DS 4440 Midterm Review

2. $W^I = {}_{2}\overset{3}{\begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 0 \end{bmatrix}}$

(A) $h = {}_{1}\underset{x}{\overset{2}{\begin{bmatrix} -2 & 1 \end{bmatrix}}}\;{}_{2}\overset{3}{\begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 0 \end{bmatrix}} = {}_{1}\begin{bmatrix} 1 & -7 & 2 \end{bmatrix}$

$h = \begin{bmatrix} 1 & 0 & 2 \end{bmatrix}$

(B) $x = b = 2\;\overset{2}{\begin{bmatrix} -2 & 1 \\ 3 & 5 \end{bmatrix}}\;{}_{2}\overset{3}{\begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 0 \end{bmatrix}}$

$= {}_{2}\overset{3}{\begin{bmatrix} 1 & 0 & 2 \\ 18 & 0 & 0 \end{bmatrix}}$

(C) $W^0 = {}_{3}\overset{1}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}$ $\quad \hat{y} = hW^0 = \begin{bmatrix} 7 \\ 18 \end{bmatrix}$

$y = \begin{bmatrix} 8 \\ 11 \end{bmatrix}$ * Sorry

$$\mathcal{L} = (y-\hat{y})^2$$

$$\nabla_h \mathcal{L} = \nabla_h \hat{y} \cdot \frac{\partial}{\partial \hat{y}} \mathcal{L}$$

$$= \nabla_h \hat{y} \cdot -2(y-\hat{y})$$

$$= -\nabla_h hW^o \cdot 2 \cdot \begin{bmatrix}1\\-7\end{bmatrix}$$

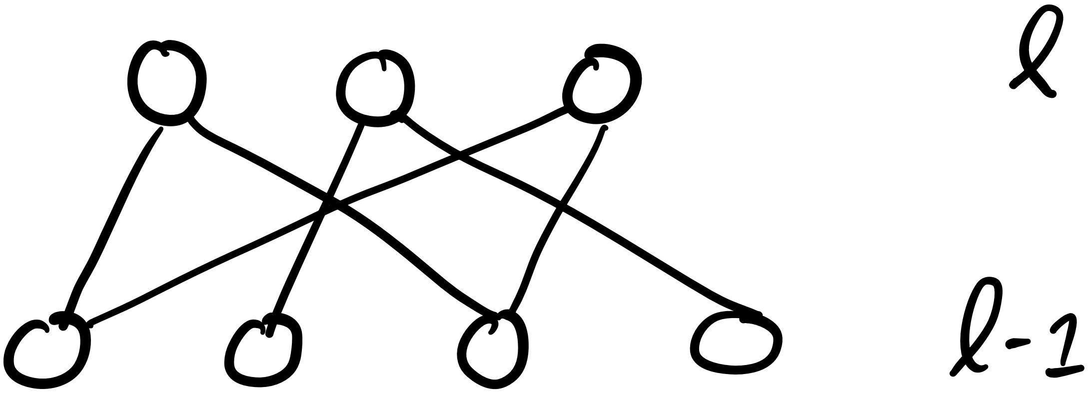
$$= -W^o \cdot \begin{bmatrix}2\\-14\end{bmatrix}$$

$$= \begin{bmatrix}-2 & -4 & -6\\ 14 & 28 & 42\end{bmatrix}$$

This differs from practice handout, which we did in lecture

3.

(A.)



(B.)

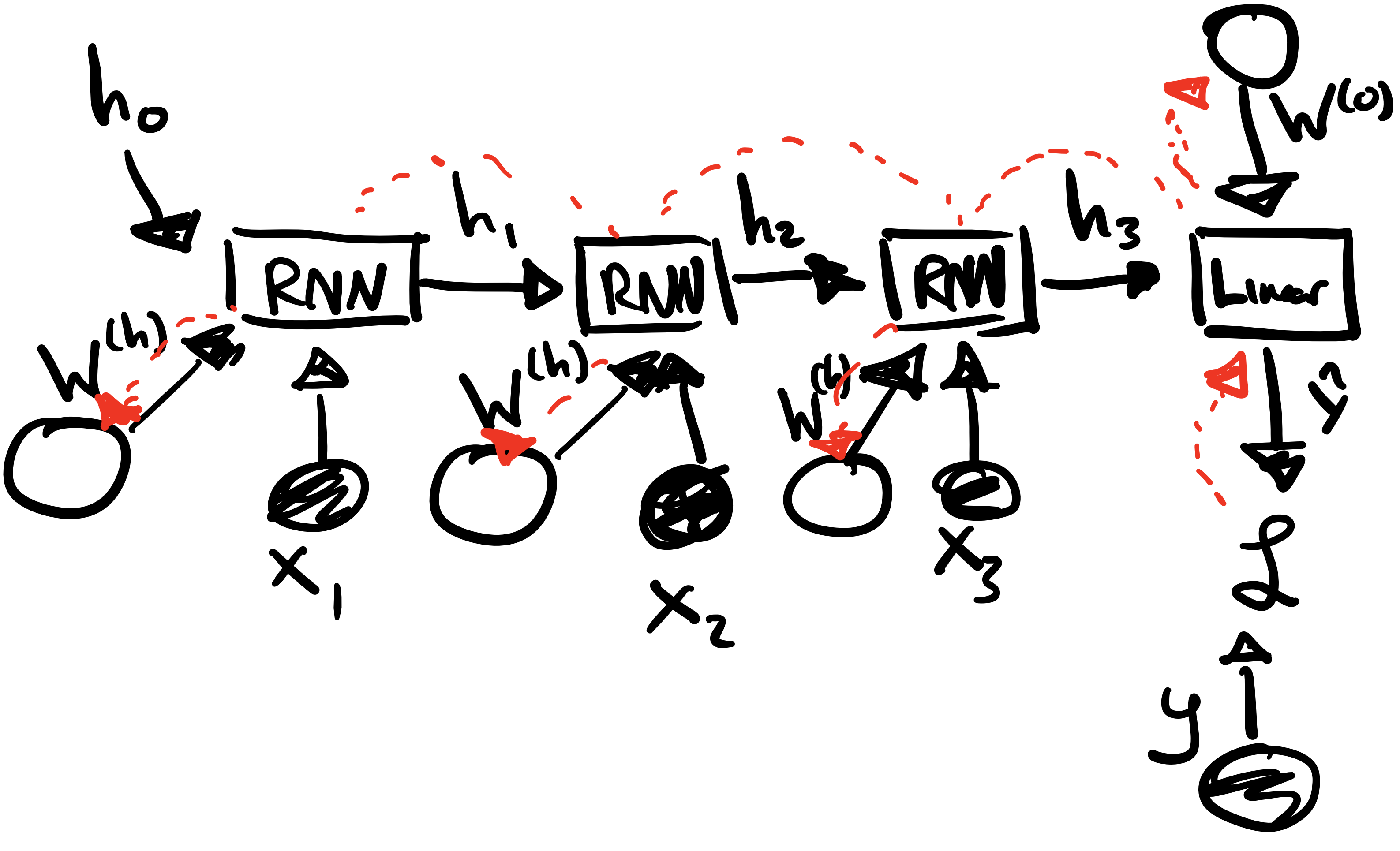
$$W^{(\ell)} = \begin{bmatrix} W_{11}^{\ell} & 0 & W_{13}^{\ell} & 0 \\ 0 & W_{22}^{\ell} & 0 & W_{24}^{\ell} \\ W_{31}^{\ell} & 0 & W_{33}^{\ell} & 0 \end{bmatrix}$$

(C.) Dropping weights is different then dropping nodes (activations) as in drop out.

4.

(A.) $$\hat{y} = \overset{\text{Sigmoid}}{\overset{\downarrow}{\sigma}}\left(\underset{\langle 1 \times d \rangle}{W^{(0)}} \cdot \underset{\langle d \times 1 \rangle}{h_3}\right)$$

(B.)



(c.)

$$\frac{\partial \mathcal{L}}{\partial W^{(h)}} = \frac{\partial \mathcal{L}}{\partial h_1} \star \frac{\partial h_1}{\partial W^{(h)}} + \frac{\partial \mathcal{L}}{\partial h_2} \star \frac{\partial h_2}{\partial W^{(h)}}$$

$$+ \frac{\partial \mathcal{L}}{\partial h_3} \star \frac{\partial h_3}{\partial W^{(h)}}$$

$$\frac{\partial \mathcal{L}}{\partial h_3} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \star \frac{\partial \hat{y}}{\partial h_3}$$

$$\frac{\partial \mathcal{L}}{\partial h_2} = \frac{\partial \mathcal{L}}{\partial h_3} \star \frac{\partial h_3}{\partial h_2}$$

$$\frac{\partial \mathcal{L}}{\partial h_1} = \frac{\partial \mathcal{L}}{\partial h_2} \star \frac{\partial h_2}{\partial h_1}$$

5. CNNs + RNNs

$$X_i = \underset{k}{\begin{bmatrix} x_{i11} \cdots x_{i1k} \\ \cdots \\ x_{im1} \cdots x_{imk} \end{bmatrix}} m$$

Input $\langle b, m, k, d \rangle$

ConvLayer

- filters cover two words + bias $= 2 \cdot d + 1$ params per filter
- Assume 32 of these =

  $64 \cdot d + 32$ params

- Assume stride 2; pass over doc i to get $\underline{\langle \frac{k \cdot m}{2}, 32 \rangle}$ outputs

- RNN w hidden weights $W^{(h)}$ $\langle 32+\overset{\text{hidden dims}}{\overset{\downarrow}{d}}, d\rangle$

- Pass over $\frac{k \cdot m}{2}$ elements per i.

- $\hat{y} = \sigma(W^{(o)} \cdot h_{last})$