Forward Backward Propagation

Dennis A. Christie

Machine Learning Course

23/02/2019

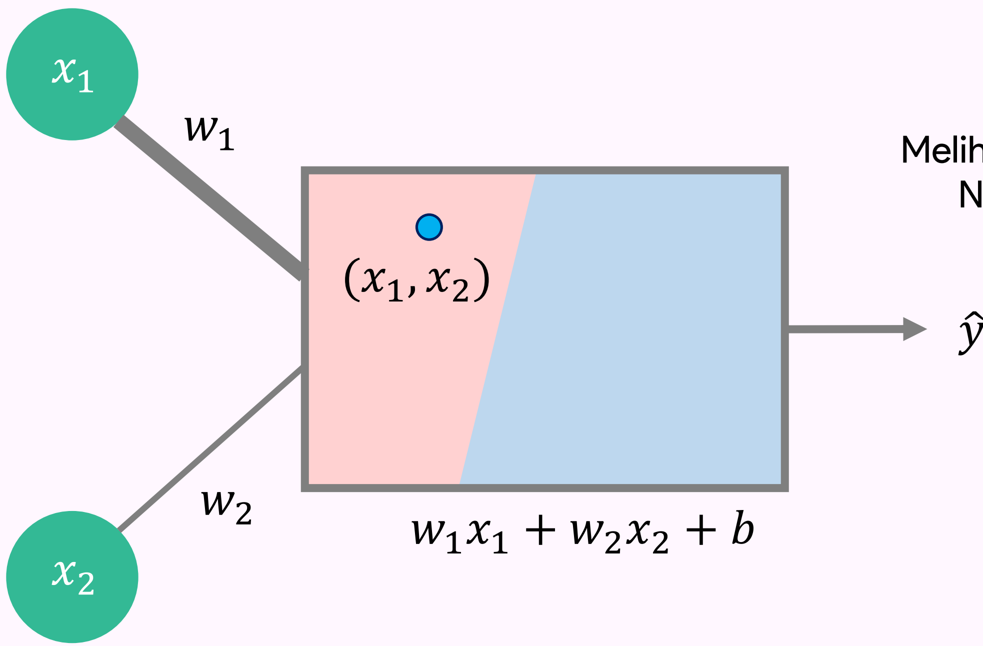
Feedforward Propagation

Tahap Pertama dalam ANN Traning

Neural Network

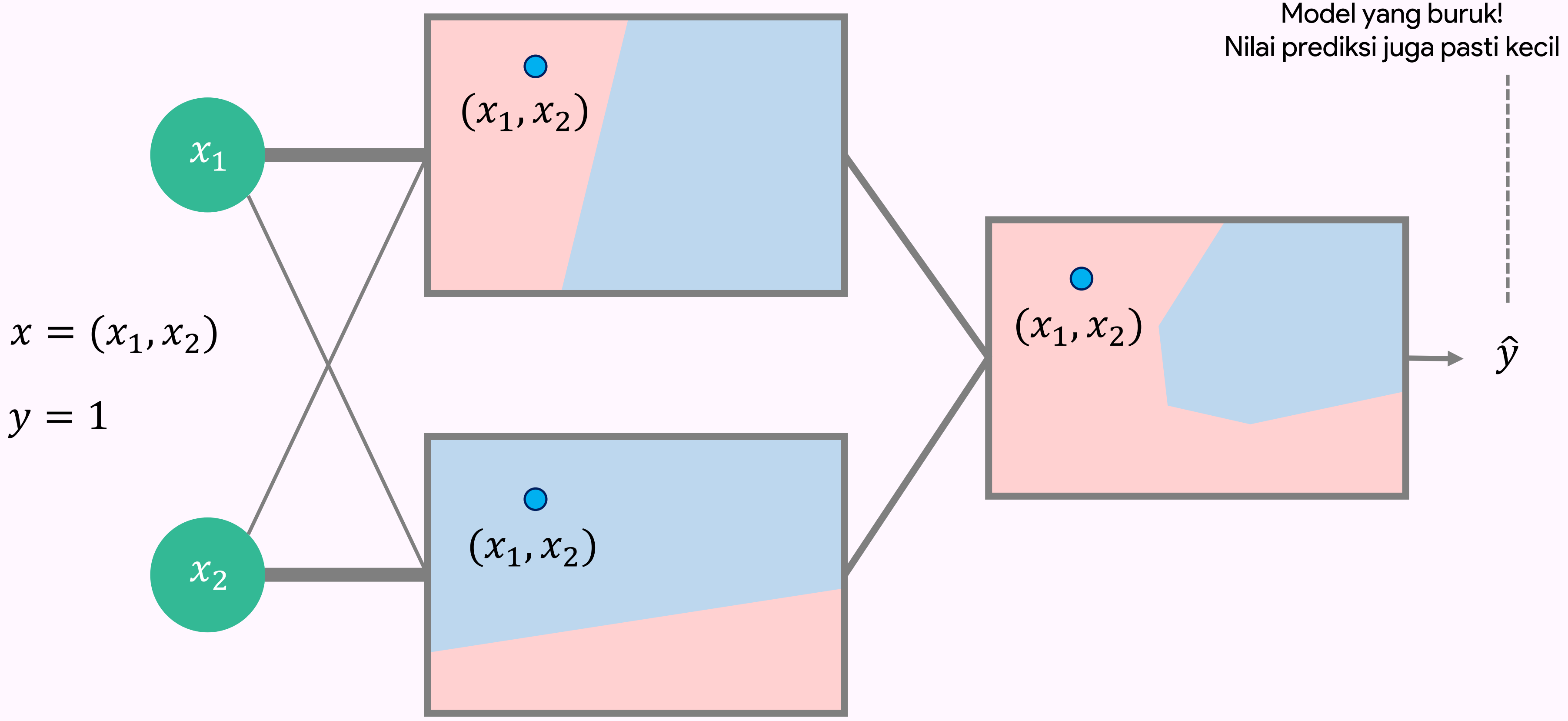
$x = (x_1, x_2)$

$y = 1$

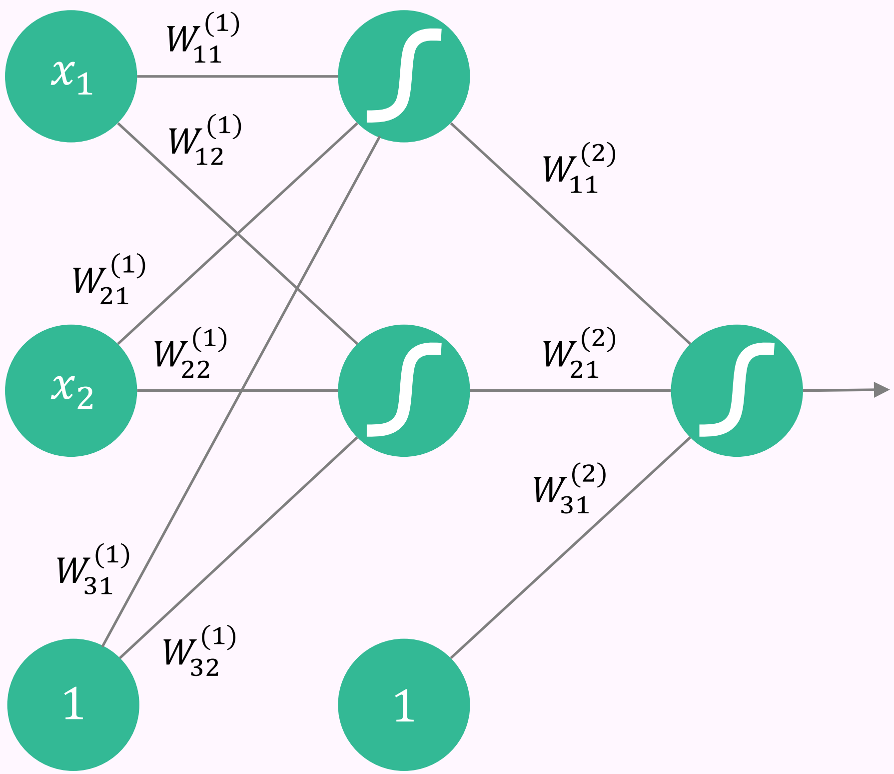


Melihat dari model yang buruk
Nilai prediksi pasti kecil

Neural Network



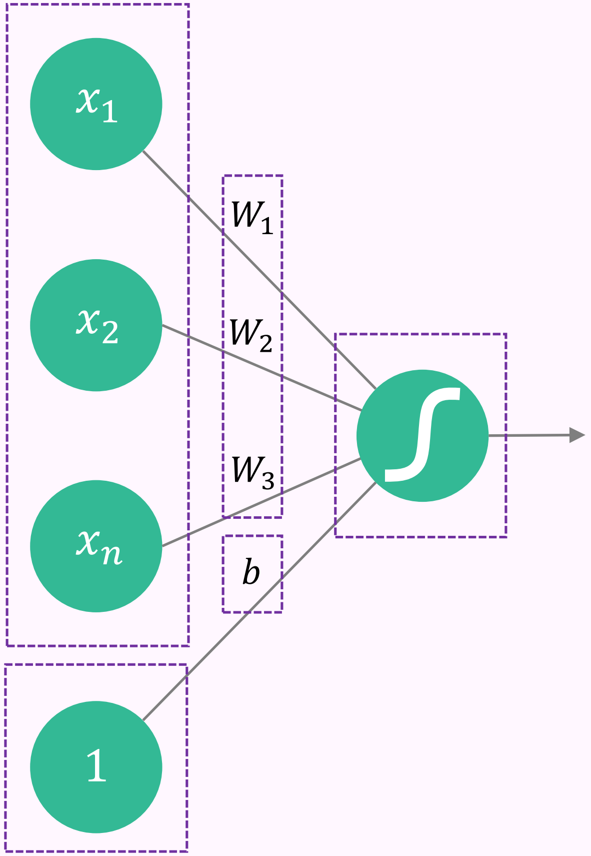
Forward Propagation



$$\sigma\left(\begin{bmatrix} W_{11}^{(1)} \\ W_{21}^{(2)} \\ W_{31}^{(3)} \end{bmatrix}^{\mathrm{T}} \sigma\left(\begin{bmatrix} W_{11}^{(1)} & W_{12}^{(1)} \\ W_{21}^{(1)} & W_{22}^{(1)} \\ W_{31}^{(1)} & W_{32}^{(1)} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}\right)\right)$$

$$\boxed{\hat{y} = \sigma \circ W^{(2)} \circ \sigma \circ W^{(1)} \circ x}$$

Feedforward 1 Layer



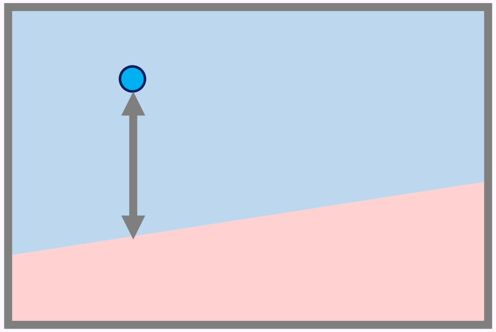
Prediction

$$\hat{y} = \sigma(\mathrm{W}\mathrm{x} + b)$$

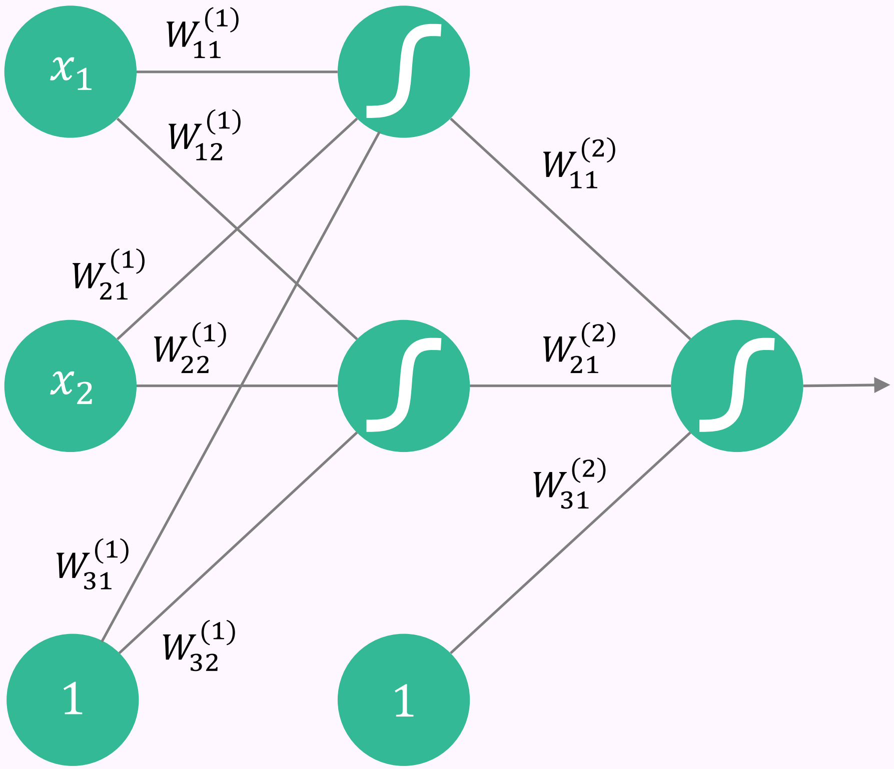
Error Function

$$\mathrm{E(W)} = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - y_i)\right)$$

Illustrasi



Feedforward Multi-Layer



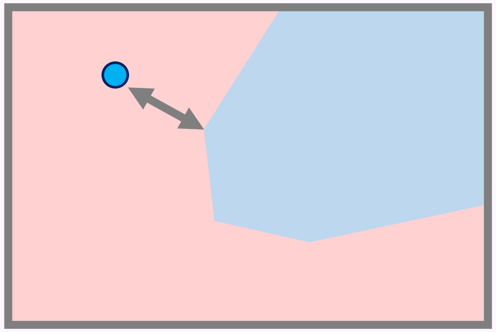
Prediction

$$\hat{y} = \sigma \circ W^{(2)} \circ \sigma \circ W^{(1)} \circ x$$

Error Function

$$\mathrm{E(W)} = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - y_i)\right)$$

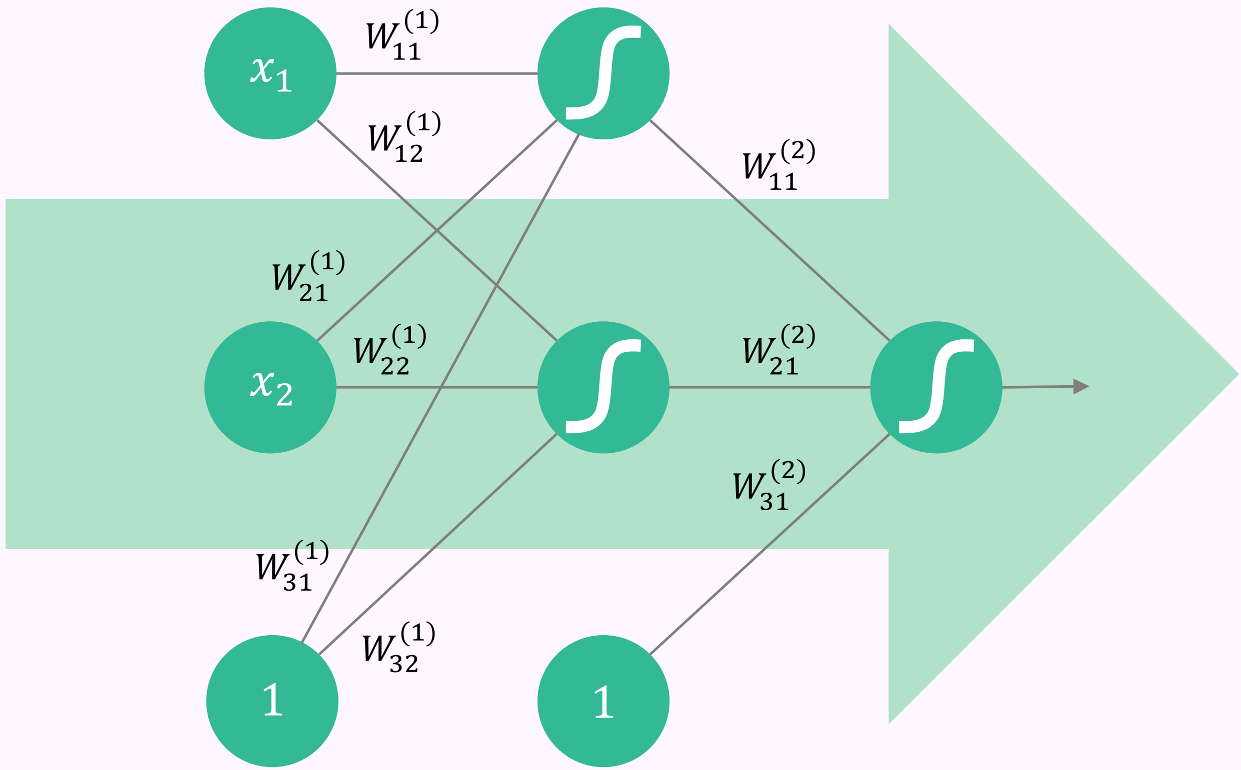
Illustrasi



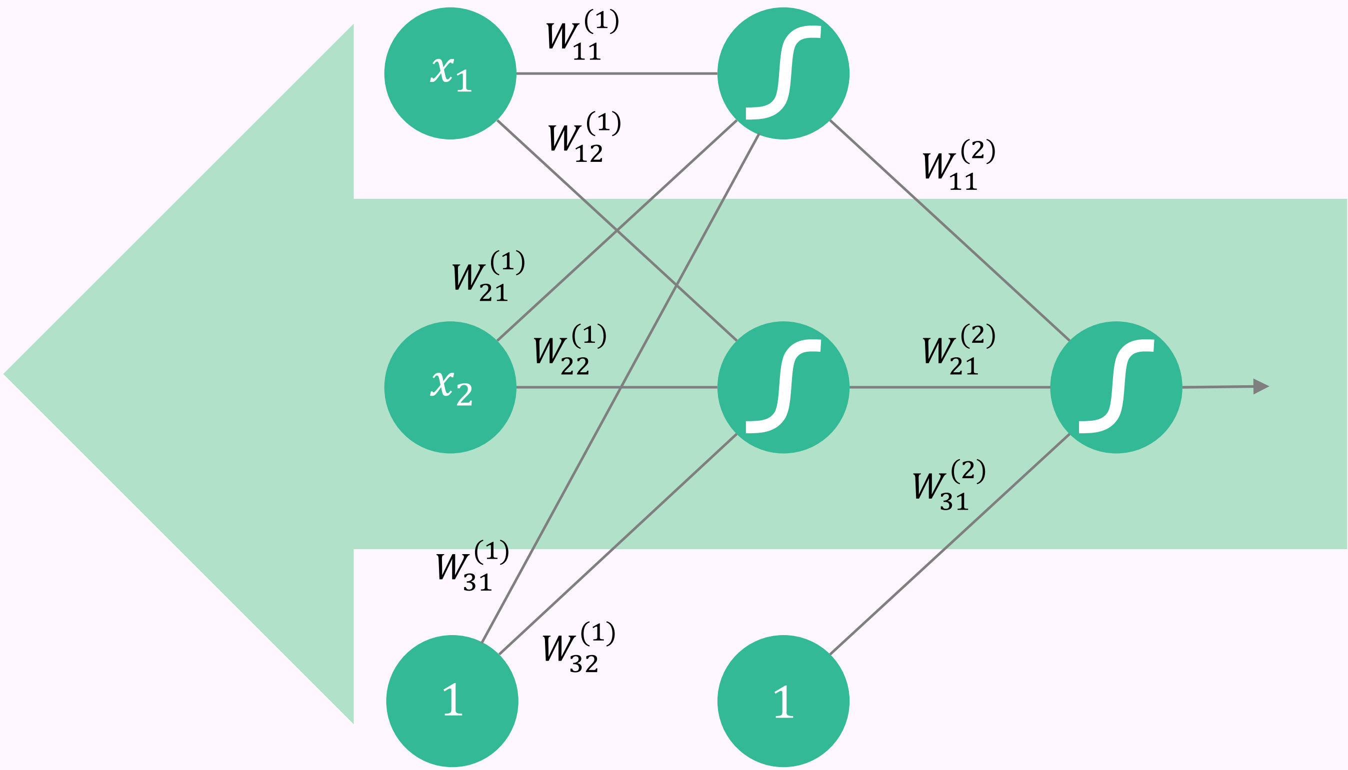
Backward Propagation

Tahap Kedua dalam ANN Traning

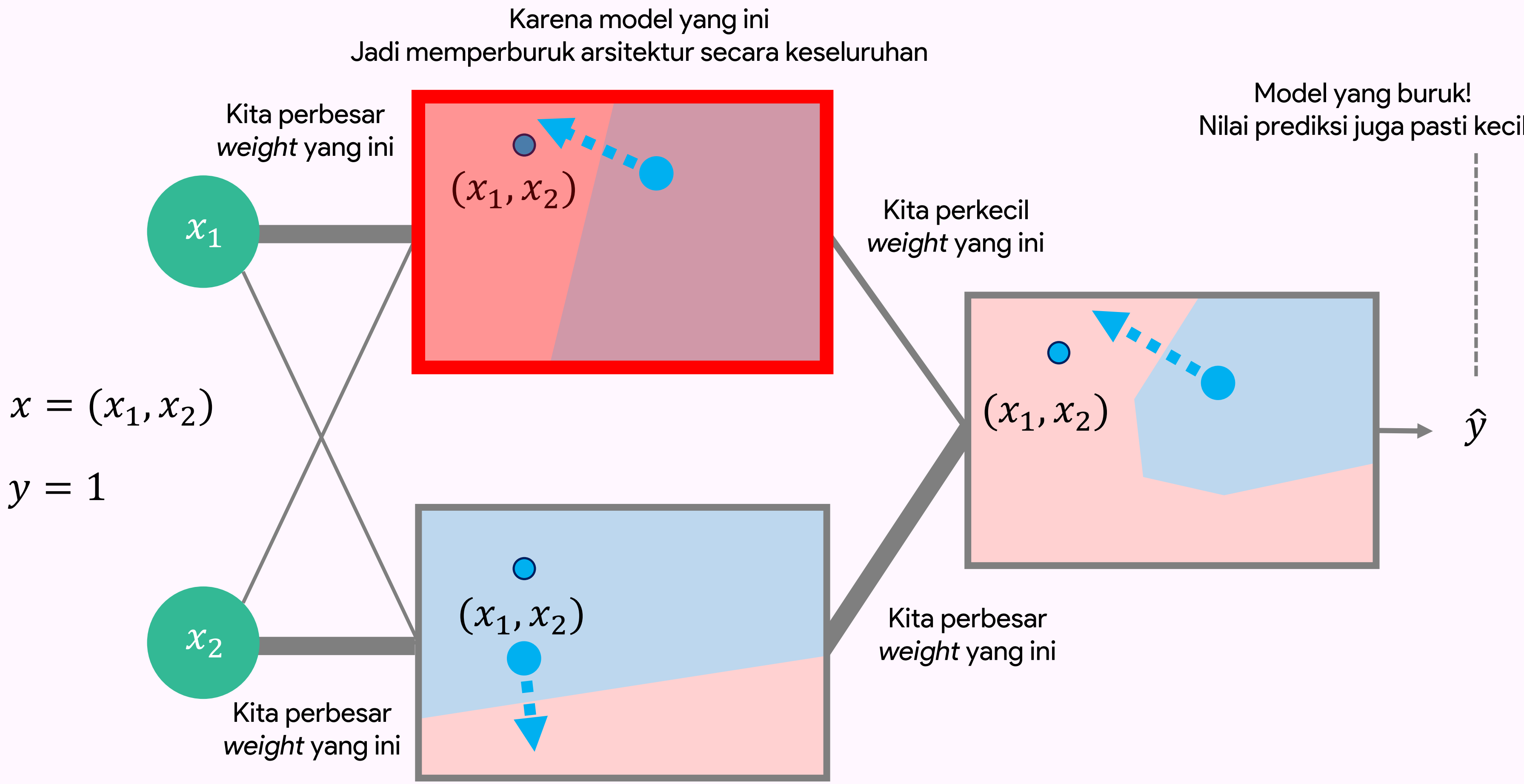
Feedforward vs Backpropagate



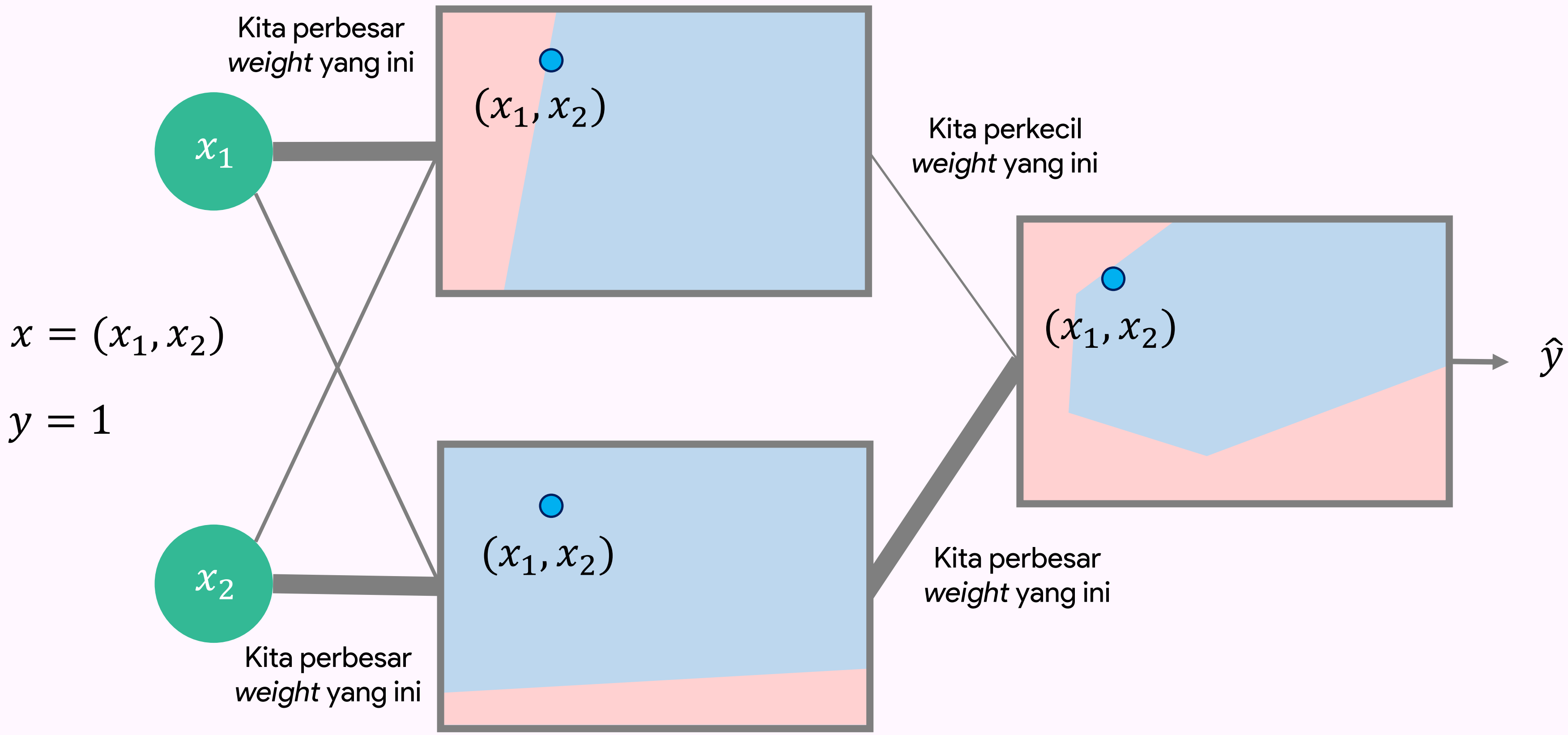
Feedforward vs Backpropagate



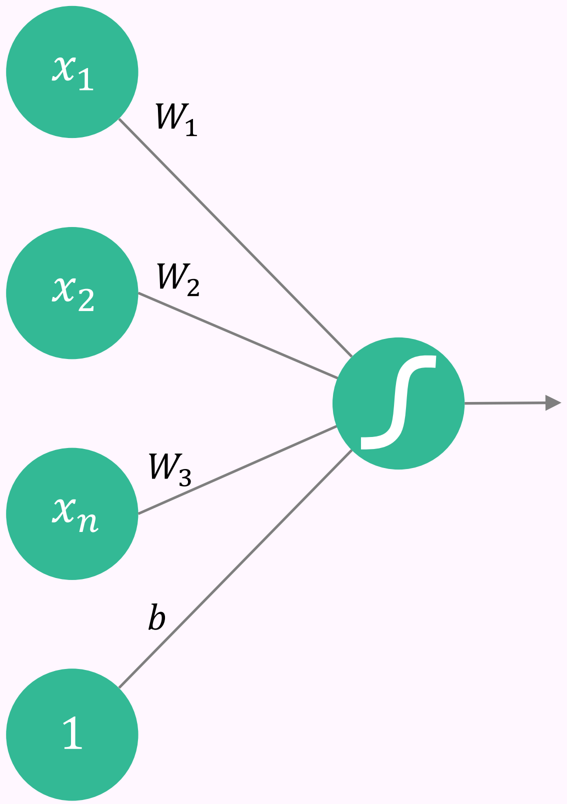
Ide dibalik Backpropagate



Ide dibalik Backpropagate



Back-propagation 1 Layer



Prediction

$$\hat{y} = \sigma(\mathrm{Wx} + b)$$

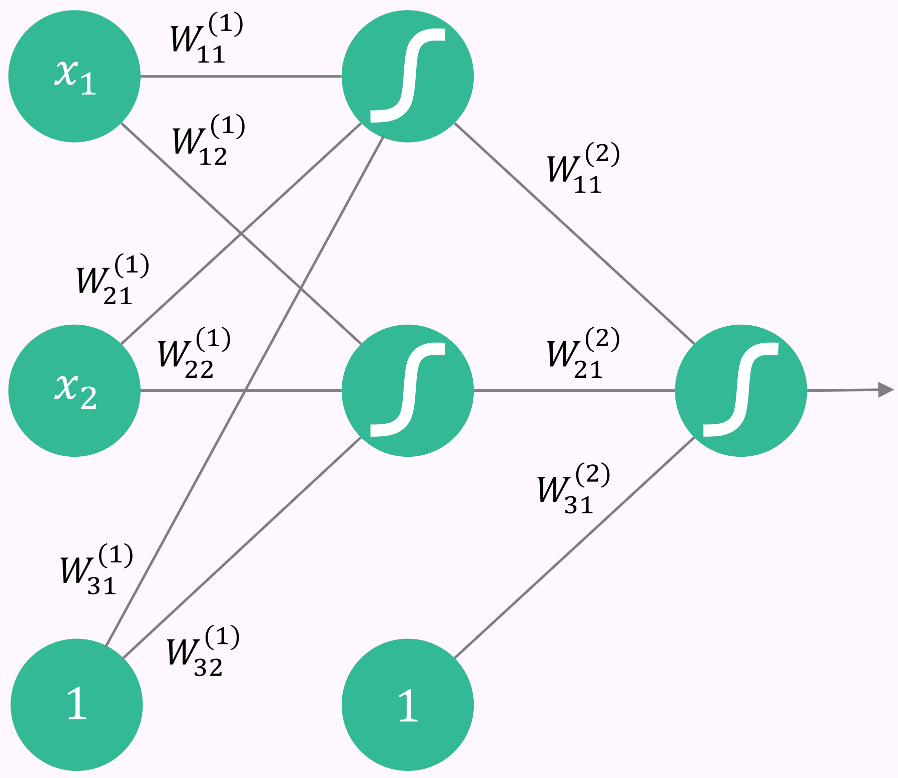
Error Function

$$\mathrm{E(W)} = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - y_i)\right)$$

Gradient dari Error Funtion

$$\nabla E = \left(\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \cdots, \frac{\partial E}{\partial w_n}, \frac{\partial E}{\partial b}\right)$$

Back-propagation n-Layer



Prediction

$$\hat{y} = \sigma \circ W^{(2)} \circ \sigma \circ W^{(1)} \circ x$$

Error Function

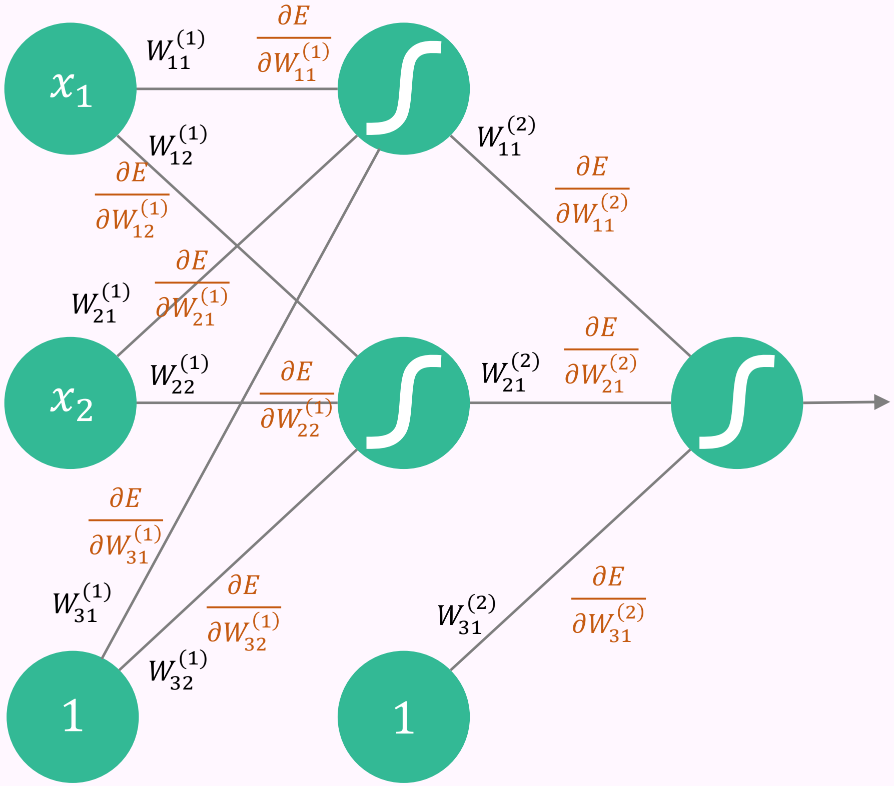
$$\mathrm{E(W)} = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - y_i)\right)$$

Gradient dari Error Funtion

$$\nabla E = \left(\cdots, \frac{\partial E}{\partial w_{ij}^{(k)}}, \cdots\right)$$

Back-propagation n-Layer

Bagaimana cara melakukan partial derivative E terhadap $W_{11}^{(1)}$?



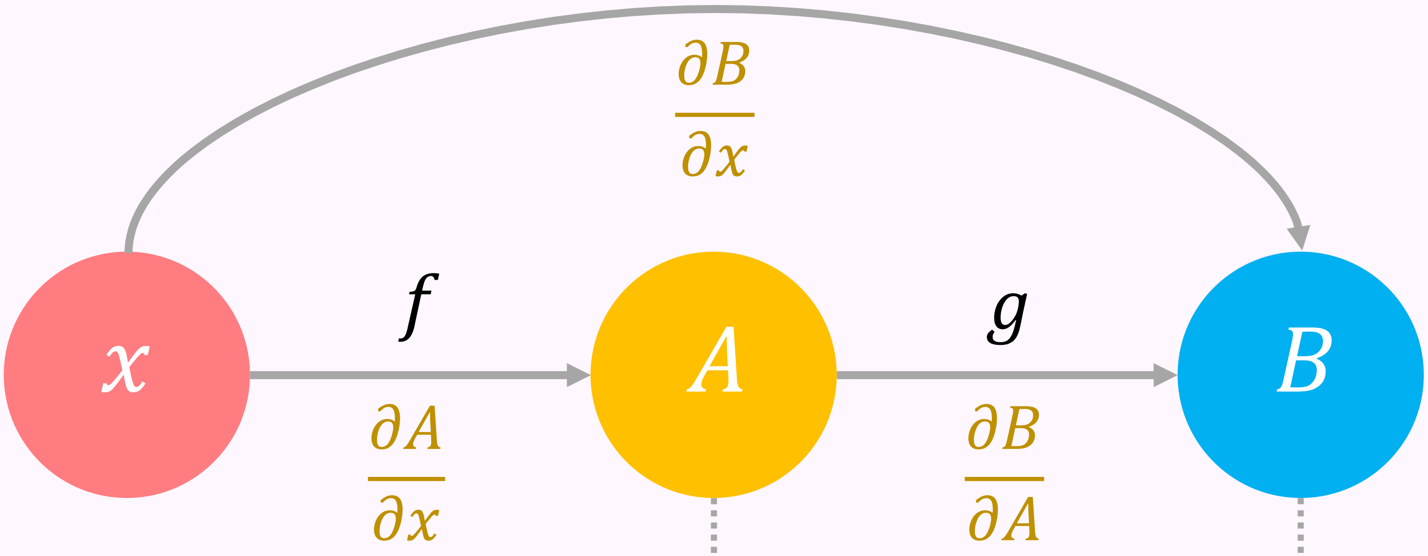
$$\hat{y} = \sigma \circ W^{(2)} \circ \sigma \circ W^{(1)} \circ x$$

$$W^{(2)} = \begin{bmatrix} W_{11}^{(1)} \\ W_{21}^{(2)} \\ W_{31}^{(3)} \end{bmatrix}^{\mathrm{T}} \qquad W^{(1)} = \begin{bmatrix} W_{11}^{(1)} & W_{12}^{(1)} \\ W_{21}^{(1)} & W_{22}^{(1)} \\ W_{31}^{(1)} & W_{32}^{(1)} \end{bmatrix}^{\mathrm{T}}$$

$$\nabla E = \left(\frac{\partial E}{\partial W_{11}^{(1)}}, \frac{\partial E}{\partial W_{12}^{(1)}}, \frac{\partial E}{\partial W_{11}^{(2)}}, \frac{\partial E}{\partial W_{21}^{(1)}}, \frac{\partial E}{\partial W_{22}^{(1)}}, \frac{\partial E}{\partial W_{21}^{(2)}}, \frac{\partial E}{\partial W_{31}^{(1)}}, \frac{\partial E}{\partial W_{32}^{(1)}}, \frac{\partial E}{\partial W_{31}^{(2)}} \right)$$

$$W_{ij}^{\prime(k)} = W_{ij}^{(k)} - \alpha \frac{\partial E}{\partial W_{ij}^{(k)}}$$

Chain Rule

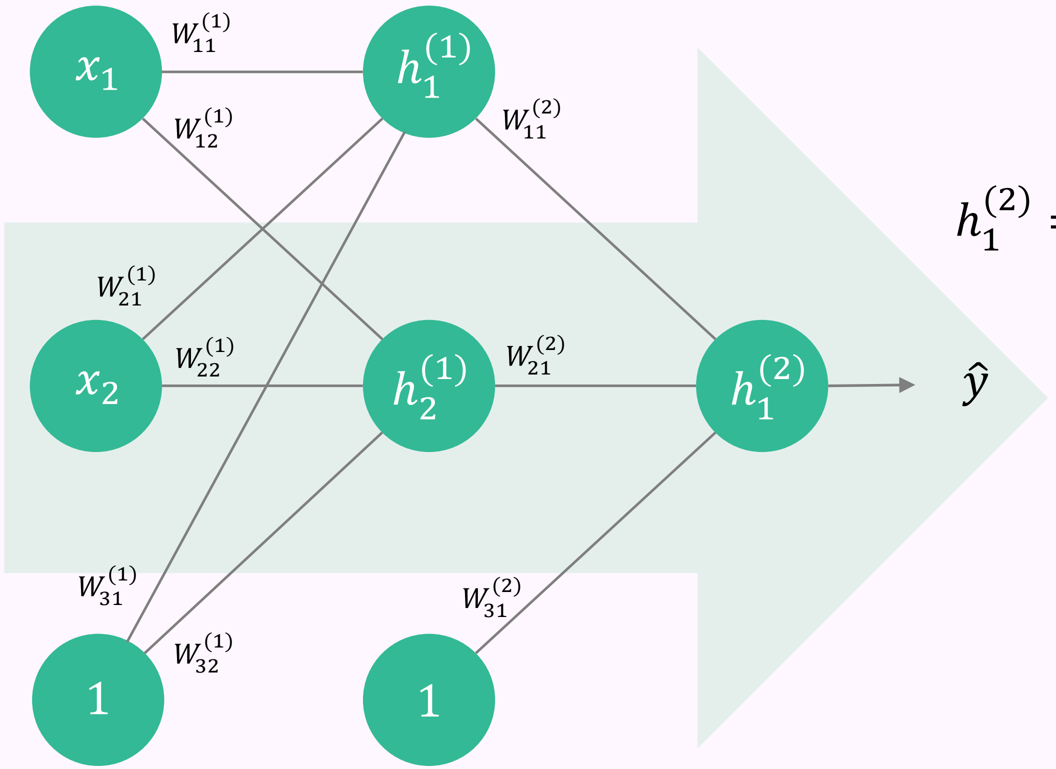


$A = f(x)$ $\qquad$ $B = g \circ f(x)$

Melakukan partial derivatif B terhadap x
Bisa melalui A terlebih dahulu

$=$

Ringkasan Algoritma



$$h_1^{(1)} = W_{11}^{(1)} x_1 + W_{21}^{(1)} x_1 + W_{31}^{(1)}$$

$$h_2^{(1)} = W_{12}^{(1)} x_1 + W_{22}^{(1)} x_1 + W_{32}^{(1)}$$

$$h_1^{(2)} = W_{11}^{(2)} \sigma(h_1^{(1)}) + W_{21}^{(2)} \sigma(h_2^{(1)}) + W_{31}^{(2)}$$

$$\hat{y} = \sigma(h_1^{(2)})$$

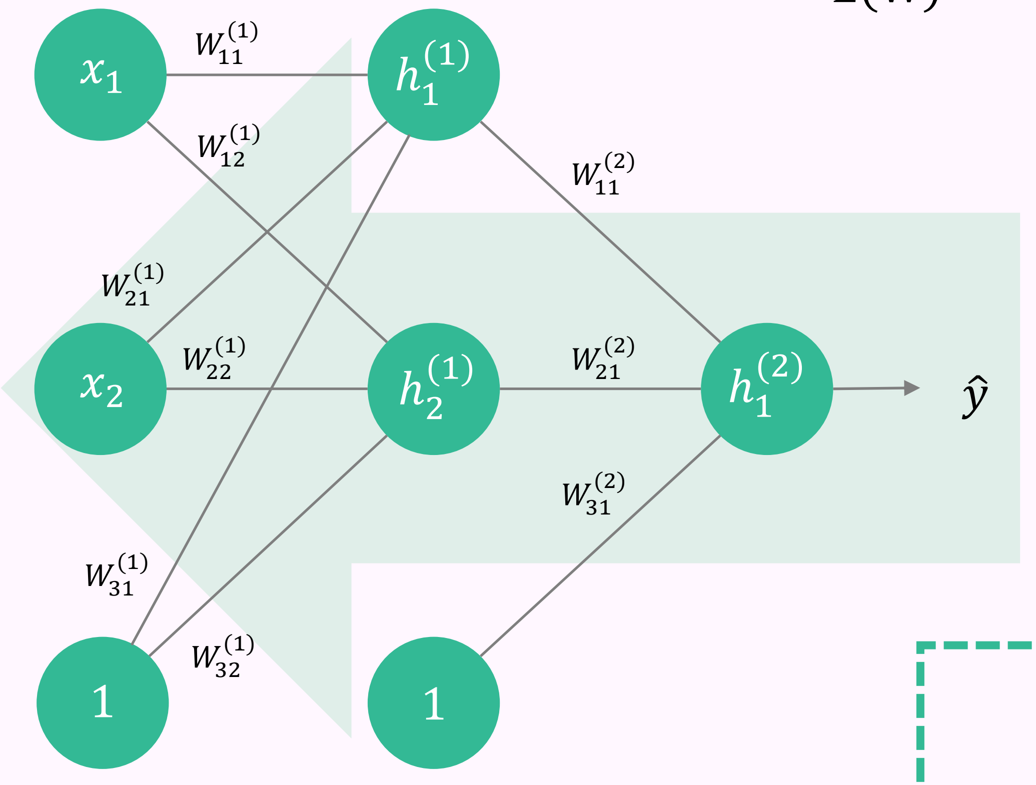
$$\hat{y} = \sigma \circ W^{(2)} \circ \sigma \circ W^{(1)} \circ x$$

$$\mathrm{E(W)} = -\frac{1}{m} \sum_{i=1}^{m} (y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - y_i))$$

Ringkasan Algoritma

$$E(W) = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - y_i)\right)$$

$$\nabla E = \left(\frac{\partial E}{\partial W_{11}^{(1)}}, \cdots, \frac{\partial E}{\partial W_{31}^{(2)}}\right)$$



Partial derivative E terhadap $W_{11}^{(1)}$ menggunakan chain-rule

$$\frac{\partial E}{\partial W_{11}^{(1)}} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_1^{(2)}} \frac{\partial h_1^{(2)}}{\partial h_1^{(1)}} \frac{\partial h_1^{(1)}}{\partial W_{11}^{(1)}}$$