Machine Learning
Lecture 19 (Week 10)

Agenda

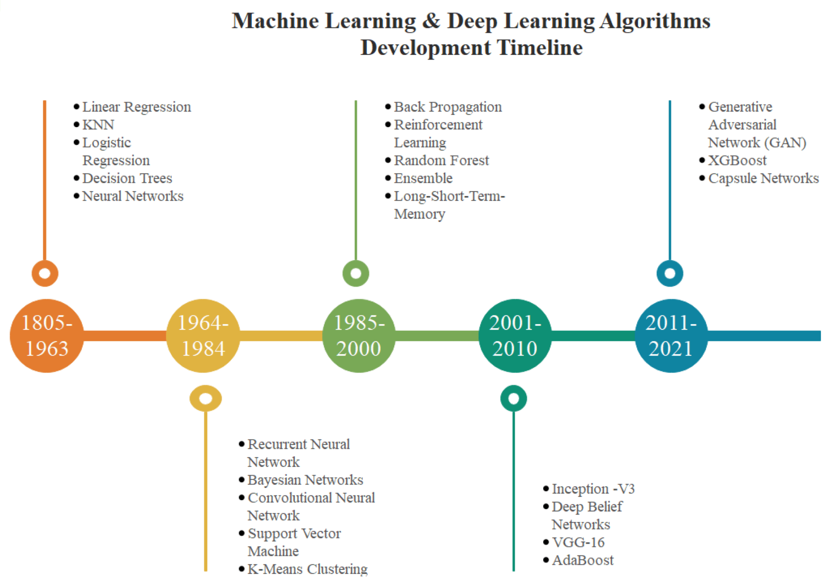
(Artifitial) Neural Networks

- Automated Differentiation
  - Computational Graphs
  - Forward, Backward, and Cross Modes
- Multilayer Perceptron (MLP)

ASIRRA

Source: `https://martin-thoma.com/captcha/`

Timeline



Source: `https://www.mdpi.com/2227-9032/10/3/541`

Neural Networks, Vol. 2, pp. 359–366, 1989
Printed in the USA.

0893-6080/89 $3.00 + .00

ORIGINAL CONTRIBUTION

Multilayer Feedforward Networks are Universal Approximators

KURT HORNIK

Technische Universität Wien

MAXWELL STINCHCOMBE AND HALBERT WHITE

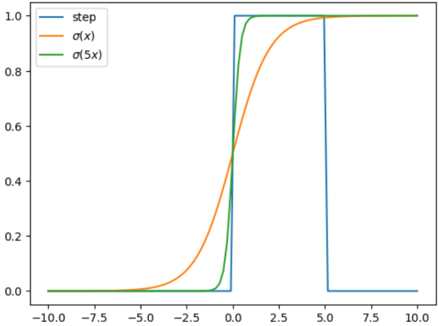
University of California, San Diego

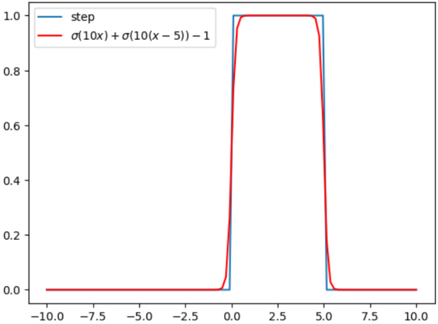
(*Received* 16 *September* 1988; *revised and accepted* 9 *March* 1989)

Abstract—*This paper rigorously establishes that standard multilayer feedforward networks with as few as one hidden layer using arbitrary squashing functions are capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficiently many hidden units are available. In this sense, multilayer feedforward networks are a class of universal approximators.*

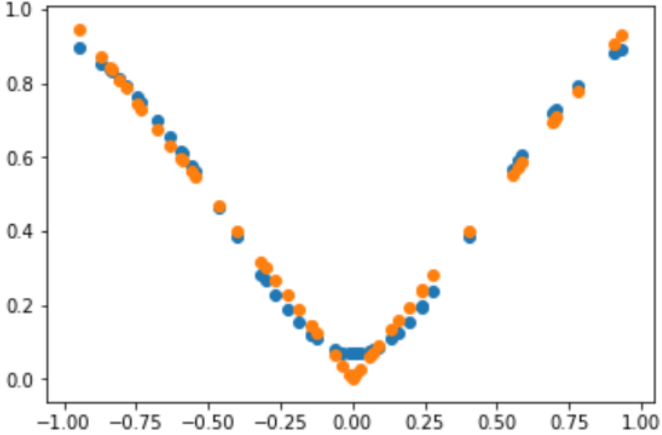
Keywords—Feedforward networks, Universal approximation, Mapping networks, Network representation capability, Stone-Weierstrass Theorem, Squashing functions, Sigma-Pi networks, Back-propagation networks.

Sigmoid Approximation





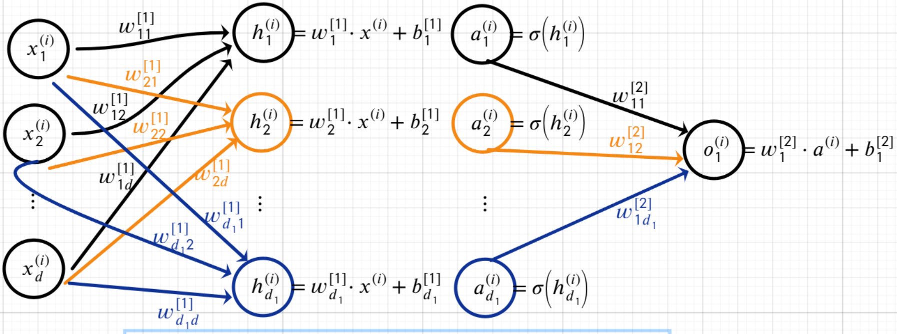
Two-layer Perceptron (one hidden layer)



Two-layer Perceptron (one hidden layer)

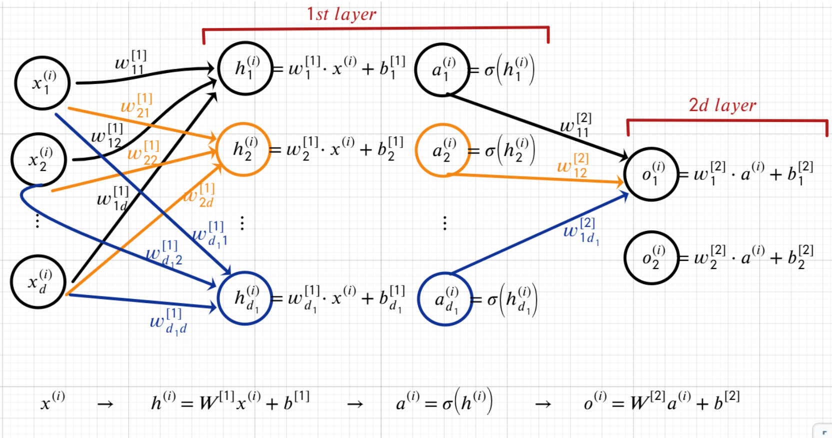
$$o_1^{(i)} = w_{11}^{[2]} a_1^{(i)} + w_{12}^{[2]} a_2^{(i)} + \ldots + w_{1d_1}^{[2]} a_{d_1}^{(i)} + b_1^{[2]}$$

$$= w_{11}^{[2]} \sigma\left(w_1^{[1]} \cdot x^{(i)} + b_1^{[1]}\right) + w_{12}^{[2]} \sigma\left(w_2^{[1]} \cdot x^{(i)} + b_2^{[1]}\right) + \ldots + w_{1d_1}^{[2]} \sigma\left(w_{d_1}^{[1]} \cdot x^{(i)} + b_{d_1}^{[1]}\right) + b_1^{[2]}$$



$$L\left(w^{[1]}, w^{[2]}, b^{[1]}, b^{[2]}\right) = \frac{1}{N} \sum_{i=1}^{N} L^{(i)}, \qquad L^{(i)} = \left(o_1^{(i)} - y^{(i)}\right)^2$$

Matrix Representation



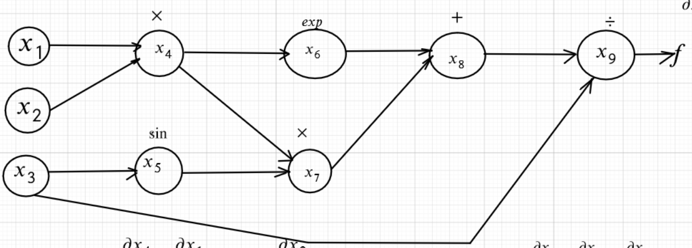
$\boxed{Forward\ mode}$

$Backward\ mode$

$Cross\ mode$

$$f(x_1,x_2,x_3)=\frac{x_1x_2\sin x_3+exp(x_1x_2)}{x_3}$$

$Computational\ Graph:$



$Goal:$

$$\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\frac{\partial f}{\partial x_3}$$

f	∇f
$O(n)$	$O(nd)$

$$\frac{\partial x_1}{\partial x_1}=1$$

$$\frac{\partial x_2}{\partial x_1}=0=\frac{\partial x_3}{\partial x_1}$$

$$\frac{\partial x_4}{\partial x_1}=\frac{\partial x_1}{\partial x_1}x_2+x_1\cdot\frac{\partial x_2}{\partial x_1}$$

$$\frac{\partial x_5}{\partial x_1}=\cos x_3\cdot\frac{\partial x_3}{\partial x_1}$$

$$\frac{\partial x_6}{\partial x_1}=e^{x_4}\cdot\frac{\partial x_4}{\partial x_1}$$

$$\frac{\partial x_7}{\partial x_1}=\frac{\partial x_4}{\partial x_1}x_5+x_4\cdot\frac{\partial x_5}{\partial x_1}$$

$$\frac{\partial x_8}{\partial x_1}=\frac{\partial x_6}{\partial x_1}+\frac{\partial x_7}{\partial x_1}$$

$$\frac{\partial x_9}{\partial x_1}=\frac{1}{x_3}\frac{\partial x_8}{\partial x_1}$$

Forward mode

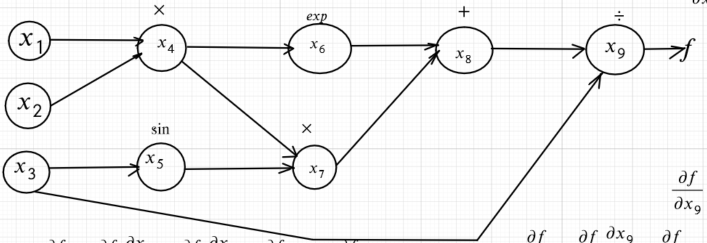
Backward mode

Cross mode

$$f(x_1,x_2,x_3)=\frac{x_1x_2\sin x_3+exp(x_1x_2)}{x_3}$$

Goal : $\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\frac{\partial f}{\partial x_3}$

Computational Graph :



f	∇f
$O(n)$	$O(nd)$
$O(n)$	$O(n)$

$$\frac{\partial f}{\partial x_9}=1$$

$$\frac{\partial f}{\partial x_8}=\frac{\partial f}{\partial x_9}\frac{\partial x_9}{\partial x_8}=\frac{\partial f}{\partial x_9}\cdot\frac{1}{x_3}$$

$$\frac{\partial f}{\partial x_6}=\frac{\partial f}{\partial x_8}\frac{\partial x_8}{\partial x_6}=\frac{\partial f}{\partial x_8}\cdot 1$$

$$\frac{\partial f}{\partial x_7}=\frac{\partial f}{\partial x_8}\frac{\partial x_8}{\partial x_7}=\frac{\partial f}{\partial x_8}\cdot 1$$

$$\frac{\partial f}{\partial x_4}=\frac{\partial f}{\partial x_6}\frac{\partial x_6}{\partial x_4}+\frac{\partial f}{\partial x_7}\frac{\partial x_7}{\partial x_4}=\frac{\partial f}{\partial x_6}e^{x_4}+\frac{\partial f}{\partial x_7}x_5$$

$$\frac{\partial f}{\partial x_5}=\frac{\partial f}{\partial x_7}\frac{\partial x_7}{\partial x_5}=\frac{\partial f}{\partial x_7}x_4$$

$$\frac{\partial f}{\partial x_1}=\frac{\partial f}{\partial x_4}\frac{\partial x_4}{\partial x_1}$$

$$\frac{\partial f}{\partial x_2}=\frac{\partial f}{\partial x_4}\frac{\partial x_4}{\partial x_2}$$

$$\frac{\partial f}{\partial x_3}=\frac{\partial f}{\partial x_5}\frac{\partial x_5}{\partial x_3}+\frac{\partial f}{\partial x_9}\frac{\partial x_9}{\partial x_3}$$

R·I·T

Computer Science @RIT