Neural Networks

EE698V - Machine Learning for Signal Processing

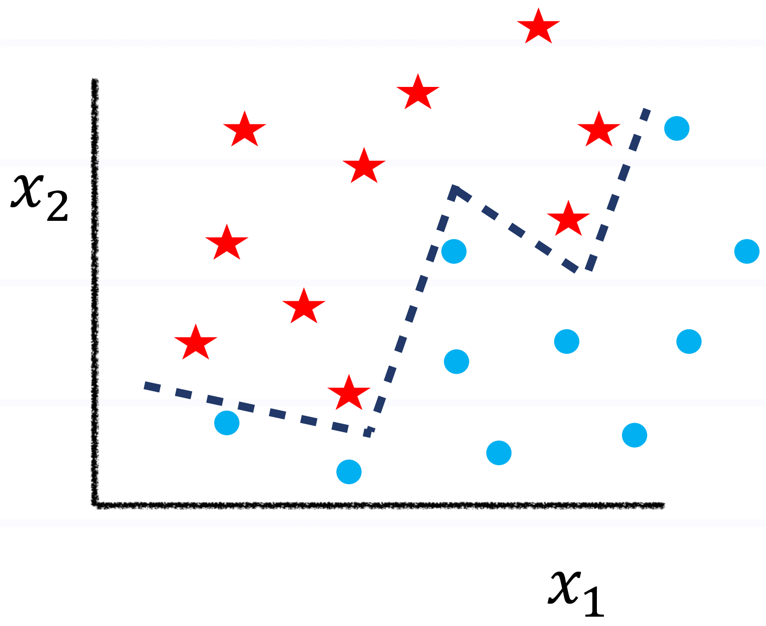
Vipul Arora

Announcements

- Quiz on Friday
  - Everything covered until ass2 (i.e., up to lecture 6)

Non-linear Regression

- Remember, non-linear models can be more powerful than linear ones



A simple non-linear model

$$\underbrace{y}_{(N_2,1)} = \underbrace{W_2}_{(N_2,N_1)} \sigma\left(\underbrace{W_1}_{(N_1,N_0)} \underbrace{x}_{(N_0,1)}\right)$$

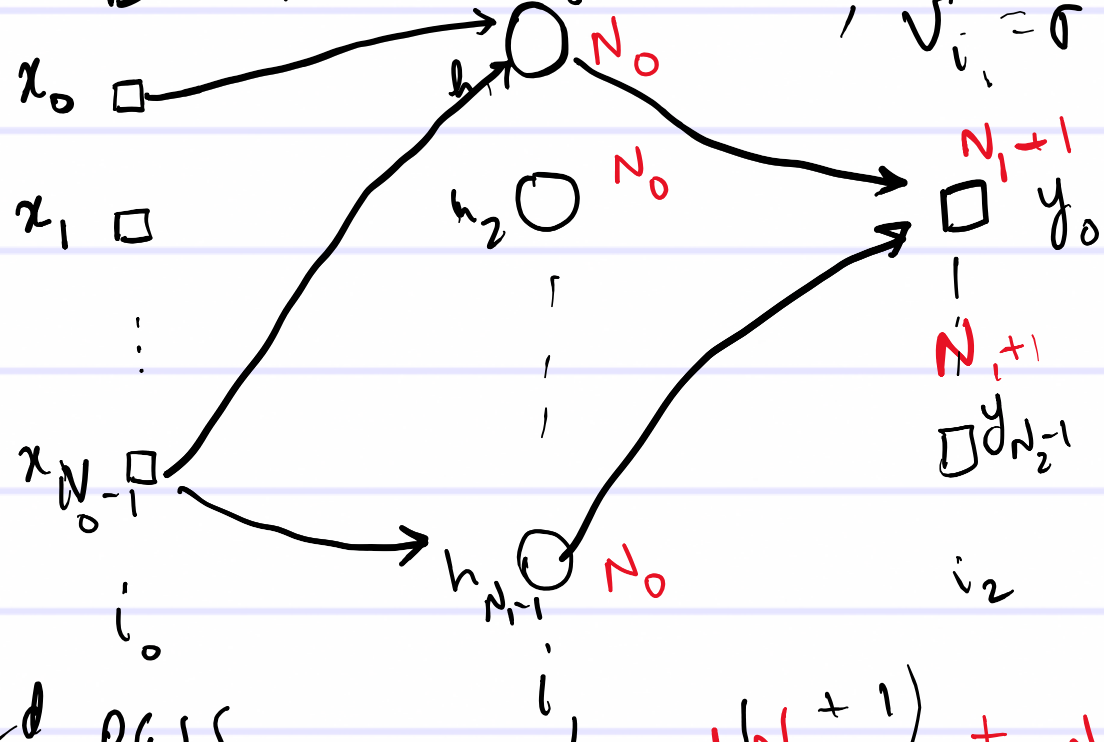
$$y_{i_2} = \sum_{i_1=1}^{N_1} w_{i_2 i_1} \sigma\left(\sum_{i_0=1}^{N_0} w_{i_1 i_0} x_{i_0}\right)$$

$$y_{i_2} = \sum_{i_1=1}^{N_1} w_{i_2 i_1} \overbrace{\sigma\left(\underbrace{\sum_{i_0=1}^{N_0} w_{i_1 i_0} x_{i_0}}_{h}\right)}^{v}$$

$$\phi_{i_0} \sim x_{i_0}$$

$$h_0 = w_{00}x_0 + w_{01}x_1 + w_{02}x_2 + \dots + w_{0N_0-1}x_{N_0-1}$$

$$h_{i_1} = \sum_{i_0} w_{i_1 i_0} x_{i_0} + w_{i_1}, \quad v_{i_1} = \sigma(h_{i_1})$$



$$y_0 = w'_{00}v_0 + w'_{01}v_1 + \dots + w'_{0N_1-1}v_{N_1-1} + w'_0$$

$$y_{i_2} = \sum_{i_1} w_{i_2 i_1} v_{i_1} + w_{i_2}$$

Forward pass

$$N_1(N_0+1) + N_2(N_1+1)$$

$N_0 = 10$
$N_1 = 100$
$N_2 = 10$

params $= 1100 + 1010$

How to find optimal model parameters?

$$E = \sum_n \sum_{i_2} \left(t_{i_2} - y_{i_2}\right)^2 \qquad y_{i_2} = \sum_{i_1} w_{i_1 i_2} v_{i_1}$$

$$\frac{\partial E}{\partial w_{i_1' i_2'}} = -\sum_n 2\left(t_{i_2} - y_{i_2}\right) \underbrace{\frac{\partial y_{i_2}}{\partial w_{i_1' i_2'}}}_{v_{i_1'}} \Bigg|_{w^{(0)}}$$

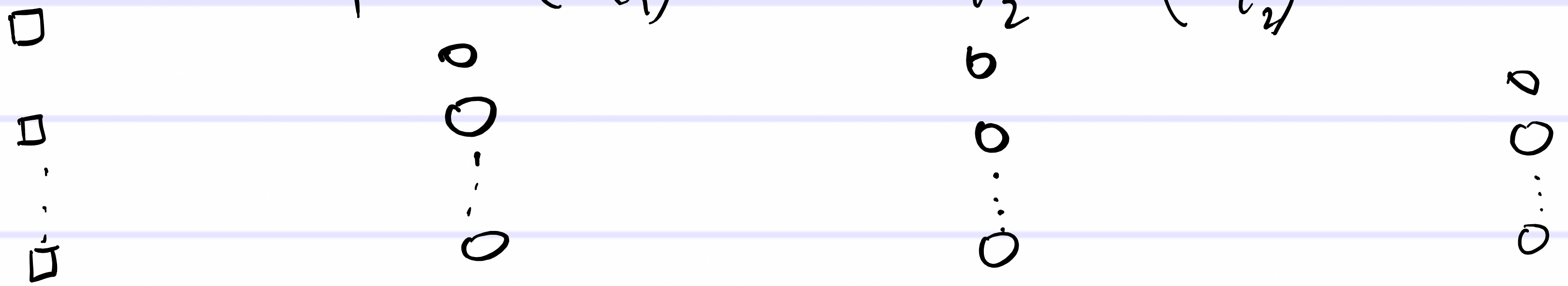
$i_1 = 0$

$i_2 = 0$

$$w_{i_1' i_2'}^{new} \leftarrow w_{i_1' i_2'}^{old} - \eta \frac{\partial E}{\partial w_{i_1' i_2'}} \Bigg|_{w^{old}}$$

x_{i_0} $\quad h_{i_1} = \sum_{i_0} W_{i_1 i_0} x_{i_0}$ $\quad h_{i_2} = \sum_{i_1} W_{i_2 i_1} v_{i_1}$ $\quad h_{i_3} = y_{i_3} = \sum_{i_2} W_{i_3 i_2} v_{i_2}$

$v_{i_1} = \sigma(h_{i_1})$ $\quad v_{i_2} = \sigma(h_{i_2})$



$$E^{(n)} = \sum_{i_3} \left(t_{i_3}^{(n)} - y_{i_3}^{(n)}\right)^2 \cdot \frac{1}{2}$$

this is error for single sample n. Sum over all training samples finally.

$$E = \sum_n E^{(n)}$$

To update weights we need gradients.

$$w_{ij}^{(new)} \leftarrow w_{ij}^{(old)} - \eta \left.\frac{\partial E^{(n)}}{\partial w_{ij}}\right|_{w^{(old)}}$$

i) For $w_{i_3 i_2}$:-

omit (n) for brevity

$$\frac{\partial E^{\cancel{(n)}}}{\partial w_{i_3' i_2'}} = \frac{\cancel{2}}{\cancel{2}} \left(t_{i_3'} - y_{i_3'}\right) \underbrace{\left(\frac{-\partial y_{i_3'}}{\partial w_{i_3' i_2'}}\right)}_{= -v_{i_2'}}$$

ii) for $w_{i_2 i_1}$:-

$$y_{i_3} = \sum_{i_2} w_{i_3 i_2} v_{i_2}$$

$$\frac{\partial E}{\partial w_{i_2' i_1'}} = \sum_{i_3} \frac{2}{2}\left(t_{i_3} - y_{i_3}\right) \underbrace{\left(-\frac{\partial y_{i_3}}{\partial w_{i_2' i_1'}}\right)}_{= w_{i_3 i_2'} \left(\frac{\partial v_{i_2'}}{\partial w_{i_2' i_1'}}\right)}$$

$$v_{i_2'} = \sigma\left(h_{i_2'}\right)$$

$$\therefore \frac{\partial v_{i_2'}}{\partial w_{i_2' i_1'}} = \sigma\left(h_{i_2'}\right)\left(1 - \sigma\left(h_{i_2'}\right)\right) \underbrace{\frac{\partial h_{i_2'}}{\partial w_{i_2' i_1'}}}_{= v_{i_1'}}$$

Consider

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\therefore \frac{\partial \sigma(x)}{\partial x} = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \sigma(x)\left(1 - \sigma(x)\right)$$

$$\frac{\left(1+e^{-x}\right) - 1}{\left(1+e^{-x}\right)^2}$$

$$= \frac{1}{1+e^{-x}} - \frac{1}{\left(1+e^{-x}\right)^2}$$

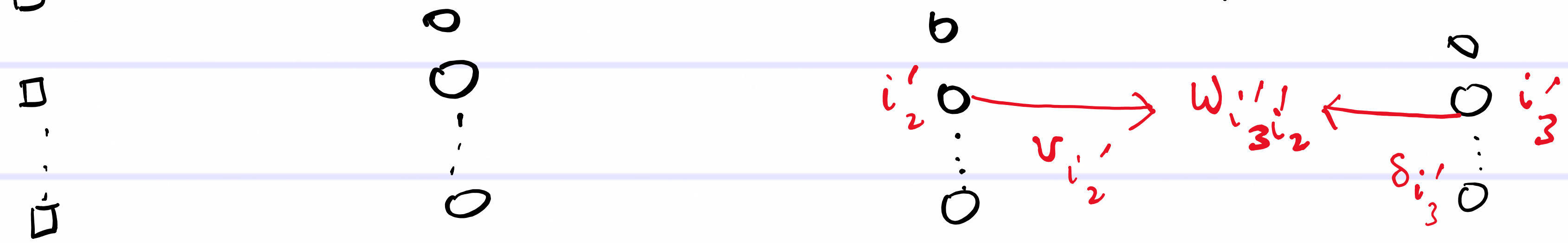
Back Propagation Algorithm

x_{i_0}

$$h_{i_1} = \sum_{i_0} w_{i_1 i_0} x_{i_0} \qquad v_{i_1} = \sigma(h_{i_1})$$

$$h_{i_2} = \sum_{i_1} w_{i_2 i_1} v_{i_1} \qquad v_{i_2} = \sigma(h_{i_2})$$

$$h_{i_3} = y_{i_3} = \sum_{i_2} w_{i_3 i_2} v_{i_2}$$

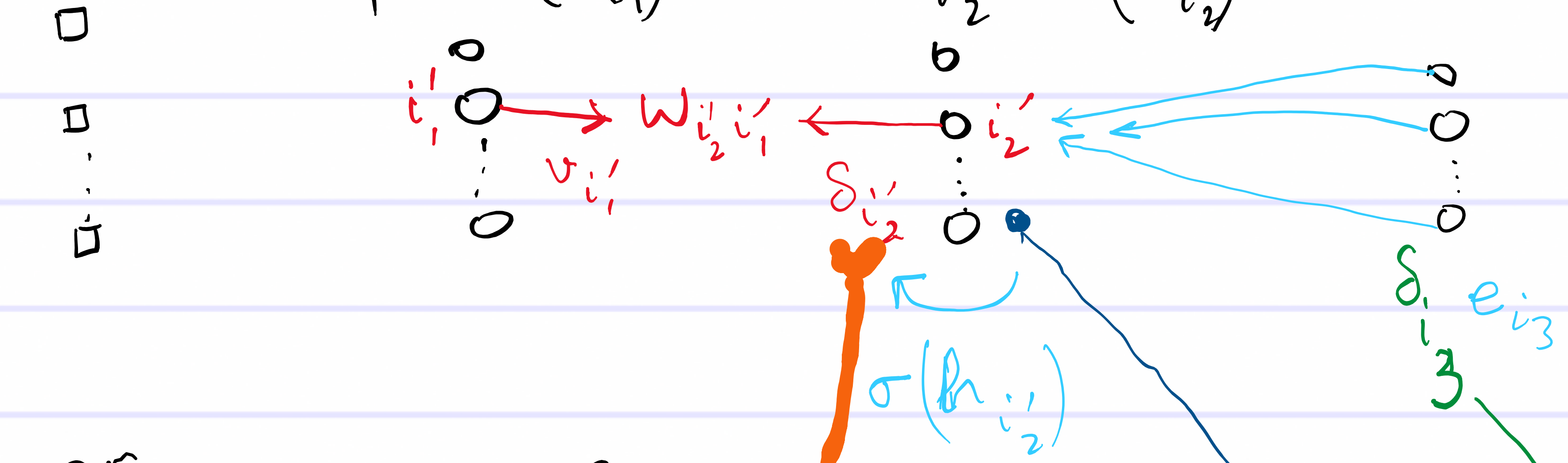


$$e_{i_3'} = -(t_{i_3'} - y_{i_3'})$$

$$\frac{\partial E}{\partial w_{i_3' i_2'}} = v_{i_2'} \cdot \underbrace{\delta_{i_3'}}_{= e_{i_3'}}$$

x_{i_0} $\qquad h_{i_1} = \sum_{i_0} w_{i_1 i_0} x_{i_0}$ $\qquad h_{i_2} = \sum_{i_1} w_{i_2 i_1} v_{i_1}$ $\qquad h_{i_3} = y_{i_3} = \sum_{i_2} w_{i_3 i_2} v_{i_2}$

$v_{i_1} = \sigma(h_{i_1})$ $\qquad v_{i_2} = \sigma(h_{i_2})$



$$\frac{\partial E}{\partial w_{i_2' i_1'}} = v_{i_1'} \cdot \underbrace{\delta_{i_2'}}$$

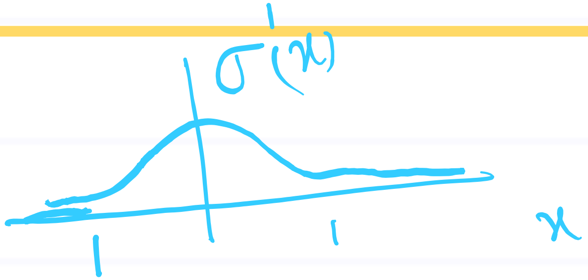
$$\delta_{i_2'} = \left(\sigma(h_{i_2'})\,(1-\sigma(h_{i_2'})) \cdot \left(\sum_{i_3} w_{i_3 i_2'} \underbrace{(\delta_{i_3})}_{e_{i_3}} \right) \right)$$

References

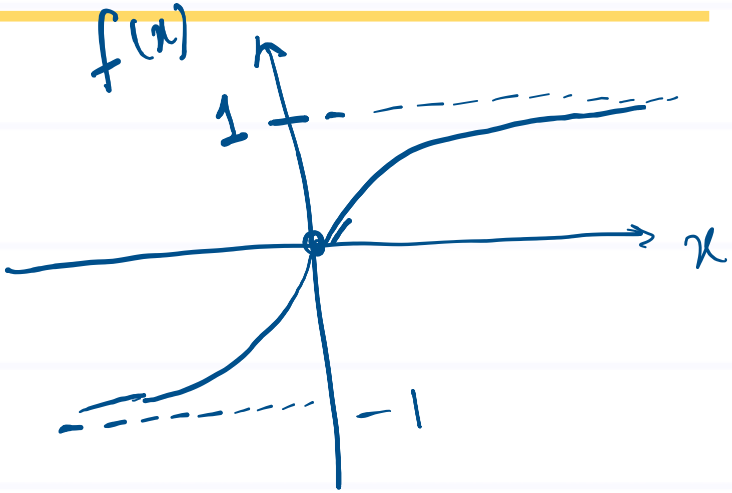
- http://www.deeplearningbook.org/contents/ml.html
- https://www.deeplearningbook.org/contents/mlp.html **(highly recommended)**
- Behera, L., & Kar, I. (2010). *Intelligent Systems and control principles and applications*. Oxford University Press, Inc.. **Chapter 2**
- PRML: Chapter 5

Non-linearities or Activation Functions

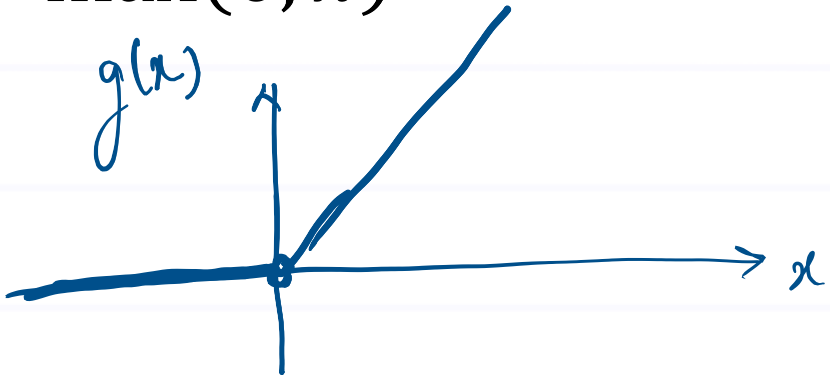
- Sigmoid



- Hyperbolic tan: $\tanh(x) = \dfrac{e^{2x}-1}{e^{2x}+1}$



- Rectified Linear Unit (ReLU): $g(x) = \max(0, x)$



- https://keras.io/activations/