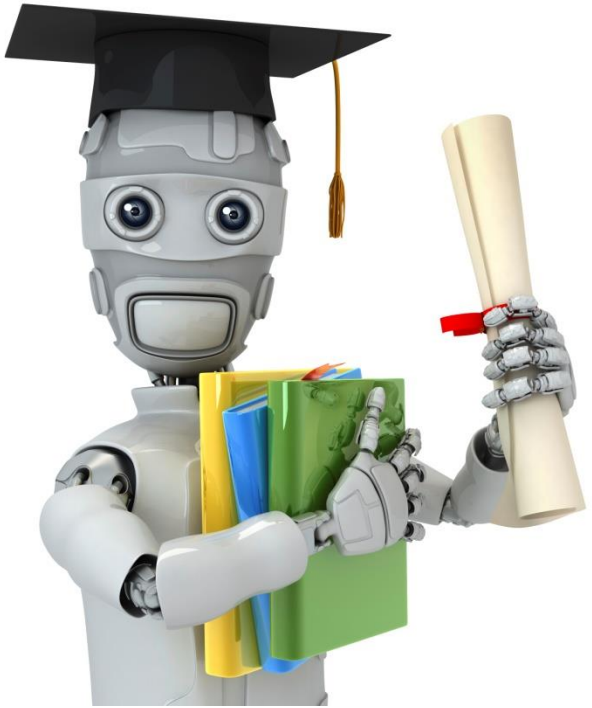
CSE 4621
Machine Learning

Lecture 7

Md. Hasanul Kabir, PhD.

Professor, CSE Department

Islamic University of Technology (IUT)



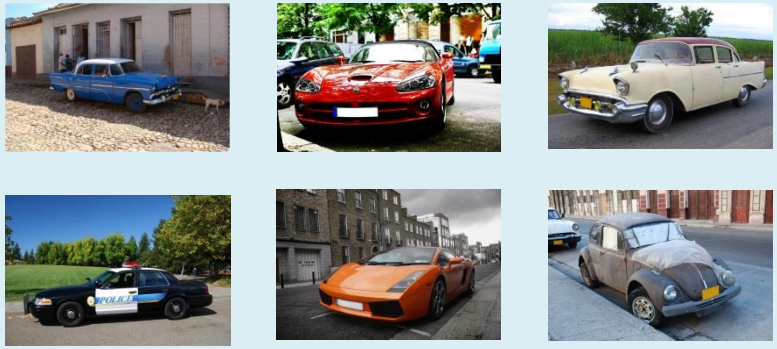
Neural Network

Introduction

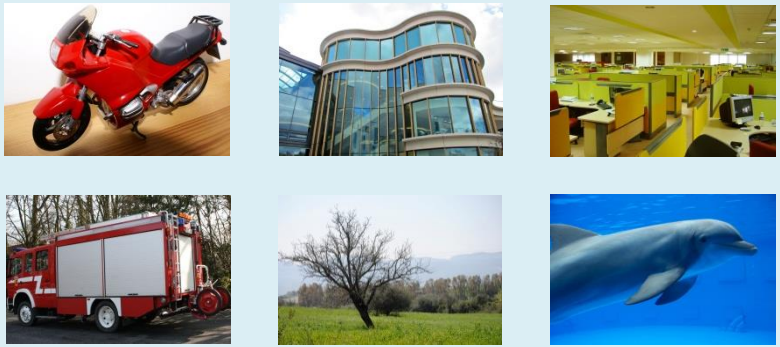
Machine Learning

Source & Special Thanks to (Coursera) Machine Learning / NN&DL Courses

Computer Vision: Car detection



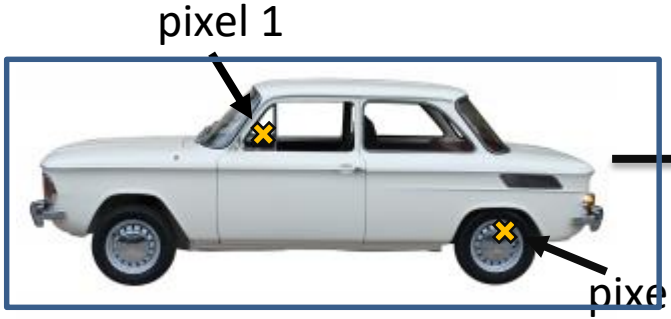
Cars



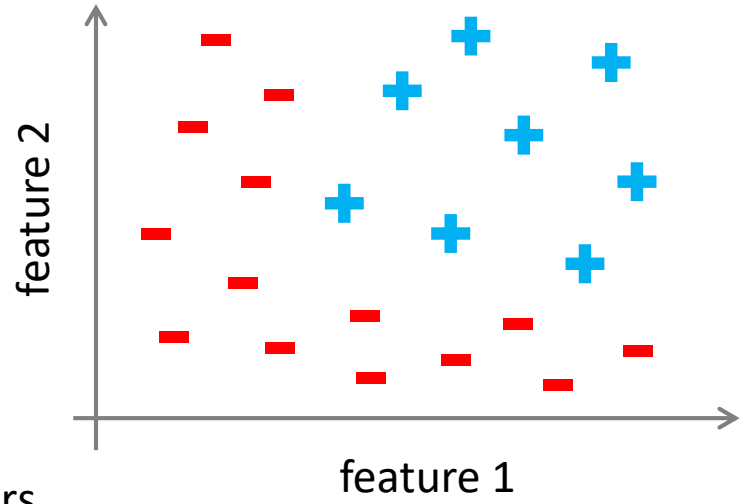
Not a car

Testing:

What is this?



Learning Algorithm



Cars

"Non"-Cars

50 x 50 pixel images→ 2500 pixels

$n = 2500$ (7500 if RGB)

$$x = \begin{bmatrix} \text{pixel 1 intensity} \\ \text{pixel 2 intensity} \\ \vdots \\ \text{pixel 2500 intensity} \end{bmatrix}$$

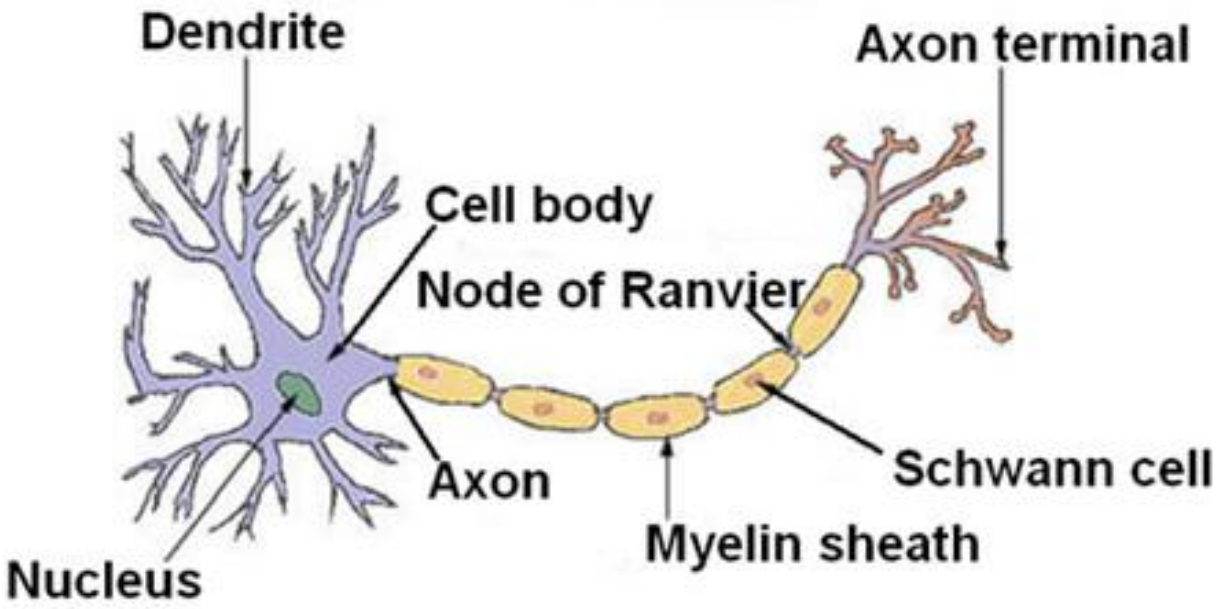
Quadratic features ($x_i \times x_j$): ≈3 million features $O(n^2)$

Neural Networks (NN)

- Origins: Algorithms that try to mimic the brain.
- Was very widely used in 80s and early 90s;
  - popularity diminished in late 90s.
- Recent resurgence: State-of-the-art technique for many applications & due to high computing devices.
- NN is another non-linear classifier
  - Decision boundary is non-linear

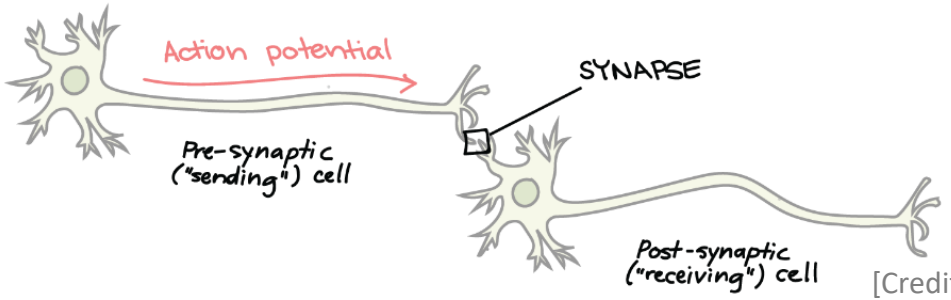
Neuron in the brain

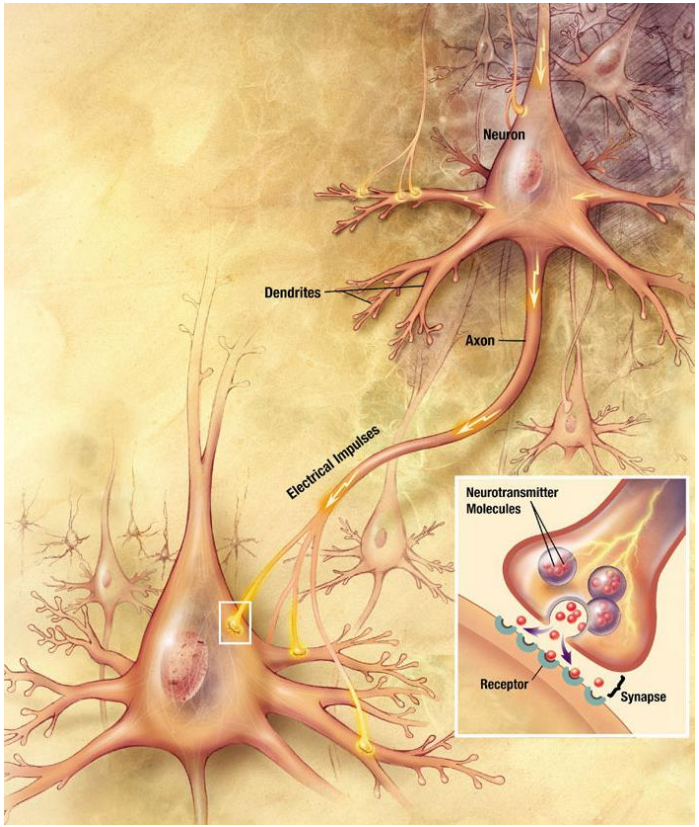
- Biological Neurons are the core components of the human brain.
- A neuron consists of a cell body, dendrites, and an axon.



Neurons in the brain

- Neurons process and transmit information to other neurons by emitting electrical signals.
- Each neuron receives input signals from its dendrites and produces output signals along its axon.
- The axon branches out and connects via synapses to dendrites of other neurons.

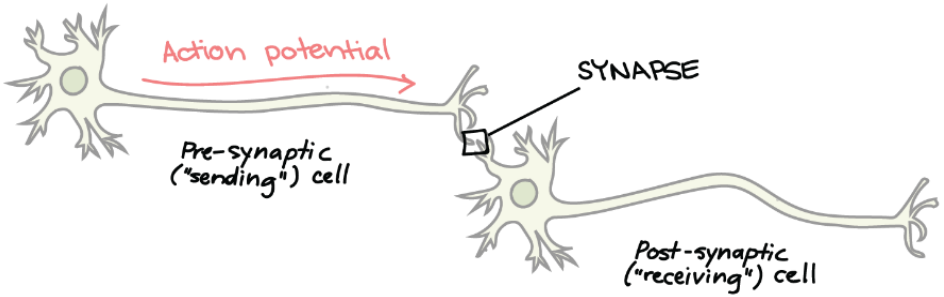


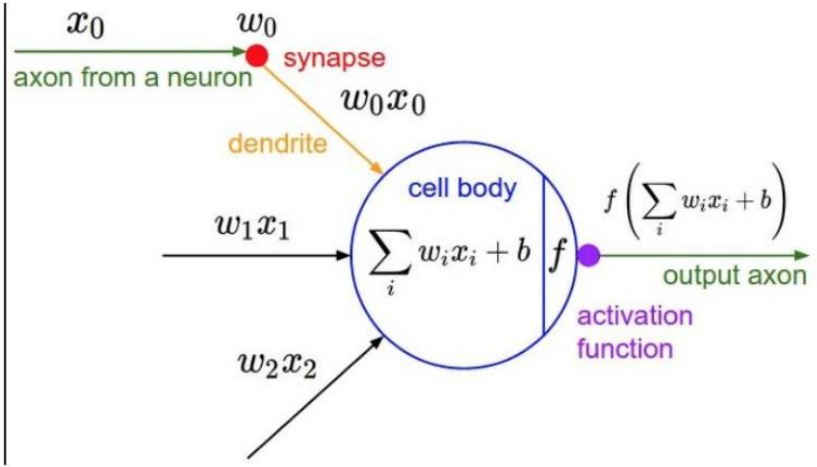


[Credit: US National Institutes of Health, National Institute on Aging]

Biological Neuron VS. Artificial Neuron

- Artificial neurons are inspired by biological neurons
- An artificial neuron has a finite number of inputs with weights associated to them, and an activation function (also called transfer function).
- The output of the neuron is the result of the activation function applied to the weighted sum of inputs.
- Artificial neurons are connected with each others to form artificial neural networks.

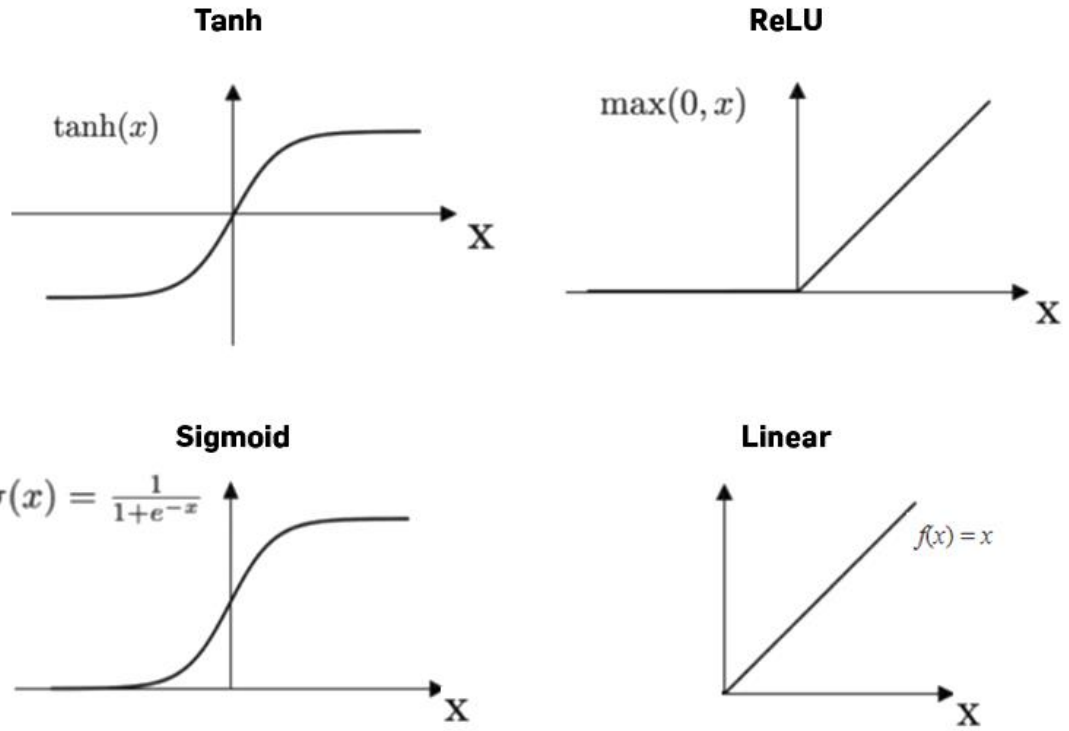




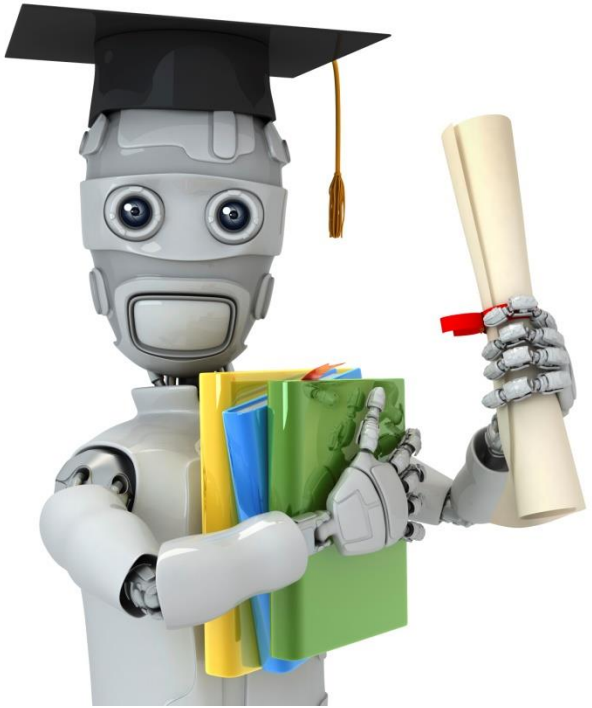
Activation Function

- Activation functions transform the weighted sum of inputs that goes into the artificial neurons.
- These functions should be non-linear to encode complex patterns of the data.
- The most popular activation functions are
  - Logistic
  - Tanh
  - Rectified Linear Unit (RELU)
  - Leaky RELU
  - Exponential Linear Unit (ELU)

Activation Function



Have a look for more: https://en.wikipedia.org/wiki/Activation_function



Machine Learning

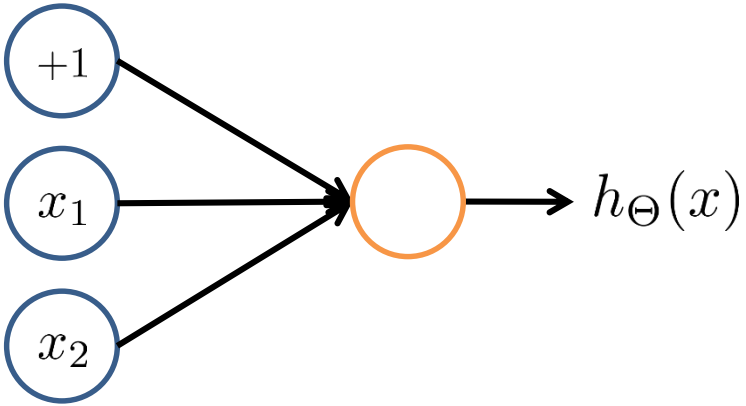
Neural Networks: Representation

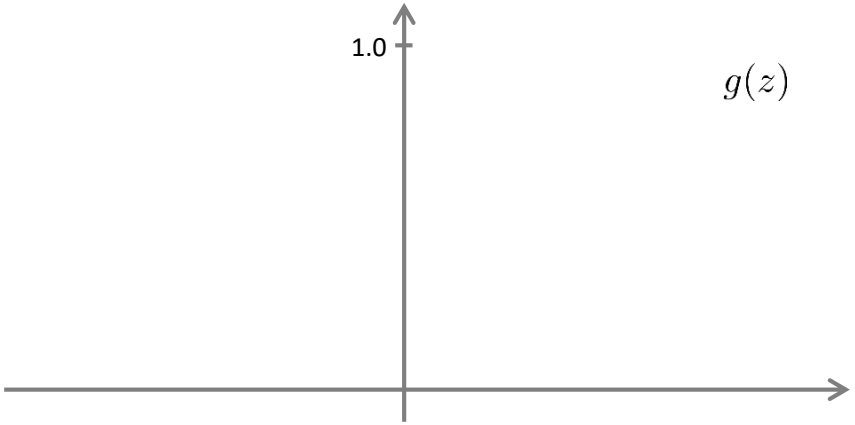
Examples and intuitions II

Simple example: AND

$x_1, x_2 \in \{0, 1\}$

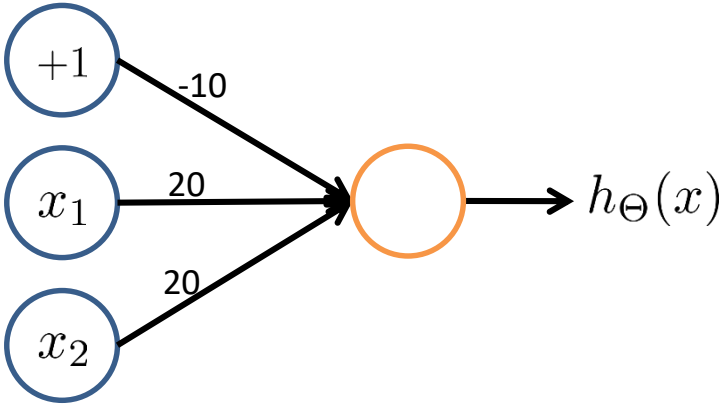
$y = x_1 \text{ AND } x_2$





x_1	x_2	$h_\Theta(x)$
0	0	
0	1	
1	0	
1	1	

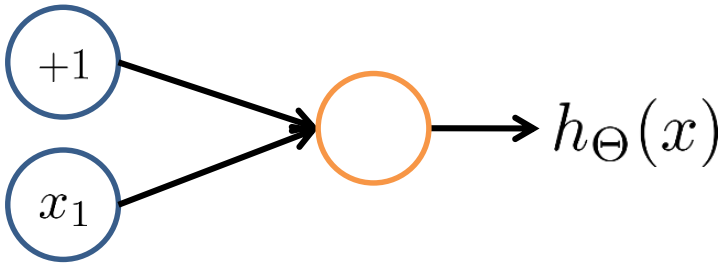
Example: OR function



x_1	x_2	$h_\Theta(x)$
0	0	
0	1	
1	0	
1	1	

x_1 AND x_2 | x_1 OR x_2

Negation:



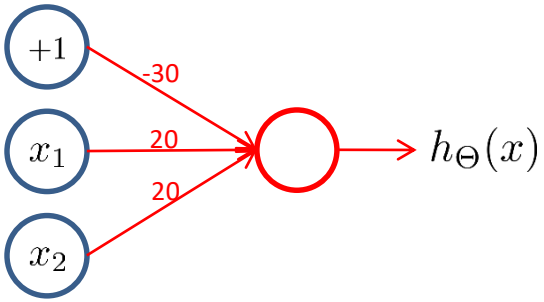
x_1	$h_\Theta(x)$
0	
1	

nor

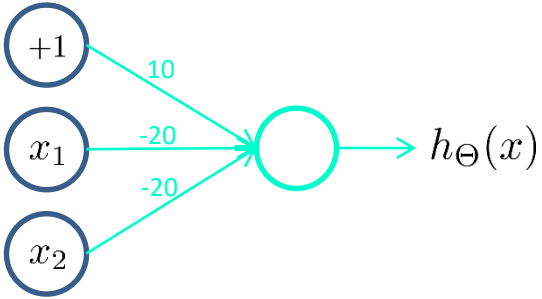
$h_\Theta(x) = g(10 - 20x_1)$

(NOT x_1) AND (NOT x_2)

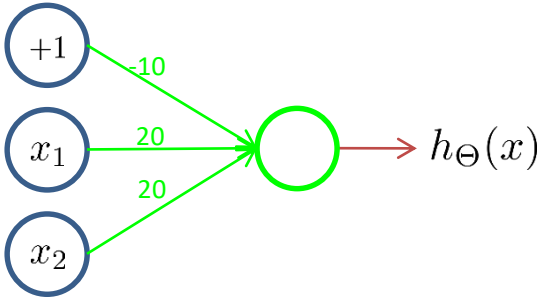
Putting it together: x_1 XNOR x_2



x_1 AND x_2

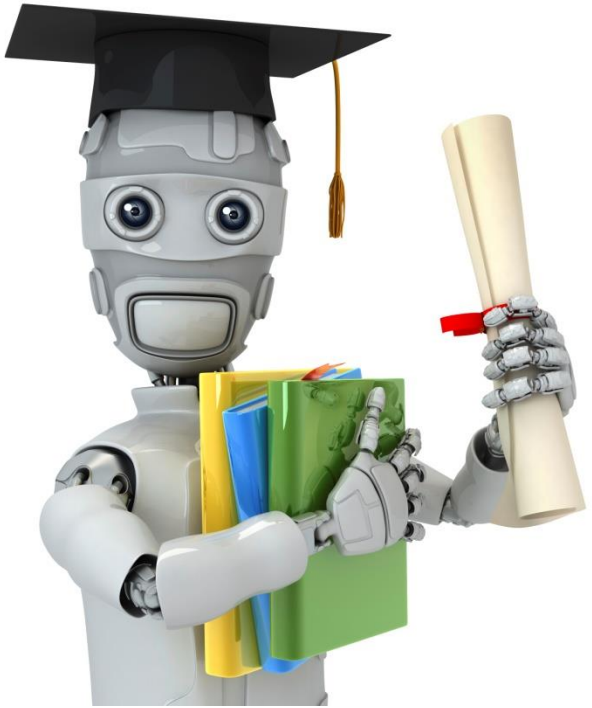


(NOT x_1) AND (NOT x_2)



x_1 OR x_2

x_1	x_2	$a_1^{(2)}$	$a_2^{(2)}$	$h_\Theta(x)$
0	0			
0	1			
1	0			
1	1			

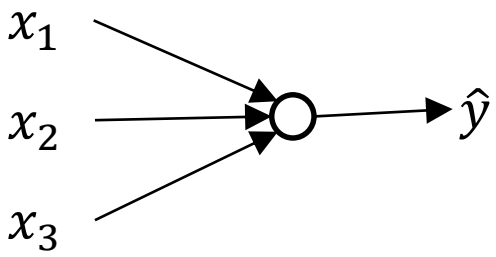


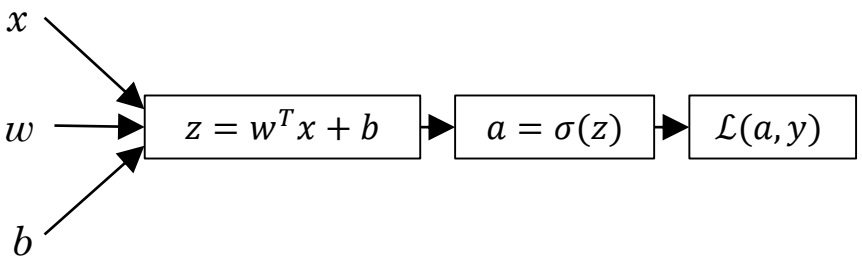
Machine Learning

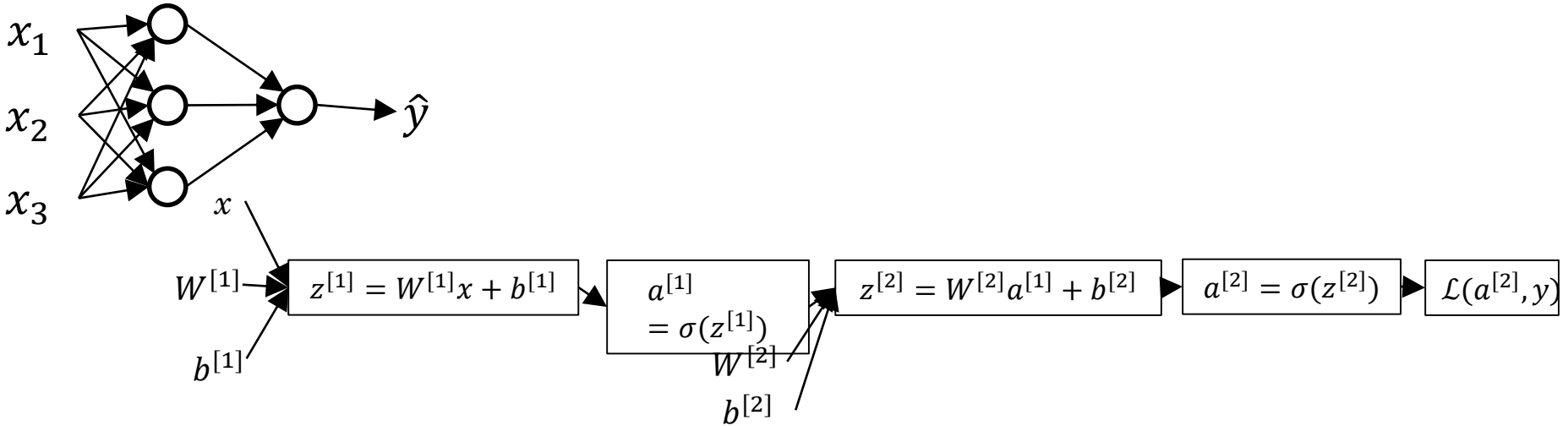
Neural Networks: Representation

Two-class classification

Define a Neural Network



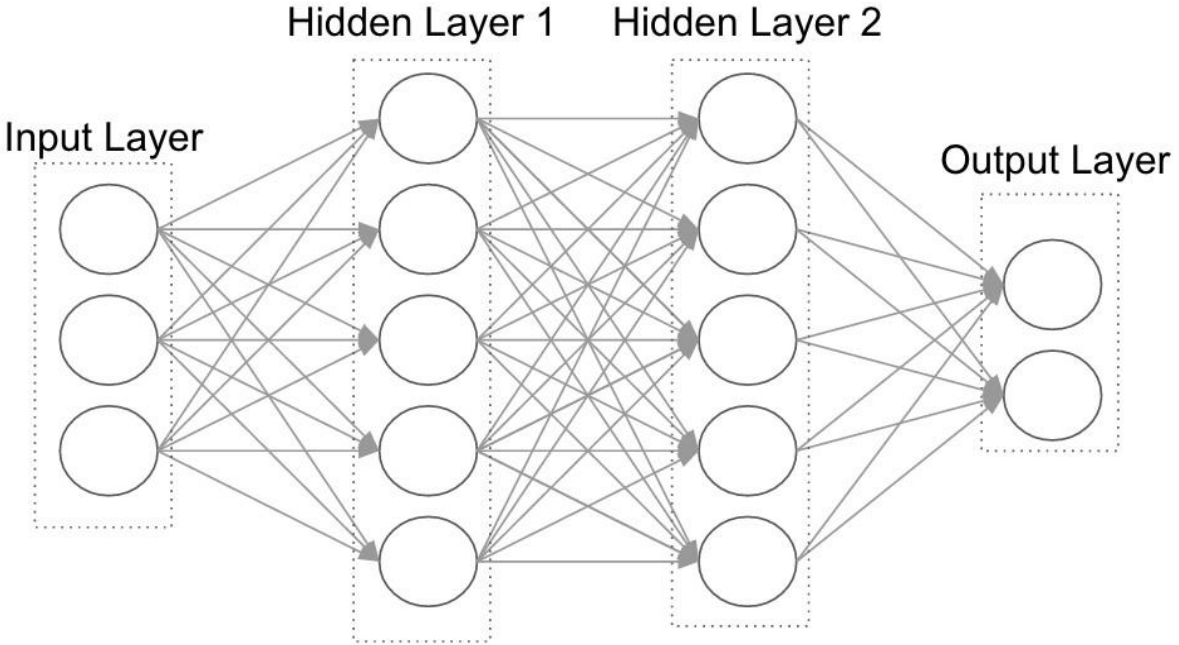




Feed Forward NN

- Feed forward Neural Networks (FFN) are the simplest form of Artificial Neural Networks (ANN).
- These networks have 3 types of layers:
  - Input layer,
  - hidden layer and
  - output layer.
- In FFN, data moves from the input layer through the hidden nodes (if any) and to the output nodes.
- “Fully-connected" means that each node is connected to all the nodes in the next layer.
- The number of hidden layers and their size are the parameters. The larger and deeper the hidden layers, the more complex patterns we can model in theory.

Multi-layer Perceptron (MLP)



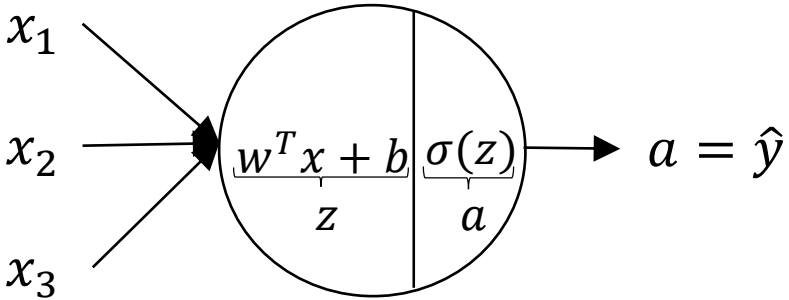
Feedforward neural network with 2 hidden layers

deeplearning.ai

One hidden layer Neural Network

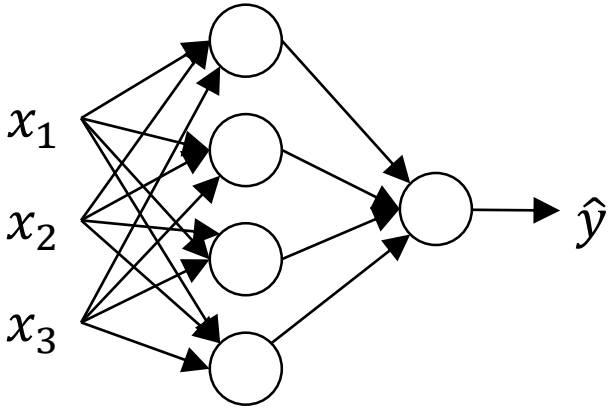
Computing a Neural Network's Output

Neural Network Representation

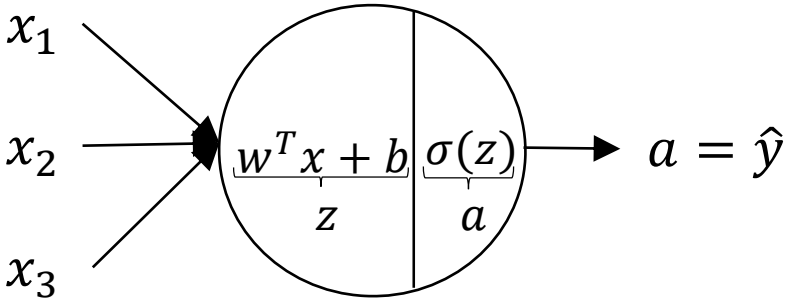


$$z = w^T x + b$$

$$a = \sigma(z)$$

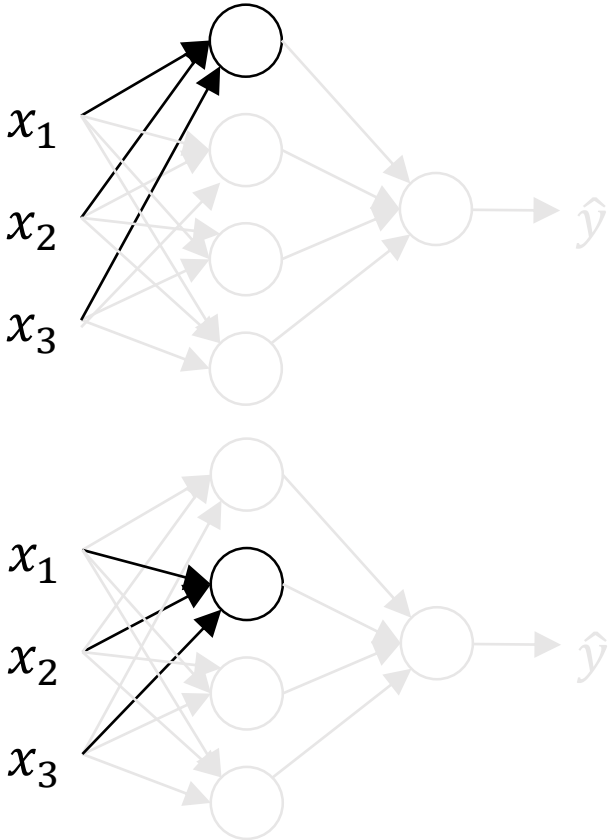


Neural Network Representation

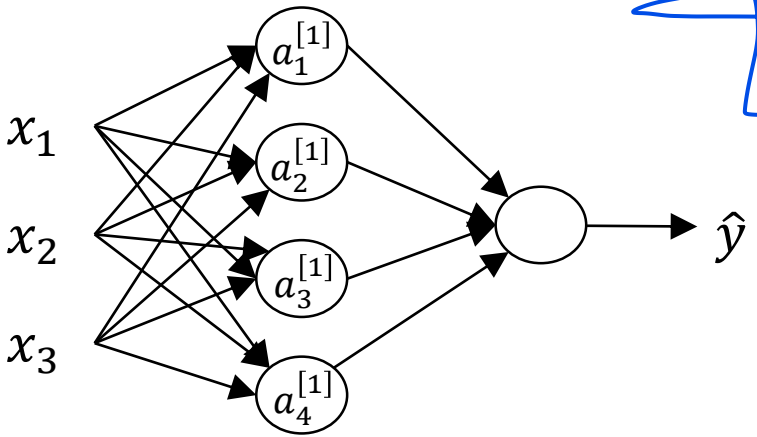


$$z = w^T x + b$$

$$a = \sigma(z)$$



Neural Network Representation



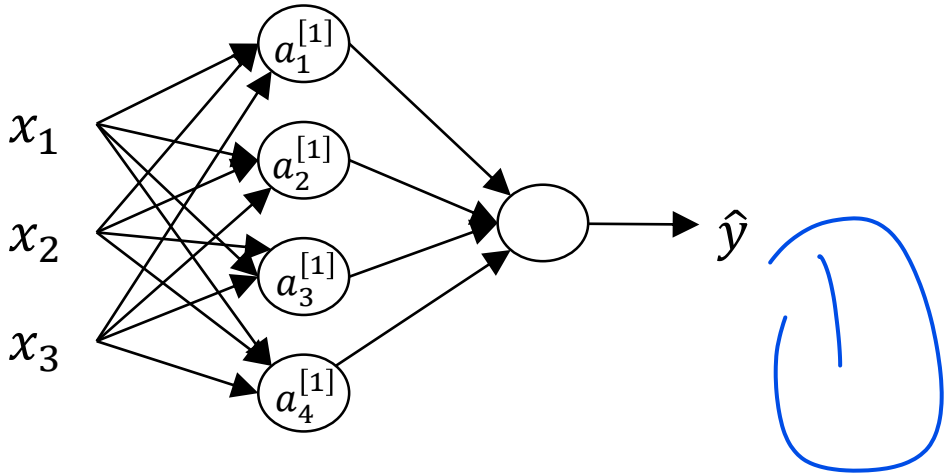
$$z_1^{[1]} = w_1^{[1]T} x + b_1^{[1]},\ a_1^{[1]} = \sigma(z_1^{[1]})$$

$$z_2^{[1]} = w_2^{[1]T} x + b_2^{[1]},\ a_2^{[1]} = \sigma(z_2^{[1]})$$

$$z_3^{[1]} = w_3^{[1]T} x + b_3^{[1]},\ a_3^{[1]} = \sigma(z_3^{[1]})$$

$$z_4^{[1]} = w_4^{[1]T} x + b_4^{[1]},\ a_4^{[1]} = \sigma(z_4^{[1]})$$

Neural Network Representation



Given input x:

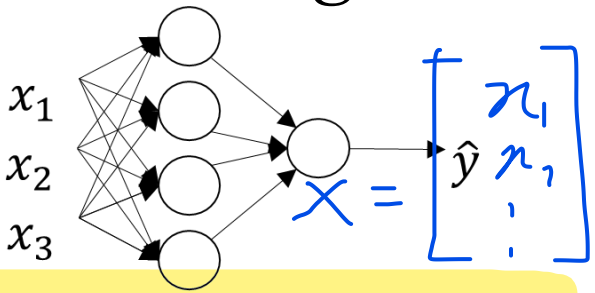
$$z^{[1]} = W^{[1]}x + b^{[1]}$$

$$a^{[1]} = \sigma(z^{[1]})$$

$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$$

$$a^{[2]} = \sigma(z^{[2]})$$

Vectorizing across multiple examples



$$X = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots & x^{(m)} \\ | & | & & | \end{bmatrix}$$

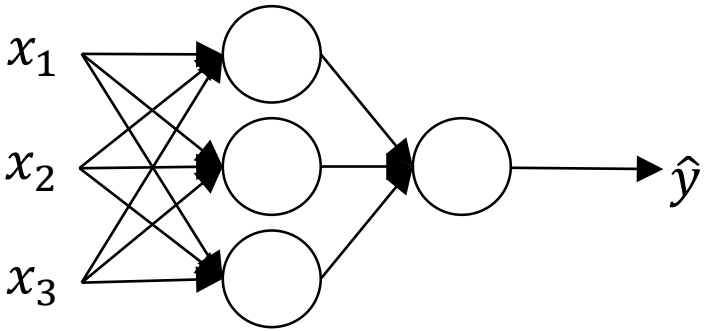
$$A^{[1]} = \begin{bmatrix} | & | & & | \\ a^{[1](1)} & a^{[1](2)} & \cdots & a^{[1](m)} \\ | & | & & | \end{bmatrix}$$

```
for i = 1 to m
```

$$z^{[1](i)} = W^{[1]} x^{(i)} + b^{[1]}$$
$$a^{[1](i)} = \sigma(z^{[1](i)})$$
$$z^{[2](i)} = W^{[2]} a^{[1](i)} + b^{[2]}$$
$$a^{[2](i)} = \sigma(z^{[2](i)})$$

$$Z^{[1]} = W^{[1]} X + b^{[1]}$$
$$A^{[1]} = \sigma(Z^{[1]})$$
$$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$$
$$A^{[2]} = \sigma(Z^{[2]})$$

Why Activation function?



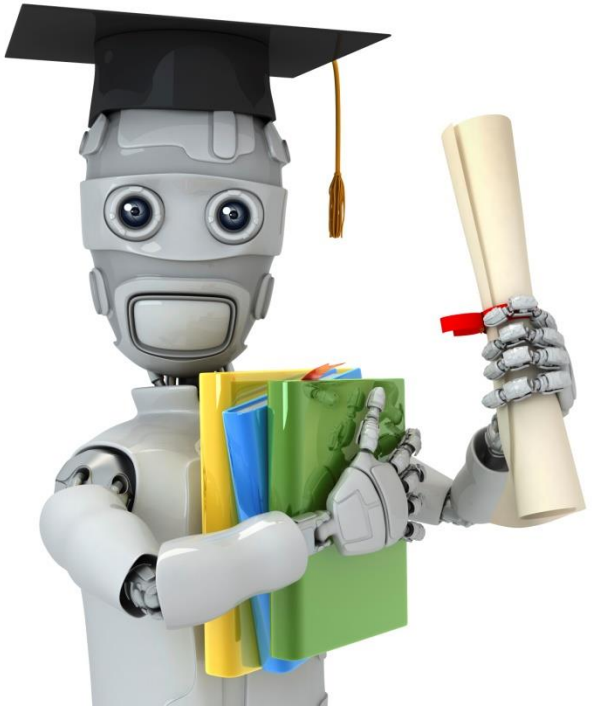
Given x:

$$z^{[1]} = W^{[1]}x + b^{[1]}$$

$$a^{[1]} = g^{[1]}(z^{[1]})$$

$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$$

$$a^{[2]} = g^{[2]}(z^{[2]})$$



Machine Learning

Neural Networks: Representation

Multi-class classification

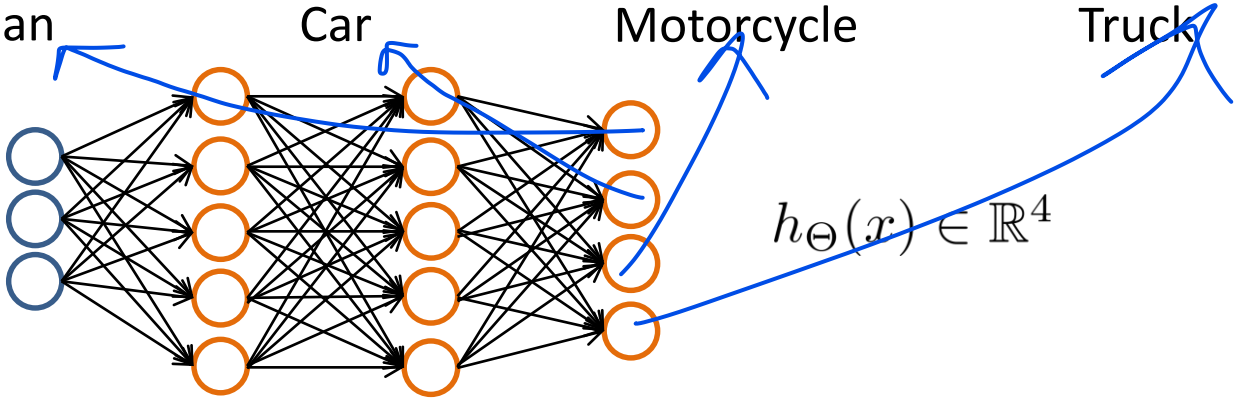
Multiple output units: One-vs-all.

Pedestrian

Car

Motorcycle

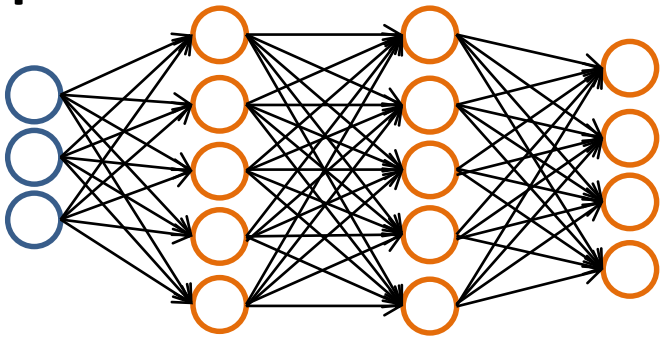
Truck



Want $h_\Theta(x) \approx \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, etc.

when pedestrian when car when motorcycle

Multiple output units: One-vs-all.



$h_\Theta(x) \in \mathbb{R}^4$

Want $h_\Theta(x) \approx \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, etc.

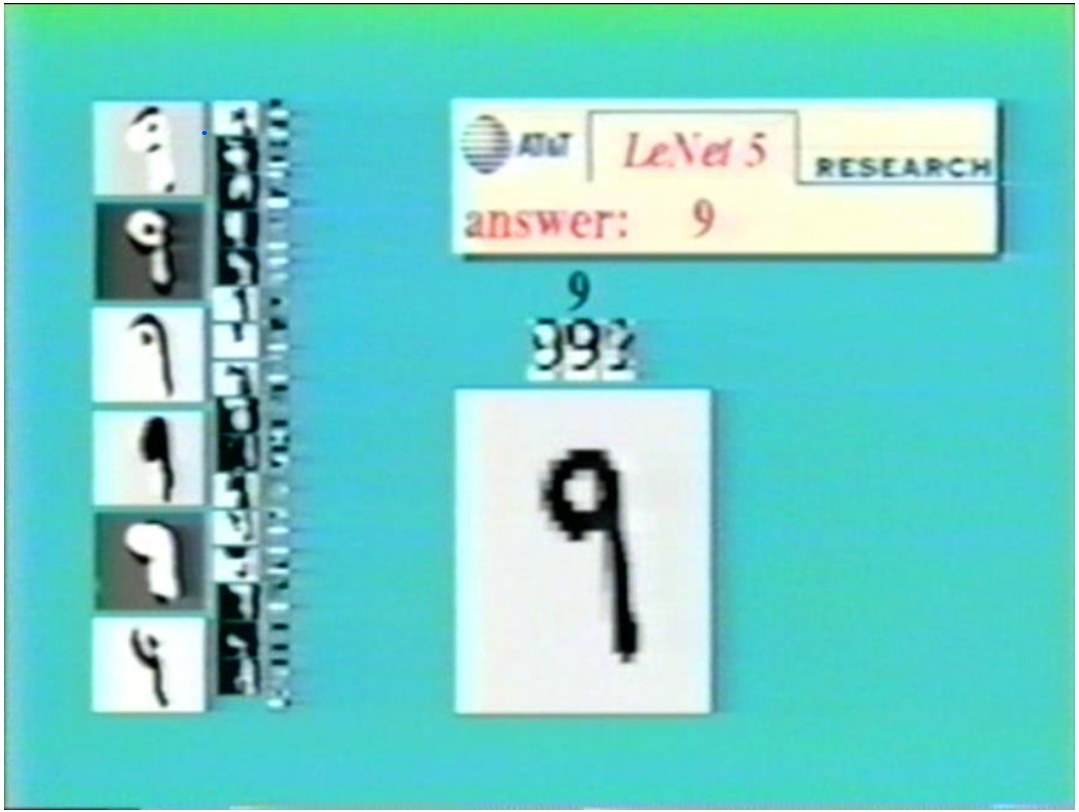
when pedestrian when car when motorcycle

Training set: $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})$

$y^{(i)}$ one of $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

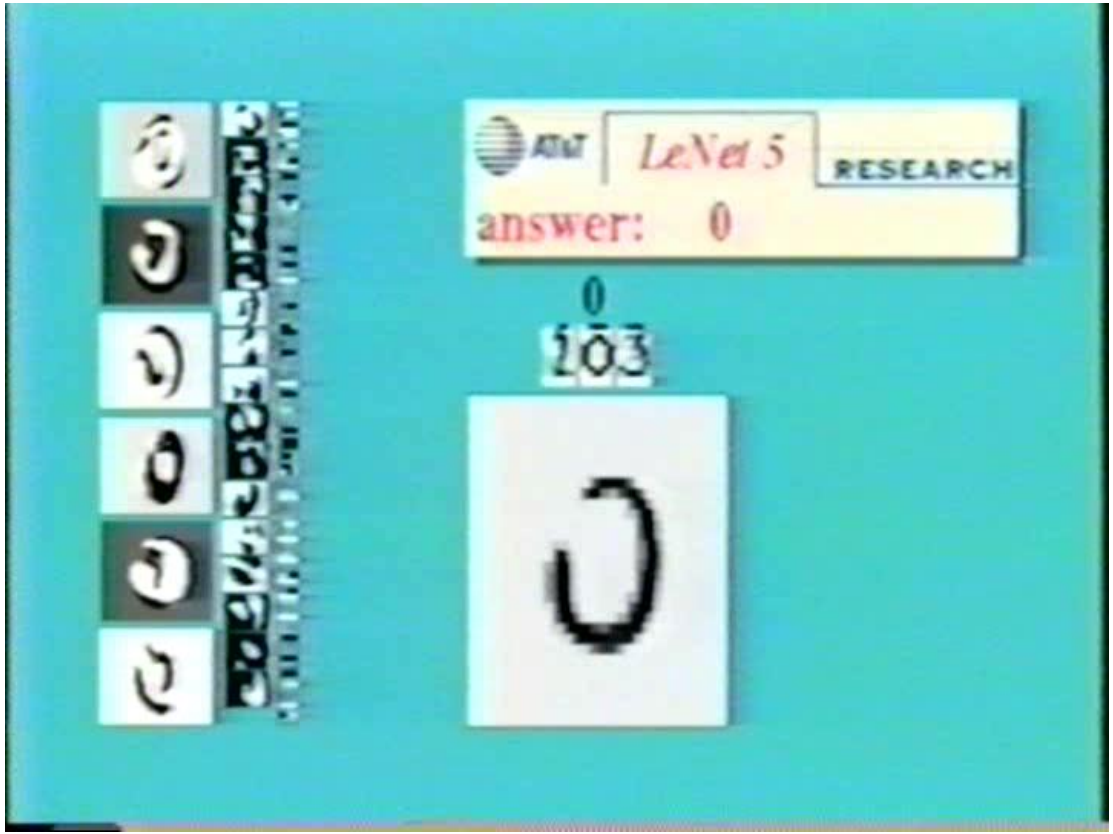
pedestrian car motorcycle truck

Handwritten digit classification with Convolutional NN (CNN)



[Courtesy of Yann LeCun]

Handwritten digit classification with ConvNet



[Courtesy of Yann LeCun]