

Deep Feedforward Networks

Lecture slides for Chapter 6 of *Deep Learning*

www.deeplearningbook.org

Ian Goodfellow

2016-09-26

Solving XOR

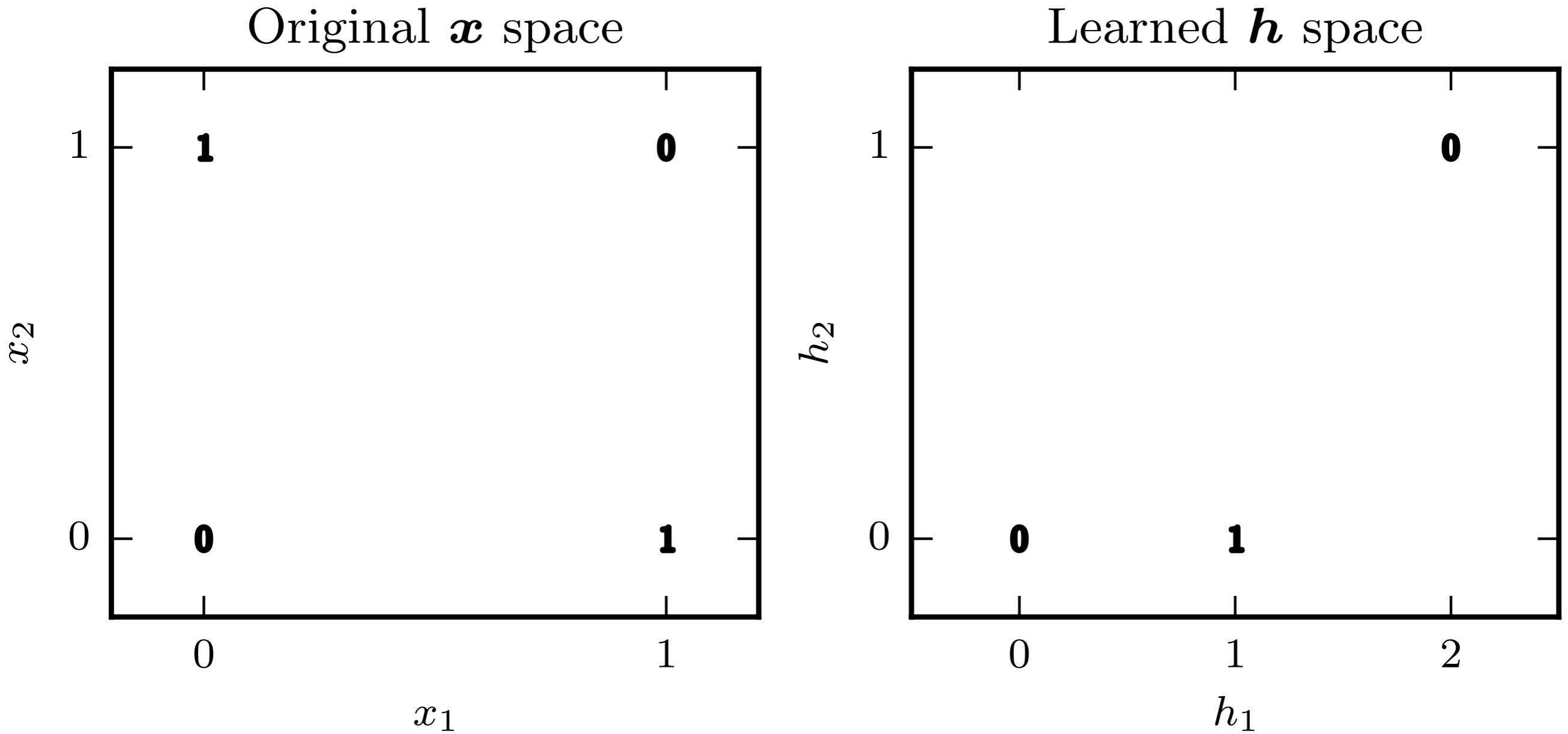


Figure 6.1

Network Diagrams

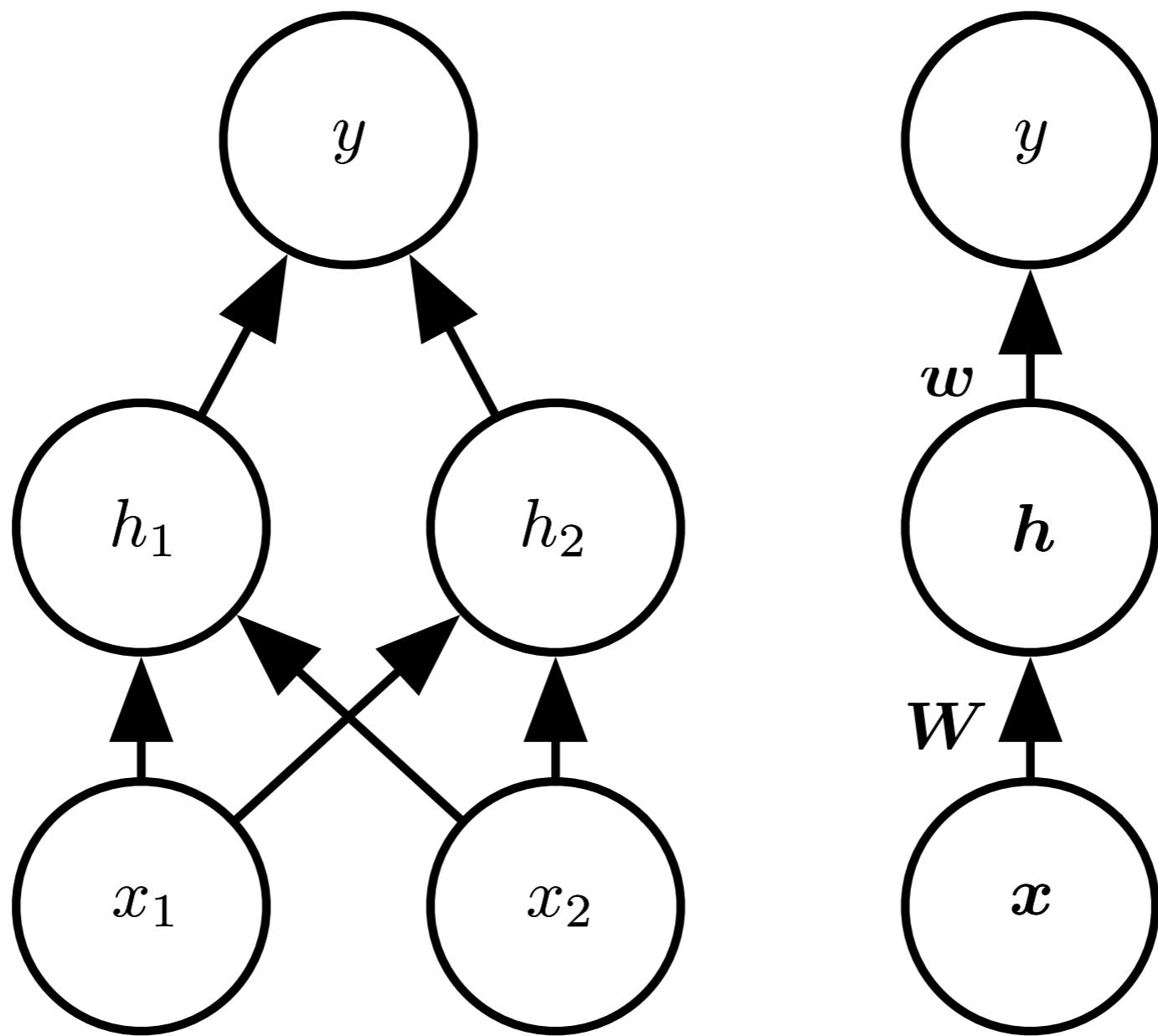


Figure 6.2

Rectified Linear Activation

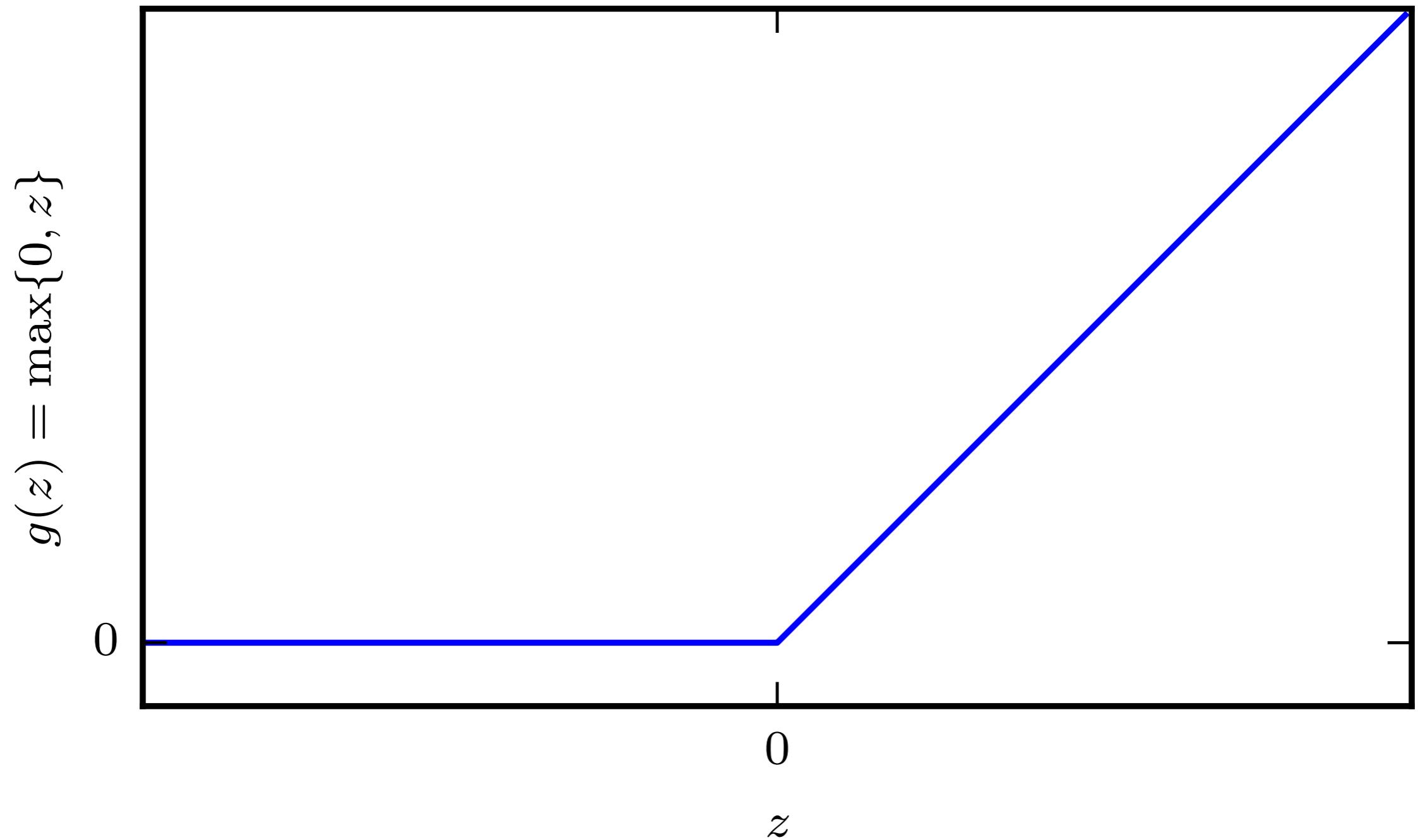


Figure 6.3

Mixture Density Outputs

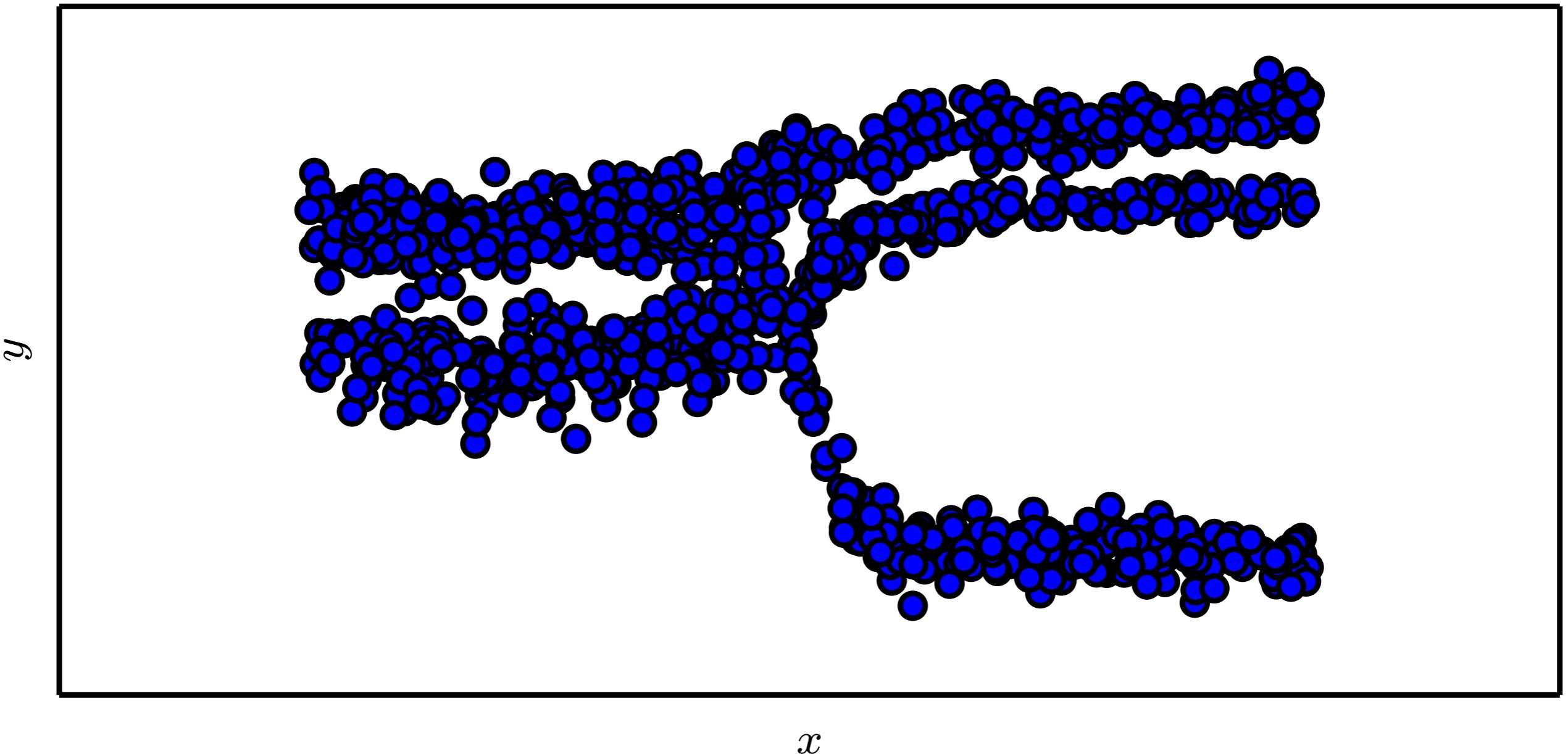


Figure 6.4

Exponential Advantage of Depth

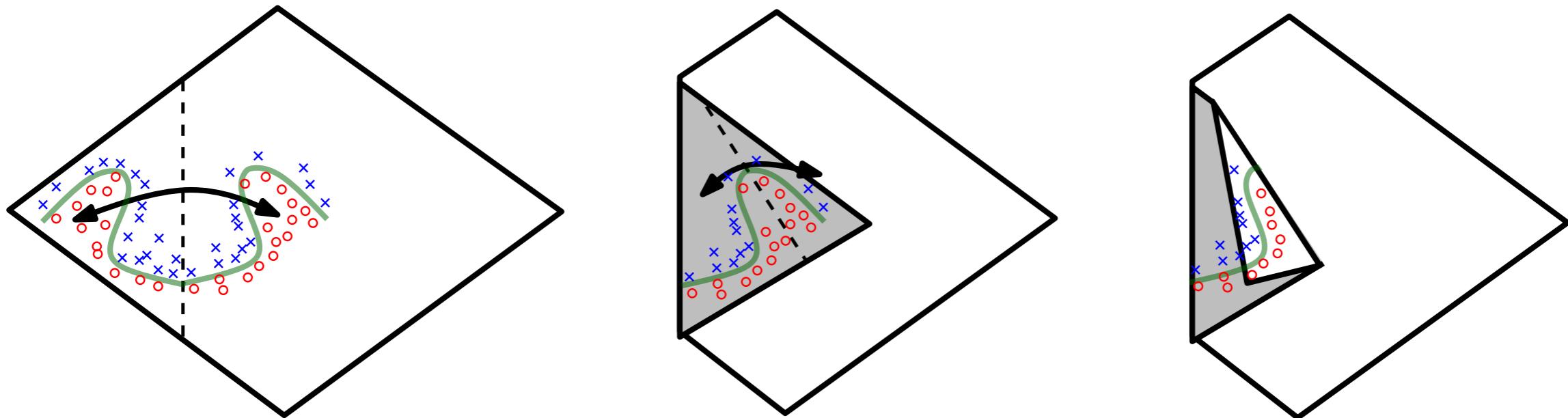


Figure 6.5

Better Generalization with Greater Depth

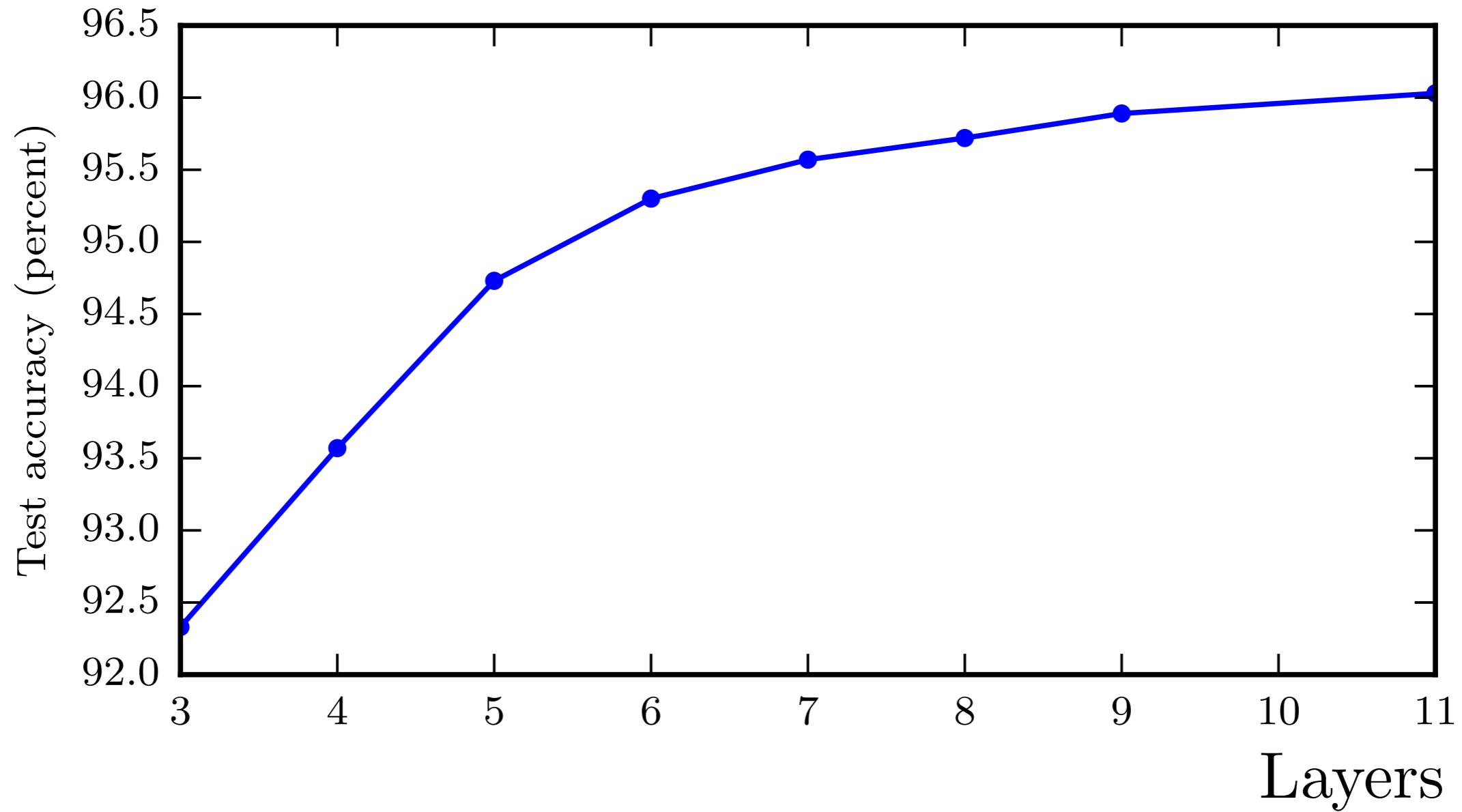


Figure 6.6

(Goodfellow 2016)

Large, Shallow Models Overfit More

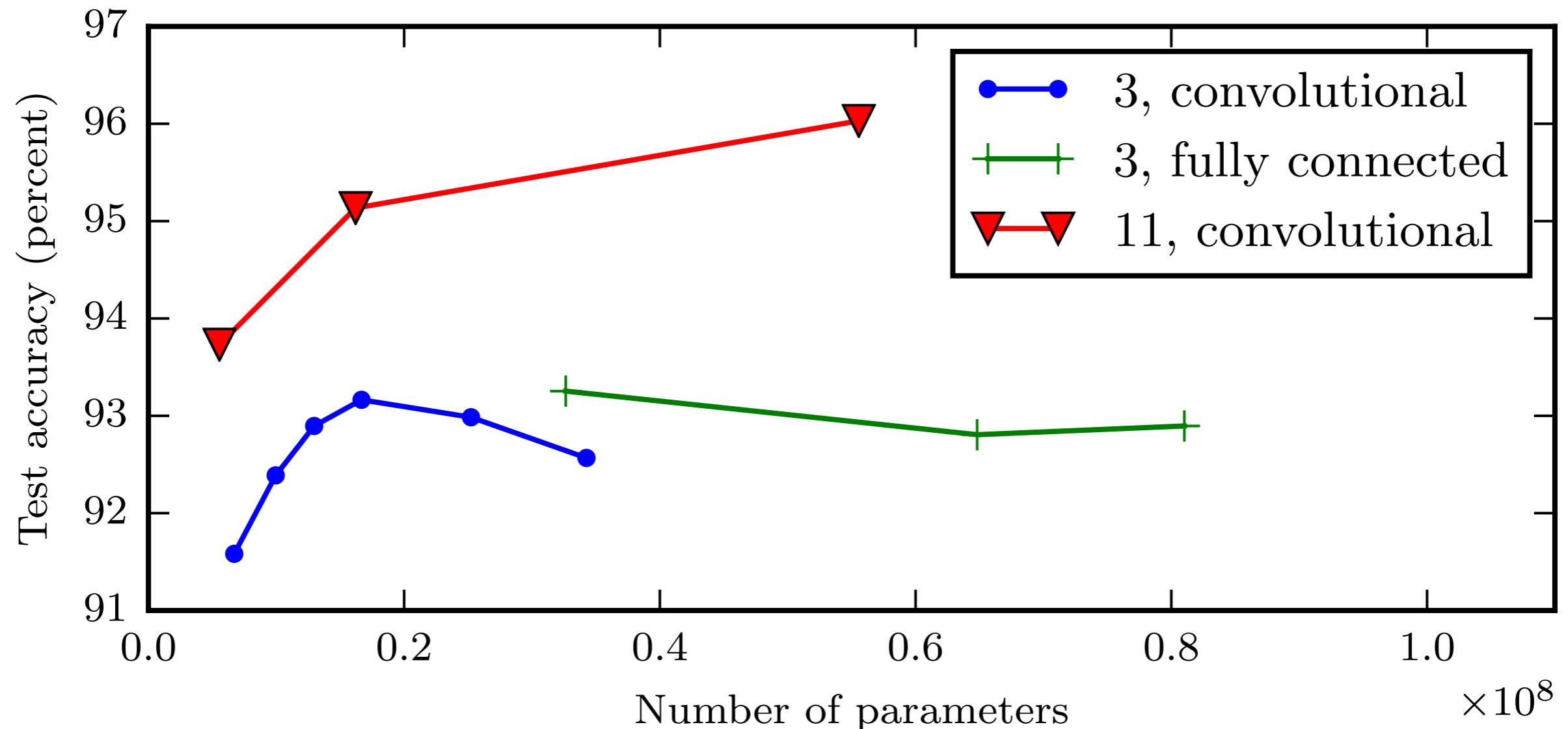
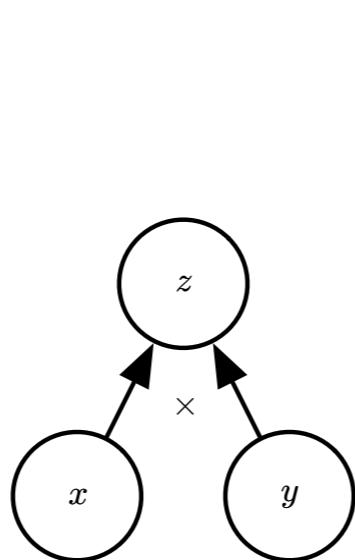


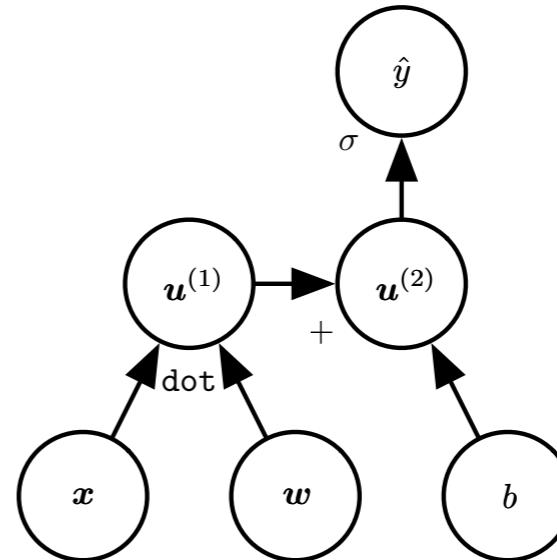
Figure 6.7

(Goodfellow 2016)

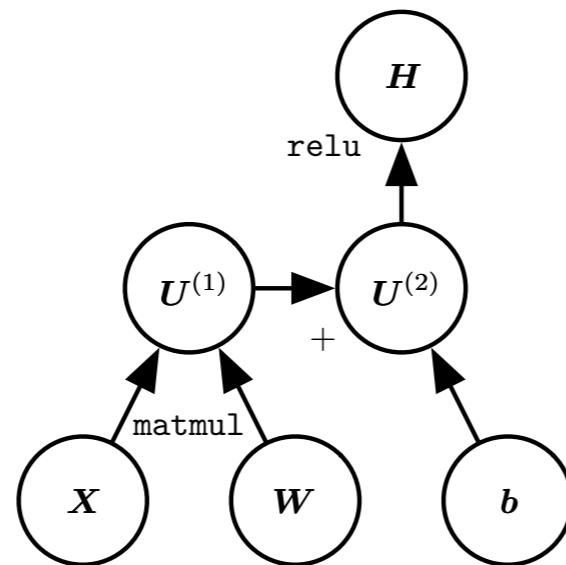
Computation Graphs



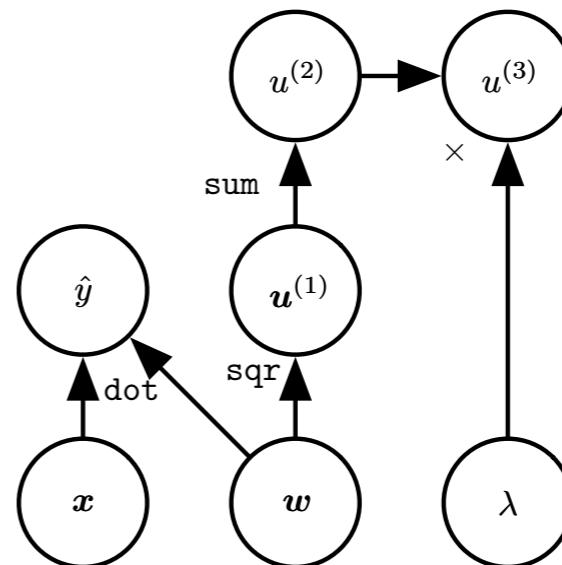
(a)



(b)



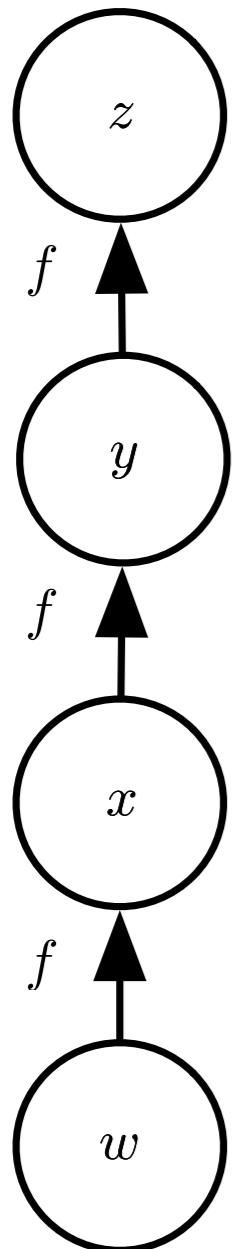
(c)



(d)

Figure 6.8

Repeated Subexpressions



$$\frac{\partial z}{\partial w} \tag{6.50}$$

$$= \frac{\partial z}{\partial y} \frac{\partial y}{\partial x} \frac{\partial x}{\partial w} \tag{6.51}$$

$$= f'(y) f'(x) f'(w) \tag{6.52}$$

$$= f'(f(f(w))) f'(f(w)) f'(w) \tag{6.53}$$

Figure 6.9

Symbol-to-Symbol Differentiation

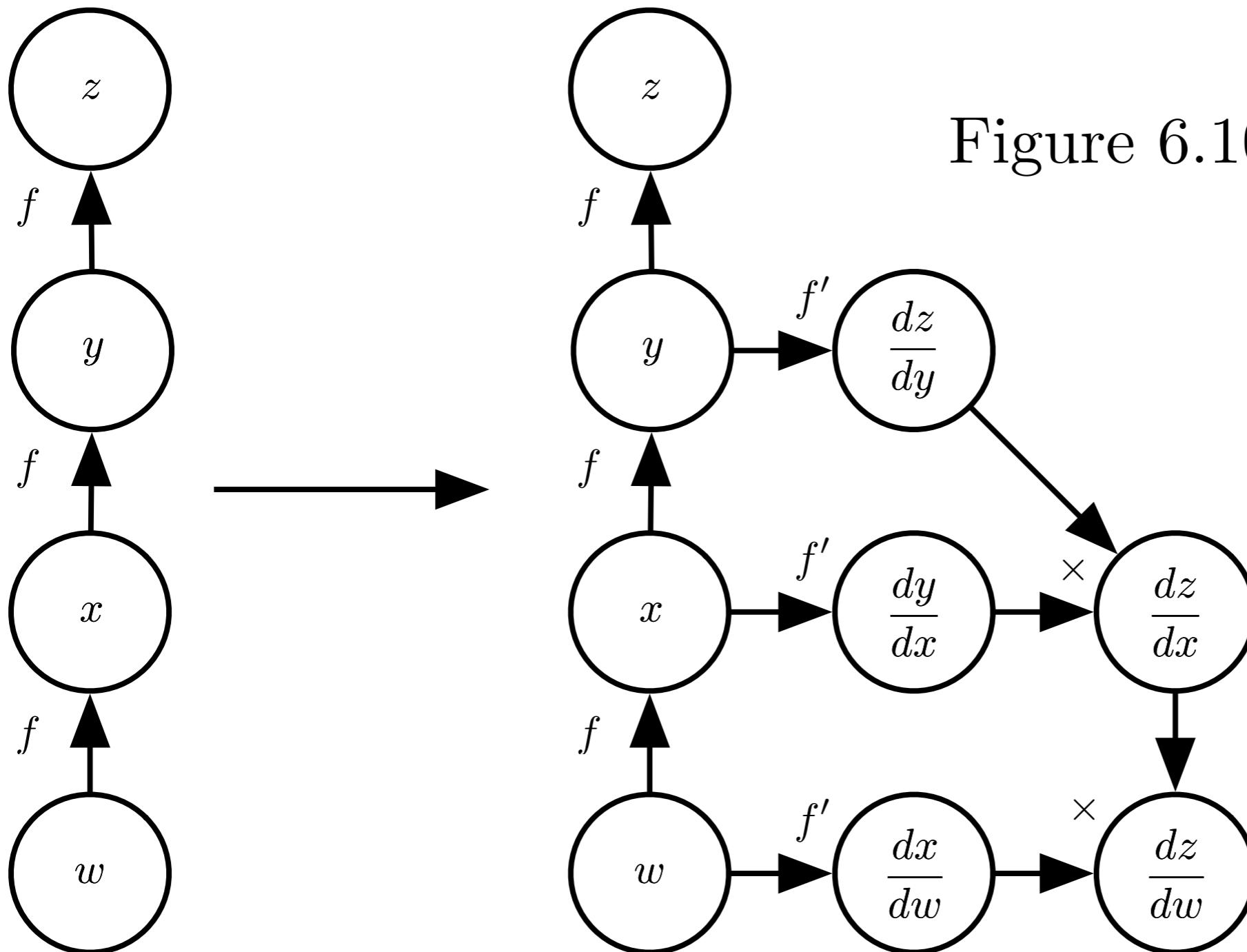


Figure 6.10

Neural Network Loss Function

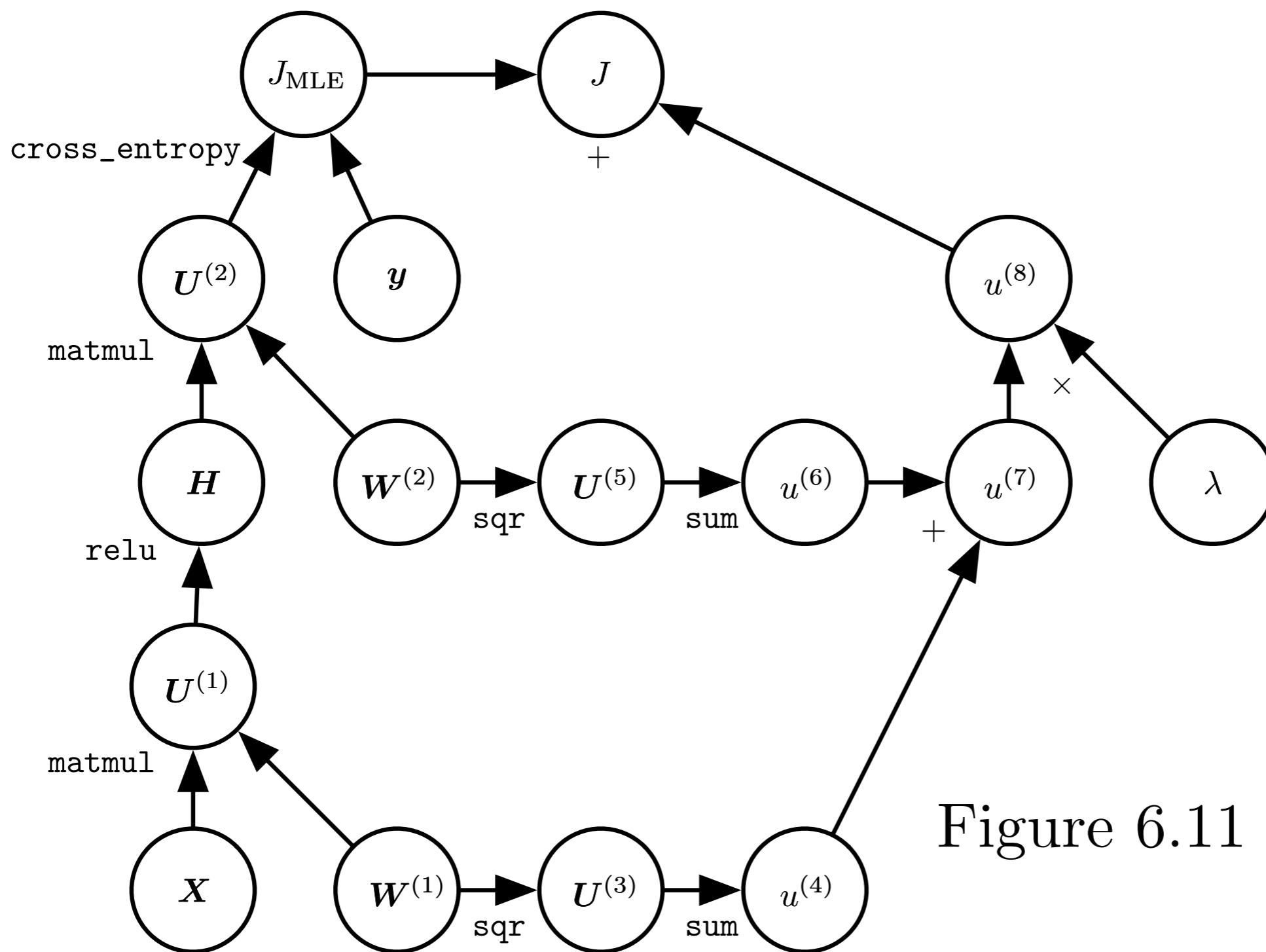


Figure 6.11

Hessian-vector Products

$$H\mathbf{v} = \nabla_{\mathbf{x}} \left[(\nabla_{\mathbf{x}} f(\mathbf{x}))^\top \mathbf{v} \right]. \quad (6.59)$$