

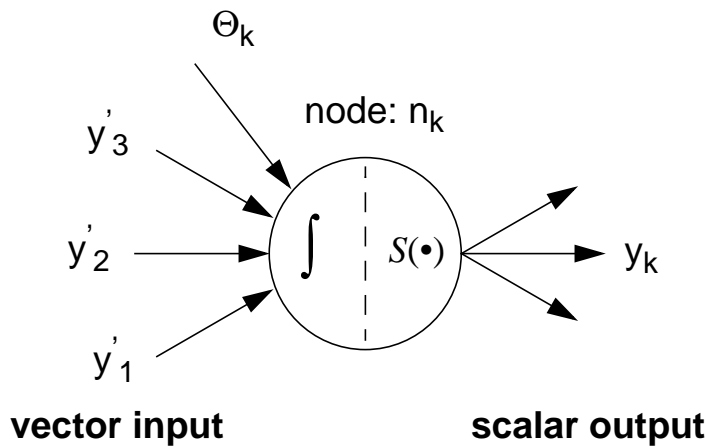
## The Artificial Neural Network (ANN)

- ❑ Premise: complex computational operations can be implemented by massive integration of individual components
- ❑ Topology and interconnections are key: in many ANN systems, spatial relationships between nodes have some physical relevance
- ❑ Properties of large-scale systems: ANNs also reflect a growing body of theory stating that large-scale systems built from a small unit need not simply mirror properties of a smaller system (contrast fractals and chaotic systems with digital filters)

### Why Artificial Neural Networks?

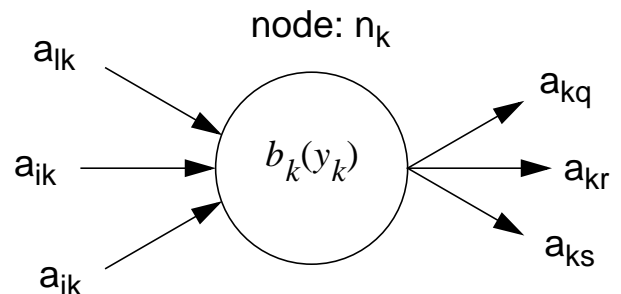
- ❑ Important physical observations:
  - The human central nervous system contains  $10^{11} - 10^{14}$  nerve cells, each of which interacts with  $10^3 - 10^4$  other neurons
  - Inputs may be excitatory (promote firing) or inhibitory

### The Artificial Neuron — Nonlinear



$$y_k \equiv S\left(\sum_{n=1}^N w_{ki} y'_i - \theta_k\right)$$

### The HMM State — Linear



$$\alpha(y_1^{t+1}, j) = \alpha(y_1^t, i) a(j|i) b(y(t+1)|j)$$

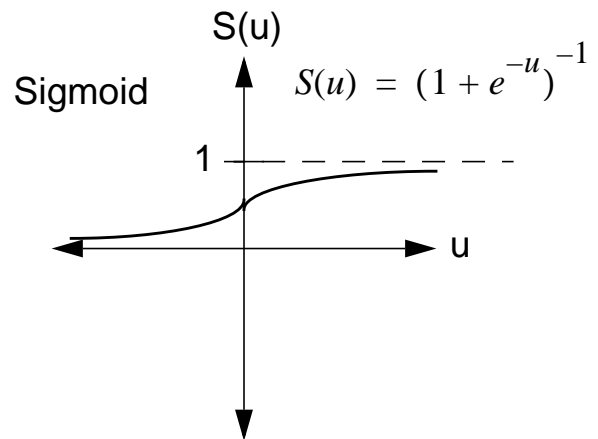
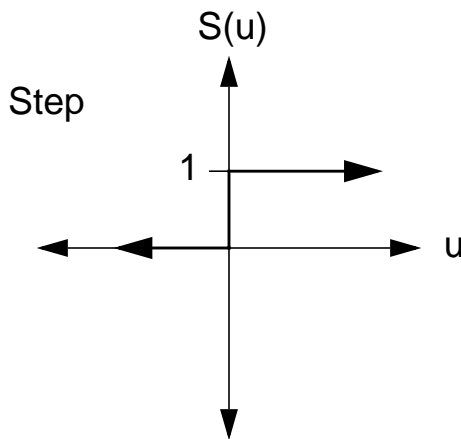
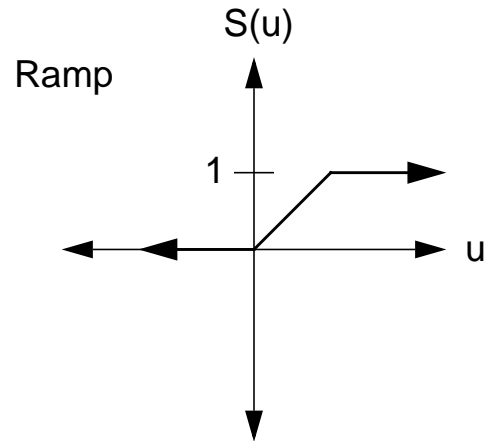
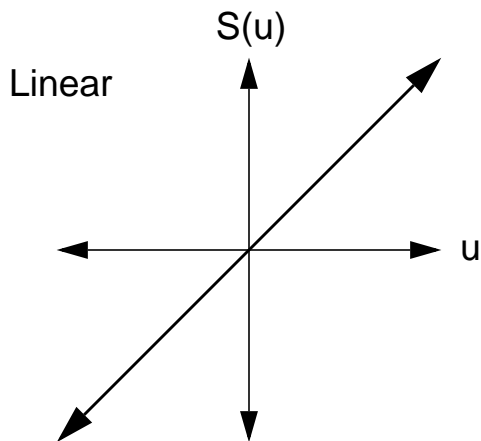


## Typical Thresholding Functions — A Key Difference

The input to the thresholding function is a weighted sum of the inputs:

$$u_k \equiv \mathbf{w}_k^T \mathbf{y}'$$

The output is typically defined by a nonlinear function:



Sometimes a bias is introduced into the threshold function:

$$y_k \equiv S(\mathbf{w}_k^T \mathbf{y}' - \theta_k) = S(u_k - \theta_k)$$

This can be represented as an extra input whose value is always -1:

$$y'_{N+1} = -1 \quad w_{k, N+1} = \theta_k$$

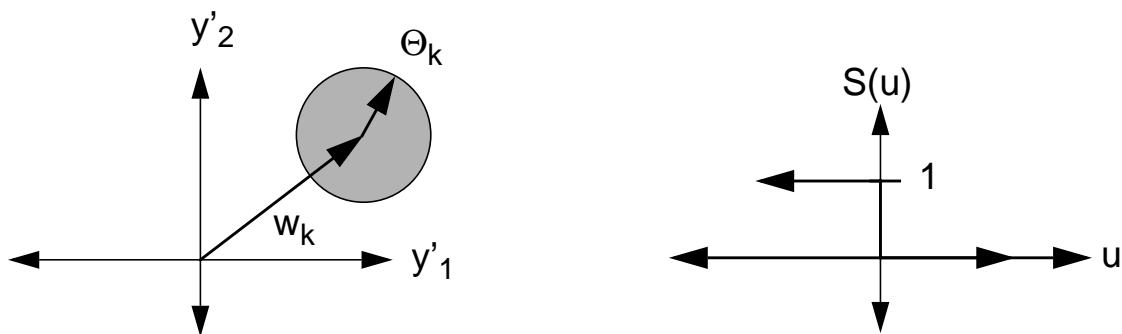
## Radial Basis Functions

Another popular formulation involves the use of a Euclidean distance:

$$y_k = S\left(\sqrt{\sum_{i=1}^N (w_{ik} - y'_i)^2} - \theta_k\right) = S(\|\mathbf{w}_k - \mathbf{y}'\|_2 - \theta_k)$$

Note the parallel to a continuous distribution HMM.

This approach has a simple geometric interpretation:



Another popular variant of this design is to use a Gaussian nonlinearity:

$$S(u) = e^{-u^2}$$

What types of problems are such networks useful for?

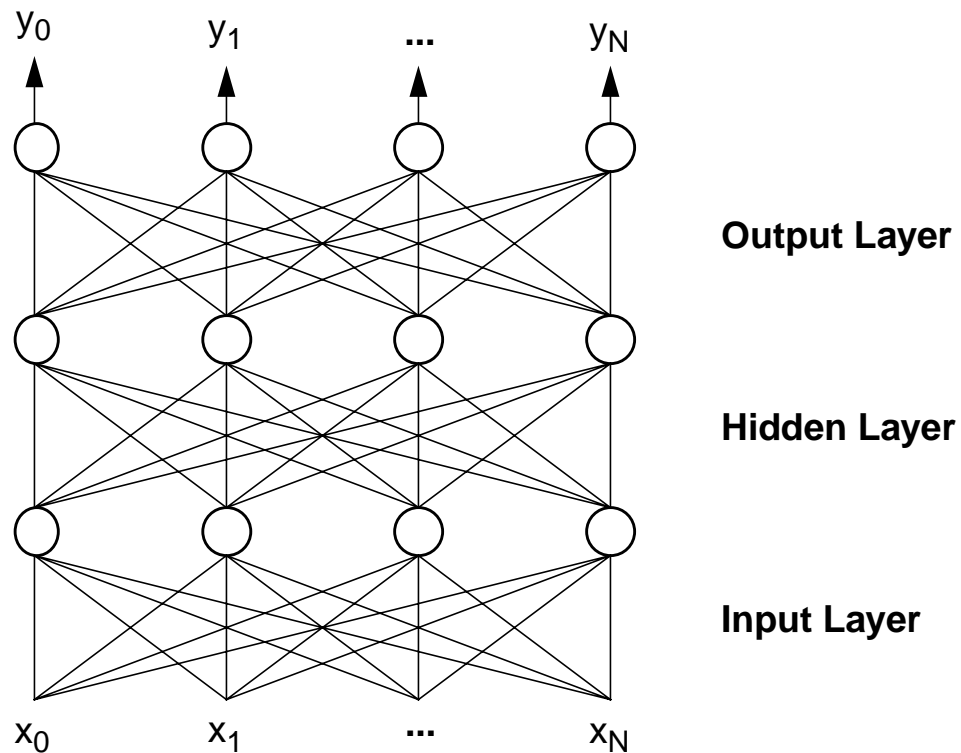
- pattern classification (N—way choice; vector quantization)
- associative memory (generate an output from a noisy input; character recognition)
- feature extraction (similarity transformations; dimensionality reduction)

We will focus on multilayer perceptrons in our studies. These have been shown to be quite useful for a wide range of problems.

## Multilayer Perceptrons (MLP)

This architecture has the following characteristics:

- Network segregated into layers:  $N_i$  cells per layer,  $L$  layers
- feedforward, or nonrecurrent, network (no feedback from the output of a node to the input of a node)



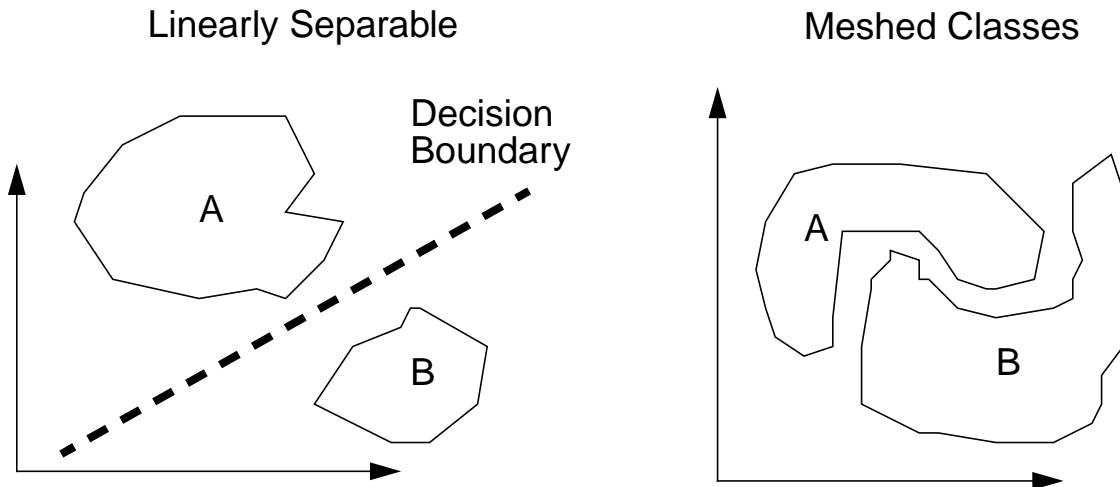
An alternate formulation of such a net is known as the learning vector quantizer (LVQ) — to be discussed later.

The MLP network, not surprisingly, uses a supervised learning algorithm. The network is presented the input and the corresponding output, and must learn the optimal weights of the coefficients to minimize the difference between these two.

The LVQ network uses unsupervised learning — the network adjusts itself automatically to the input data, thereby clustering the data (learning the boundaries representing a segregation of the data). LVQ is popular because it supports discriminative training.

## Why Artificial Neural Networks?

- An ability to separate classes that are not linearly separable:



A three-layer perceptron is required to determine arbitrarily-shaped decision regions.

- Nonlinear statistical models

The ANN is capable of modeling arbitrarily complex probability distributions, much like the difference between VQ and continuous distributions in HMM.

- Context-sensitive statistics

Again, the ANN can learn complex statistical dependencies provided there are enough degrees of freedom in the system.

### Why not Artificial Neural Networks? (The Price We Pay...)

- Difficult to deal with patterns of unequal length
- Temporal relationships not explicitly modeled

And, of course, both of these are extremely important to the speech recognition problem.