2.1. Examples and non-examples of multi-layer perceptrons $\boxed{1}$

Are the following functions f multi-layer perceptrons with the specified activation function ρ?

(a)	$f(x) = \max\{\lvert x\rvert, 0\}$,	$\rho(x) = \max\{x, 0\}$,
(b)	$f(x) = x^2$,	$\rho(x) = x$,
(c)	$f(x) = x$	$\rho(x) = \max\{x, 0\}$.

2.2. Function spaces of multi-layer perceptrons $\boxed{1}$

Every multi-layer perceptron with fixed input dimension d, fixed output dimension 1, fixed number of layers L, and fixed activation function $\rho : \mathbb{R} \to \mathbb{R}$ is a function $f : \mathbb{R}^d \to \mathbb{R}$. Do these functions f form a vector space?

2.3. Parameter-dependence of multi-layer perceptrons $\boxed{1}$

Consider the single-layer perceptron

$$f(x) = w\rho(ax + b),$$

where x, w, a, b are some real numbers, and $\rho(x) = \max\{0, x\}$ is the rectified linear unit. Draw f as a function of x, then as a function of w, then as a function of a, and then as a function of b, each time setting the other variables to some constant of your choice.

2.4. Defining and plotting multi-layer perceptrons $\boxed{1}$

Follow the instructions in the Jupyter notebook.

2.5. Random initialization of network coefficients $\boxed{1}$

Follow the instructions in the Jupyter notebook.

2.6. Learning a sine function $\boxed{1}$

Follow the instructions in the Jupyter notebook.

You can achieve a total of $\boxed{6}$ points for this problem set.

Author: Zhang Tianyu.

Collaborator: Wang Yiran

2.1. **Examples and non-examples of multi-layer perceptrons** $\boxed{1}$

Are the following functions f multi-layer perceptrons with the specified activation function ρ?

(a)	$f(x) = \max\{\lvert x\rvert, 0\}$,	$\rho(x) = \max\{x, 0\}$,
(b)	$f(x) = x^2$,	$\rho(x) = x$,
(c)	$f(x) = x$	$\rho(x) = \max\{x, 0\}$.

A multilayer perceptron with $L \in \mathbb{Z}^+$ layers & activation function could be represented as: $f = T_L \circ \rho \circ T_{L-1} \circ \cdots \circ \rho \circ T_1$, where all Ts are affine function

(a). Yes $f(x) = T_2 \circ \rho \circ T_1(x)$, where:

$T_1(x) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} x$, $T_2(x) = (1 \quad 1)x$. proved with construction

(b). No. since $\rho(x)$ is linear, and all Ts are required to be linear (affine)

$f(x)$ is a linear combination of linear functions, which is also linear.

It could not be quadratic ($f(x) = x^2$)

(c). Yes. $f(x) = T_2 \circ \rho \circ T_1(x)$, where:

$T_1(x) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} x$, $T_2(x) = (1 \quad -1)x$. proved with construction.

2.2. **Function spaces of multi-layer perceptrons** $\boxed{1}$

Every multi-layer perceptron with fixed input dimension d, fixed output dimension 1, fixed number of layers L, and fixed activation function $\rho : \mathbb{R} \to \mathbb{R}$ is a function $f : \mathbb{R}^d \to \mathbb{R}$. Do these functions f form a vector space?

We make use of the fact that the set of real valued function defined on some non-empty set S (in this case $S = \mathbb{R}^d$) $F: S \to \mathbb{R}$ is a vector space. let the set of aborementioned functions as V

① closure under addition:

let $f_1, f_2 \in V$, $f_1 = T_L^1 \circ \rho \circ \cdots \circ \rho \circ T_1^1$, $f_2 = T_L^2 \circ \rho \circ \cdots \circ \rho \circ T_1^2$

, where all T^1's and T^2's are in matrix format: $T_i^j(x) = A_i^j x + b_i^j$

define $g = T_L \circ \rho \circ \cdots \circ \rho \circ T_1$, where for $k = 1, 2, \cdots, L-1$.

$T_k = \begin{pmatrix} A_k^1 & 0 \\ 0 & A_k^2 \end{pmatrix} \cdot \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} A_k^1 \cdot x^1 \\ A_k^2 \cdot x^2 \end{pmatrix}$, $T_L = (1 \quad 1)\begin{pmatrix} A_L^1 & 0 \\ 0 & A_L^2 \end{pmatrix} \cdot \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}$

So that $g(x) = f_1(x) + f_2(x) \in V$.

② Closure under scalar multiplication

let $f \in V$: $f = T_L \circ \rho \circ \cdots \circ \rho \circ T_1$, constant be c.

define $g = T_L' \circ \rho \circ \cdots \circ \rho \circ T_1'$, for $i = 1, 2, \cdots, L$. let $T_i = Ax + b$

then set $T_i' = Ax + cb$, so that $g(x) = c \cdot f(x) \in V$.

Hence, V is a subspace of $F: \mathbb{R}^d \to \mathbb{R}$, thus a vector space

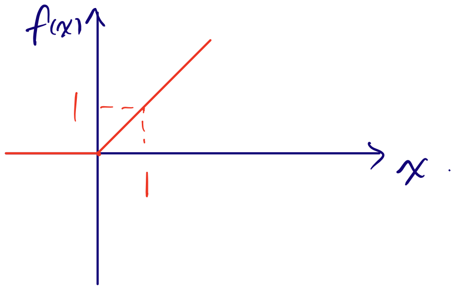
2.3. **Parameter-dependence of multi-layer perceptrons** $\boxed{1}$

Consider the single-layer perceptron

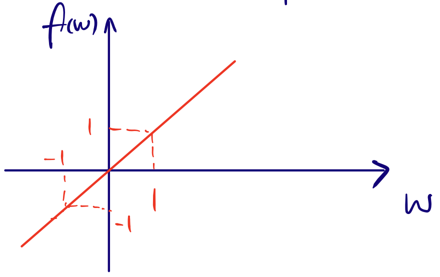
$$f(x) = w\rho(ax + b),$$

where x, w, a, b are some real numbers, and $\rho(x) = \max\{0, x\}$ is the rectified linear unit. Draw f as a function of x, then as a function of w, then as a function of a, and then as a function of b, each time setting the other variables to some constant of your choice.

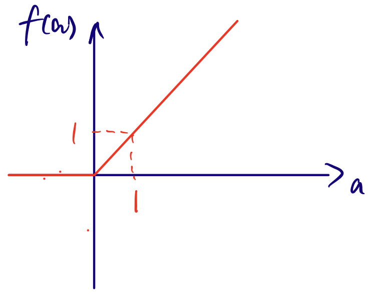
f as a function of x: let $w=1$, $b=0$, $a=1$



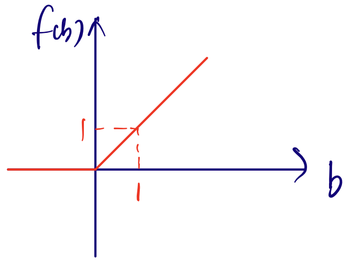
f as a function of w: let $a=1$, $b=1$, $x=0$



f as a funtion of a, let $w=1$, $x=1$, $b=0$



f as a funtion of b. let $x=0$, $a=0$, $w=1$



2.4. **Defining and plotting multi-layer perceptrons** $\boxed{1}$

Follow the instructions in the Jupyter notebook.

2.5. **Random initialization of network coefficients** $\boxed{1}$

Follow the instructions in the Jupyter notebook.

2.6. **Learning a sine function** $\boxed{1}$

Follow the instructions in the Jupyter notebook.

github.com/zty022/MH3520_Mathematics_of_Deep_Learning/blob/main/02_problemset.ipynb