machine learning (ML):

data. → neural network: computer vision. natural language processing

scientific computing:

solve differential equations in physics and engineering

heat equation:
$$
\begin{cases}
\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2}. & 0 < x < 1. \quad t > 0 \\
u(x. 0) = \sin(2\pi x) \\
u(0. t) = 0. \quad u(1. t) = 0. & t > 0
\end{cases}
$$

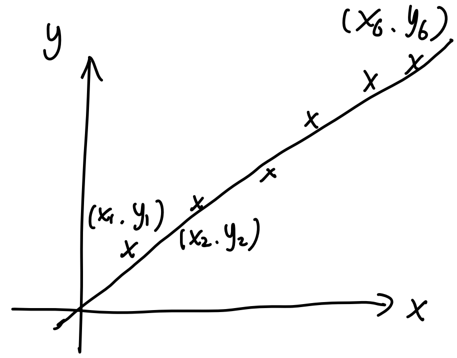
scientific machine learning:

use ML to solve problems in scientific computing.

Linear regression:

Given $\{(x_i, y_i)\}_{i=1}^{N}$. find linear relationship between y and x:

$$y = mx + b$$



find out m and b such that

$$L = \sum_{i=1}^{n} (mx_i + b - y_i)^2 \quad \text{is minimized}$$

solution: $$L(m, b) = \sum_{i=1}^{n} (mx_i + b - y_i)^2$$

$$\frac{\partial L}{\partial m} = 0, \quad \frac{\partial L}{\partial b} = 0$$

$$\frac{\partial L}{\partial m} = \sum_{i=1}^{n} 2(mx_i + b - y_i) \cdot x_i = 2 \cdot \left(m \sum_{i=1}^{n} x_i^2 + b \cdot \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i\right)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} 2(mx_i + b - y_i) = 2 \cdot \left(m \sum_{i=1}^{n} x_i + nb - \sum_{i=1}^{n} y_i\right)$$

$$\begin{cases} 2\left(m \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i\right) = 0 \\ 2\left(m \sum_{i=1}^{n} x_i + nb - \sum_{i=1}^{n} y_i\right) = 0 \end{cases} \Rightarrow \boxed{b = \frac{\sum_{i=1}^{n} y_i - m \sum_{i=1}^{n} x_i}{n}}$$

$$m \sum_{i=1}^{n} x_i^2 + \frac{\sum_{i=1}^{n} y_i - m \sum_{i=1}^{n} x_i}{n} \cdot \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i = 0$$

$$n \cdot \left(m \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i\right) + \left(\sum_{i=1}^{n} y_i - m \sum_{i=1}^{n} x_i\right) \cdot \sum_{i=1}^{n} x_i = 0$$

$$m\cdot\left(n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\right)=n\cdot\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\cdot\sum_{i=1}^{n}y_i$$

$$\boxed{m=\frac{n\sum_{i=1}^{n}x_iy_i-\left(\sum_{i=1}^{n}x_i\right)\cdot\left(\sum_{i=1}^{n}y_i\right)}{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}}$$

calculus: minimize of function $L(m.b)$

① $\frac{\partial L}{\partial m}=0.\quad \frac{\partial L}{\partial b}=0$

② Hessian matrix is positive definite

$$H=\begin{pmatrix}\frac{\partial^2 L}{\partial m^2} & \frac{\partial^2 L}{\partial m\partial b}\\ \frac{\partial^2 L}{\partial m\partial b} & \frac{\partial^2 L}{\partial b^2}\end{pmatrix}$$ (homework)

Generalization: Given $\{(x_1^{(i)}.\ x_2^{(i)}.\ y^{(i)})\}_{i=1}^{n}$. find out linear relationship between y and $x_1.\ x_2$:

$$y=m_1x_1+m_2x_2+b$$

find out $m_1.\ m_2.\ b$ such that

$$L(m_1.\ m_2.\ b)=\sum_{i=1}^{n}\left(m_1x_1^{(i)}+m_2x_2^{(i)}+b-y^{(i)}\right)^2$$

is minimized.

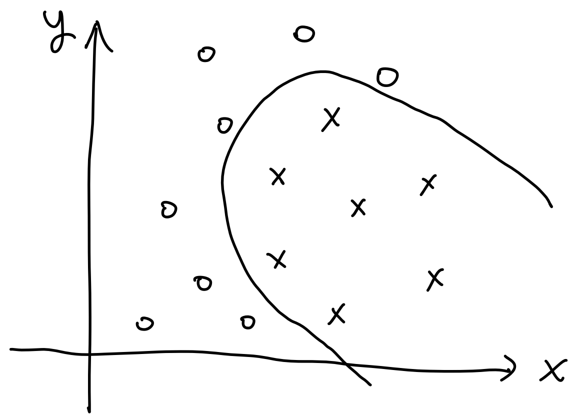
Solution: homework

classification problem:

find out a mapping (function)

that take any points (x, y) in $\mathbb{R}^2$

and returns either A or B



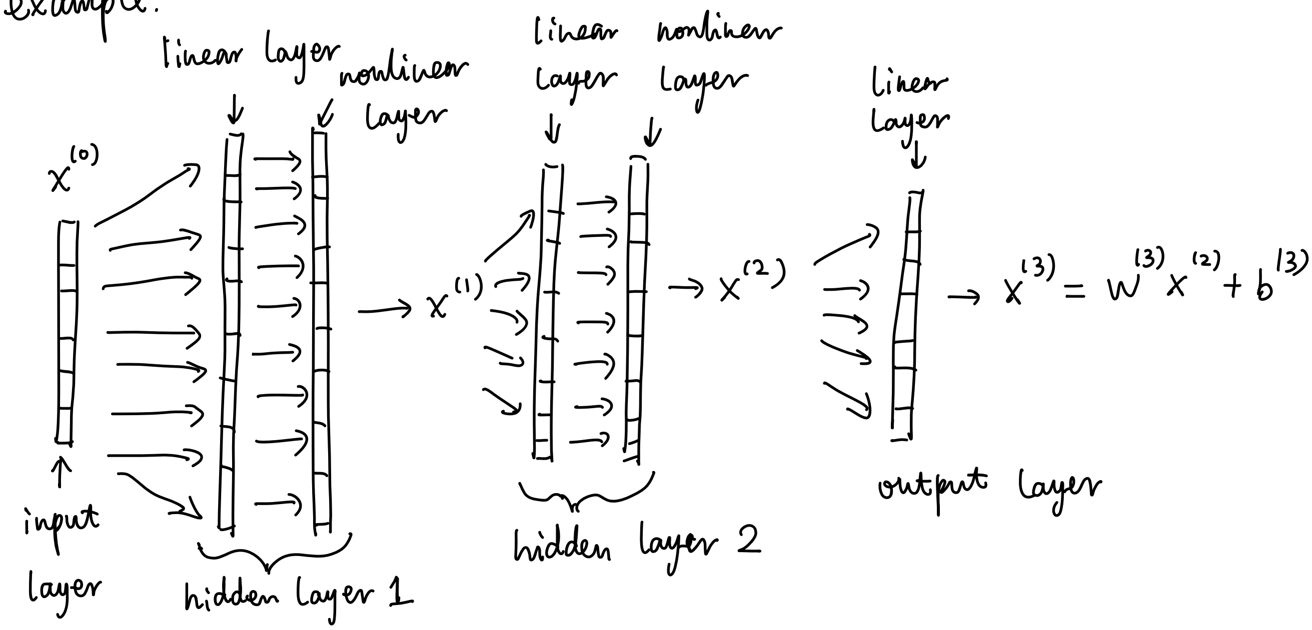
circle: category A

cross: category B

Feedforward neural network (FNN):

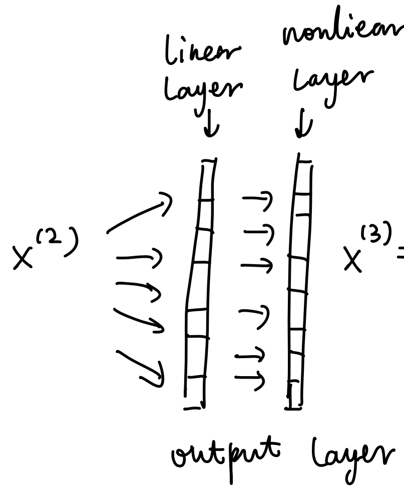
(or mutilayer perceptron (MLP))

example:



$$\sigma\left(W^{(1)} x^{(0)} + b^{(1)}\right) = x^{(1)} \qquad \sigma\left(W^{(2)} x^{(1)} + b^{(2)}\right) = x^{(2)}$$

activation function　　weight　　bias



$$x^{(3)} = \sigma\left(w^{(3)} x^{(2)} + b^{(3)}\right)$$

$\uparrow$ postprocessing function

$$x^{(0)}$$
$$\downarrow$$
$$x^{(1)} = \sigma\left(w^{(1)} x^{(0)} + b^{(1)}\right)$$
$$\downarrow$$
$$x^{(2)} = \sigma\left(w^{(2)} x^{(1)} + b^{(2)}\right)$$
$$\downarrow$$
$$x^{(3)} = w^{(3)} x^{(2)} + b^{(3)}$$

$$\Rightarrow \underbrace{x^{(3)}}_{\text{output}} = w^{(3)} \cdot \sigma\left(w^{(2)} \sigma\left(w^{(1)} \underbrace{x^{(0)}}_{\text{input}} + b^{(1)}\right) + b^{(2)}\right) + b^{(3)}$$

Remarks:

(1) If we take $\sigma(x) = x$, then

$$x^{(3)} = w^{(3)}\left(w^{(2)}\left(w^{(1)} x^{(0)} + b^{(1)}\right) + b^{(2)}\right) + b^{(3)} \rightarrow \text{linear function}$$

(2) activation functions:

(*) sigmoid function: $\underline{\underline{\sigma(x)}} = \frac{1}{1+e^{-x}}$ $\leftarrow$ scalar x

(*) ReLU function: $\sigma(x) = \begin{cases} x. & \text{if } x \geq 0 \\ 0. & \text{if } x < 0. \end{cases}$

(*) Tanh function: $\sigma(x) = \tanh(x)$.

(3) $x^{(3)} = W^{(3)} \cdot \sigma\left(W^{(2)} \cdot \sigma\left(W^{(1)} x^{(0)} + b^{(1)}\right) + b^{(2)}\right) + b^{(3)}$

take input $x^{(0)} \in \mathbb{R}^2$.

$W^{(1)} \in \mathbb{R}^{3\times 2}$. $b^{(1)} \in \mathbb{R}^3$ $\longrightarrow$ number of neurons is 3 in the hidden layer 1

$$\sigma(\underbrace{\overset{\overset{3\times 2}{\uparrow}}{W^{(1)}} \overset{\overset{2}{\uparrow}}{x^{(0)}} + \overset{\overset{3}{\uparrow}}{b^{(1)}}}_{3})$$

$\sigma(z)$. $z = (z_1. z_2. z_3)$

$\sigma(z) = (\sigma(z_1). \ \sigma(z_2). \ \sigma(z_3))$

$W^{(2)} \in \mathbb{R}^{5\times 3}$. $b^{(2)} \in \mathbb{R}^5$. $\longrightarrow$ number of neurons is 5 in the hidden layer 2.

$W^{(3)} \in \mathbb{R}^{2\times 5}$. $b^{(3)} \in \mathbb{R}^2$.

$x^{(3)} \in \mathbb{R}^2$.

Now we have a FNN $f: \mathbb{R}^2 \to \mathbb{R}^2$.

$$f(x) = \sigma\left(W^{(3)} \cdot \sigma\left(W^{(2)} \cdot \sigma\left(W^{(1)} x + b^{(1)}\right) + b^{(2)}\right) + b^{(3)}\right)$$

Find parameters $\underbrace{W^{(1)}, W^{(2)}, W^{(3)}, b^{(1)}, b^{(2)}, b^{(3)}}$ such that

$f(x)$ is close to $(1, 0)$, if $x \in A$

$f(x)$ is close to $(0, 1)$, if $x \in B$

Then, given a new point $x \in \mathbb{R}^2$, we can classify it

if $f_1(x) > f_2(x)$, then we say that $x \in A$

if $f_1(x) < f_2(x)$, then we say that $x \in B$