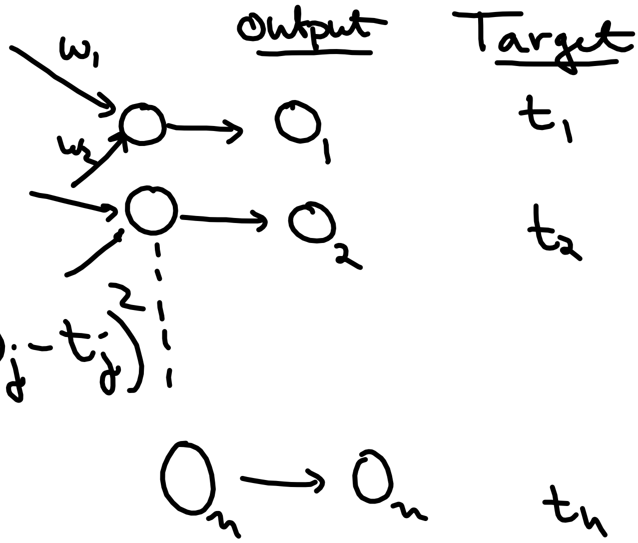
Training



Error

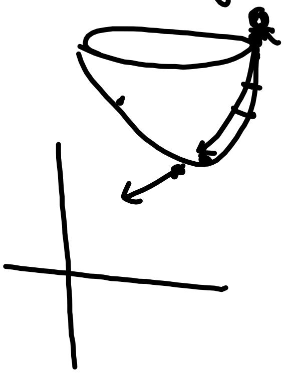
$$E = \frac{1}{n}\sum_{j=1}^{n}(O_j - t_j)^2$$

$$E(\vec{w})$$

weights

Gradient Descent

$f(x,y)$



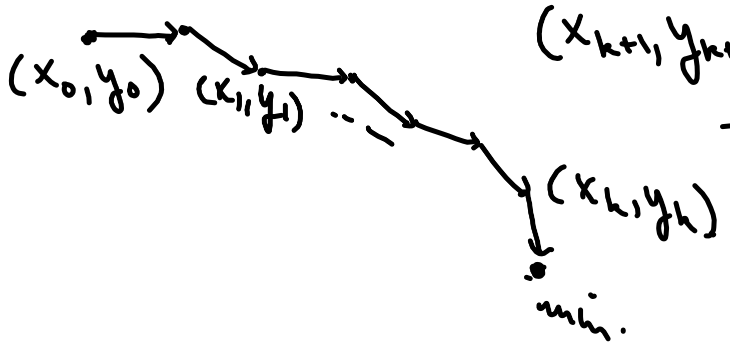
the direction of steepest descent =

$$-\frac{\left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}}.$$

Identify the minimum by when the gradient is $\vec{0}$.

Gradient Descent

$\equiv$ moving one small step at a time in the direction of the gradient.

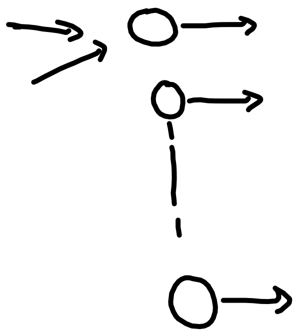


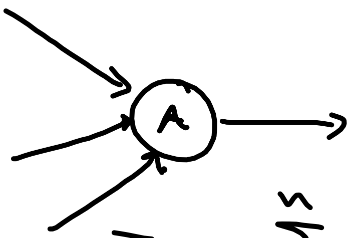
$$(x_{k+1}, y_{k+1}) = (x_k, y_k) - \eta \left\langle \frac{\partial f}{\partial x}(x_k, y_k), \frac{\partial f}{\partial y}(x_k, y_k) \right\rangle$$

Back Propagation

an algorithm to calculate the gradient of the error with respect to the weights.

Single Layer





$$O_A = \phi(I_A)$$

$$I_A = \sum_{i=1}^{n} \underline{w_i}\, x_i$$

$$E = \frac{1}{2}\left\{ (\underline{\underline{O_A}} - t_A)^2 + (\underline{\underline{O_B}} - t_B)^2 \cdots \right\}$$

depends on the weights for B.

$$\frac{\partial E}{\partial w_i} = ?$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \frac{1}{2}\left(O_A - t_A\right)^2 \qquad I_A = \sum_{i=1}^{n} w_i x_i$$

$$\boxed{O_A = \phi(I_A)}$$

<u>Chain Rule</u>

$$= \frac{2}{2}\left(O_A - t_A\right) \frac{\partial}{\partial w_i} O_A$$

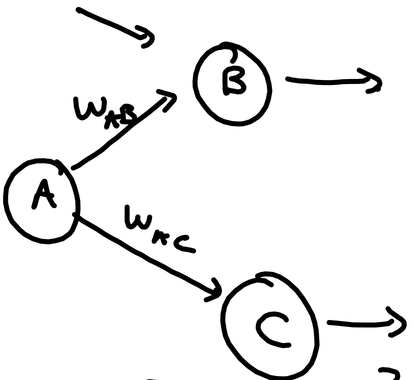
$$= \left(O_A - t_A\right) \frac{\partial O_A}{\partial I_A} \frac{\partial I_A}{\partial w_i}$$

$$= \left(O_A - t_A\right) \phi'(I_A)\, x_i$$

$$\frac{\partial E}{\partial w_i} = (O_A - t_A)\,\underbrace{\phi(I_A)(1-\phi(I_A))}_{\phi'(I_A)}\;x_i$$

$$= \boxed{(O_A - t_A)\,O_A(1-O_A)}\,x_i$$

$$w_i^{(k+1)} = w_i^{(k)} - \eta\,(O_A - t_A)\,O_A(1-O_A)\,x_i$$



$O_B - t_B$

$O_C - t_C$

$O_B = \phi(I_B)$ $\sum\{\text{weights} \times \text{outputs}\}$

$= \cdots\cdots + w_{AB} O_A.$

$$E = \frac{1}{2}\left\{(O_B - t_B)^2 + (O_C - t_C)^2\right\}$$

$$\frac{\partial E}{\partial w_{AB}} = \frac{2}{2}(O_B - t_B)\frac{\partial O_B}{\partial w_{AB}}$$

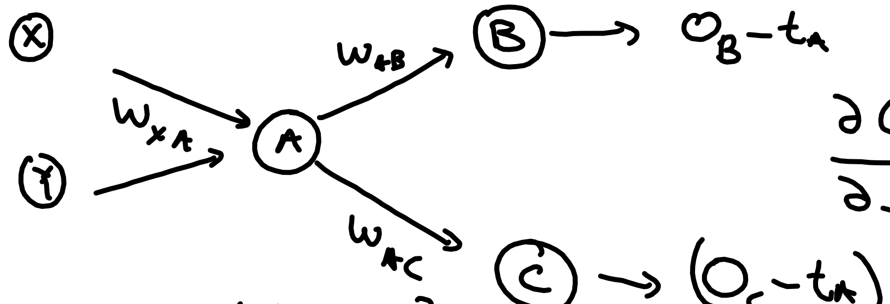
$$= (O_B - t_B)\,\phi'(I_B)\,\partial I_B / \partial w_{AB}$$

$$= (O_B - t_B)\,\phi'(I_B)\,O_A$$

$$= (O_B - t_B)\,\underbrace{\phi(I_B)}_{O_B}\,(1-\underbrace{\phi(I_B)}_{O_B})\,O_A.$$

$$= (O_B - t_B)\,O_B(1-O_B)\,O_A$$

$$w_{AB}^{new} = w_{AB}^{old} - \eta\,(O_B - t_B)\,O_B(1-O_B)\,O_A$$



$$E = \frac{1}{2}\left((O_B - t_B)^2 + (O_C - t_C)^2 \right).$$

$$\frac{\partial E}{\partial W_{XA}} = \frac{\partial E}{\partial O_A} \cdot \frac{\partial O_A}{\partial I_A} \cdot \underbrace{\frac{\partial I_A}{\partial W_{XA}}}_{O_X}$$

$$\frac{\partial O_A}{\partial I_A} = \frac{\partial \phi(I_A)}{\partial I_A}$$

$$= \phi'(I_A)$$

$$= \phi(I_A)(1 - \phi(I_A))$$

$$= O_A(1 - O_A)$$

$$\frac{\partial E}{\partial O_A} = \frac{\partial}{\partial O_A}\frac{1}{2}\left\{ (O_B - t_B)^2 + (O_C - t_C)^2 \right\}$$

$$= (O_B - t_B)\underbrace{\frac{\partial O_B}{\partial O_A}} + (O_C - t_C)\frac{\partial O_C}{\partial O_A} = \phi'(I_C)\,w_{AC} \quad O_C(1-O_C)\,w_{AC}$$

$$\frac{\partial \phi(I_B)}{\partial O_A} = \phi'(I_B)\,w_{AB}$$

$$O_B(1-O_B)\,w_{AB}$$

$$\frac{\partial E}{\partial O_A} = (O_B - t_B)\,O_B(1-O_B)\,w_{AB} + (O_C - t_C)\,O_C(1-O_C)\,w_{AC}$$

let $\delta_Y = (Y - t_Y)\, O_Y (1 - O_Y)$

$$\frac{\partial E}{\partial w_{XA}} = w_{AB}\,\underline{\delta_B} + w_{AC}\,\underline{\delta_C}$$