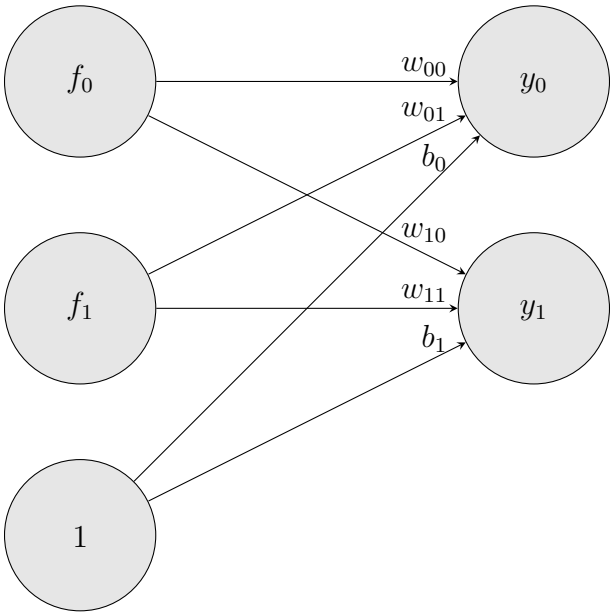
Fully Connected Layer Backpropagation

Maik Schmidt

14. März 2017



Forward:

$$y_i^l = \sum_{j=0}^{\substack{(l-1)\\ \text{layer size}\\ -1}} w_{i,j}^l \cdot f_j^{l-1} + b_i^l$$

$$f_i^l = h(y_i^l)$$

Backward:

$$\delta_i^l = \begin{cases} \frac{\partial h(y_i^l)}{\partial y_i^l} \cdot \frac{\partial l(y_i^l)}{\partial y_i^l} & \text{if } l = \#\text{ of layers} - 1 \\ \frac{\partial h(y_i^l)}{\partial y_i^l} \cdot \sum_{j=0}^{(l+1)\text{layer size}-1} w_{j,i}^{l+1} \cdot \delta_j^{l+1} & \text{else} \end{cases}$$

Gradient step:

$$w_{i,j}^l = w_{i,j}^l - \alpha \cdot \delta_i^l \cdot f_j^{l-1}$$

$$b_i^l = b_i^l - \alpha \cdot \delta_i^l$$