

# Neural Networks

## Required reading:

- Bishop Chapter 5, especially 5.1, 5.2, 5.3, and 5.5 through 5.5.2

## Optional reading:

- Neural nets: Mitchell chapter 4

Machine Learning 10-701

Tom M. Mitchell

Center for Automated Learning and Discovery

Carnegie Mellon University

September 28, 2006

## Artificial Neural Networks to learn $f: X \rightarrow Y$

- $f$  might be non-linear function
- $X$  (vector of) continuous and/or discrete vars
- $Y$  (vector of) continuous and/or discrete vars

- Represent  $f$  by network of threshold units
- Each unit is a logistic function

$$\text{unit output} = \frac{1}{1 + \exp(w_0 + \sum_i w_i x_i)}$$

- MLE: train weights of all units to minimize sum of squared errors of network function

# Connectionist Models

---

Consider humans:

- Neuron switching time  $\sim .001$  second
- Number of neurons  $\sim 10^{10}$
- Connections per neuron  $\sim 10^{4-5}$
- Scene recognition time  $\sim .1$  second
- 100 inference steps doesn't seem like enough

→ much parallel computation

Properties of artificial neural nets (ANN's):

- Many neuron-like threshold switching units
- Many weighted interconnections among units
- Highly parallel, distributed process