Foundations of Machine Learning
AI2000 and AI5000

FoML-23
Neural Networks - UAT

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

So far in FoML

- Intro to ML and Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Supervised learning
    a. Linear Regression with basis functions (regularization, model selection)
    b. Bias-Variance Decomposition (Bayesian Regression)
    c. Decision Theory - three broad classification strategies
        - Probabilistic Generative Models - Continuous & discrete data
        - (Linear) Discriminant Functions - least squares solution, Perceptron
        - Probabilistic Discriminative Models - Logistic Regression

Neural Networks - II

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

Neural Networks are universal approximators

Universal Approximation Theorem

- Can represent any continuous function ($f : R^m \rightarrow R^n$) on a compact area, to any desired approximation ($|g(x) - f(x)| < \varepsilon$) with a linear combination of sigmoid neurons

Universal Approximation Theorem

- In other words, NN with a single hidden layer can be used to approximate any continuous function to a desired precision

Universal Approximation Theorem

Math. Control Signals Systems (1989) 2: 303–314

Mathematics of Control, Signals, and Systems

Approximation by Superpositions of a Sigmoidal Function*

G. Cybenko†

Neural Networks, Vol. 4, pp. 251–257, 1991
Printed in the USA.

0893-6080/91 $3.00 + .00

ORIGINAL CONTRIBUTION

Approximation Capabilities of Multilayer Feedforward Networks

KURT HORNIK
Technische Universität Wien, Vienna, Austria

Universal Approximation Theorem

Theorem 0.1 (UAT, [Cyb89, Hor91]). *Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a non-constant, bounded, and continuous function. Let I_m denote the m-dimensional unit hypercube $[0,1]^m$. The space of real-valued continuous functions on I_m is denoted by $C(I_m)$. Then, given any $\varepsilon > 0$ and any function $f \in C(I_m)$, there exist an integer N, real constants $v_i, b_i \in \mathbb{R}$ and real vectors $w_i \in \mathbb{R}^m$ for $i = 1, \ldots, N$, such that we may define:*

$$F(\boldsymbol{x}) = \sum_{i=1}^{N} v_i \sigma\left(\boldsymbol{w}_i^T \boldsymbol{x} + b_i\right) = \boldsymbol{v}^\intercal \sigma\left(\boldsymbol{W}^\intercal \boldsymbol{x} + \boldsymbol{b}\right)$$

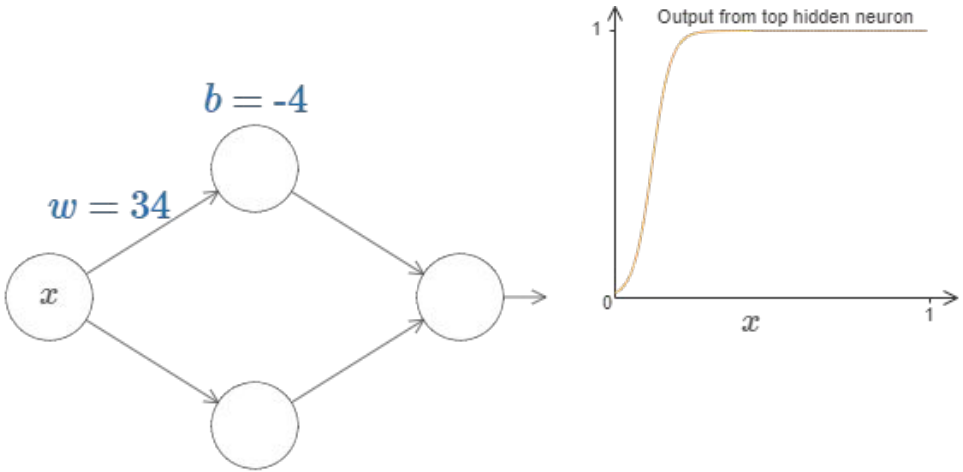
as an approximate realization of the function f; that is,

$$|F(\boldsymbol{x}) - f(\boldsymbol{x})| < \varepsilon$$

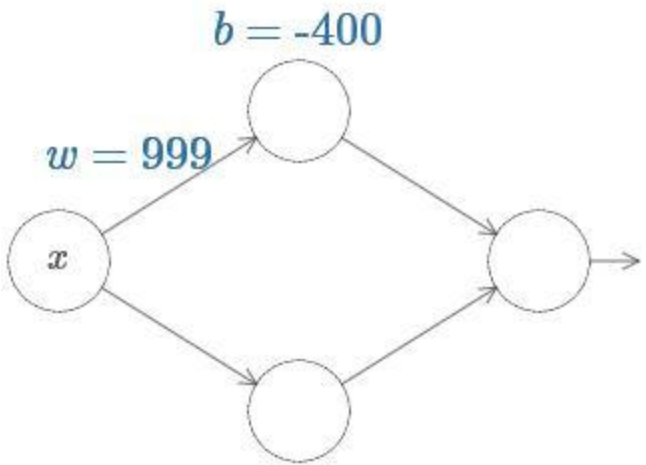
for all $\boldsymbol{x} \in I_m$.

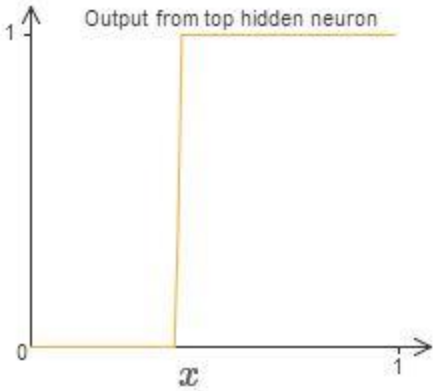
Visual proof with one i/p & one o/p and Sigmoid activation

Universality with one i/p & one o/p

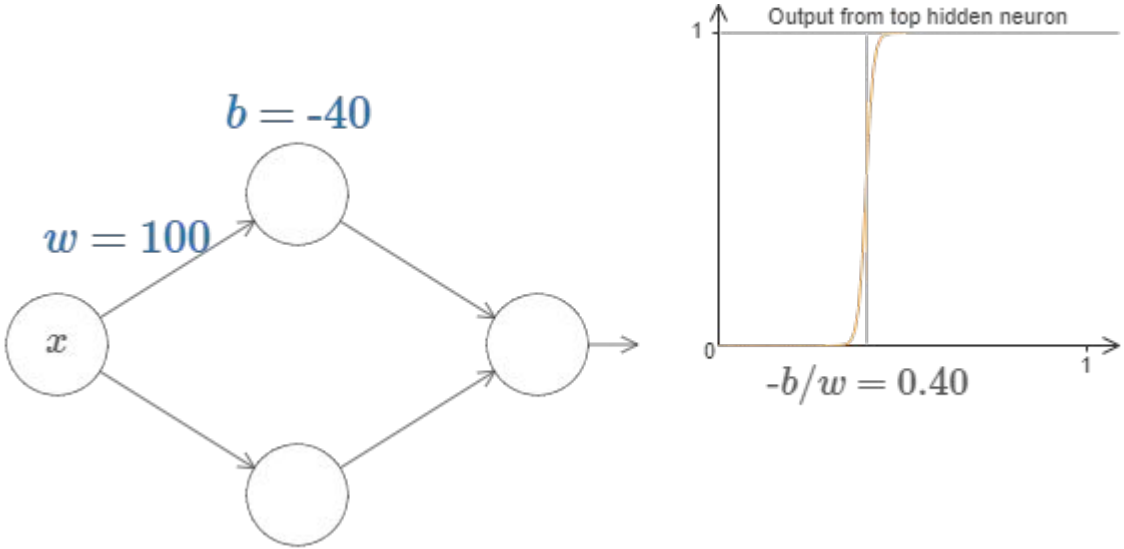


Universality with one i/p & one o/p

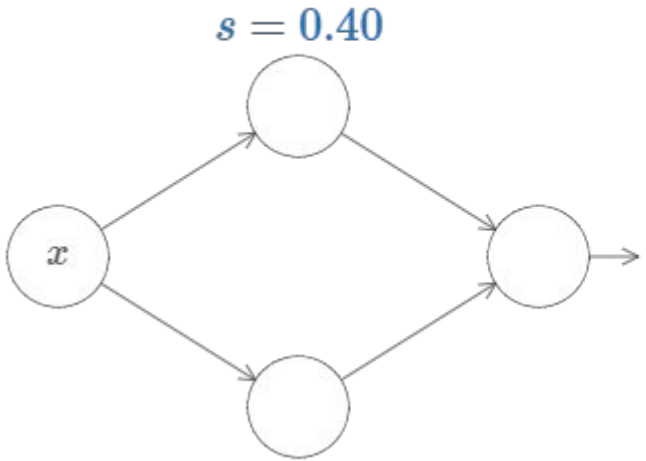


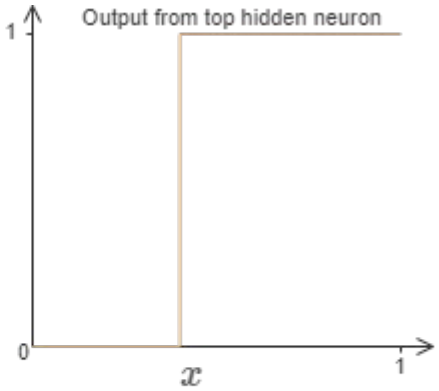


Universality with one i/p & one o/p

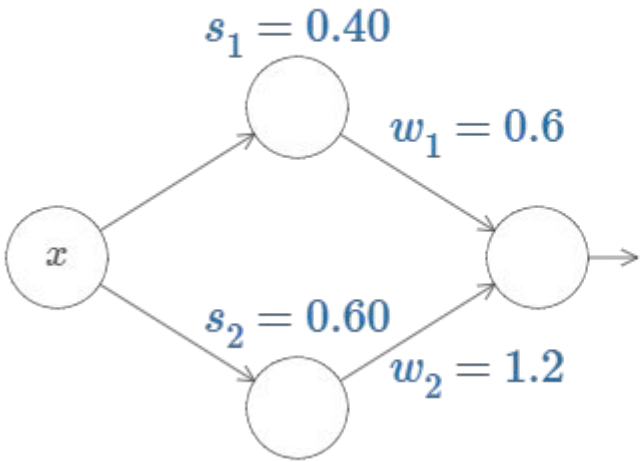


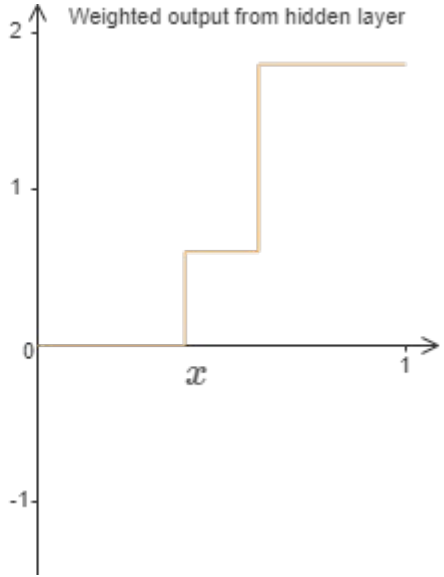
Universality with one i/p & one o/p



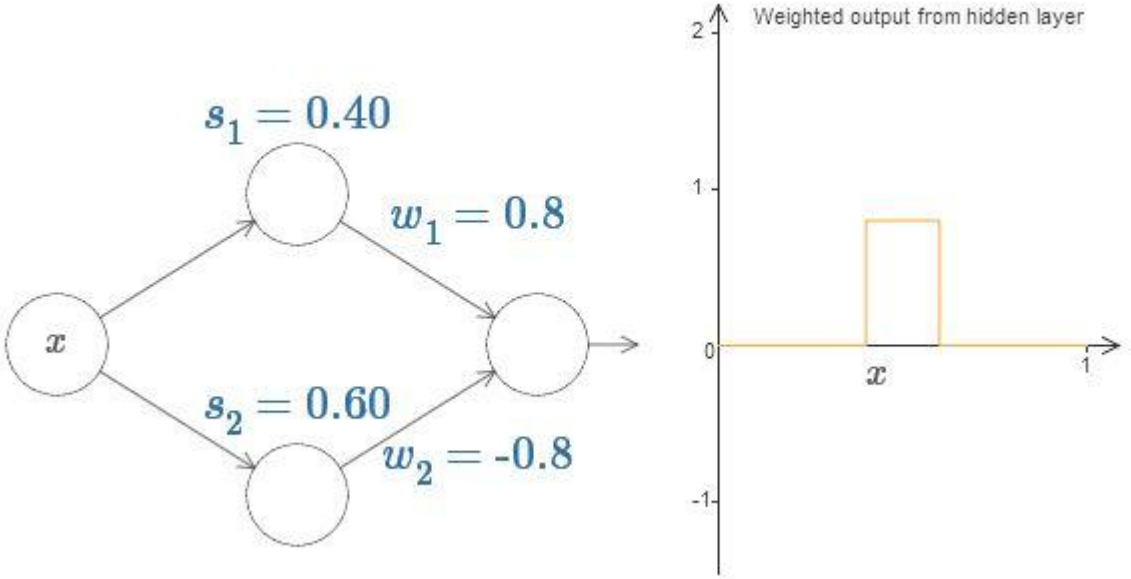


Universality with one i/p & one o/p

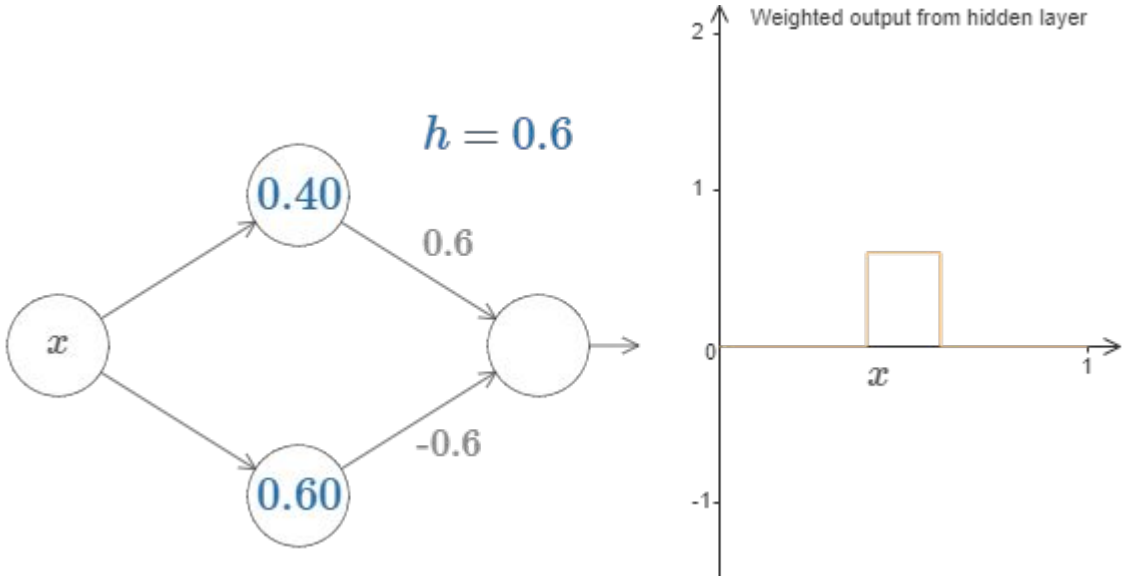




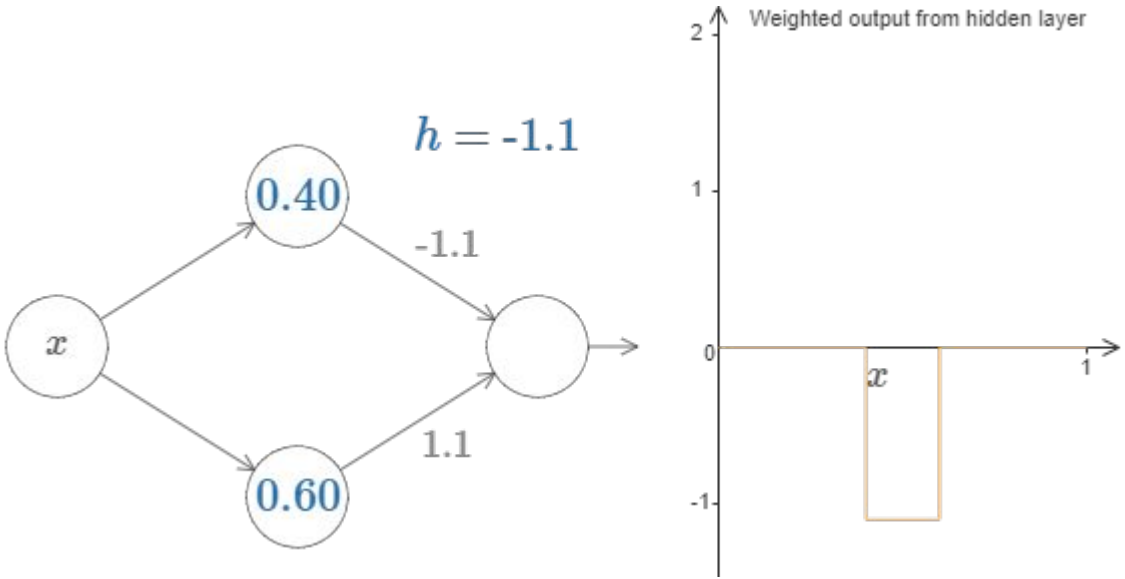
Universality with one i/p & one o/p



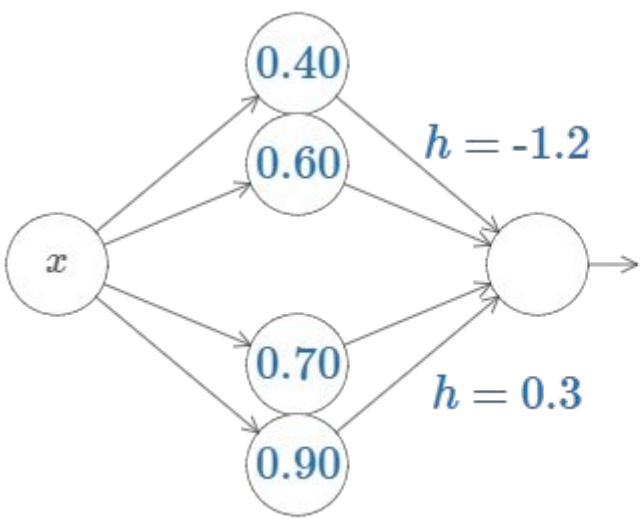
Universality with one i/p & one o/p

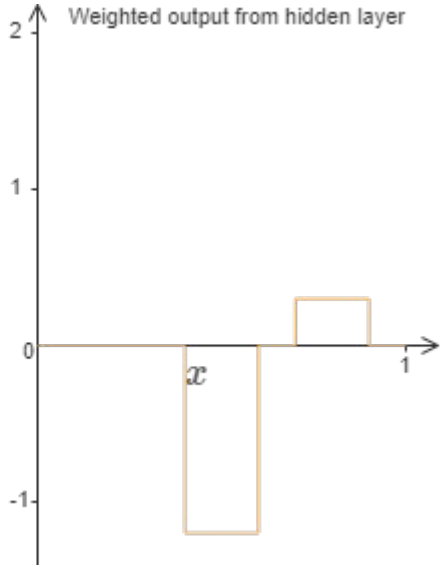


Universality with one i/p & one o/p

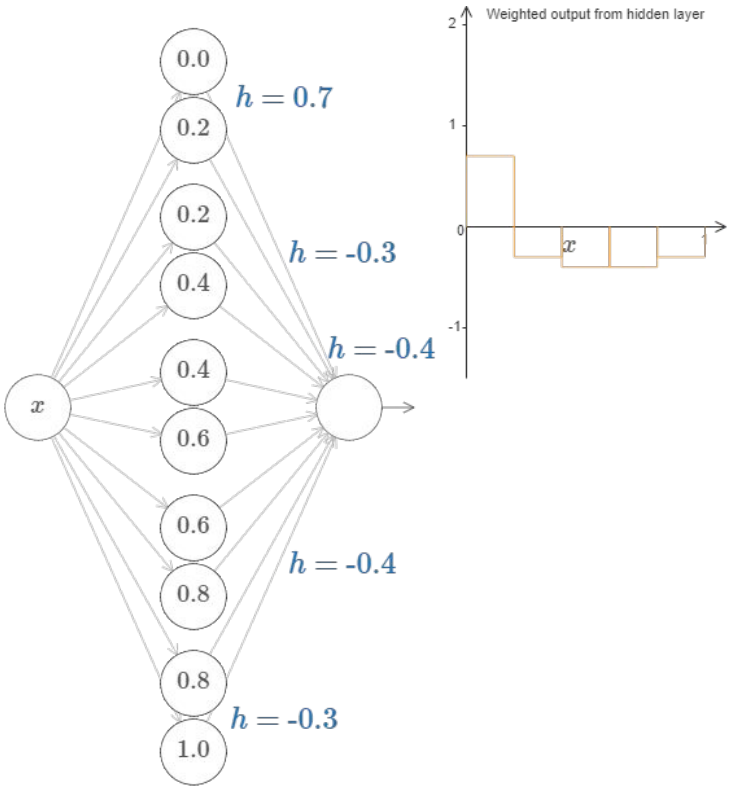


Universality with one i/p & one o/p

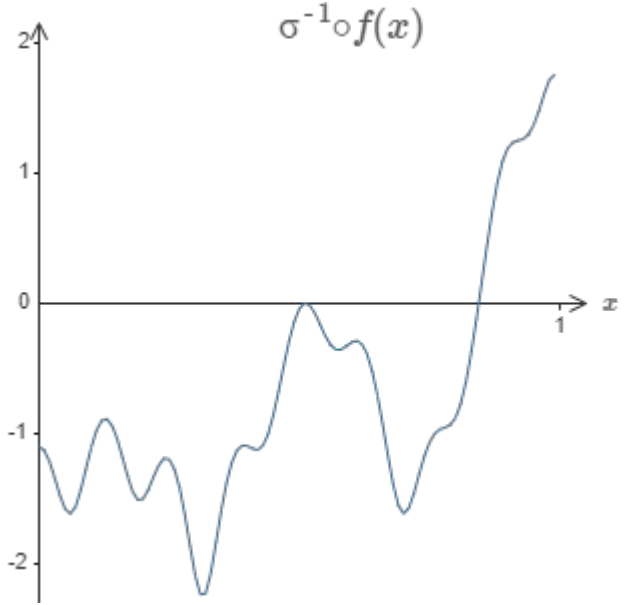




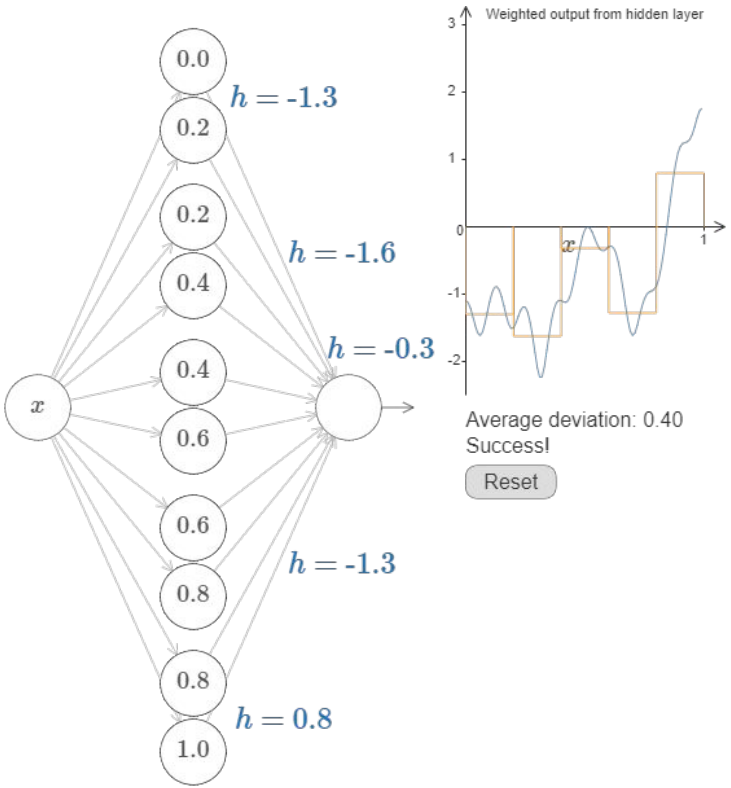
Universality with one i/p & one o/p



Universality with one i/p & one o/p



Universality with one i/p & one o/p

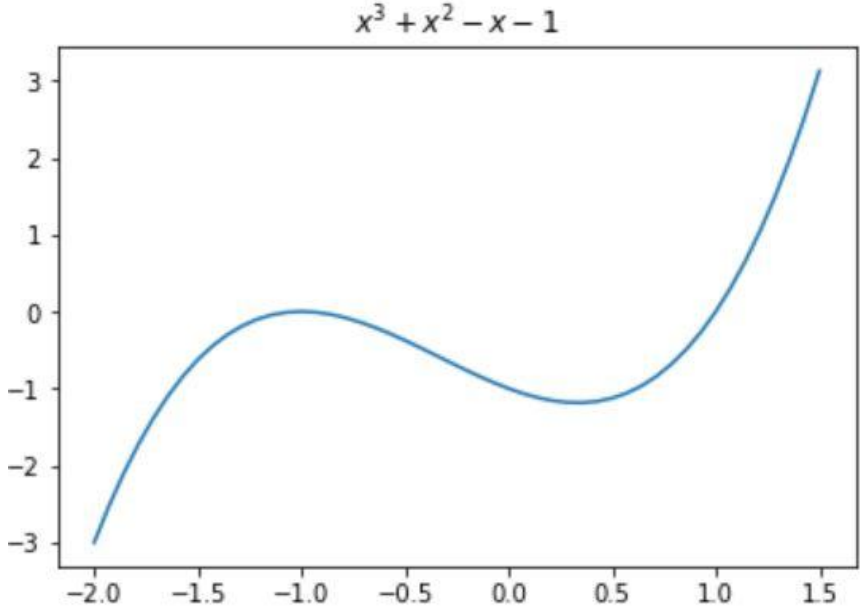


భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

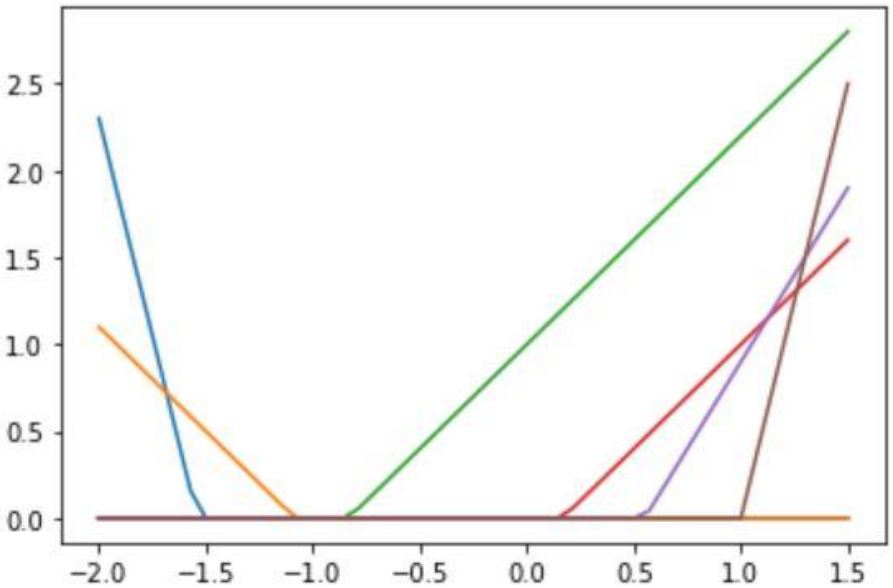
With ReLU activation

Universality with one i/p & one o/p



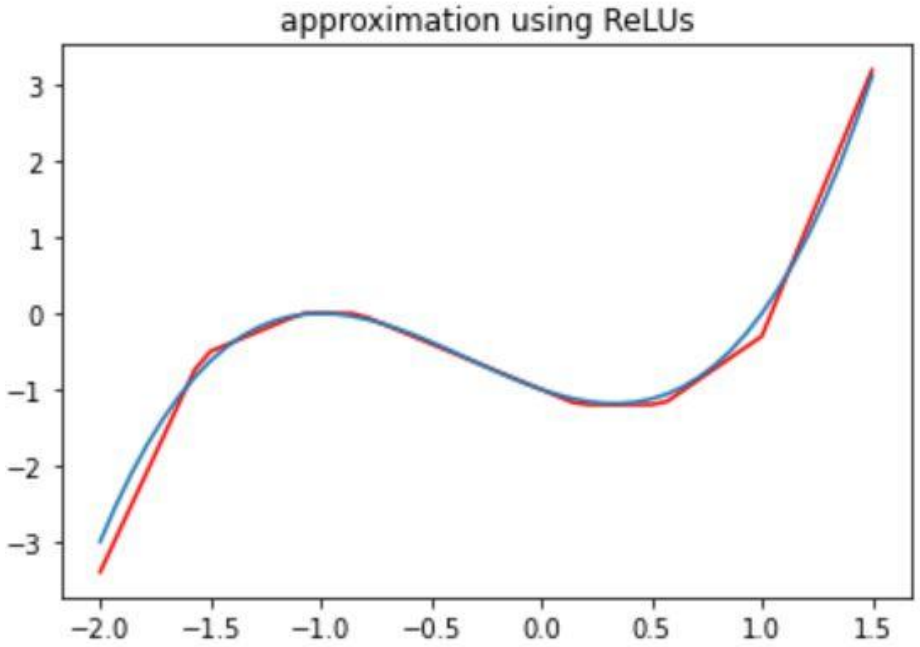
Universality with one i/p & one o/p

$$
\begin{aligned}
n_1 &= ReLU(-5x - 7.7) \\
n_2 &= ReLU(-1.2x - 1.3) \\
n_3 &= ReLU(1.2x + 1) \\
n_4 &= ReLU(1.2x - 0.2) \\
n_5 &= ReLU(2x - 1.1) \\
n_6 &= ReLU(5x - 5)
\end{aligned}
$$



Universality with one i/p & one o/p

$$
\begin{aligned}
n_1 &= ReLU(-5x - 7.7) \\
n_2 &= ReLU(-1.2x - 1.3) \\
n_3 &= ReLU(1.2x + 1) \\
n_4 &= ReLU(1.2x - 0.2) \\
n_5 &= ReLU(2x - 1.1) \\
n_6 &= ReLU(5x - 5)
\end{aligned}
$$



Next
Backpropagation