Aufgabe 10

1. Linear multi-layer network

single-layer network $y(x) = b + w_x^e$

(a) a two-stage hypothesis class is.

$$y(x) = b_2 + w_2(b_1 + w_1 x)$$
$$= \tilde{w}_x + \tilde{b}$$

⇒ Here is the problem: we are still just using a normal linear classifer.
The apparent increase in complexity by concatenating does not give us any additional representational power

(b) $y(x) = b_2 + w_2 \cdot g(b_1 + w_1 x)$ ; $g(a) = \tanh(a)$

we use $g(a) = \tanh(a)$ (a non-linear activation function) at each node.
the two-layer network of the previous example, became equ. to have $y(x)_{(b)}$ and now is better.

with a single hidden layer is a universal function approximator, can approximate any function over the input argument (not very useful)

2. (a) $\sigma(a) = \frac{1}{1+e^{-a}}$.

$$\sigma'(a) = \left(\frac{1}{1+e^{-a}}\right)' = \frac{e^{-a}}{(1+e^{-a})^2} = \frac{1+e^{-a}-1}{(1+e^{-a})^2}$$

$$= \frac{1}{(1+e^{-a})}\left(1 - \frac{1}{(1+e^{-a})}\right)$$

$$= \sigma(a)\,(1-\sigma(a))$$

(b) ReLU. for $a \in R_{\neq 0}$.

$$f(x) = \begin{cases} 0 & \text{for } x<0 \\ x & \text{for } x \geq 0 \end{cases}, \quad f'(x) = \begin{cases} 0 & \text{for } x<0 \\ 1 & \text{for } x \geq 0 \end{cases}, \quad \text{Range} \in [0, \infty)$$

$\Rightarrow$ A function is differentiable at a particular point if there exist left derivatives and right derivatives and both the derivatives are equal at that point. ReLU is differentable at all point except 0. the left derivative at $z=0$ is 0. and right is 1.

(c) ELU

$$h(a) = \begin{cases} a & a \geq 0 \\ e^{a}-1 & a<0 \end{cases}, \quad h'(a) = \begin{cases} 1 & a \geq 0 \\ e^{a} & a<0 \end{cases}$$

when $a=0$. left $(a \geq 0)$ is 1.

same as right $(a<0)$ is $e^{0}=1$.

$\Rightarrow$ continuously differentiable in $a=0$.

(d)

$$\tanh(a) = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}}, \quad \sigma(x) = \frac{1}{1+e^{-x}}$$

so we have $1-\sigma(x) = \sigma(-x) \iff 1 - \frac{1}{1+e^{-x}} = \frac{1}{1+e^{x}}$

so. $\tanh(a) = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}} = \frac{e^{a}+e^{-a}-2e^{-a}}{e^{a}+e^{-a}} = 1 + \frac{-2e^{-a}}{e^{a}+e^{-a}} = 1 - \frac{2}{e^{2a}+1}$

$= 1 - 2\sigma(-2a) = 1 - 2(1-\sigma(2a)) = 1 - 2 + 2\sigma(2a) = 2\sigma(2a) - 1.$

3. weight space symmetries

(a) $\tanh(a) = \frac{\sinh(a)}{\cosh(a)} = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}}$

so. $-\tanh(a) = \frac{e^{-a}-e^{a}}{e^{a}+e^{-a}}$

$\left.\begin{array}{l}\tanh(-a) = \frac{e^{-a}-e^{a}}{e^{a}+e^{-a}}\end{array}\right\} \Rightarrow \tanh(-a) = -\tanh(a)$

(b) $f(x, w_1, b_1, w_2, b_2) = b_2 + w_2 \tanh(b_1 + w_1 x)$

$f(x, -w_1, -b_1, -w_2, b_2) = b_2 + (-w_2)\tanh(-b_1 - w_1 x)$

$= b_2 - w_2 \cdot (-\tanh(b_1 + w_1 x))$ (sei $-(b_1 + w_1 x) = a$ and use $\tanh(-a) = -\tanh(a)$)

$= b_2 + w_2 \tanh(b_1 + w_1 x)$

$\Rightarrow f(x, w_1, b_1, w_2, b_2) = f(x, -w_1, -b_1, -w_2, b_2)$

(c) while $P^{-1} = P^{T}$, $PP^{T} = I$

$\Rightarrow P^{-1}_{ij} = P^{T}_{ij} = P_{ij}$

(d) $Q = I - 2\frac{vv^{T}}{v^{T}v}$

sei $A_1 = \begin{pmatrix} 1 & 0 & \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ $5A_1 = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 5 \end{pmatrix} = 5 \cdot A_1 = 5 \cdot A_1$

(e)

① $f(x, w_1, b_1, w_2, b_2)$ same as $f(x, -w_1, -b_1, -w_2, b_2)$

② $f(x, w, b, w_2, -b_2)$ same as $f(+x, w, b, -w_2, b_2)$

~~same as $f(x, w, b$~~

③ $f(x, w_1, b_1, -w_2, -b_2)$ same as $f(x, -w_1, -b_1, -w_2, -b_2)$

same as ①.

for $\tanh(b_1 + w_1 x)$ we have $-(b_1, w_1, x)$ same as $(-b_1, -w_1, x)$

or $(-b_1, w_1, -x)$

for $f$ we have $(b_2, w_2, \tan(a))$ same as $(-b_2, -w_2, \tan(a))$ and $(-b_2, w_2, -\tan(a))$

$(-b_2, w_2, \tan(a))$ same as $(b_2, -w_2, \tan(a))$ and $(b_2, w_2, -\tan(a))$

zusammen $2 \cdot 3! \cdot 3! = 72$.