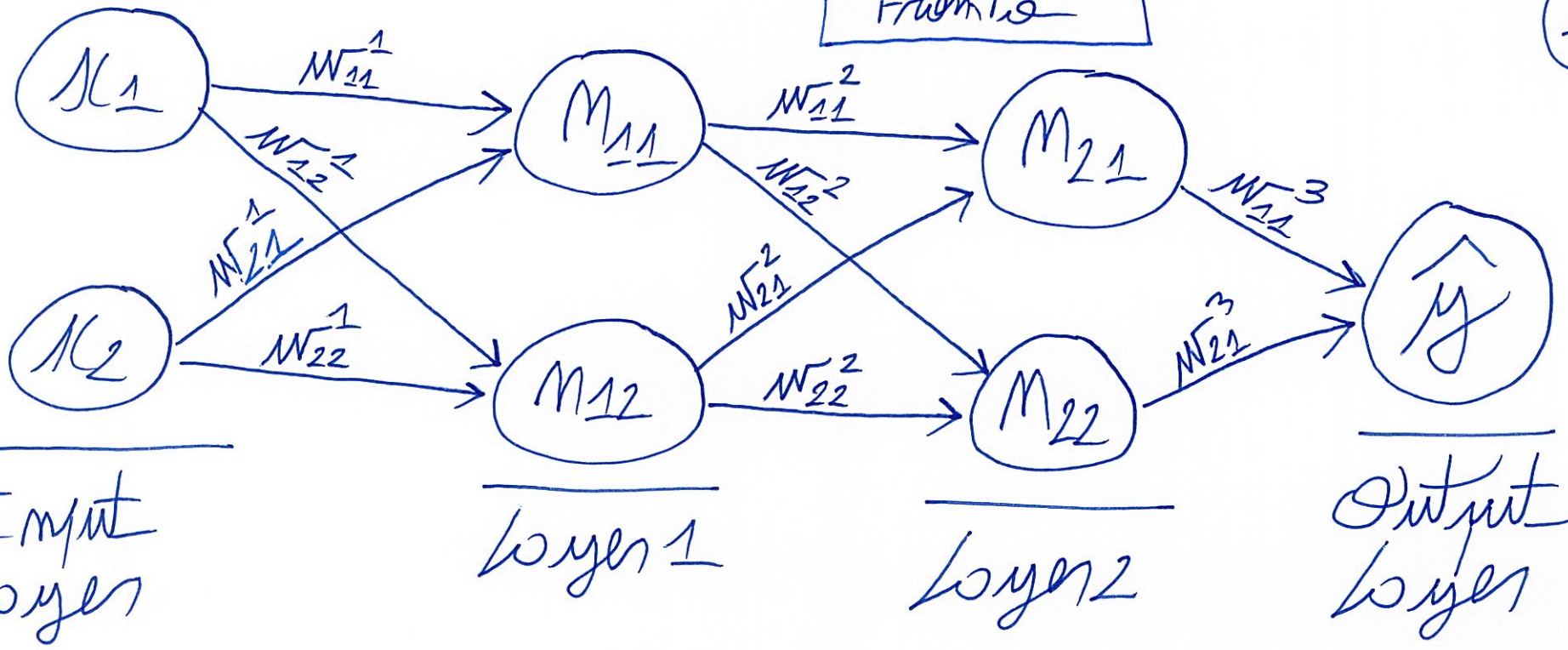
BACK PROPAGATION

W Next Layer From To



$$\hat{y} = w_{11}^3 m_{21} + w_{21}^3 m_{22}$$

$$m_{21} = w_{11}^2 m_{11} + w_{21}^2 m_{12}$$

$$m_{lj} = \sum_{k=1}^{2} w_{kj}^{l} m_{l-1\,k}$$

No activation functions
No bias terms
Not a problem, that example can easily be generalised

Loss function $\ell(\hat{y}) = (\hat{y} - y)^2$

$\frac{d\ell}{d\hat{y}} = 2(\hat{y} - y) = 2e$ (e = error term)

$\frac{d\ell}{dm_{21}} = \frac{d\hat{y}}{dm_{21}} \frac{d\ell}{d\hat{y}} = w_{11}^3 (2e)$

$\frac{d\ell}{dm_{22}} = \frac{d\hat{y}}{dm_{22}} \frac{d\ell}{d\hat{y}} = w_{21}^3 (2e)$

$\frac{d\ell}{dm_{11}} = \frac{d\hat{y}}{dm_{11}} \frac{d\ell}{d\hat{y}} = \frac{d\hat{y}}{dm_{11}} (2e) = \left[\frac{d\hat{y}}{dm_{21}} \frac{dm_{21}}{dm_{11}} + \frac{d\hat{y}}{dm_{22}} \frac{dm_{22}}{dm_{11}}\right] (2e)$

$\hat{y}\left[m_{21}(m_{11}, m_{12}),\ m_{22}(m_{11}, m_{12})\right]$ (differential calculus)

$\frac{d\ell}{dm_{11}} = \left(w_{11}^3 w_{11}^2 + w_{21}^3 w_{12}^2\right) (2e)$

Differential calculus

$$\ell = L(u, v)$$

$$\ell(x) = L(u(x), v(x))$$

$$\frac{d\ell}{dx} = \frac{\partial L}{\partial u}\frac{du}{dx} + \frac{\partial L}{\partial v}\frac{dv}{dx}$$

The differential is linear, variations add

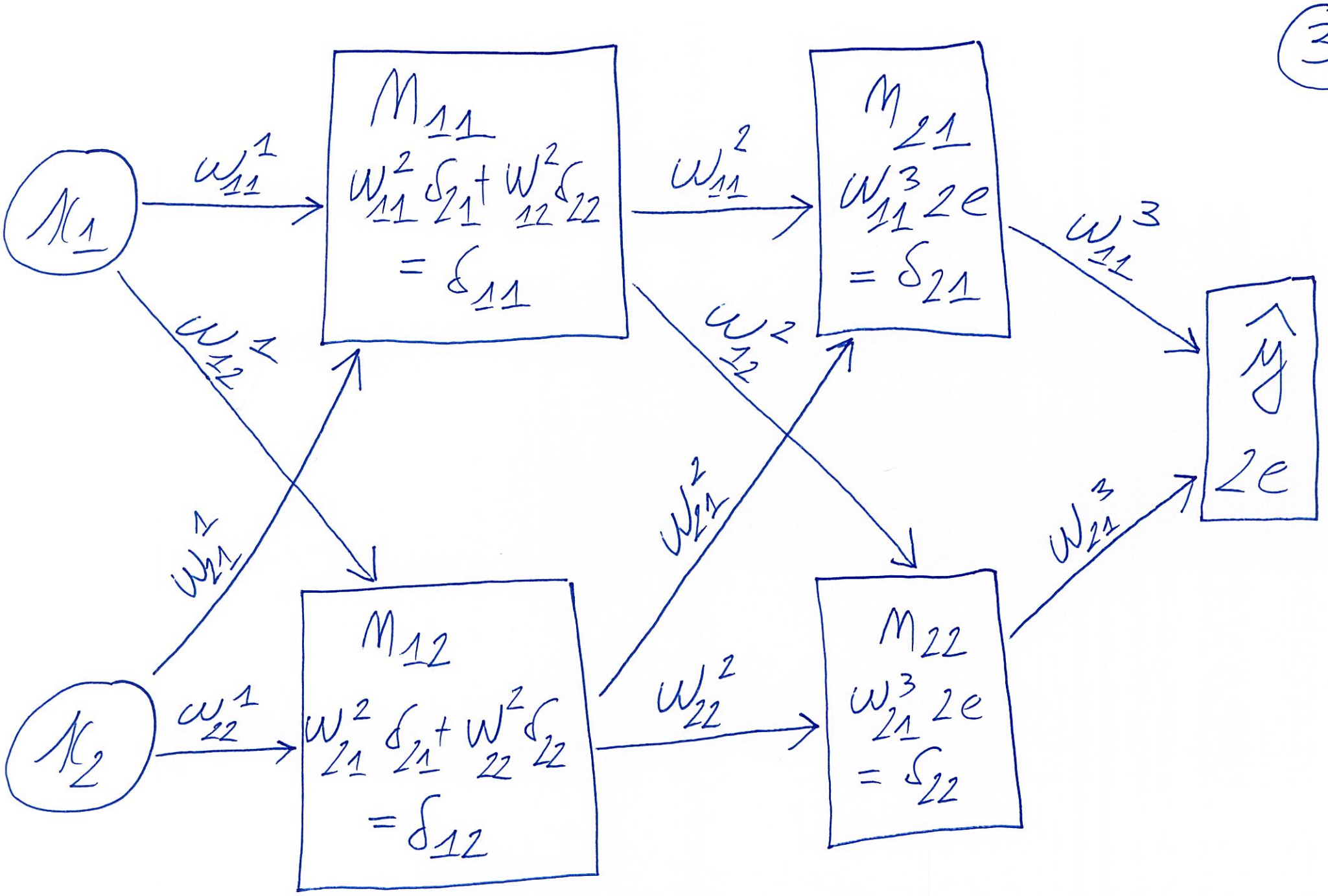
Example $\ell(x) = \dfrac{u(x)}{v(x)}$

$$L(u, v) = \frac{u}{v}$$

$$\frac{d\ell}{dx} = \frac{\partial L}{\partial u}u'(x) + \frac{\partial L}{\partial v}v'(x)$$

$$\frac{d\ell}{dx} = \frac{u'(x)}{v(x)} - \frac{u(x)}{v^2(x)}v'(x)$$

$$\frac{d\ell}{dx} = \frac{(u'(x)v(x) - v'(x)u(x))}{v^2(x)}$$


x_1
x_2
w_{11}^1
w_{12}^1
w_{21}^1
w_{22}^1
M_{11}
$w_{11}^2 \delta_{21} + w_{12}^2 \delta_{22}$
$= \delta_{11}$
M_{12}
$w_{21}^2 \delta_{21} + w_{22}^2 \delta_{22}$
$= \delta_{12}$
w_{11}^2
w_{12}^2
w_{21}^2
w_{22}^2
M_{21}
$w_{11}^3 2e$
$= \delta_{21}$
M_{22}
$w_{21}^3 2e$
$= \delta_{22}$
w_{11}^3
w_{21}^3
$\hat{y}$
$2e$

$$\frac{\partial e}{\partial w_{11}^3} = \frac{\partial \hat{y}}{\partial w_{11}^3} \frac{\partial e}{\partial \hat{y}} = m_{21}\, 2e$$

$$\frac{\partial e}{\partial w_{21}^3} = \frac{\partial \hat{y}}{\partial w_{21}^3} \frac{\partial e}{\partial \hat{y}} = m_{22}\, 2e$$

$$\frac{\partial e}{\partial w_{11}^2} = \frac{\partial m_{21}}{\partial w_{11}^2} \frac{\partial e}{\partial m_{21}} = m_{11}\, s_{21}$$

$$\frac{\partial e}{\partial w_{12}^2} = \frac{\partial m_{22}}{\partial w_{12}^2} \frac{\partial e}{\partial m_{22}} = m_{11}\, s_{22}$$

$$\frac{\partial e}{\partial w_{11}^1} = \frac{\partial m_{11}}{\partial w_{11}^1} \frac{\partial e}{\partial m_{11}} = 1/1\, \epsilon_{11}$$

— — — —

The gradient is identical very easily using the networks values m_{ij} and s_{ij}!