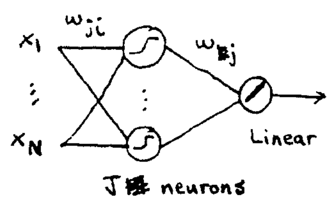
Approximation Theory – Universal Approximation – Two-layer sigmoid network

23 May 1990 Stone – Weierstrass Theorem – Networks

Neil E Cotter

ex: Two-Layer sigmoid network fails to satisfy S-W thm:



$$f(x_1, \ldots, x_N) = \sum_{j=1}^{J} w_j \frac{1}{1+e^{-(\sum_{i=1}^{N} w_{ji} x_i + w_{j0})}}$$

$$\mathcal{F} = \left\{ f(x_1, \ldots, x_N) = \sum_{j=1}^{J} w_j \frac{1}{1+e^{-(\sum_{i=1}^{N} w_{ji} x_i + w_{j0})}} \qquad w_j, w_{ji}, w_{j0} \in \mathbb{R} \right\}$$

I) Identity $\quad f(\vec{x}) = 2 \dfrac{1}{1+e^{-0}} = 1 \in \mathcal{F}$ ✓ $\quad w_{j=1} = 1$, $w_{j\neq 1} = 0$

all 1^{st} layer weights $= 0$

II) Separability $\quad \overset{\vec{x}_1}{(x_1, \ldots, x_N)_1} \neq \overset{\vec{x}_2}{(x_1, \ldots, x_N)_2}$

then some entry is different $\quad x_{i1} \neq x_{i2}$

Then $f(\vec{x}) = \dfrac{1}{1+e^{-x_i}} \in \mathcal{F}$ separates $\vec{x}_1, \vec{x}_2$ since is *monotonic in* x_i entry of vectors.

Note: For separability we need at least one func in $\mathcal{F}$ *for each dim of space. Here, we have a different* $f(\vec{x})$ *for each* x_i.

III) Closure $\quad f(x) = \displaystyle\sum_{j=1}^{J} w_j \frac{1}{1+e^{-(\sum_i w_{ji} x_i + w_{j0})}} \qquad g(x) = \sum_{k=1}^{K} w_k \frac{1}{1+e^{-(\sum_i w_{ki} x_i + w_{k0})}}$

Additive $\quad af(x) + bg(x) = \displaystyle\sum_{j}^{J} a w_j \frac{1}{1+e^{-\sum_i w_{ji} x_i + w_{j0}}} + \sum_{k}^{K} b w_k \frac{1}{1+e^{-\sum_i w_{ki} x_i + w_{k0}}}$

can be written as single Σ of J+K terms $\in \mathcal{F}$ ✓

Multiplicative ~~fails~~: we terms of form $\dfrac{1}{1+e^{-\sum_i \sim}} \dfrac{1}{1+e^{-\sum_i \sim}}$

$$= \frac{1}{1+e^{-\sum_i} + e^{-\sum_i} + e^{-(\sum_i \sim) - (\sum_i \sim)}}$$

cannot be written as sum of terms of form $\dfrac{1}{1+e^{-\Sigma \sim}}$.

Note: Even though $fg \notin \mathcal{F}$ & S-W fails, we can approx fg quite well and 2-layer net can approx. arbitrary funcs satisfactorily.