbitrary function.

The grid tensor of $f$ w.r.t. the template vectors $x_1, x_2, \dots, x_k \in \mathcal{X}$ is the tensor $\mathcal{A}(f) \in \mathbb{R}^{\underbrace{k \times k \times \dots \times k}_{T \text{ times}}}$ is defined by

$$\mathcal{A}(f)_{i_1 i_2 \dots i_T} = f(x_{i_1}, x_{i_2}, \dots, x_{i_T})$$

↳ ex $T = 2$, template vectors $x_1, x_2, x_3$

$$\mathcal{A}(f) \in \mathbb{R}^{3 \times 3} \qquad \mathcal{A}(f) = \begin{bmatrix} f(x_1, x_1) & f(x_1, x_2) & f(x_1, x_3) \\ f(x_2, x_1) & f(x_2, x_2) & f(x_2, x_3) \\ f(x_3, x_1) & f(x_2, x_3) & f(x_3, x_1) \end{bmatrix}$$

Lemma 2: Let $f:\mathcal{X}^T \to \mathbb{R}$. For any template vectors $x_1, x_2, \dots, x_n \in X$ the grid tensor $A(f)$ is s.t.

$$sep_t(f) \geq rank\left([\![A(f)]\!]_{(1,\dots,t)(t+1,\dots,T)}\right)$$

Thm Consider an RAC with $n$ neurons and $L$ layers. $\mathcal{X} = \mathbb{R}^d$

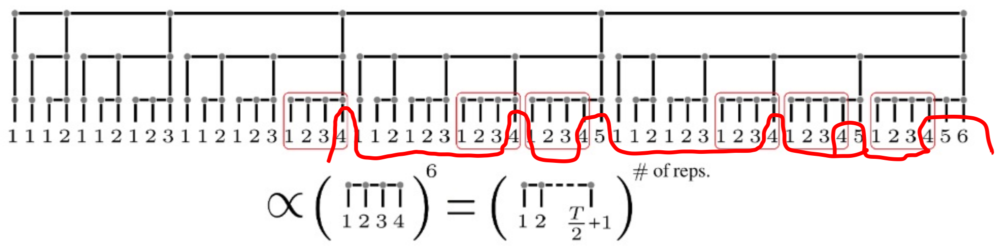
Assume $n \leq d$

Shallow RAC $(L=1)$ : $sep_{T/2}(f) = n$ (with P 1)

DEEP RAC $(L=2)$ : $sep_{T/2}(f) \gtrsim \frac{4^n}{4\sqrt{\pi(n-1)}}$

exp. dependency n

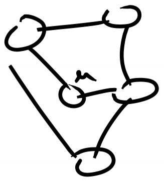
L=3



ENTANGLEMENT ENTROPY.

Explainable Natural Language Processing with Matrix Product States

Jirawat Tangpanitanon[1,2,†], Chanatip Mangkang[3,†], Pradeep Bhadola[4], Yuichiro Minato[5], Dimitris G. Angelakis[6,7], Thiparat Chotibut[3,*]



$f_u: x_{v_1}, \dots, x_{v_{|V|}} \longmapsto \mathbb{R}$

$sep_{\{u\}, V\setminus\{u\}}(f_u)$

$sep_{\{1\text{-Hop}(u)\}, V\setminus 1\text{-Hop}\}}(f_u)$

WRITE STYLUS LAB

WACOM Tablet.