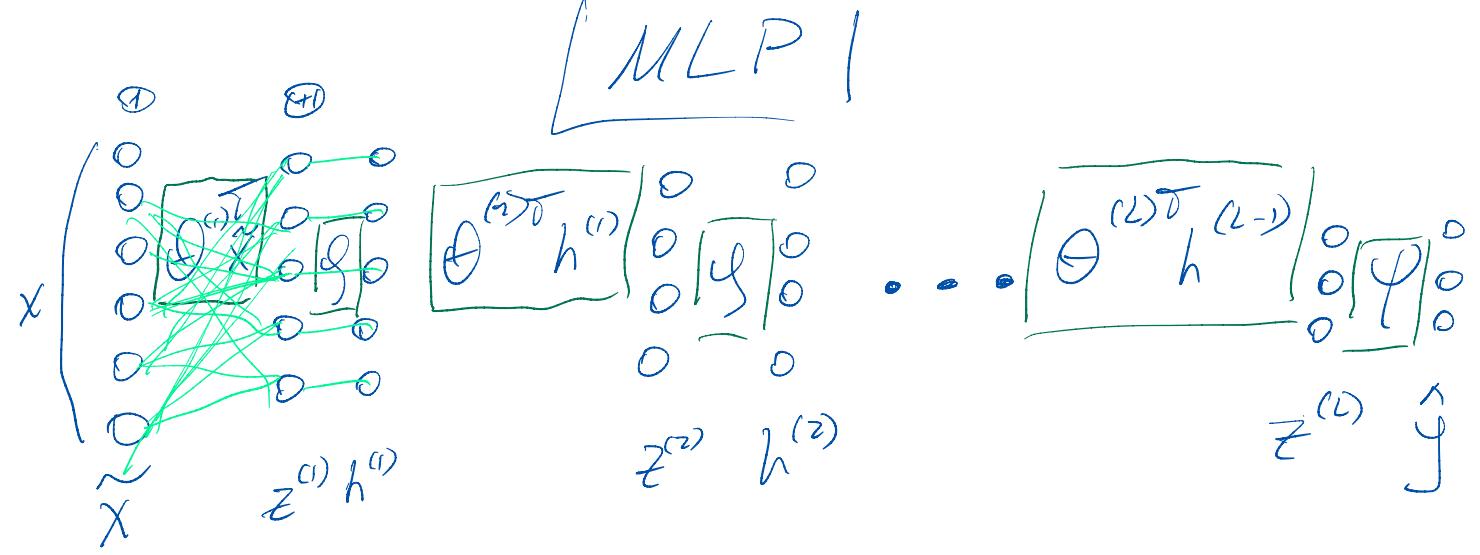


MLP |



$$z^{(1)} = \Theta^{(1)T} \tilde{x} \quad h^{(1)} = \varphi(z^{(1)})$$

$$z^{(2)} = \Theta^{(2)T} h^{(1)} \quad h^{(2)} = \varphi(z^{(2)})$$

$$z^{(L)} = \Theta^{(L)T} h^{(L-1)} \quad \hat{y} = \psi(h^{(L-1)})$$

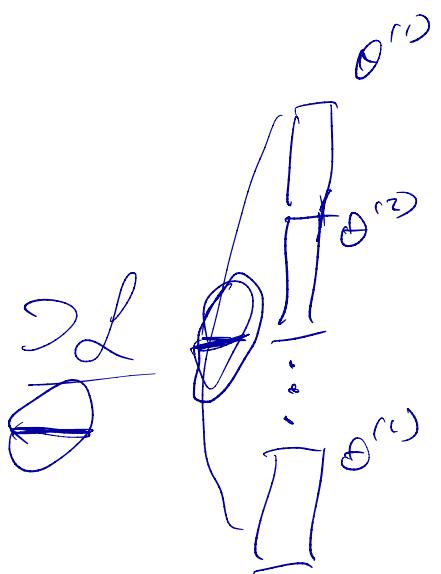
$\mathcal{T}: \{x_i, y_i\}$

$x_i \in \mathcal{X}$
 $y_i \in \mathcal{Y}$

$$\mathcal{L}(\Theta, \{x_i, y_i\}) = \begin{matrix} \text{MSE}(\Theta, \mathcal{G}) \\ \text{BCE}(\Theta, \mathcal{G}) \\ \text{CE}(\Theta, \mathcal{G}) \end{matrix}$$

$$\mathcal{J} = \mathcal{L}(\Theta^{(1)}, \Theta^{(2)}, \Theta^{(3)}, \dots, \{x_i, y_i\})$$

$$\frac{\partial \mathcal{J}}{\partial \Theta^{(1)}}; \quad \frac{\partial \mathcal{J}}{\partial \Theta^{(2)}};$$



Algorithmus zur Anwendung

① $\theta_0, \{, T, B, C(\epsilon)$

② for t in $1 \dots T$:

a) $b(f) \in \mathcal{F}_B$

b) $\theta_t = \theta_{t-1} - \eta \nabla_{\theta} \mathcal{L}(b(f), \theta_{t-1})$

c) if $C(\epsilon) < \epsilon_{top}$

$$\frac{\partial F(f(x), g(x))}{\partial x} = \frac{\partial F}{\partial f} \frac{\partial f}{\partial x} + \frac{\partial F}{\partial g} \frac{\partial g}{\partial x}$$

$$\frac{\partial f(g(x))}{\partial x} = \frac{\partial f}{\partial g} \frac{\partial g}{\partial x}$$

$$f = ((a+c) \cdot (b-d))^2 = 24^2 = 576$$

| 5 7 3
 a c
 b d

$$\begin{aligned}
 e &= \frac{a+c}{a+c} = 6 & b-d &= g = 4 \\
 h &= \frac{e \cdot g}{e \cdot g} = 24 & h^2 &= 144 \\
 f &= 576
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial f}{\partial a} &= \frac{\partial f}{\partial h} \frac{\partial h}{\partial e} \frac{\partial e}{\partial a} \\
 \frac{\partial f}{\partial a} &= 2h \cdot g \cdot 1 =
 \end{aligned}$$

$$= 48 \cdot 4 \cdot 1 = 192$$

$$\begin{aligned}
 f(h) &= h^2 \\
 h(e) &= eg \\
 e(a) &= a+c
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial f}{\partial h} &= 2h \\
 \frac{\partial h}{\partial e} &= g \\
 \frac{\partial e}{\partial a} &= 1
 \end{aligned}$$

$$\hat{y} = F_{NN}(\theta, x)$$

$$\frac{\partial L}{\partial \theta}$$

$$\frac{\partial L}{\partial g}$$

$$\frac{\partial L}{\partial x}$$

$$L(\hat{y}, x, \theta)$$

$$\frac{\partial \hat{y}}{\partial \theta}$$

$$\frac{\partial \hat{y}}{\partial x}$$

$$\hat{y}_i = \mu_i = I(\theta^T x_i)$$

θ, x_i - bias of error, logit

$$L = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

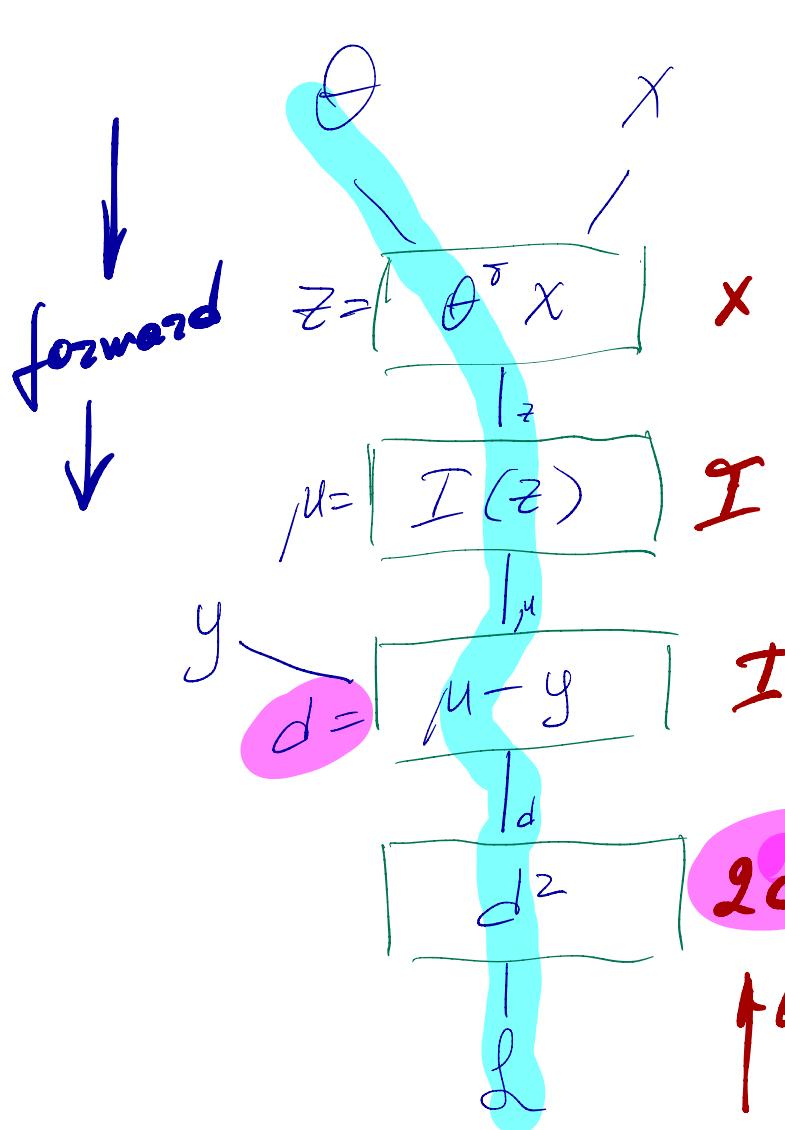
$$\frac{\partial L}{\partial \theta} - ?$$

$$x \in \mathbb{R}^D$$

$$\theta \in \mathbb{R}^D$$

$$\frac{\partial L}{\partial \theta} = \vec{e}_0 \cdot \frac{\partial L}{\partial \theta_0} + \vec{e}_1 \cdot \frac{\partial L}{\partial \theta_1} + \dots + \vec{e}_D \cdot \frac{\partial L}{\partial \theta_D}$$

$$\frac{\partial L}{\partial \theta} \in \mathbb{R}^D$$



$$L = (y - \mu)^2$$

$$\frac{\partial L}{\partial \theta} - ?$$

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial d} \cdot \frac{\partial d}{\partial \mu} \cdot \frac{\partial \mu}{\partial z} \cdot \frac{\partial z}{\partial \theta}$$

$$\frac{\partial \theta^T x}{\partial \theta} = x$$

$$\frac{\partial L}{\partial \theta} = 2d \cdot I \cdot I \cdot x$$

2d

backward