kptan86@gmail.com

List of Topics

Machine learning

Formal Definition:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E." - Mitchell, T. (1997).

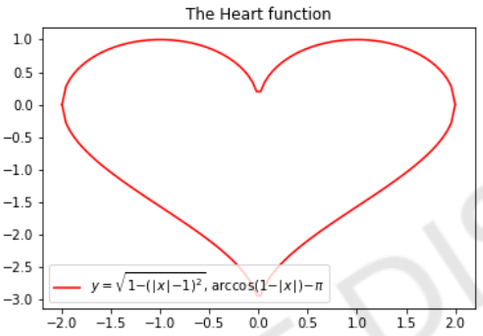
Loose Definition:

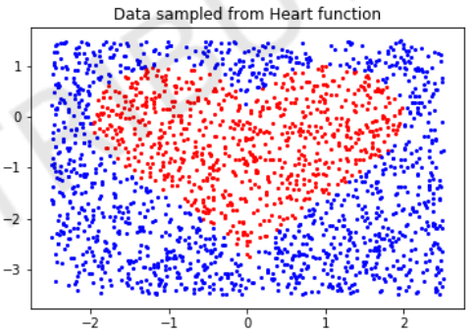
"Approximating the underlying function to a target task by synthesis of basis functions of simple form."

Classifier	basis function
Decision trees	Step function
Nearest neighbours	Delta function
Naive Bayes, mixture models	Gaussian
Neural networks	Logistic function

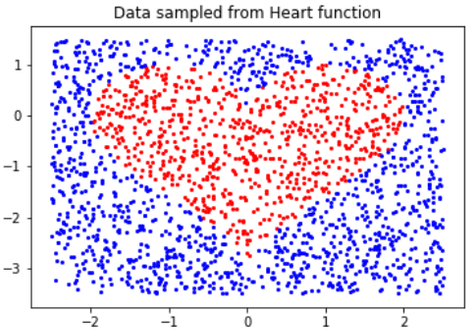
kptan86@gmail.com

Function Approximation demonstration



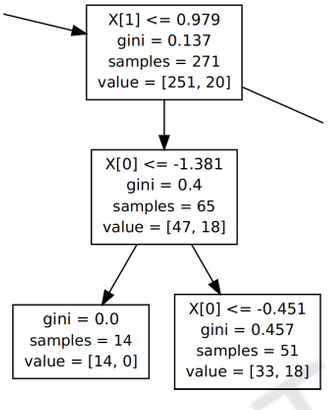


Function Approximation demonstration

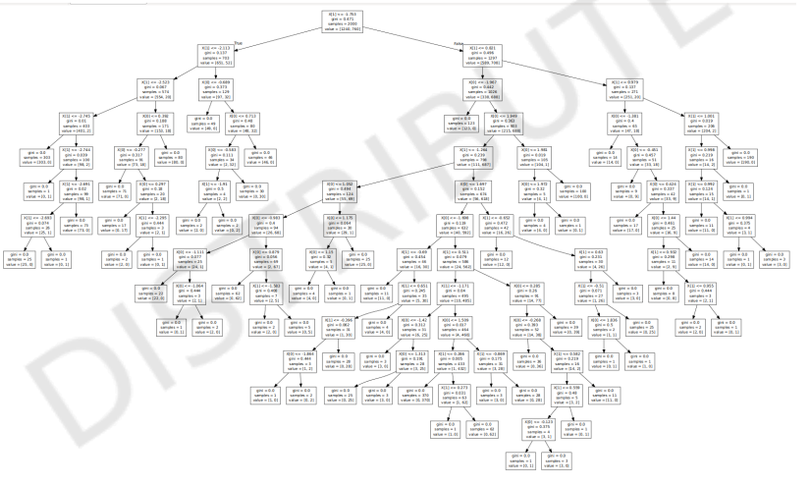


Function Approximation demonstration

Example step functions



Decision tree approximating Heart function



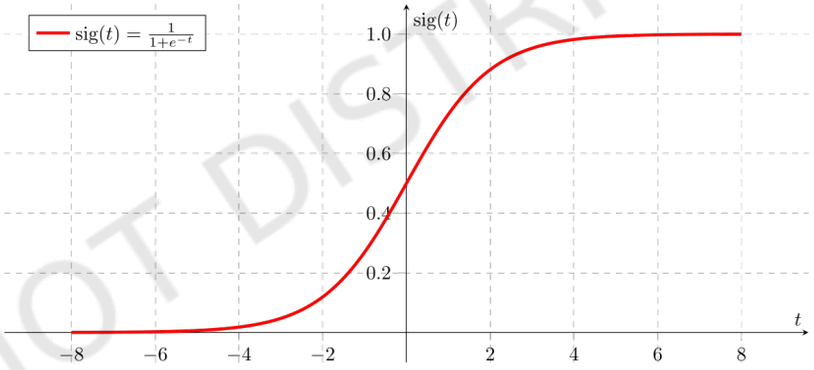
DO NOT DISTRIBUTE

kptan86@gmail.com

Logistic Regression

$$y = (1 + \exp(-\{\sum_{\forall i} w_i x_i + b_i\}))^{-1}$$

Objective: Find optimal $\vec{w_i}, \vec{b_i}$ over trainining set $(\vec{X}, \vec{Y})$



Stacking Logistic Regression

- 2-layer:

$$y = (1 + \exp(-\{\sum_{\forall j} w_j \boxed{(1 + \exp(-\{\sum_{\forall i} w_i x_i + b_i\}))^{-1}} + b_j\}))^{-1}$$

Objective: Find optimal $\vec{w_i}, \vec{w_j}, \vec{b_i}, \vec{b_j}$ over trainining set $(\vec{X}, \vec{Y})$

This is already a neural network !

Keep Stacking

- many-layer:

$$y = (1 + \ldots \boxed{(1 + \exp(-\{\sum_{\forall j} w_j \boxed{(1 + \exp(-\{\sum_{\forall i} w_i x_i + b_i\}))^{-1}} + b_j\}))^{-1} \ldots})^{-1}$$

Objective: Find optimal $\vec{w_i}, \vec{w_j}, \ldots, \vec{b_i}, \vec{b_j}, \ldots$ over trainining set $(\vec{X}, \vec{Y})$

This is a *deep* neural network

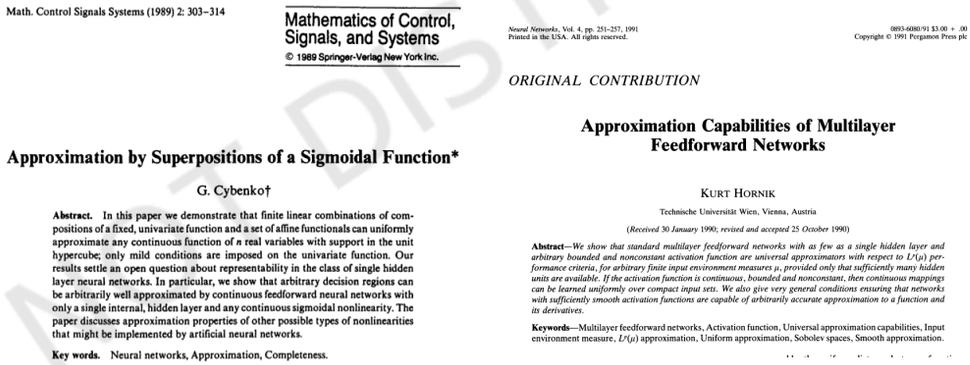
DO NOT DISTRIBUTE

kptan86@gmail.com

Universal function approximator

"... the universal approximation theorem states that **a feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions on compact subsets of** R^n, under mild assumptions on the activation function."

~ Cybenko, G. 1989; Balázs Csanád Csáji 2001



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

Mathematics of Control, Signals, and Systems
© 1989 Springer-Verlag New York Inc.

Approximation by Superpositions of a Sigmoidal Function*

G. Cybenko†

Abstract. In this paper we demonstrate that finite linear combinations of compositions of a fixed, univariate function and a set of affine functionals can uniformly approximate any continuous function of n real variables with support in the unit hypercube; only mild conditions are imposed on the univariate function. Our results settle an open question about representability in the class of single hidden layer neural networks. In particular, we show that arbitrary decision regions can be arbitrarily well approximated by continuous feedforward neural networks with only a single internal, hidden layer and any continuous sigmoidal nonlinearity. The paper discusses approximation properties of other possible types of nonlinearities that might be implemented by artificial neural networks.

Key words. Neural networks, Approximation, Completeness.

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

0893-6080/91 $3.00 + .00
Copyright © 1991 Pergamon Press plc

ORIGINAL CONTRIBUTION

Approximation Capabilities of Multilayer Feedforward Networks

KURT HORNIK

Technische Universität Wien, Vienna, Austria

(*Received* 30 *January* 1990; *revised and accepted* 25 *October* 1990)

Abstract—*We show that standard multilayer feedforward networks with as few as a single hidden layer and arbitrary bounded and nonconstant activation function are universal approximators with respect to* $L^p(\mu)$ *performance criteria, for arbitrary finite input environment measures* μ, *provided only that sufficiently many hidden units are available. If the activation function is continuous, bounded and nonconstant, then continuous mappings can be learned uniformly over compact input sets. We also give very general conditions ensuring that networks with sufficiently smooth activation functions are capable of arbitrarily accurate approximation to a function and its derivatives.*

Keywords—Multilayer feedforward networks, Activation function, Universal approximation capabilities, Input environment measure, $L^p(\mu)$ approximation, Uniform approximation, Sobolev spaces, Smooth approximation.

Discussion:

why not just use universal function approximator all the times?

DO NOT DISTRIBUTE