T6.

Part 1 RNN and review of FFL

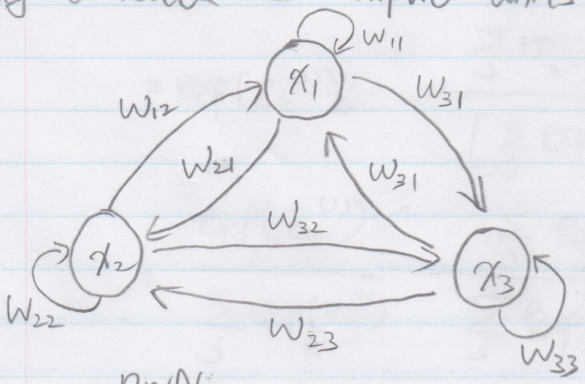
ex1 circular shift register

$001 \xrightarrow{\text{shift one places to the left}} 010 \longrightarrow 100$

$010 \longrightarrow 100 \longrightarrow 001$

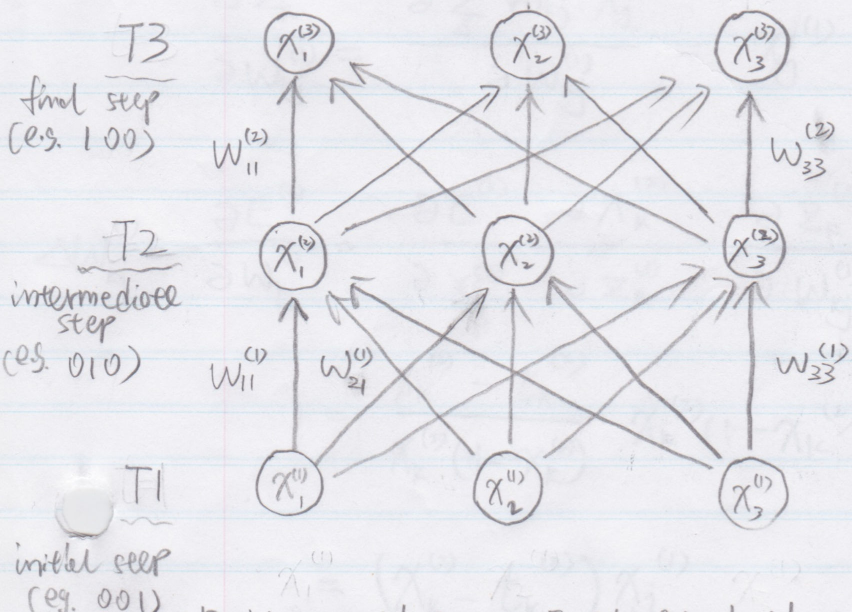
$100 \longrightarrow 001 \longrightarrow 010$

Fully connected 3 input units, no hidden unit



RNN → FFN

Time T:



NB: We evaluate FFN at **both** T2 and T3

Forward pass in T1 → T2

$$z_k^{(1)} = \sum_j w_{kj}\, x_j^{(1)}$$

$$x_k^{(2)} = \frac{\exp(z_k^{(1)})}{\sum_j \exp(z_j^{(1)})}$$

$$E^{(2)} = -\sum_{k=1}^{3} t_k^{(2)} \log x_k^{(2)}$$

Back-prop T2 → T1:

$$\frac{\partial E^{(2)}}{\partial w_{kj}^{(1)}} = \frac{\partial E^{(2)}}{\partial x_k^{(2)}} \frac{\partial x_k^{(2)}}{\partial z_k^{(1)}} \frac{\partial z_k^{(1)}}{\partial w_{kj}^{(1)}}$$

$$\frac{\partial E^{(2)}}{\partial x_k^{(2)}} = -\frac{\partial \sum_j t_j^{(2)} \log x_j^{(2)}}{\partial x_k^{(2)}}$$

$$= -\frac{\partial\, t_k^{(2)} \log x_k^{(2)}}{\partial x_k^{(2)}} - \frac{\partial (1-t_k^{(2)}) \log(1-x_k^{(2)})}{\partial x_k^{(2)}}$$

$$= -\frac{t_k^{(2)}}{x_k^{(2)}} + \frac{1-t_k^{(2)}}{1-x_k^{(2)}}$$

$$= \frac{x_k^{(2)} - t_k^{(2)}}{x_k^{(2)}(1-x_k^{(2)})}$$

Back-prop cont'd

$$\frac{\partial x_k^{(2)}}{\partial z_k^{(1)}} = \frac{\partial \exp(z_k^{(1)})}{\partial z_k^{(1)}} \frac{\partial x_k^{(2)}}{\partial \exp(z_k^{(1)})}$$

$$= \exp(z_k^{(1)}) \cdot \frac{\partial \exp(z_k^{(1)})\left[\sum_j \exp(z_j^{(1)})\right]^{-1}}{\partial \exp(z_k^{(1)})}$$

$$= \exp(z_k^{(1)}) \left\{ \left[\sum_j \exp(z_j^{(1)})\right]^{-1} + \exp(z_k^{(1)})(-1)\left[\sum_j \exp(z_j^{(1)})\right]^{-2} \right\}$$

$$= \exp(z_k^{(1)}) \cdot \frac{\sum_j \exp(z_j^{(1)}) - \exp(z_k^{(1)})}{\left[\sum_j \exp(z_j^{(1)})\right]^2}$$

$$= \frac{\exp(z_k^{(1)})}{\sum_j \exp(z_j^{(1)})} \frac{\sum_{j \neq k} \exp(z_j^{(1)})}{\sum_j \exp(z_j^{(1)})}$$

$$= x_k^{(2)} (1 - x_k^{(2)})$$

$$\frac{\partial z_k^{(1)}}{\partial w_{kj}^{(1)}} = \frac{\partial \sum_j w_{kj}^{(1)} x_j^{(1)}}{\partial w_{kj}^{(1)}} = x_j^{(1)}$$

$$\Delta W_{kj}^{(1)} = \frac{\partial E^{(2)}}{\partial W_{kj}^{(1)}} = \frac{\partial E^{(2)}}{\partial x_k^{(2)}} \frac{\partial x_k^{(2)}}{\partial z_k^{(1)}} \frac{\partial z_k^{(1)}}{\partial W_{kj}^{(1)}}$$

$$= \frac{x_k^{(2)} - t_k^{(2)}}{x_k^{(2)}(1 - x_k^{(2)})} x_k^{(2)}(1 - x_k^{(2)}) \, x_j^{(1)}$$

$$= (x_k^{(2)} - t_k^{(2)}) x_j^{(1)}$$

Similarly.

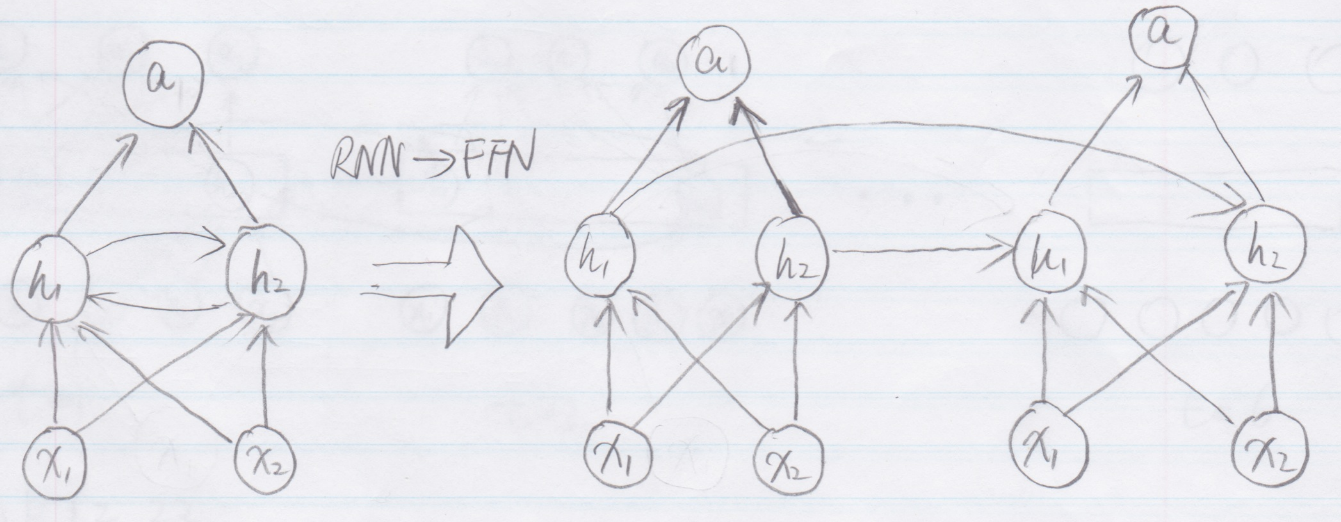
$$\Delta W_{kj}^{(2)} = (x_k^{(3)} - t_k^{(3)})\, x_j^{(2)}.$$

Update $W_{kj}^{(1)}$ and $W_{kj}^{(2)}$ with combined gradient

$$W_{kj}^{(1)*} = W_{kj}^{(1)} - \varepsilon(\Delta W_{kj}^{(1)} + \Delta W_{kj}^{(2)})$$

$$W_{kj}^{(2)*} = W_{kj}^{(2)} - \varepsilon(\Delta W_{kj}^{(1)} + \Delta W_{kj}^{(2)})$$

ex2. Simple conversion of a RNN with hidden units



$t = 1$

ex3. sequence completion (high level overview)

LETTER	A	B	C	D	E
Hidden Encoding	12	23	31	21	13

25 training sequences:

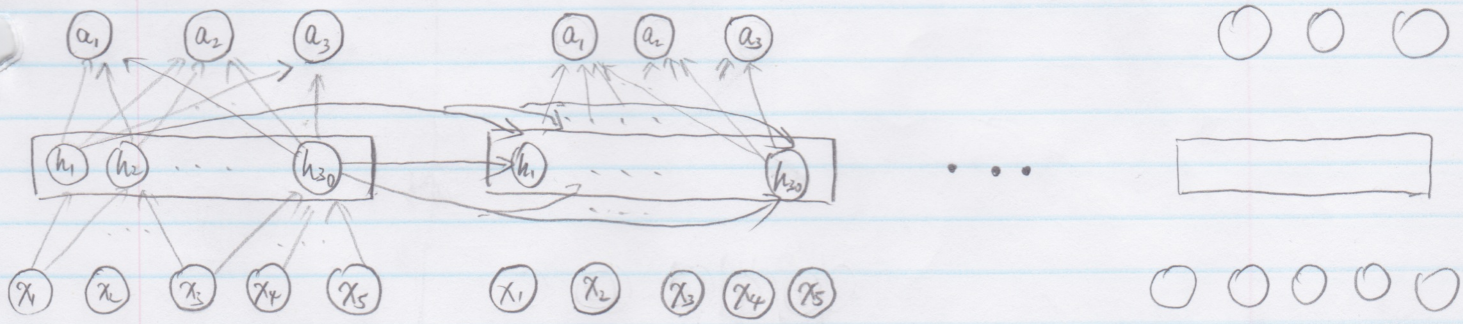
AA1212 . AB1223 . AC1231 . . .

BA2312

. . .

⋮

Task: Given the first two letters, predict the remaining sequenc



t=1 t=2 t=6

e.g. AB12 23.

	I/O unit:	A	B	C	D	E	a_1	a_2	a_3
							Ideal output		
time:	t=1	✓	×	×	×	×	—	—	—
	2	×	✓	×	×	×	—	—	—
	3	×	×	×	×	×	✓	×	×
	4	×	×	×	×	×	×	✓	×
	5	×	×	×	×	×	×	✓	×
	6	×	×	×	×	×	×	×	✓

Forward pass at Time t:

$$y_k^{(t)} = \sum_{kj} W_{kj}^{(t)} x_j + \sum_{kl} W_{kl}^{(t-1)} h_l^{(t-1)} + b_k^{(t)}$$

$$h_k^{(t)} = \frac{\exp(y_k^{(t)})}{1 - \exp(y_j^{(t)})}$$

$$Z_s^{(t)} = \sum_{sk} W_{sk} h_k^{(t)} + b_s^{(t)}$$

$$a_s^{(t)} = \frac{\exp(Z_s^{(t)})}{\sum_{s=1}^{3} \exp(Z_s^{(t)})}$$

$$E^{(t)} = -\sum_s t_s^{(t)} \log a_s^{(t)}$$

Back-prop to get $\frac{\partial E}{\partial W_{kj}^{(t)}}$, $\frac{\partial E}{\partial W_{kl}^{(t)}}$, $\frac{\partial E}{\partial W_{sk}^{(t)}}$, $\frac{\partial E}{\partial b_{s/k}^{(t)}}$

Update:

$$\Delta W_{kj} = \sum_t \frac{\partial E^{(t)}}{\partial W_{kj}^{(t)}}$$

$$\Delta W_{kl} = \sum_t \frac{\partial E^{(t)}}{\partial W_{kl}^{(t)}}$$

$$\Delta W_{sk} = \sum_t \frac{\partial E^{(t)}}{\partial W_{sk}^{(t)}}$$

$$\Delta b_{s/k} = \sum_t \frac{\partial E^{(t)}}{\partial b_{s/k}}$$