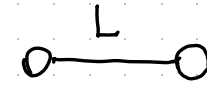


MSE



N data points



y
↑
output of network

$L(y)$ loss

$$y \in \mathbb{R}^{N \times 1}$$

↑
 N size of dataset

$$L(y) = \frac{1}{N} \sum_{i=0}^{N-1} (y_i - y_i^{\text{true}})^2$$

$$\frac{\partial L}{\partial y_j} = \frac{2}{N} (y_j - y_j^{\text{true}})$$

$$\frac{\partial L}{\partial y} = \frac{2}{N} (y - y^{\text{true}})^T \in \mathbb{R}^{1 \times N}$$