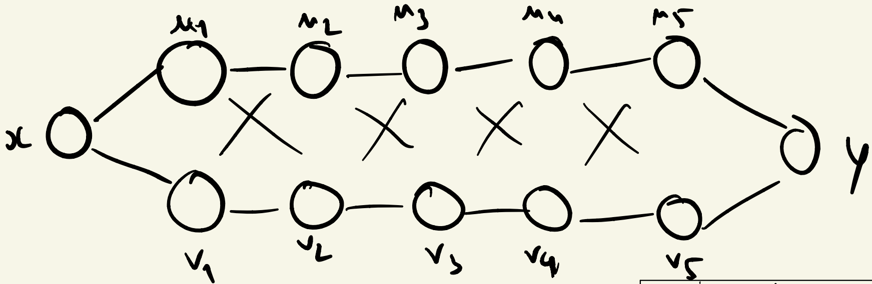

$$\frac{\partial y}{\partial x} = \frac{\partial(u_5 \cdot v_5)}{\partial x} = u_5 \frac{\partial v_5}{\partial x} + v_5 \frac{\partial u_5}{\partial x}$$

$$\frac{\partial u_5}{\partial x} = \frac{\partial(\sin(u_4+v_4))}{\partial x} = \frac{\partial(u_4+v_4)}{\partial x} \cos(u_4+v_4) = \frac{\partial(u_4+v_4)}{\partial x} v_5$$

$$\frac{\partial v_5}{\partial x} = \frac{\partial(\cos(u_4+v_4))}{\partial x} = \frac{\partial(u_4+v_4)}{\partial x}(-\sin(u_4+v_4)) = \frac{\partial(u_4+v_4)}{\partial x}(-u_5)$$

$$\frac{\partial u_4}{\partial x} = \frac{\partial(\sin(u_3+v_3))}{\partial x} = \frac{\partial(u_3+v_3)}{\partial x} \cos(u_3+v_3) = \frac{\partial(u_3+v_3)}{\partial x} v_4$$

$$\frac{\partial v_4}{\partial x} = \frac{\partial(u_3+v_3)}{\partial x}(-u_4) \qquad \frac{\partial u_3}{\partial x} = \frac{\partial(u_2+v_2)}{\partial x} v_3 \qquad \frac{\partial v_3}{\partial x} = -\frac{\partial(u_2+v_2)}{\partial x} u_3$$

$$\frac{\partial u_2}{\partial x} = \frac{\partial(u_1+v_1)}{\partial x} v_2 \qquad \frac{\partial v_2}{\partial x} = -\frac{\partial(u_1+v_1)}{\partial x} u_1 \qquad \frac{\partial u_1}{\partial x} = \frac{\partial(\sin(x))}{\partial x} = \cos(x) = v_1$$

$$\frac{\partial v_1}{\partial x} = \frac{\partial(\cos(x))}{\partial x} = -\sin(x) = -u_1$$

	valor	$\frac{\partial}{\partial x}$
x	1	1
u_1	$\sin(1) = 0.8414056$	$v_1 = 0.5403$
v_1	$\cos(1) = 0.5403023$	$-u_1 = -0.8414709$
u_2	$\sin(u_1+v_1) = 0.9829883$	$(v_1 - u_1)\, v_2 = -0.0565854$
v_2	$\cos(u_1+v_1) = 0.18788549$	$(u_1 - v_1)\, u_2 = 0.2958043$
u_3	$\sin(u_2+v_2) = 0.92078483$	$\left(\frac{\partial u_2}{\partial x} + \frac{\partial v_2}{\partial x}\right) v_3 = 0.093310789$
v_3	$\cos(u_2+v_2) = 0.39007095$	$\left(\frac{\partial u_2}{\partial x} + \frac{\partial v_2}{\partial x}\right) u_3 = -0.220265468$
u_4	$\sin(u_3+v_3) = 0.9664052$	$\left(\frac{\partial u_3}{\partial x} + \frac{\partial v_3}{\partial x}\right) v_4 = -0.032630289$
v_4	$\cos(u_3+v_3) = 0.257023130$	$\left(\frac{\partial u_3}{\partial x} + \frac{\partial v_3}{\partial x}\right) u_4 = 0.1226896686$
u_5	$\sin(u_4+v_4) = 0.940271984$	$\left(\frac{\partial u_4}{\partial x} + \frac{\partial v_4}{\partial x}\right) v_5 = 0.030658387$
v_5	$\cos(u_4+v_4) = 0.34042414$	$\left(\frac{\partial u_4}{\partial x} + \frac{\partial u_4}{\partial x}\right) u_5 = -0.0846803195$
y	$u_5 \cdot v_5 = 0.32009129$	$u_5 \cdot \frac{\partial v_5}{\partial x} + v_5 \cdot \frac{\partial u_5}{\partial x} = -0.069186$

$J(w,b) = \frac{1}{2}(y-z)^2 \quad , y = \sigma(wx+b)$

$J(w,b) = \frac{1}{2}(y-z)^2$

$\nabla J(w,b) = \left[\frac{\partial J}{\partial w} \quad \frac{\partial J}{\partial b} \right]$

$\frac{\partial J}{\partial w} = (y-z)\frac{\partial y}{\partial w} \qquad \frac{\partial J}{\partial b} = (y-z)\frac{\partial y}{\partial b}$

$$\frac{\partial y}{\partial w} = \frac{\partial(\sigma(wx+b))}{\partial w} = \frac{\partial(wx+b)}{\partial w}(y(1-y)) =$$

$$= xy(1-y)$$

$$\frac{\partial y}{\partial b} = \frac{\partial(wx+b)}{\partial b}(y(1-y)) = y(1-y)$$

$$\nabla J(w,b) = \begin{bmatrix} xy(y-z)(1-y) \\ y(y-z)(1-y) \end{bmatrix}$$

$$\|\nabla J\| = \sqrt{(xy(y-z)(1-y))^2 + (y(y-z)(1-y))^2} =$$

$$= \sqrt{(1+x^2)(y(y-z)(1-y))^2} =$$

$$= \sqrt{(1+x^2)(y-z)} \; y(1-y)$$

$$= \sqrt{1+x^2} \; \underbrace{[y(1-y)]} \sqrt{(y-z)^2}$$

$$y \in [0,1] \Rightarrow y(1-y) \in [0, \tfrac{1}{4}]$$

$$y(1-y) \leq \frac{1}{4}$$

$$\|\nabla J\| \leq \frac{1}{4}\sqrt{(1+x^2)}\,|y-z|$$

Given the boundness of $y \in [0,1]$:

$z > y \Rightarrow |y-z| \leq 1+|z|$, where $|y-z| = |z|$ when $y = \emptyset$ and $z > 1$

$z < y \Rightarrow |y-z| \leq 1+|z|$, where $|y-z| = 1+|z|$ when $y = 1$ and $z < \emptyset$

Thus, $1+|z| \geq |y-z|$, $\forall z \in \mathbb{R}$

$$\|\nabla J\| \leq \frac{1}{4}\sqrt{(1+x^2)}\,(1+|z|)$$

$$w_{n+1} = w_n - \eta \left.\frac{\partial J}{\partial w}\right|_{w=w_n, b=b_n} = w_n - \eta x_n y_n (y_n - z)(1-y_n)$$

$$b_{n+1} = b_n - \eta \left.\frac{\partial J}{\partial b}\right|_{w=w_n, b=b_n} = w_n - \eta y_n (y_n - z)(1-y_n)$$

$$\hat{y}_k = \frac{e^{a_k}}{\sum e^{a_j}} \Rightarrow \frac{e^{a_k+c}}{\sum e^{a_j+c}} =$$

$$= \frac{e^{a_k} e^c}{\sum e^{a_j} e^c} =$$

$$= \frac{e^{a_k} \cancel{e^c}}{\cancel{e^c} \sum e^{a_j}} = \hat{y}_k$$

$$\hat{y}_k = e^{a_k - \max_j(a_j)}$$

Given a MLP with linear activation $s(x) = x$ and L Layers, with $W^{(i)} + b^{(i)}$ being the weight matrix and bias vector at each layer i:

$a^{(i)} = s(W^{(i)} a^{(i-1)} + b^{(i)})$, $a^{(0)}$ is the input

As such, the output would be:

$$a^{(L)} = s(W^{(L)} a^{(L-1)} + b^{(L)})$$

Given that $s(x) = x$:

$$a^{(L)} = W^{(L)} a^{(L-1)} + b^{(L)} =$$

$$= W^{(L)}\left(W^{(L-1)}\left(W^{(L-2)}\left(\cdots\left(W^{(1)}a^{(0)} + b^{(1)}\right)\cdots\right) + b^{(L-2)}\right) + b^{(L-1)}\right) + b^{(L)}$$

As such:

$$W_{new} = W^{(L)} W^{(L-1)} W^{(L-2)} \cdots W^{(1)}$$

$$b_{new} = W^{(L)}\left(W^{(L-1)}\left(W^{(L-2)}\left(\cdots\left(W^{(2)}b^{(1)} + b^{(2)}\right)\cdots\right) + b^{(L-2)}\right) + b^{(L-1)}\right) + b^{(L)}$$

With this, we have:

$$a^{(L)} = W_{new} a^{(0)} + b_{new}$$

which represents a linear model.

$\mu = 0, \ \sigma \ll \emptyset$

For $s(x) = \frac{1}{1+e^{-x}}$

Using taylor's expansion when

$$s''(x) = \frac{e^{-x}}{(1+e^{-x})^2} - \frac{2e^{-x}}{(1+e^{-x})^3} = \frac{1}{2^2} - \frac{2}{2^3} =$$

$$= \frac{1}{4} - \frac{1}{4} = \emptyset$$

$$s(x) \approx s(0) + x s'(0) + \frac{s''(0)}{2} x^2 + \cdots =$$

$$\simeq \frac{1}{2} + \frac{x}{4} = \frac{x+2}{4}$$, behaves linearly

For $s(x) = \tanh(x)$

$s'(x) = 1 - \tanh^2(x)$

$s''(x) = -2\tanh(x)(1 - \tanh^2(x))$

$s(0) = \emptyset$

$s'(0) = 1$

$s''(0) = \emptyset$

$$s(x) = s(0) + x s'(0) + \frac{x^2}{2} s''(0) + \cdots$$

$$\approx \emptyset + x + \emptyset = x$$ behaves linearly

For $s(x) = \mathrm{relu}(x)$

$$s'(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$s''(x) = \begin{cases} \emptyset & \text{if } x > 0 \\ \emptyset & \text{if } x < 0 \end{cases}$$

$$s(x) = s(0) + x s'(0) + s''(0)\frac{x^2}{2} + \cdots$$

$$= \begin{cases} \emptyset + x + \emptyset + \cdots & \text{if } x > 0 \\ \emptyset + \emptyset + \emptyset + \cdots & \text{if } x < 0 \end{cases} = \left\{ \begin{matrix} x \text{ if } x > 0 \\ 0 \text{ if } x < 0 \end{matrix} \right\} \text{non linear}$$

For $s(x) = \mathrm{selu}(x)$

$$s'(x) = \begin{cases} \lambda & \text{if } x > 0 \\ \alpha e^x & \text{if } x < 0 \end{cases}$$

$$s''(x) = \begin{cases} 0 & \text{if } x > 0 \\ \alpha e^x & \text{if } x < 0 \end{cases}$$

$$s(x) = s(0) + x s'(0) + s''(0)\frac{x^2}{2} + \cdots$$

$$= \begin{cases} 0 + \lambda x + 0 + \cdots & \text{if } x > 0 \\ 0 + \alpha x + \alpha\frac{x^2}{2} + \cdots & \text{if } x < 0 \end{cases} = \left\{ \begin{matrix} \lambda x \text{ if } x > 0 \\ \alpha\left(x + \frac{x^2}{2} + \cdots\right) \text{ if } x < 0 \end{matrix} \right\} \text{non-linear}$$