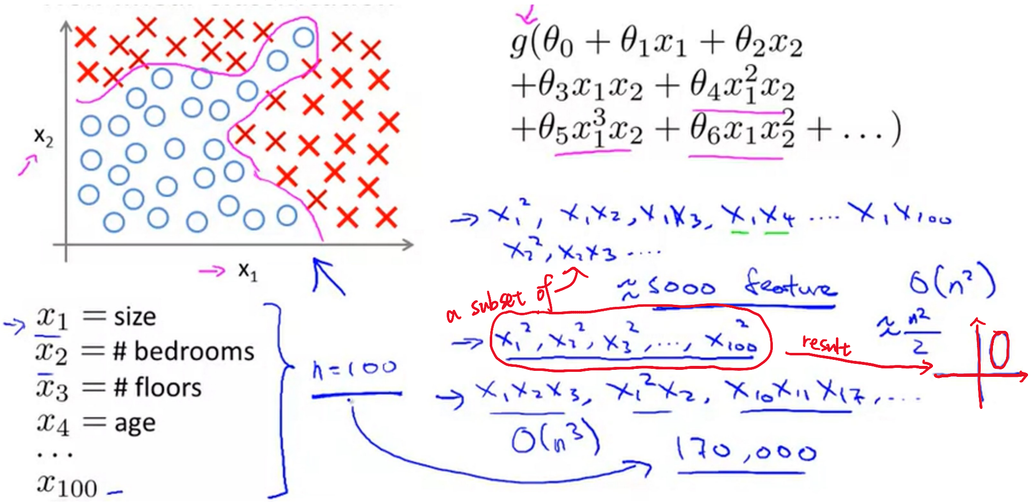
Week 4

- **Motivations:**

  a. Non-linear Hypotheses:



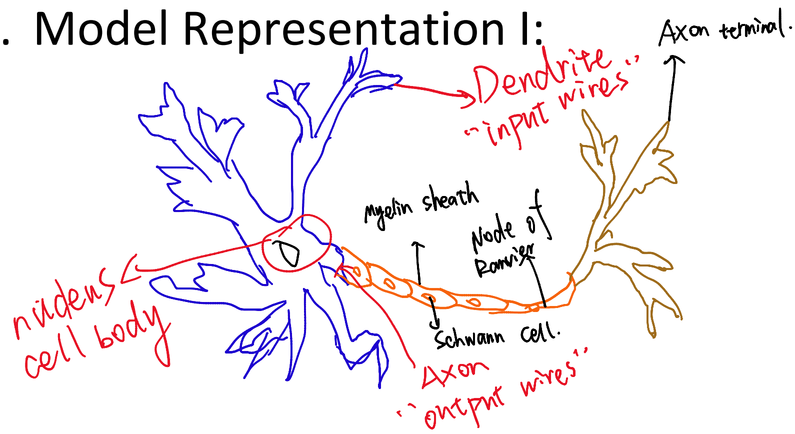
It's not a good way to just add quadratic or cubic features when n is large to fit non-linear hypotheses.

  b. Neurons and the Brain:

Origins: Algorithms that try to mimic the brain. We need to do is figure out some approximation or to whatever the barin's learning algorithm is and implement. Then the brain learned itself how to process these different types of data.

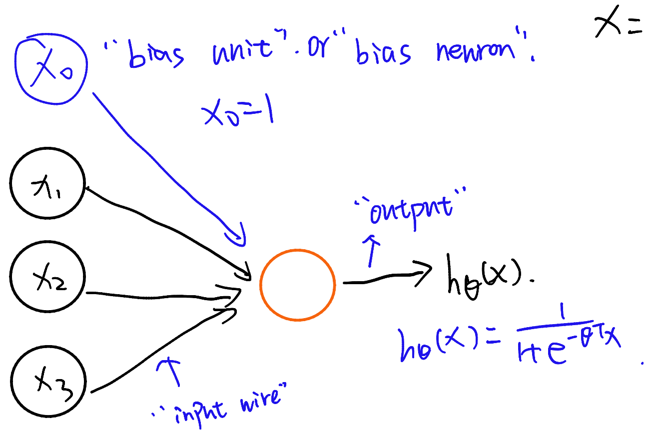
- **Neural Networks:**

  a. Model Representation I:



If neuron want to send some message, axon open wire and connect to the dendrites of second neuron.

logistic unit.

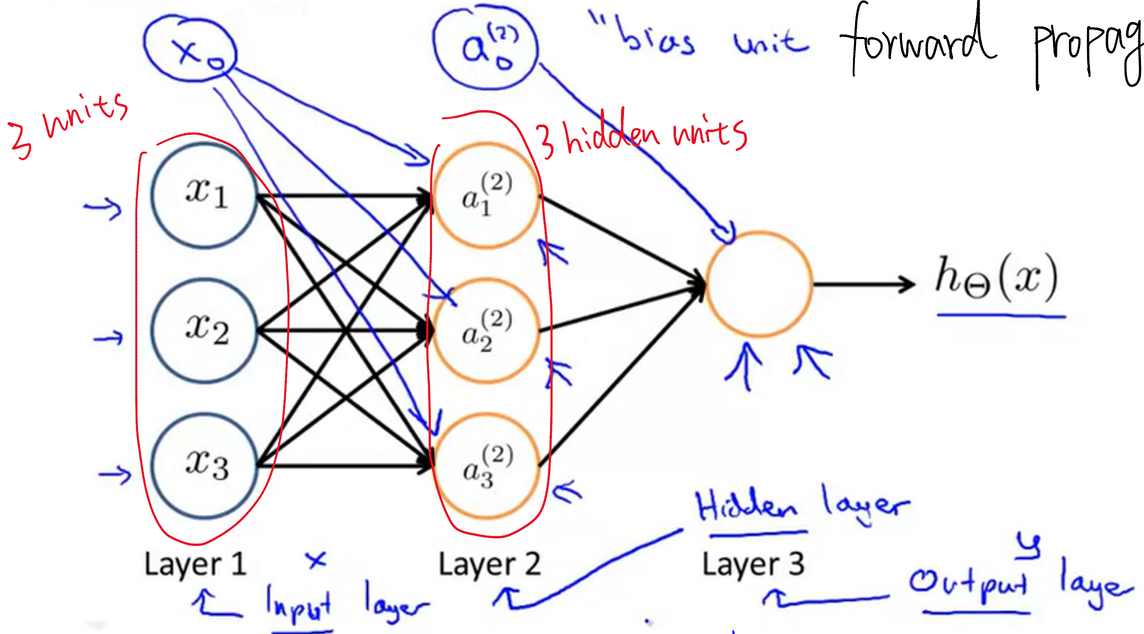


$$x=\begin{bmatrix}x_0\\x_1\\x_2\\x_3\end{bmatrix}\quad \theta=\begin{bmatrix}\theta_0\\\theta_1\\\theta_2\\\theta_3\end{bmatrix}$$

"weight". = "parameters".

Sigmoid(logistic)activation function.

forward propagation.



$a_i^{(j)}$ = "activation" of unit i in layer j. compute the output.

$\Theta^{(j)}$ = matrix of weights controlling function mapping from layer j to layer $j+1$.

$\Theta^{(1)} \in \mathbb{R}^{3\times 4}$. $a_1^{(2)} = (z_1^{(2)})$ layer 2.

$$a_1^{(2)} = g(\Theta_{10}^{(1)}x_0 + \Theta_{11}^{(1)}x_1 + \Theta_{12}^{(1)}x_2 + \Theta_{13}^{(1)}x_3)$$
$$a_2^{(2)} = g(\Theta_{20}^{(1)}x_0 + \Theta_{21}^{(1)}x_1 + \Theta_{22}^{(1)}x_2 + \Theta_{23}^{(1)}x_3)$$
$$a_3^{(2)} = g(\Theta_{30}^{(1)}x_0 + \Theta_{31}^{(1)}x_1 + \Theta_{32}^{(1)}x_2 + \Theta_{33}^{(1)}x_3)$$
$$h_\Theta(x) = a_1^{(3)} = g(\Theta_{10}^{(2)}a_0^{(2)} + \Theta_{11}^{(2)}a_1^{(2)} + \Theta_{12}^{(2)}a_2^{(2)} + \Theta_{13}^{(2)}a_3^{(2)})$$

layer 2 to 3.

If network has s_j units in layer j, s_{j+1} units in layer $j+1$, then $\Theta^{(j)}$ will be of dimension $s_{j+1} \times (s_j + 1)$.

$S_{j+1} \times (S_j + 1)$

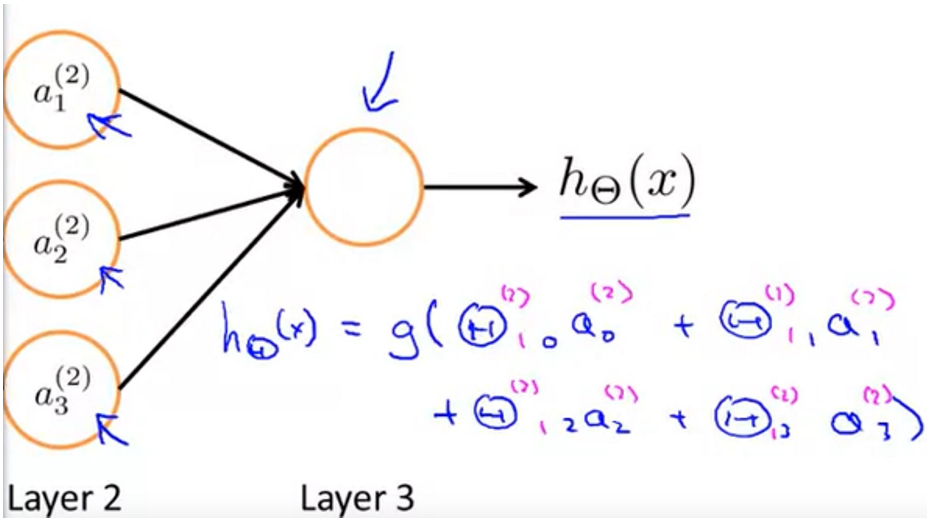
a. Model Representation II:

like last part. $a_1^{(2)}=g(z_1^{(2)})$ $a_2^{(2)}=g(z_2^{(2)})$ $a_3^{(2)}=g(z_3^{(2)})$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \hookrightarrow \mathbb{R}^4 \qquad z^{(2)} = \begin{bmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{bmatrix} \hookrightarrow \mathbb{R}^3 \qquad z^{(2)} = \Theta^{(1)} \overset{a^{(1)}}{x}. \quad a^{(2)} = \boxed{g(z^{(2)})} \hookrightarrow \mathbb{R}^3 \to \text{we get } a_1^{(2)}, a_2^{(2)}, a_3^{(2)}.$$

$a^{(1)} = x.$

Add $a_0^{(2)} = 1$ $\quad z^3 = \Theta^{(2)} a^{(2)} \quad h_\Theta(x) = a^{(3)} = g(z^{(3)}).$



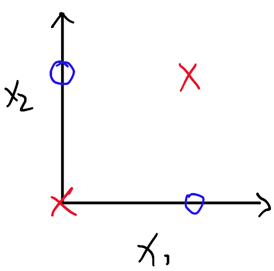
If we just look at these two layers, it looks like a logistic regression model.(just the input x1,x2,x3 was insteaded by a1, a2, a3.)

$$a^{(j)} = g(z^{(j)}) \qquad z^{(j+1)} = \Theta^{(j)} a^{(j)} \qquad h_\Theta(x) = a^{(j+1)} = g(z^{(j+1)}).$$

- **Applications:**

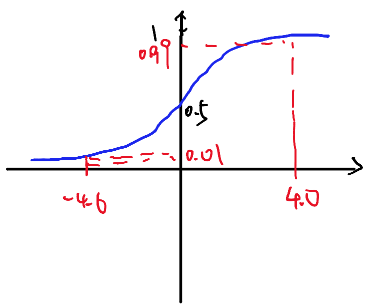
  a. Examples and Intuitions I:

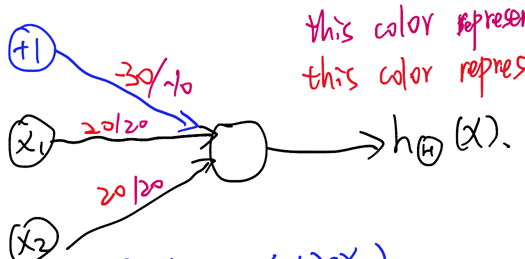
x_1, x_2 are binary (0 or 1).



$y = x_1 \text{ XOR } x_2$
$= x_1 \text{ XNOR } x_2$
$= \text{NOT } (x_1 \text{ XOR } x_2).$ $\Rightarrow$ we want to fit this XOR function.

Simple example: AND.
$x_1, x_2 \in \{0, 1\}$ $\quad y = x_1 \text{ AND } x_2$





this color represent OR
this color represent AND.

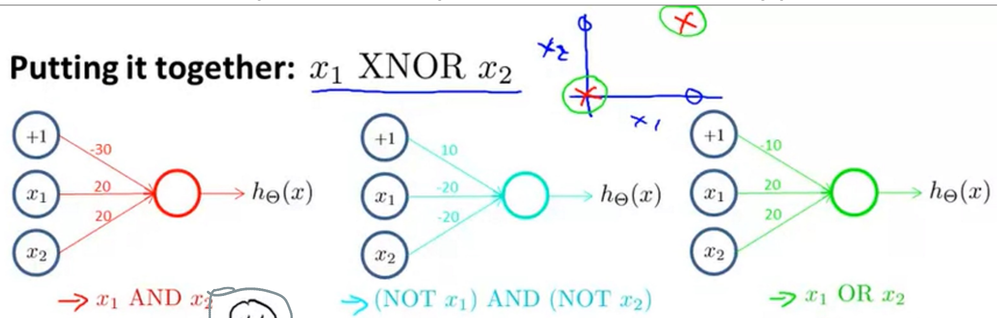
$h_\Theta(x) = g(-30+20x_1+20x_2)$

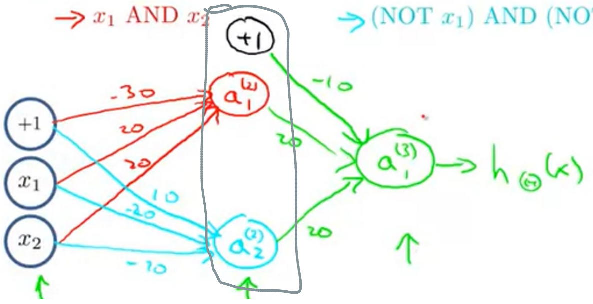
x_1	x_2	$h_\Theta(x)$
0	0	≈ 0
0	1	≈ 0
1	0	≈ 0
1	1	≈ 1.

a. Examples and Intuitions II:

How to compute complex non-linear hypothesis.

Putting it together: $x_1 \text{ XNOR } x_2$



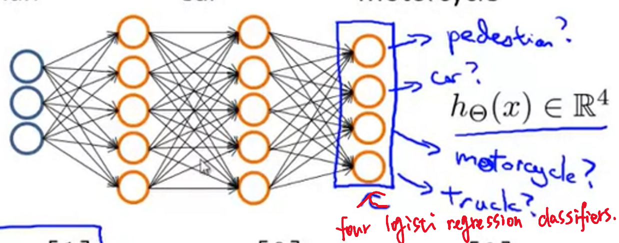


x_1	x_2	$a_1^{(2)}$	$a_2^{(2)}$	$h_\Theta(x)$
0	0	0	1	1
0	1	0	0	0
1	0	0	0	0
1	1	1	0	1

b. Multiclass 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

previously. $y \in \{1,2,3,4\}$ now.

$y^{(i)}$ is one of $\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$. $\rightarrow h_\Theta(x^{(i)}) \approx y^{(i)}$

$\mathbb{R}^4$

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

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

pedestrian car motorcycle truck