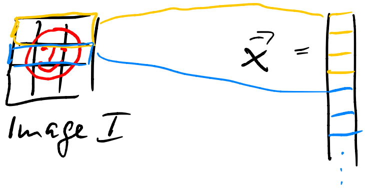
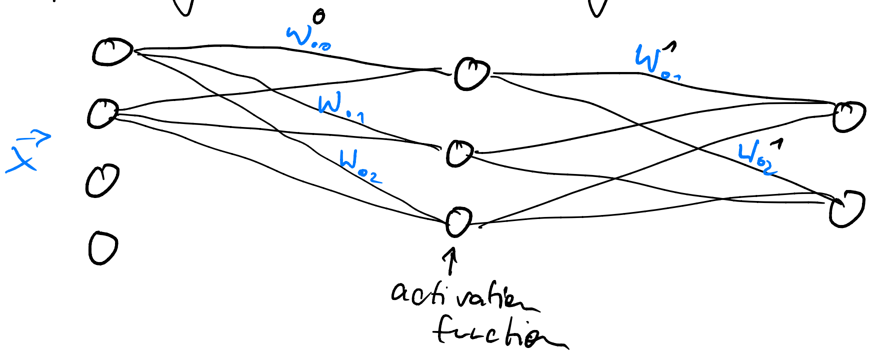
Input $\vec{x}$



Classifier $\underbrace{\{face, car\}}_{Y}$

$Y = [1\ 0] \ldots$

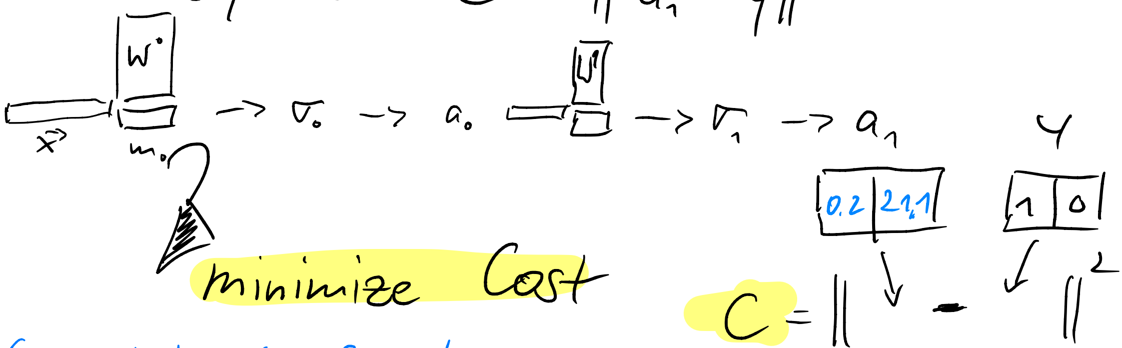
Input Layer Hidden Layer Output layer



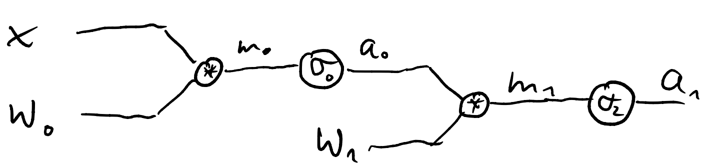
$$\vec{x} \cdot W_0 = m_0 \qquad \vec{a}_0 \cdot W_1 = m_1$$

$$\sigma_0(m_0) = a_0 \qquad \sigma_1(m_1) = a_1$$

Loss / Cost $\quad C = \| a_1 - Y \|^2$

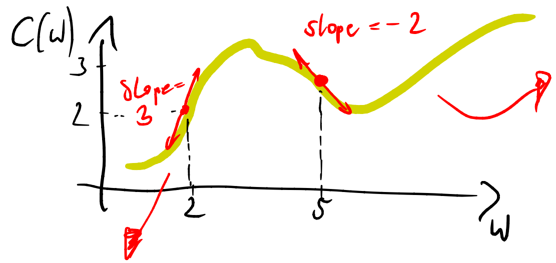


Computational Graph



$C(w) \rightarrow$ Find direction of change

$$\frac{\partial C(w)}{\partial w} \Rightarrow \text{slope}$$



$\frac{\partial C(w)}{\partial w}$ is negative when $C(w)$ is decreasing

$\frac{\partial C(w)}{\partial w}$ is positive when $C(w)$ is increasing

Goal: Minimize $C(w)$

Learning rate

if we set $w = w - \alpha \cdot \text{slope}$

at points above $C(w)$ will decrease

$1.85 = w_{new} = 2 - \alpha \cdot 3$ $\qquad \alpha = 0.05$

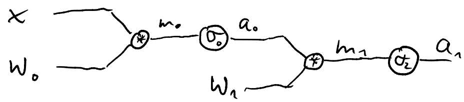
$5.1 = w_{new} = 5 - \alpha \cdot (-2)$

$\vec{x} \cdot W_0 = m_0$ $\qquad$ $\vec{a}_0 \cdot W_1 = m_1$

$\sigma_0(m_0) = a_0$ $\qquad$ $\sigma_1(m_1) = a_1$

$C = \| a_1 - y \|^2$

Computational graph



Chain Rule

$$\frac{\partial C(w)}{\partial w}$$

partial derivative W_0:

$$\boxed{\frac{\partial C}{\partial W_0}} = \frac{\partial C}{\partial a_1} \cdot \frac{\partial a_1}{\partial m_1} \cdot \frac{\partial m_1}{\partial a_0} \cdot \frac{\partial a_0}{m_0} \cdot \frac{\partial m_0}{\partial W_0}$$

partial derivative W_1:

$$\boxed{\frac{\partial C}{\partial W_1}} = \frac{\partial C}{\partial a_1} \cdot \frac{\partial a_1}{\partial m_1} \cdot \frac{\partial m_1}{\partial W_1}$$

$$\frac{\partial C}{\partial a_1} = 2(a_1 - y) \qquad \frac{\partial a_1}{\partial m_1} = \sigma'(m_1) \qquad \frac{\partial m_1}{\partial a_0} = W_1$$

$$\frac{\partial a_0}{\partial m_0} = \sigma_0'(m_0) \qquad \frac{\partial m_0}{\partial W_0} = x$$

$$\frac{\partial m_1}{\partial W_1} = \boxed{a_0}$$ → is computed in forward pass → should be cached