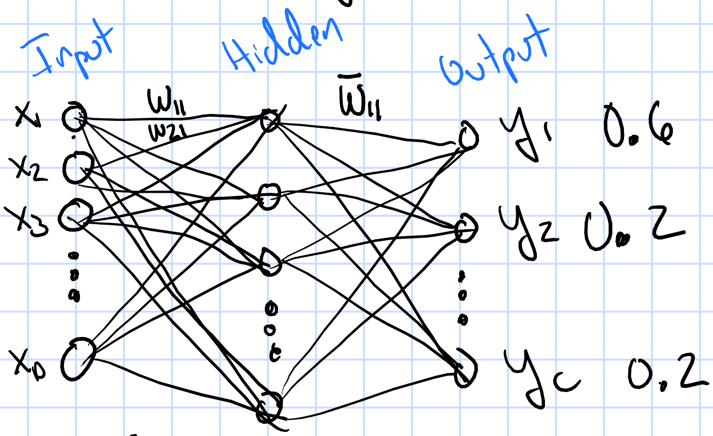
Neural Nets (Part I)

An artificial neural network (ANN) is a highly interconnected network of information processing elements that "mimic" the connectivity and functionality of the human brain.



Definition

$$D = \{x(n), d(n)\}_{n=1}^{N}$$

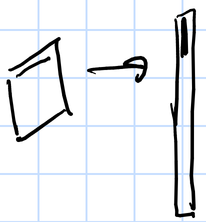
input (feature vector)

desired output

e.g. $d(n) \in \{1, 0\}$

$d(n) \in \{1, 0\}^C$

$d(n) \in \mathbb{R}^K$



$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

We can calculate the error of each output:

$$e_j(n) = d_j(n) - y_j(n) \qquad (2)$$

The instantaneous error energy @ the j^{th} neuron

$$E_j(n) = \frac{1}{2} e_j^2(n)$$

x o

o x

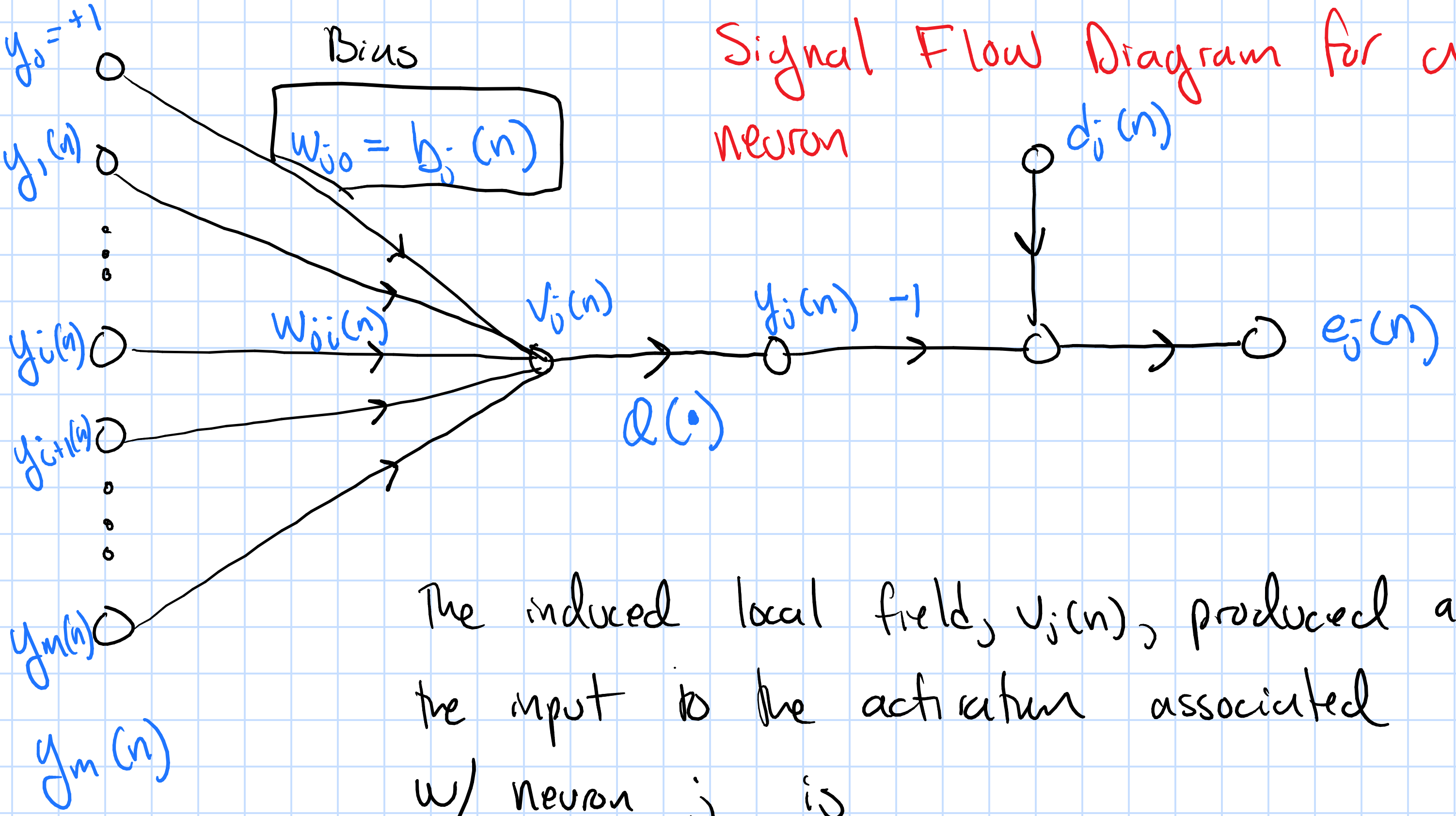
Total error

$$E(n) = \sum_{j=1}^{C} E_j(n) = \frac{1}{2} \sum_{j=1}^{C} e_j^2(n) \qquad (1)$$

Overall samples

$$E_{AV}(N) = \frac{1}{N} \sum_{n=1}^{N} E(n) = \frac{1}{2N} \sum_{n=1}^{N} \sum_{j=1}^{C} e_j^2(n)$$

Signal Flow Diagram for a neuron



The induced local field, $v_j(n)$, produced at the input to the activation associated w/ neuron j is

$$v_j(n) = \sum_{i=1}^{m} w_{ji}(n)\, y_i(n) \qquad (4)$$

The signal at the neuron is

$$y_j(n) = \varphi(v_j(n)) \qquad (3)$$

We need to determine an update rule for the weights

$$\frac{\partial E}{\partial w_{ji}}$$

<u>Recall</u>: $w(t+1) = w(t) + \Delta w(t)$

$$\Delta w(t) \rightarrow -\eta \frac{\partial E}{\partial w}$$

The chain rule from calculus I

$$\frac{\partial E}{\partial w_{ji}(n)} = \frac{\partial E}{\partial e_j(n)} \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial v_j(n)} \cdot \frac{\partial v_j(n)}{\partial w_{ji}(n)}$$

Diff (1) wrt $e_j(n)$

$$\frac{\partial E}{\partial e_j(n)} = e_j(n)$$

Diff (2) wrt $y_j(n)$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1$$

Diff (3) wrt $v_j(n)$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi_j'(v_j(n))$$

Diff (4) wrt $w_{ji}(n)$

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n)$$

$$\frac{\partial E}{\partial w_{ji}(n)} = -e_j(n)\,\varphi_j'(v_j(n))\,y_i(n) \rightarrow \Delta w_{ji}(n) = -\eta \frac{\partial E}{\partial w_{ji}}$$