Introduction

What is an artificial neuron?

- An artificial neuron is a model of a real neuron.
- Practical realizations of models of real neurons may be made in one of the following forms:
  - Electronic circuit models of real neurons.
  - Computational models of real neurons.
- Computational neuron models are computer software implementations of mathematical models of real neurons.
- Basic types of computational neuron models:
  - Biophysical neuron models.
    Software realizations of detailed mathematical models that are used to study the behavior of real neurons.
  - Engineering neuron models.
    Software realizations of simple (abstract) models that are meant for use to build information processing systems.

Note: The scope of this set of lecture notes is restricted to the coverage of the principles and the applications of a relatively small set of engineering neuron models.

Activation Functions

Outline

Introduction

Activation Functions

- Monotonic Activation Functions
  - Basic Concept and Types of Monotonic Activation Functions
  - Hard-threshold Monotonic Activation Functions
  - Two-parameter Piecewise Linear Activation Functions
  - Sigmoidal Monotonic Activation Functions
- Radial Basis Functions
  - Basic Concept and Types of Radial Basis Functions

Basic concept and types of monotonic activation functions (I)

- **Basic concept:** A monotonic activation function is a real function which has the following characteristics:
  - **It is a non-decreasing function.**
    Its value either increases or stays constant as the value of the input (or excitation) variable x increases.
  - **It is bounded both below and above.**
    It has a lower limit L and an upper limit H. The upper limit is 1, whereas the lower limit is either 0 or -1. The activation function is said to be unipolar if $L = 0$ or bipolar if $L = -1$.
  - **It defines a dichotomy by means of a parameter θ.**
    It divides its domain (or input space) into two regions: The LOW and the HIGH regions. The value of the function is LOW for values of x below θ and HIGH for values of x above θ. We therefore refer to θ as the transition threshold excitation level.

Basic concept and types of monotonic activation functions (II)

- **Most popular types of monotonic activation functions:**
  1. The hard-threshold monotonic activation functions.
  2. The two-parameter piecewise linear monotonic activation functions.
  3. The sigmoidal monotonic activation functions.

Definition of a hard-threshold monotonic activation function

- **Basic concept:**
  A hard-threshold monotonic activation function $g_{\mathrm{d}}(x;\theta)$ is a piecewise constant (or pure binary) decision function: Its value is either L or H.
- **Mathematical definition:**

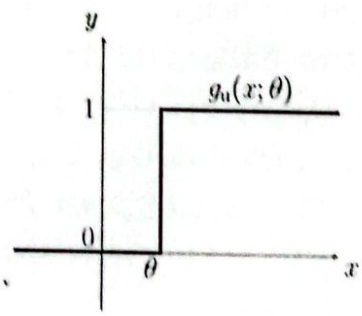
$$g_{\mathrm{d}}(x;\theta) \triangleq \begin{cases} \mathrm{L} & \text{if } x \leqslant \theta, \\ \mathrm{H} & \text{if } x > \theta. \end{cases}$$

- **Families:**
  There are two families of monotonic activation functions, namely:
  - The Heaviside (or unit function) family of activation functions.
    The Heaviside family of activation functions is a family of unipolar monotonic activation functions.
  - The signum (or sign function) family of activation functions.
    The signum family of activation functions is a family of bipolar activation functions.

The Heaviside and the signum functions

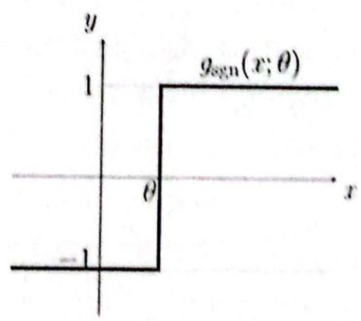
- A Heaviside (or unit) activation function:

$$g_{\mathrm{u}}(x;\theta) \triangleq \begin{cases} 0 & \text{if } x \leqslant \theta, \\ 1 & \text{if } x > \theta. \end{cases}$$



- A signum (or sign) activation function:

$$g_{\mathrm{sgn}}(x;\theta) \triangleq \begin{cases} -1 & \text{if } x \leqslant \theta, \\ 1 & \text{if } x > \theta. \end{cases}$$



- *Remark:*

$$g_{\mathrm{sgn}}(x;\theta) = 2g_{\mathrm{u}}(x;\theta) - 1.$$

Outline

Introduction

Activation Functions

Monotonic Activation Functions

Hard-threshold Monotonic Activation Functions

Two-parameter Piecewise Linear Activation Functions

Sigmoidal Monotonic Activation Functions

Radial Basis Functions

Two-parameter piecewise linear activation functions (I)

- **Basic concept**
  A two-parameter piecewise linear monotonic activation function $g_{\text{pwl}}(x; \theta_1, \theta_2)$ is an activation function which has the following properties:
  - It takes the value L for values of x less than the transition threshold θ_1.
  - It increases linearly for values of x between θ_1 and the shape parameter θ_2.
  - It takes the value the H for values of x greater than θ_2.
- **Rationale**
  Compared to hard-threshold monotonic activation functions two-parameter piecewise linear monotonic activation functions have the following characteristics which facilitate better approximation functional dependence on input data:
  - Their shapes can be adapted to the data.
  - They have well-behaved derivatives that facilitate fine tuning of network parameters during learning by error correction.

Two-parameter piecewise linear activation functions (II)

- They form the simplest class of monotonic activation functions with well-behaved derivatives.
- Mathematical definition

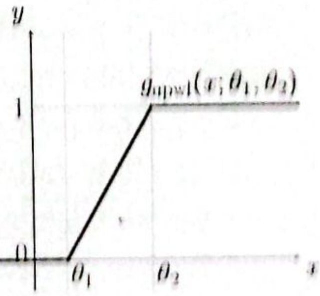
$$g_{\mathrm{pwl}}(x;\theta_1,\theta_2) \triangleq \begin{cases} L & \text{if } x < \theta_1, \\ \frac{(H-L)}{(\theta_2-\theta_1)}(x-\theta_1)+L & \text{if } \theta_1 \leqslant x \leqslant \theta_2, \\ H & \text{if } x > \theta_2, \end{cases}$$

- Basic types
  - The family of unipolar two-parameter piecewise linear (UPWL) activation functions.
  - The family of bipolar two-parameter piecewise linear (BPWL) activation functions.

The unipolar and the bipolar families of piecewise linear activation functions

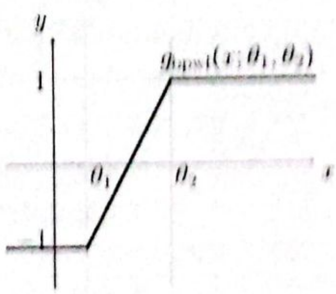
- A unipolar piecewise linear activation (UPWL) function:

$$g_{\mathrm{upwl}}(x;\theta_1,\theta_2) \triangleq \begin{cases} 0 & \text{if } x < \theta_1, \\ \frac{(x-\theta_1)}{(\theta_2-\theta_1)} & \text{if } \theta_1 \leqslant x \leqslant \theta_2, \\ 1 & \text{if } x > \theta_2. \end{cases}$$



- A bipolar piecewise linear (BPWL) activation function:

$$g_{\mathrm{bpwl}}(x;\theta_1,\theta_2) \triangleq \begin{cases} -1 & \text{if } x < \theta_1, \\ \frac{2(x-\theta_1)}{(\theta_2-\theta_1)}-1 & \text{if } \theta_1 \leqslant x \leqslant \theta_2, \\ 1 & \text{if } x > \theta_2. \end{cases}$$



Exercises on piecewise linear activation functions

Exercise 1

Determine the derivatives of the unipolar and the bipolar activation functions. Are the derivatives related in any way to the hard-threshold monotonic activation functions?

Exercise 2

Determine the type of function that a unipolar piecewise linear monotonic activation function degenerates to in the limit as the upper excitation threshold θ_2 approaches the lower excitation threshold θ_1.

Exercise 3

Determine the type of function that a bipolar piecewise linear monotonic activation function degenerates to in the limit as the upper excitation threshold θ_2 approaches the lower excitation threshold θ_1.

Sigmoidal monotonic activation functions (I)

- **Basic concept**
  A sigmoidal monotonic activation function $g_{\text{sig}}(x; \beta, \theta)$ is an infinitely smooth S-shaped function which has a shape factor $\beta (> 0)$. Such a function has the following characteristics:
  - Its value increases monotonically with the excitation x.
  - Its value saturates towards H as x approaches $+\infty$.
  - Its value saturates towards L as x approaches $-\infty$.
  - It has a smooth n-th order derivative for any $n \geqslant 1$.
- **Rationale**
  - A sigmoidal activation function features an S-shape that may be adapted with a high degree of flexibility to any pattern recognition or function approximation problem.
  - The nonlinearity and the differentiability of a sigmoidal activation function may be exploited to facilitate the fine tuning of connection and PE parameters during learning.

Sigmoidal monotonic activation functions (II)

- Types and examples
  - Unipolar sigmoidal activation functions:
    - The logistic sigmoid function.
    - The error function:
      The error function is the integral of the Gaussian (or normal) probability density function.

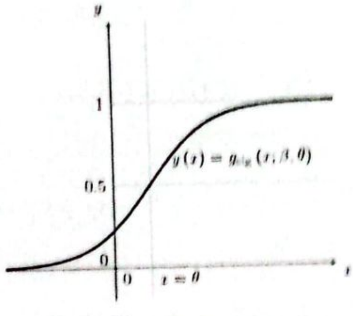
$$\operatorname{errf}(x;\sigma,\mu) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(v-\mu)^2}{2\sigma^2}\right) dv$$

  - Bipolar sigmoidal activation functions:
    - The hyperbolic tangent function.

The logistic sigmoid and the hyperbolic tangent activation functions

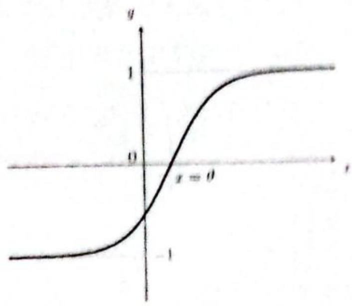
- A logistic sigmoid activation function:

$$g_{\text{sig}}(x;\beta,\theta) \triangleq \frac{1}{1+e^{(-\beta(x-\theta))}}.$$



- A hyperbolic tangent activation function:

$$g_{\text{ht}}(x;\beta,\theta) \triangleq \tanh(\beta(x-\theta)) = \frac{e^{\beta(x-\theta)} - e^{-\beta(x-\theta)}}{e^{\beta(x-\theta)} + e^{-\beta(x-\theta)}}.$$



Exercises on the logistic sigmoid and the hyperbolic tangent activation functions (I)

Exercise 4

1. Differentiate the logistic sigmoid activation function.
2. Express the derivative of the sigmoid activation function in terms of some familiar function.
3. Formulate the expression of the value of the derivative of the logistic sigmoid activation function at the excitation threshold.
4. Using the expression established in within 3, infer the type of function that the logistic sigmoid activation function degenerates to as the value of the function's shape factor is made
   4.1 increasingly small.
   4.2 increasingly large.

Exercises on the logistic sigmoid and the hyperbolic tangent activation functions (II)

Exercise 5

1. Derive the expression for the derivative of the hyperbolic tangent activation function.
2. Formulate the expression of the value of the derivative of the hyperbolic tangent activation function at the excitation threshold.
3. Using the expression established in within 2, infer the type of function that the hyperbolic tangent activation function degenerates to as the value of the function's shape factor is made
   3.1 increasingly small.
   3.2 increasingly large.

Radial Basis Activation Functions

Basic concept and common types of radial basis functions

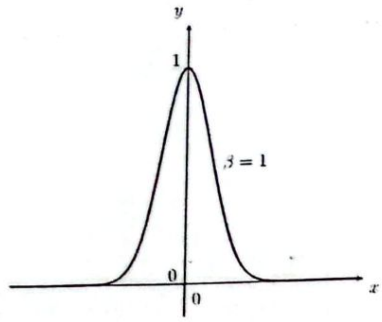
- **Basic concept**
  - A radial basis function (RBF) is a real function which exhibits radial symmetry.
  - A real function $f(x)$ has radial symmetry if the following condition holds for all values of θ
$$f(x+\theta) = f(x-\theta).$$
  - $\varphi(x;\beta)$ denotes a radial basis activation function with a shape factor β.
- **Importance**
  - RBFs are widely used in function approximation and data interpolation.
  - RBFs are easy to adapt to the resolution of nonlinearly separable patterns in the input space.
- **Common types**
  The following are some of the most commonly used types of RBFs:
  - Gaussian RBFs.
  - Multiquadric RBFs.
  - Inverse multiquadric RBFs.
  - Inverse quadric RBFs.

The Gaussian and the inverse quadric activation functions

- A Gaussian activation function:

$$\varphi(x;\beta) \triangleq e^{(-\beta x^2)}$$

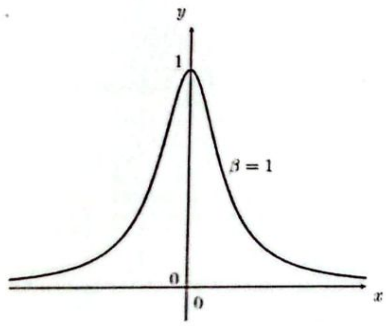
where $\beta > 0$.



- An inverse quadric activation function:

$$\varphi(x;\beta) \triangleq \frac{1}{1+\beta x^2}$$

where $\beta > 0$.

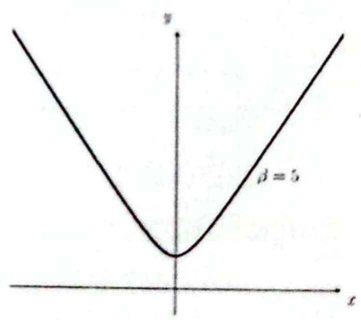


The multiquadric and the inverse multiquadric activation functions

- A multiquadric activation function:

$$\varphi(x;\beta) \triangleq \sqrt{1+(\beta x)^2}$$

where $\beta > 0$.



- An inverse multiquadric activation function:

$$\varphi(x;\beta) \triangleq \frac{1}{\sqrt{1+(\beta x)^2}}$$

where $\beta > 0$.