CS182 HW4.

1.(a). $c\,d_{in}\,d_{out}^2 + c\,d_{in}\,d_{out} = 2c\,d_{in}\,d_{out}^2$

(b). $2c\,d_{out}\,d_{in}^2 \rightarrow$ updated runtime.

$\frac{d_{out}}{d_{in}} >> 1$ $\Rightarrow$ $d_{in}d_{out}^2 >> d_{out}d_{in}^2$

Updated runtime : compute W^TW, $W(W^TW)$, $p(W) = \frac{3}{2}W - \frac{1}{2}W(W^TW)$

significantly faster. since we avoid compute $d_{out} \cdot d_{out}$.

2.(a). maximize dynamic range and coverage of the representable values

(b). $Var(y_i) \approx \sum_{j=1}^{d_{in}} Var(cW_{ij}x_j) = c^2 \sum_{j=1}^{d_{in}} Var(W_{ij})\,Var(x_j)$

$W_{ij} \sim N(0,1) \Rightarrow Var(W_{ij}) = 1$

$RMS(x) = 1 \Rightarrow \sqrt{\frac{1}{d_{in}} \sum Var(x_j)} = 1 \quad \therefore \sum_{j=1}^{d_{in}} Var(x_j) = d_{in}$

$RMS(y) = 1 \Rightarrow Var(y_i) \approx 1$

$\therefore c^2 d_{in} = 1$

$\therefore c = \frac{1}{\sqrt{d_{in}}}$

(c). $RMS(\Delta y) = \frac{\|\Delta y\|_2}{\sqrt{d_{out}}} \le \frac{c\,\|\Delta W\|_2\,\|x\|_2}{\sqrt{d_{out}}} = c\,\|\Delta W\|_2\,RMS(x)$

$\because RMS(x) = 1$

$\therefore RMS(\Delta y) \le c\,\|\Delta W\|_2$

We want $RMS(\Delta y) \le 1$ with $RMS(x) = 1$, $c = \frac{1}{\sqrt{d_{in}}}$

$\therefore \|\Delta W\|_2 \le \sqrt{d_{in}}$

(d). $\Delta W = \alpha\,\mathrm{sign}(\nabla_W L)$

$\|\Delta W\|_2 = \|\alpha\,\mathrm{sign}(\nabla_W L)\|_2 = |\alpha| \cdot \|\mathrm{sign}(\nabla_W L)\|_2 = |\alpha| \cdot \sqrt{d_{in}d_{out}}$

$\|\Delta W\|_2 \le \sqrt{d_{in}}$

$|\alpha| \cdot \sqrt{d_{out}d_{in}} \le \sqrt{d_{in}} \Rightarrow |\alpha| \le \frac{1}{\sqrt{d_{out}}}$

$\therefore \alpha = \frac{1}{\sqrt{d_{out}}}$

(e). $\Delta W = \alpha \cdot UV^T$, $U \in R^{d_{out} \times r}$, $V \in R^{d_{in} \times r}$

$\|UV^T\|_2 = 1$

$\therefore \|\Delta W\|_2 = \|\alpha UV^T\|_2 = |\alpha|$

Through (c): $\|\Delta W\|_2 \le \sqrt{d_{in}}$

$\therefore |\alpha| \le \sqrt{d_{in}}$

$\therefore \alpha = \sqrt{d_{in}}$

(f). $\nabla_{x_n} L = \left(\frac{\partial x_{n+1}}{\partial x_n}\right)^T \nabla_{x_{n+1}} L = (c_n W_n)^T \nabla_{x_{n+1}} L = c_n W_n^T \nabla_{x_{n+1}} L$

If $c_n = 1$:

$\nabla_{x_n} L = W_n^T \nabla_{x_{n+1}} L$. intermediate backpropagated gradients can be ignored.

(g).

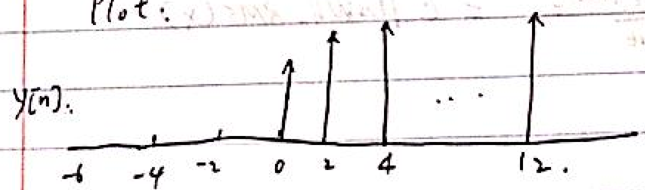
3.(a). $n<0$, $y[n]=0$.

$y[n] = \sum_{-\infty}^{\infty} x[n-i]h[i] = \sum_{-\infty}^{\infty} x[i]h[n-i] = \sum_{i=0}^{L-1} h[n-i]$

$y[0] = 1$, $y[1] = \frac{3}{2}$, $y[2] = 2 - \frac{1}{4} = \frac{7}{4}$, ...

$y[7] = 2 - (\frac{1}{2})^7$, ...

Plot:



3/(b). $y_2(n) = h(n) * x_2(n) = h(n) * x(n-N) = y(n-N)$

3/(c). $y = \begin{bmatrix} 40 & 40 & 40 \\ 40 & 40 & 40 \\ 40 & 40 & 40 \end{bmatrix}$

3/(d). i: $\begin{bmatrix} 19 & 28 & 32 & 36 & 29 \\ 30 & 40 & 40 & 40 & 30 \\ 30 & 40 & 40 & 40 & 30 \\ 30 & 40 & 40 & 40 & 30 \\ -49 & -68 & -72 & -76 & -59 \end{bmatrix}$, ii: $\begin{bmatrix} 19 & 32 & 29 \\ 34 & 40 & 34 \\ 49 & 52 & 59 \end{bmatrix}$

4/(a). $K^2FC + F$

output shape: $W_{out} \times H_{out} \times F$. , $W_{out} = \lfloor \frac{W-K+2P}{S} \rfloor + 1$

$H_{out} = \lfloor \frac{H-K+2P}{S} \rfloor + 1$

(b). $W_{out} = \lfloor \frac{W-K+2P}{S} \rfloor + 1 = \lfloor \frac{W-K}{S} \rfloor + 1$ $(P=0)$.

$H_{out} = \lfloor \frac{H-K}{S} \rfloor + 1$

$\therefore$ shape: $W_{out} \times H_{out} \times C$.

(c). $RF_\ell = RF_{\ell-1} + (K_\ell - 1) \cdot S_{\ell-1}$

$RF_L = 1 + L(K-1)$

receptive field size of the last output: $LK - L + 1$.

d).

7/(a).

8. (a). Google Gemini

(b). Nobody

(c). 4 hours.