Backpropagation

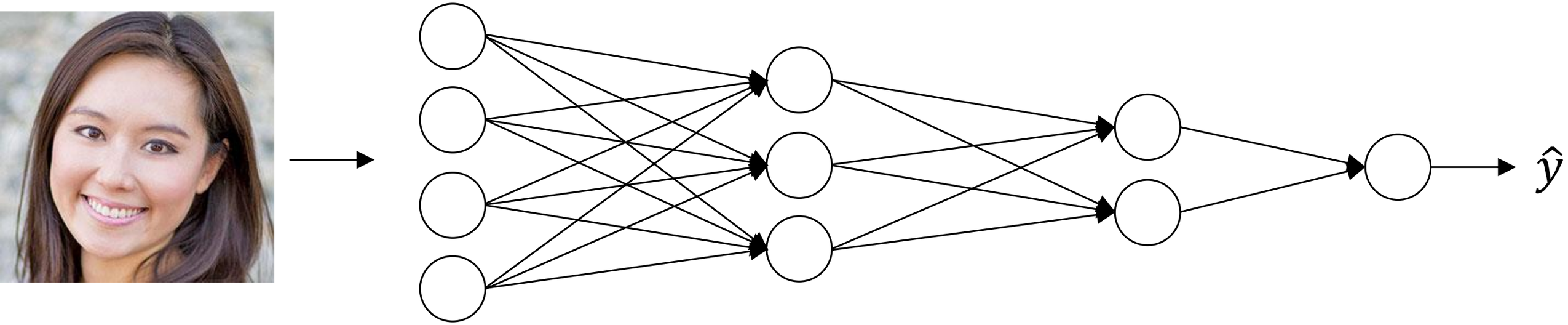
Seyoung Yun

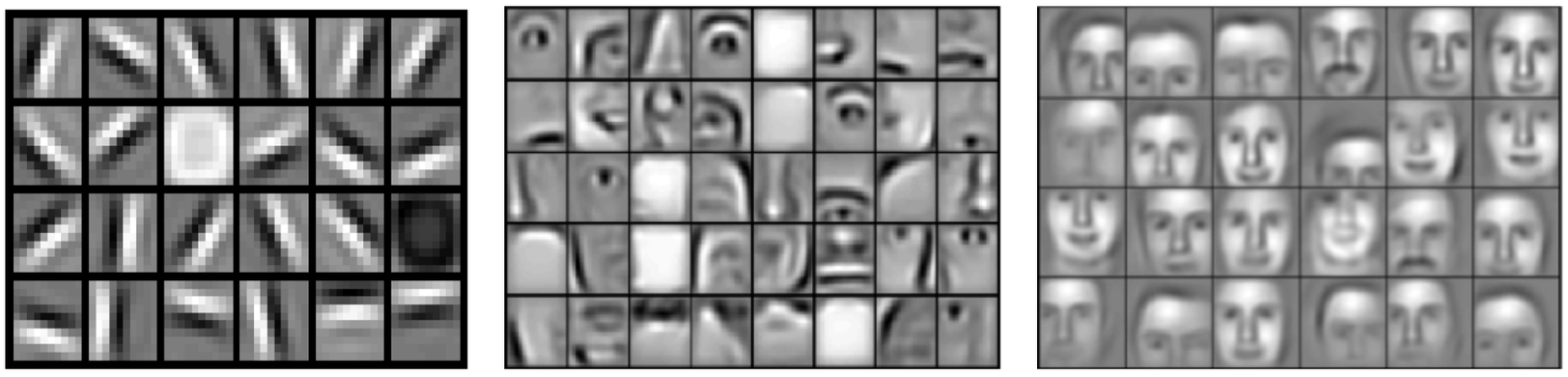
- http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture4.pdf

- http://cs231n.github.io/optimization-2/

- http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf

Intuition about deep representation

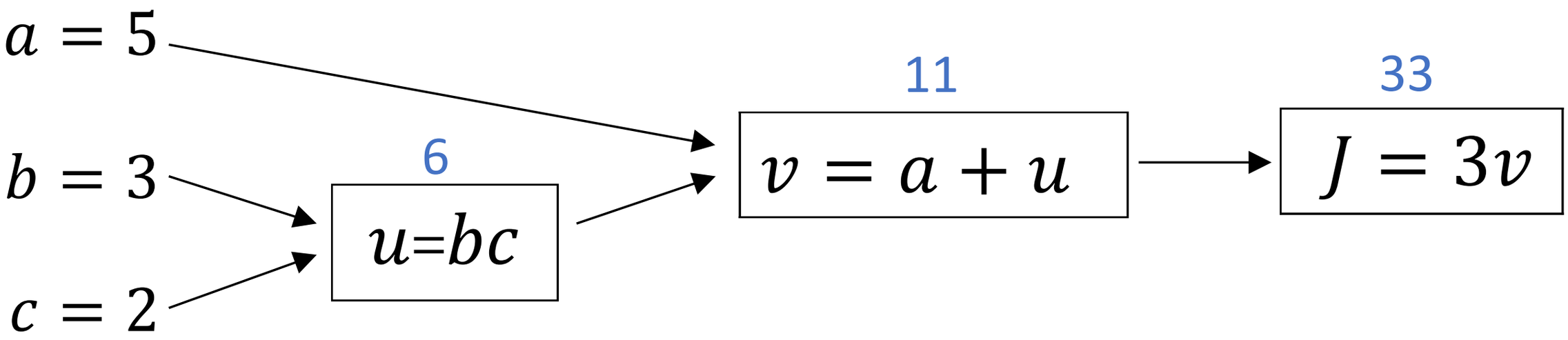




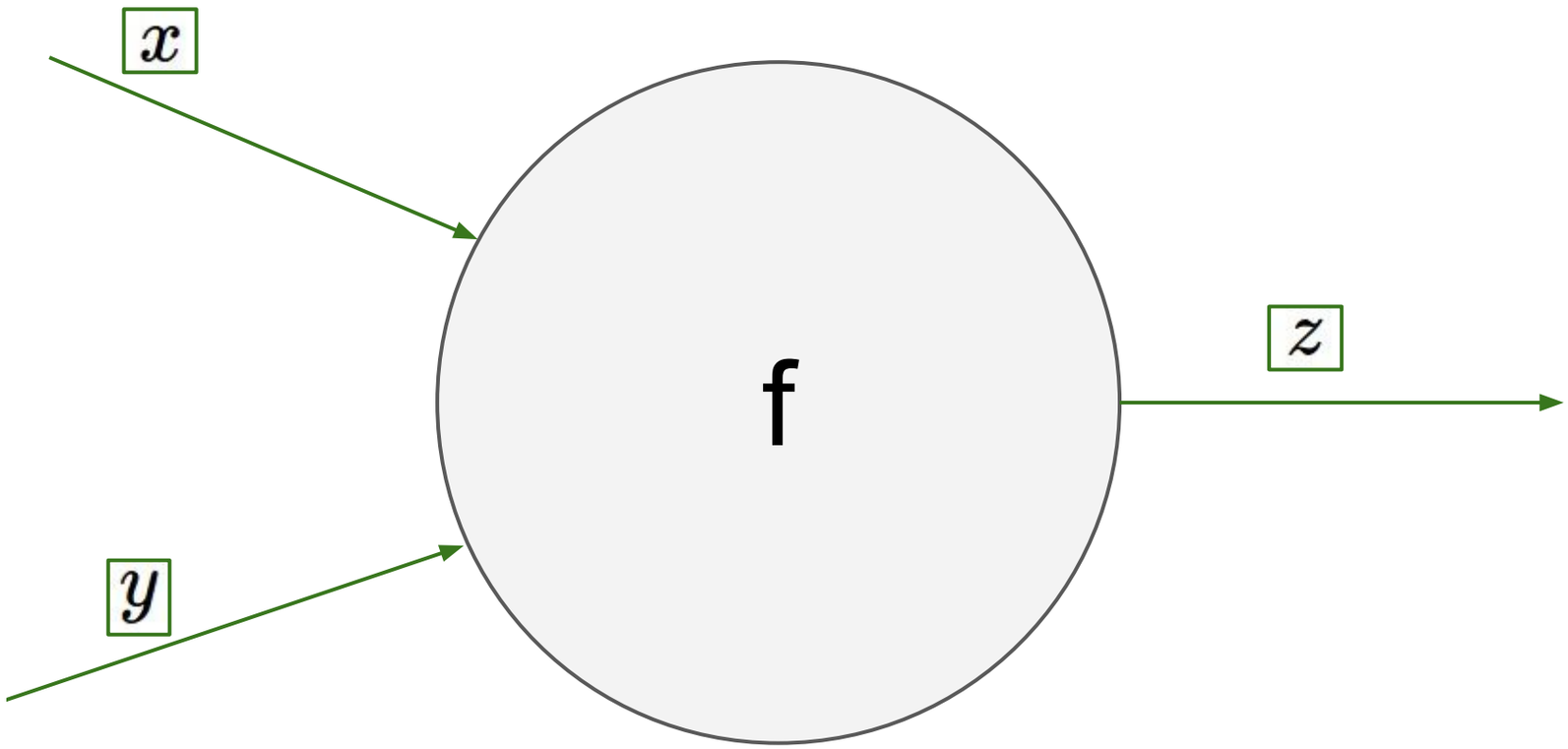
Informally: There are functions you can compute with a “small” L-layer deep neural network that shallower networks require exponentially more hidden units to compute.

Computational Graph

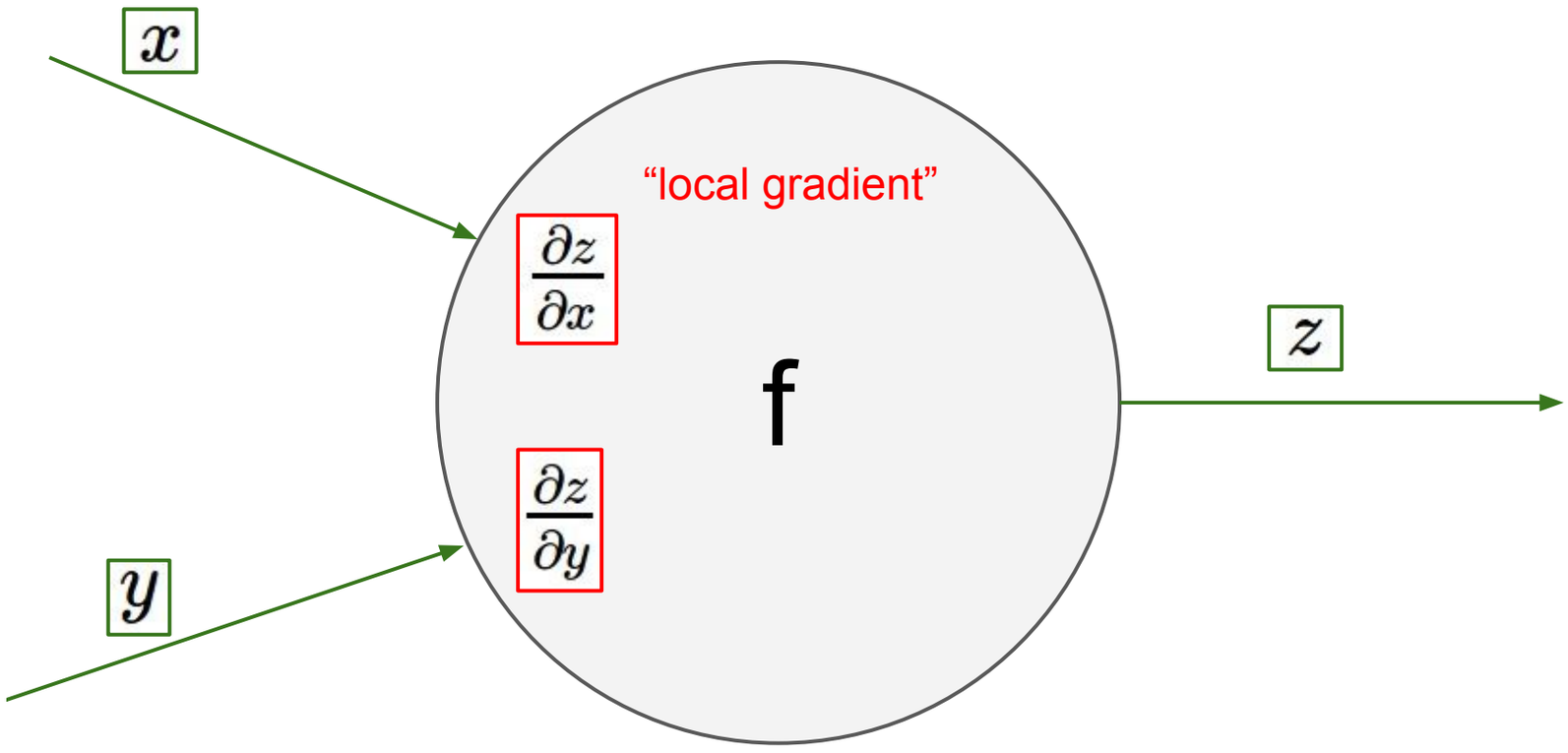
Computing derivatives



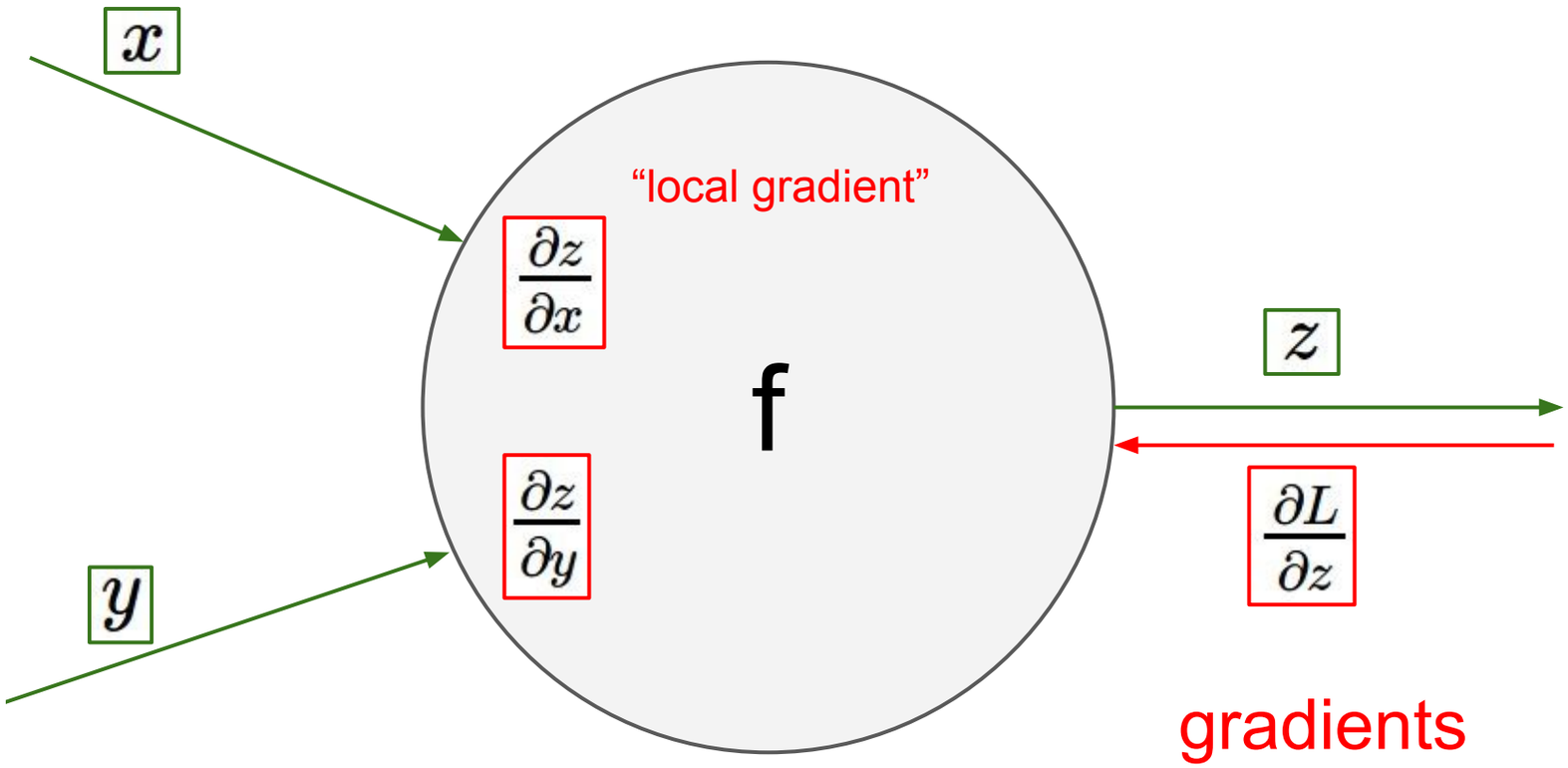
Backpropagation



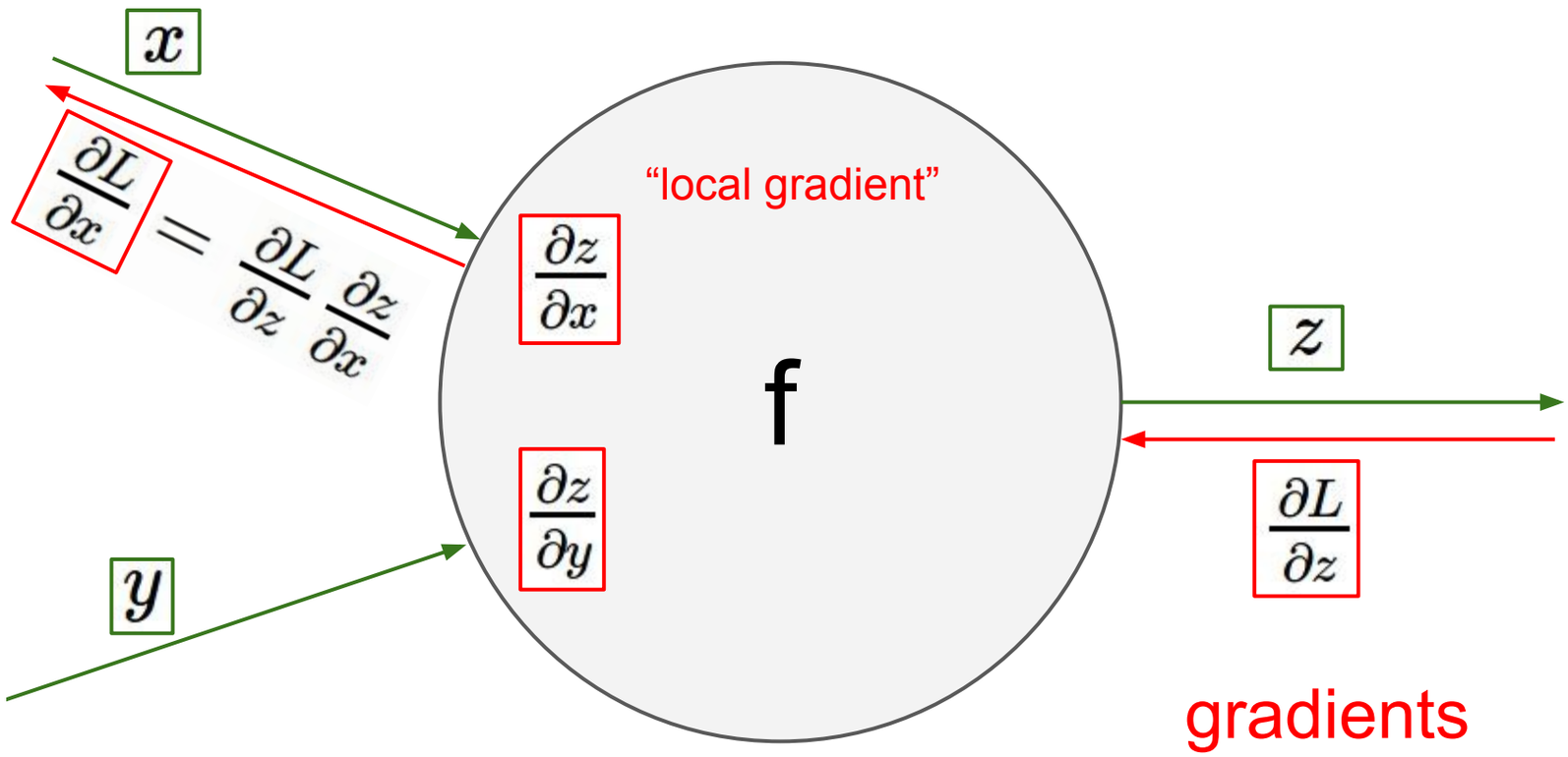
Backpropagation



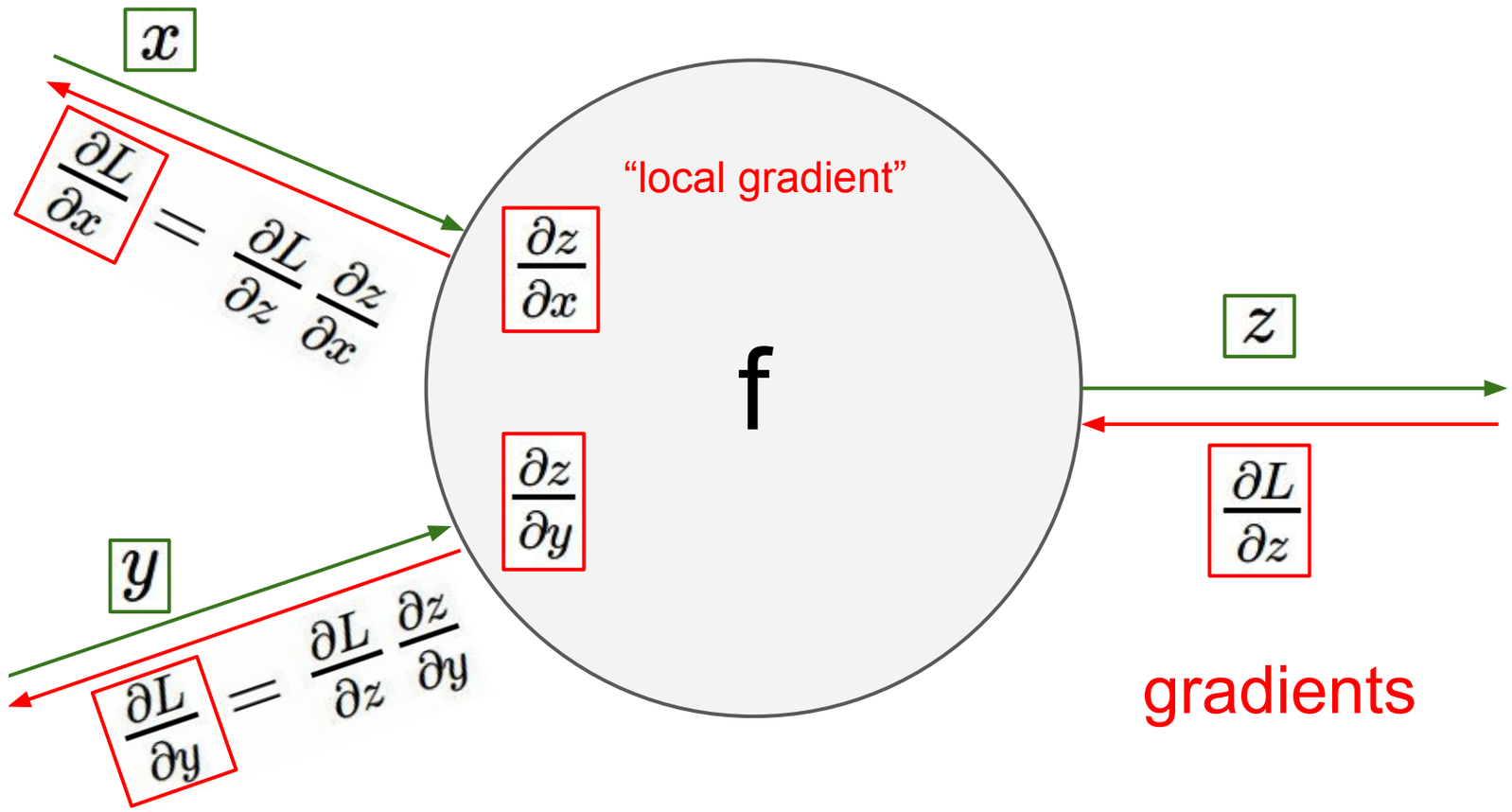
Backpropagation



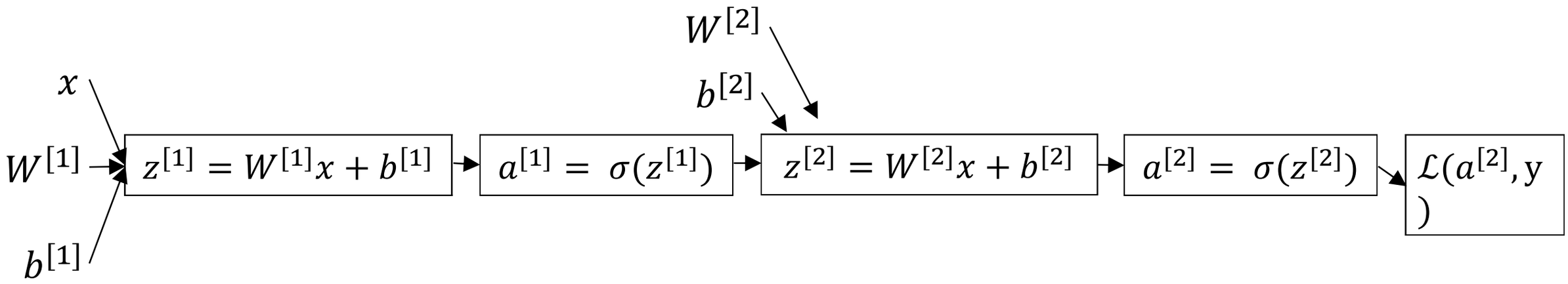
Backpropagation



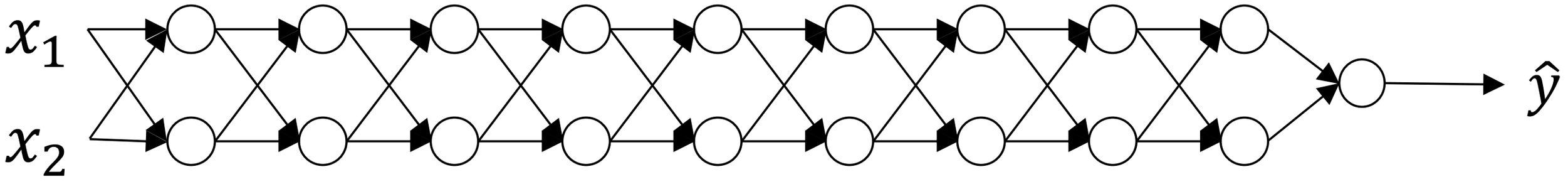
Backpropagation



Neural Net gradients



Vanishing/exploding gradients



Derivatives of activation function

- Sigmoid and tanh

Derivatives of activation function

- ReLu