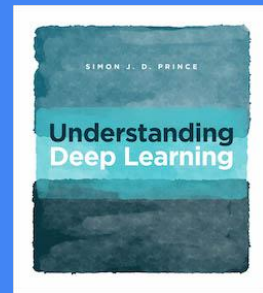
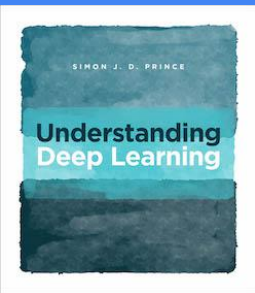


Fitting Models



Gradient Descent: Step 1

Compute the derivatives of the loss with respect to the parameters

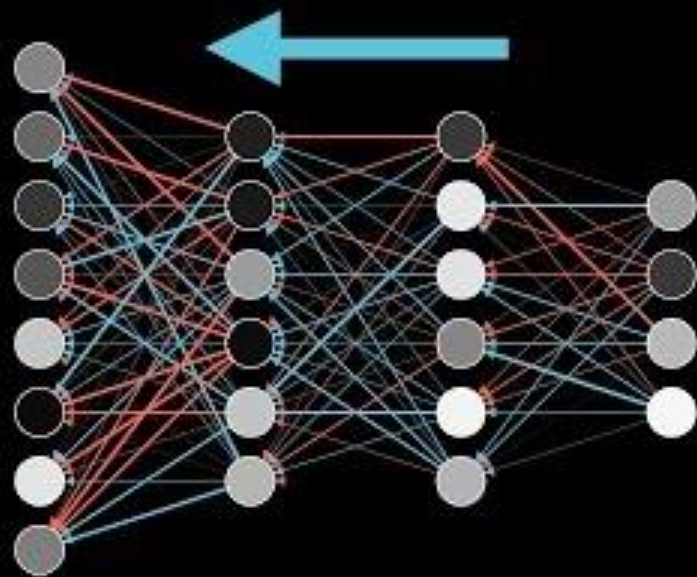
$$\frac{\partial L}{\partial \phi} = \begin{bmatrix} \frac{\partial L}{\partial \phi_0} \\ \frac{\partial L}{\partial \phi_1} \\ \vdots \\ \frac{\partial L}{\partial \phi_N} \end{bmatrix}$$

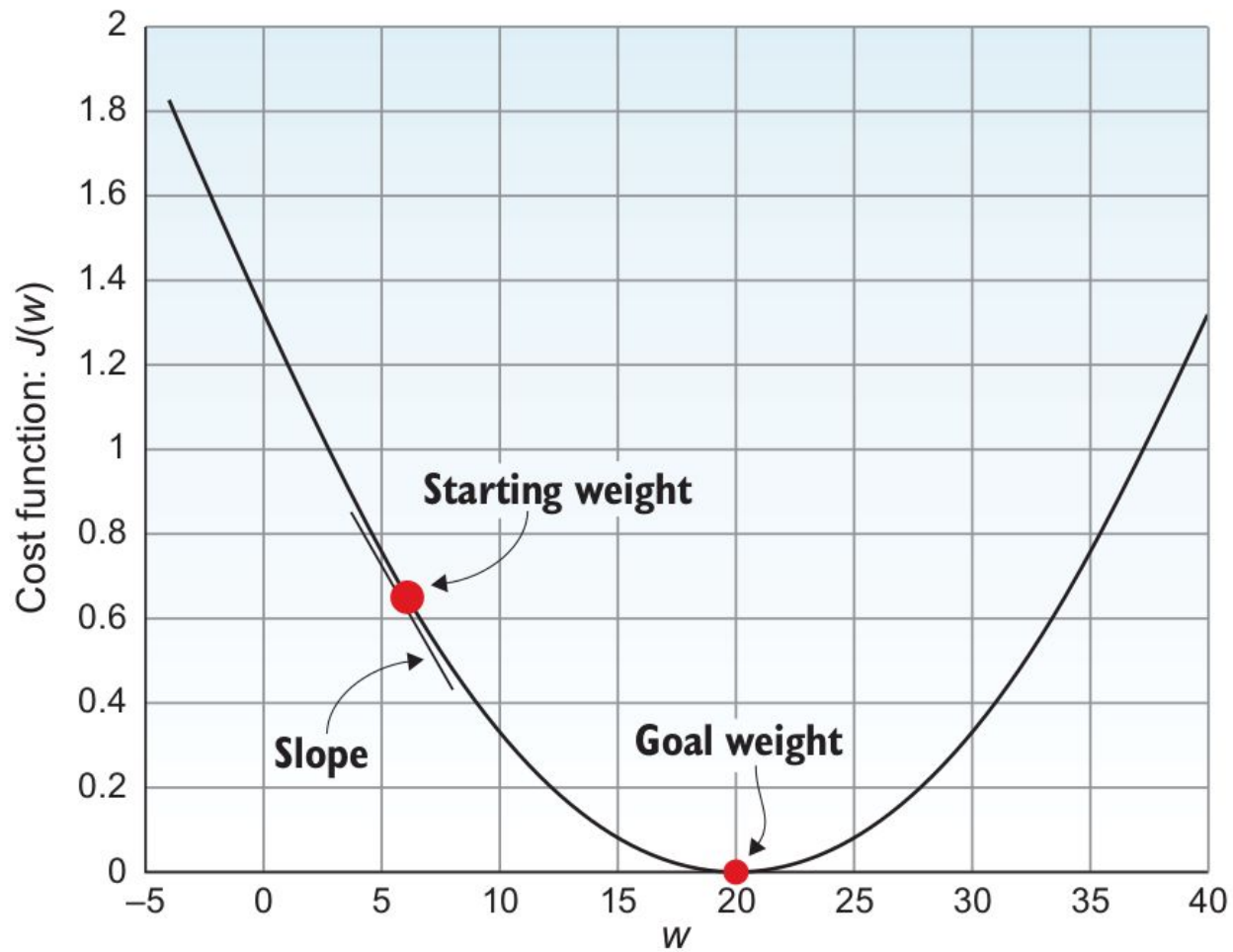
Gradient Descent: Step 1

Update the parameters according to the rule

$$\phi \longleftarrow \phi - \alpha \cdot \frac{\partial L}{\partial \phi}$$

Backpropagation





Partial derivative

$$\text{Constant Rule: } \frac{d}{dx} c = 0$$

$$\text{Constant Multiple Rule: } \frac{d}{dx} [cf(x)] = cf'(x)$$

$$\text{Difference Rule: } \frac{d}{dx} [f(x) - g(x)] = f'(x) - g'(x)$$

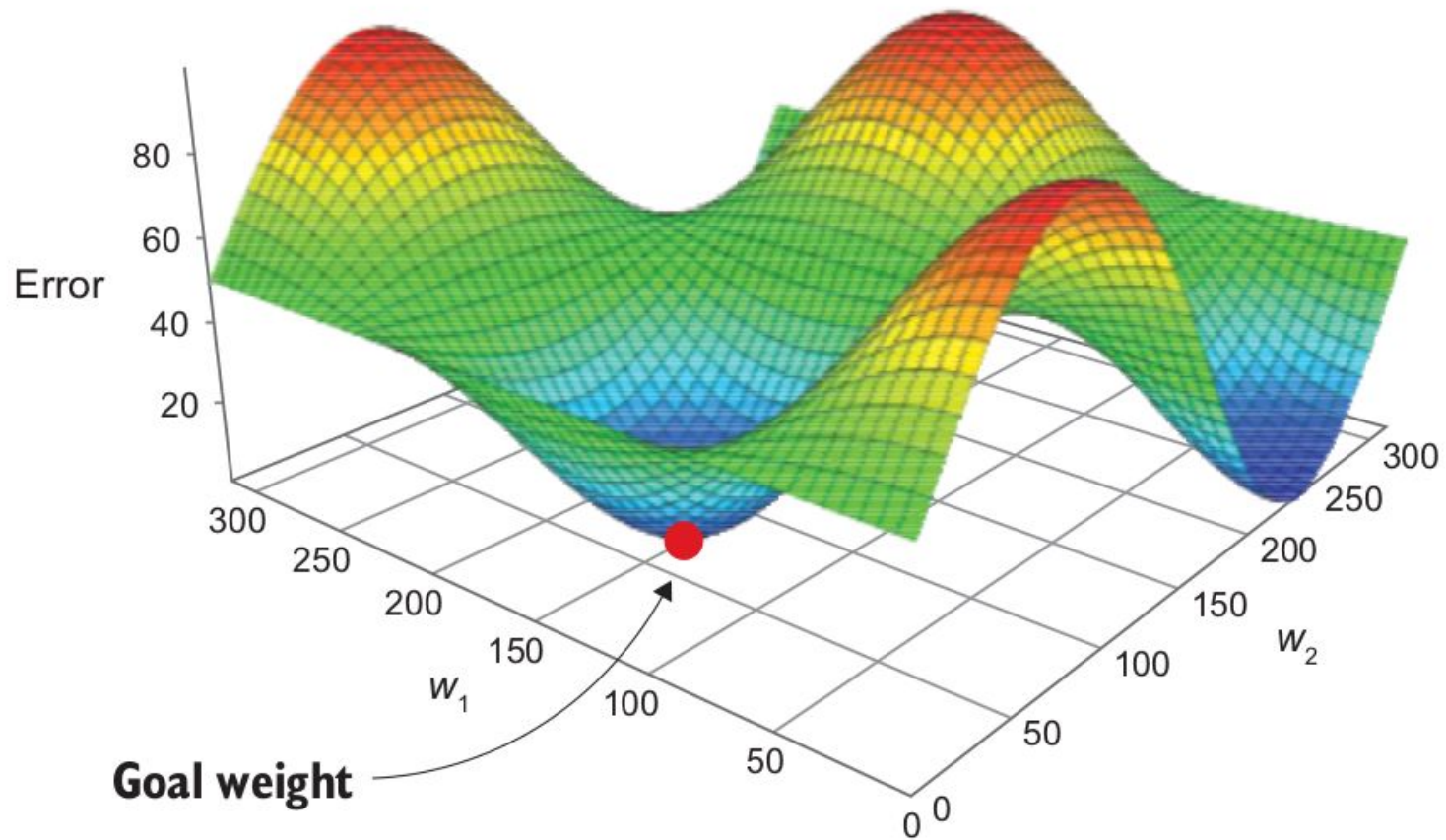
$$\text{Sum Rule: } \frac{d}{dx} [f(x) + g(x)] = f'(x) + g'(x)$$

$$\text{Product Rule: } \frac{d}{dx} [f(x)g(x)] = f(x)g'(x) + g(x)f'(x)$$

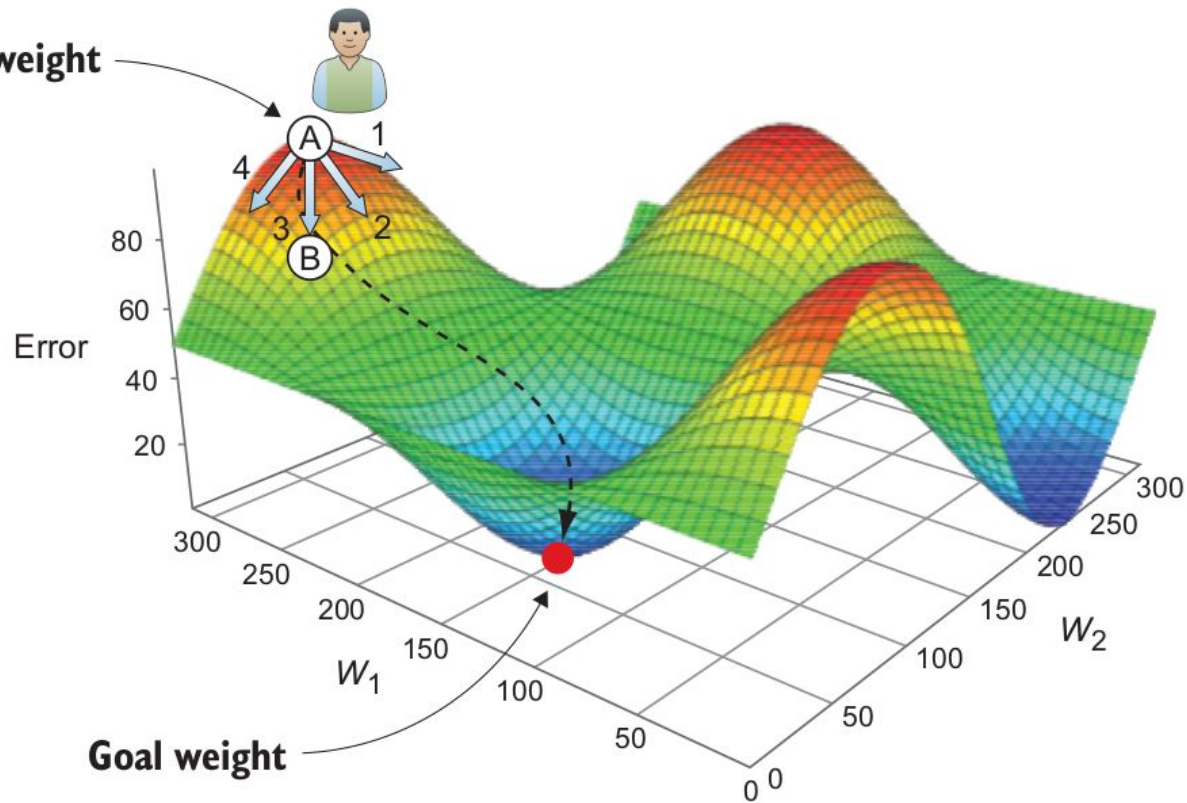
$$\text{Quotient Rule: } \frac{d}{dx} \left[\frac{f(x)}{g(x)} \right] = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$$

$$\text{Power Rule: } \frac{d}{dx} x^n = nx^{n-1}$$

$$\text{Chain Rule: } \frac{d}{dx} [f(g(x))] = f'(g(x))g'(x)$$



Starting weight



Goal weight

$$\Delta w_i = -\alpha \frac{dE}{dw_i}$$

$$w_{\text{next-step}} = w_{\text{current}} + \Delta w$$

The diagram shows the weight update equation $W_{new} = W_{old} - \alpha \left(\frac{\partial Error}{\partial W_x} \right)$. Four labels with arrows point to the components of the equation: 'Old weight' points to W_{old} , 'Derivative of error with respect to weight' points to $\frac{\partial Error}{\partial W_x}$, 'New weight' points to W_{new} , and 'Learning rate' points to α .

Old weight

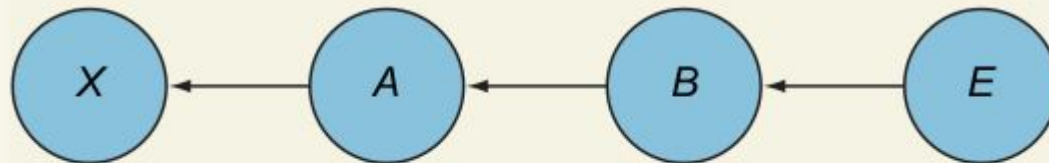
Derivative of error with respect to weight

$$W_{new} = W_{old} - \alpha \left(\frac{\partial Error}{\partial W_x} \right)$$

New weight

Learning rate

Chain Rule: $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$



$$\frac{dE}{dx} = \frac{dE}{dB} \cdot \frac{dB}{dA} \cdot \frac{dA}{dx}$$

Reading for Next Class

Chapter 7 from UDL

