XOR-problem for One Hidden Layer

(A problem posted by 戴晨洋)

Sign function

For a vector $u = (u_1, u_2, \dots, u_d) \in \mathbb{R}^d$, define the vector $\mathrm{sgn}(u) \in \mathbb{R}^d$ whose i-th position is

$$\mathrm{sgn}(u)_i = \begin{cases} 1, & \text{if } u_i > 0; \\ 0, & \text{if } u_i \leq 0, \end{cases}$$

and define $u' = (1, u_1, \dots, u_d) \in \mathbb{R}^{d+1}$.

Example

$$\mathrm{sgn}(2, -3, 0, 1, 2)' = (1, 1, 0, 0, 1, 1).$$

Hidden layer is necessary in solving XOR-problem

Proposition

There is no column vector $\Theta = (\theta_0, \theta_1, \theta_2)^T$ *that satisfies*

$$\operatorname{sgn}(x'\Theta) = \operatorname{sgn}(\theta_0 + x_1\theta_1 + x_2\theta_2) = \begin{cases} 0, & \text{if } x_1 = x_2; \\ 1, & \text{if } x_1 \neq x_2, \end{cases}$$

where $x' = (1, x_1, x_2)$ *is binary vector.*

Proof.

If such triple $(\theta_0, \theta_1, \theta_2)$ exists, then applying the four cases $(x_1, x_2) = (0,0), (1,1), (1,0), (0,1)$ into the above equation, we have

$$-\theta_0 > 0, \quad -\theta_0 - \theta_1 - \theta_2 > 0, \quad \theta_0 + \theta_1 > 0, \quad \theta_0 + \theta_2 > 0,$$

but summing these four inequalities will yield the contradiction $0 > 0$. □

XOR-problem for one hidden layer

Proposition

For any positive integer d there exists a positive integer n, a $(d+1)\times n$ matrix Θ and $(n+1)\times 1$ matrix Λ such that

$$\operatorname{sgn}(\operatorname{sgn}(x'\Theta)'\Lambda) = \begin{cases} 1, & \text{if } x_1 + \cdots + x_d \text{ is odd;} \\ 0, & \text{if } x_1 + \cdots + x_d \text{ is even,} \end{cases}$$

where $x = (x_1, \ldots, x_d)$ is a binary vector.

Proof. Let

$$n = d,\ \Theta = \begin{pmatrix} \frac{-1}{2} & \frac{-3}{2} & \cdots & -\frac{(2d+1)}{2} \\ 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix},\ \Lambda = \begin{pmatrix} -\frac{1}{2} \\ 1 \\ -1 \\ \vdots \\ (-1)^{d+1} \end{pmatrix}.$$

Continue the proof

Then the i-th position of $\mathrm{sgn}(x'\Theta)$ is

$$\begin{aligned}(\mathrm{sgn}(x'\Theta))_i =& \mathrm{sgn}(x_1 + x_2 + \cdots + x_d - i + \frac{1}{2}) \\ =& \begin{cases} 1, & \text{if } i < x_1 + x_2 + \cdots + x_d; \\ 0, & \text{if } i > x_1 + x_2 + \cdots + x_d, \end{cases}\end{aligned}$$

and

$$\mathrm{sgn}(x'\Theta)'\Lambda = -\frac{1}{2} + \begin{cases} 1, & \text{if } x_1 + x_2 + \cdots + x_d \text{ is odd}, \\ 0, & \text{if } x_1 + x_2 + \cdots + x_d \text{ is even}. \end{cases}$$

Hence

$$\mathrm{sgn}(\mathrm{sgn}(x'\Theta)'\Lambda) = \begin{cases} 1, & \text{if } x_1 + \cdots + x_d \text{ is odd}; \\ 0, & \text{if } x_1 + \cdots + x_d \text{ is even}, \end{cases}$$

□

Remark

Let $n(d)$ denote the minimum n such that the $(d+1) \times n$ matrix Θ in the above proposition exists. The proof of the above proposition implies that $n(d) \leq d$.

Problem. Determine $n(d)$.