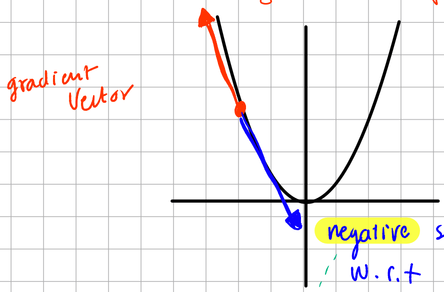
I choose to Lying Flat

Daily Reminder

beautiful things happen
when you do the work to reprogram
that negative voice in your head

To Formalize: We are approaching the training as an optimization problem in an iterative way meaning we are starting from an initial guess, we compute ΔW, we sum it up, update the weights and we're moving along in the error function.

Now, to compute the gradient to perform the minimization



negative search step w.r.t gradient vector

$$\boxed{\frac{\partial E}{\partial w} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial w}} \rightarrow$$ we have the error function derived by $W_1, W_2 \ldots$ (it is a gradient, a vector)

Label output input

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial w} = -\underbrace{(\bar{y} - y)}_{\text{Error w.r.t } y} \cdot x = -\delta \cdot x$$

Then $\Delta W = f\left(\frac{\partial E}{\partial w}\right) \longrightarrow W_{new} = W_{old} + \Delta W$

$$\Delta W = \eta \delta x$$

How to compute the analytical Formulation?

$$E = \sum_{k=1}^{R} \sum_{i=1}^{N} \frac{1}{2} \left(t_i^{(k)} - u_i^{(k)} \right)^2$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial \left(\frac{1}{2} \left(t_i^{(k)} - u_i^{(k)} \right)^2 \right)}{\partial w_{ij}}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial \left(\frac{1}{2} \left(t_i^{(K)} - u_i^{(K)} \right)^2 \right)}{\partial w_{ij}}$$

Summations are not necessary in derivatives

We apply the chain rule

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial \left(\frac{1}{2} \left(t_i - u_i \right)^2 \right)}{\partial u_i} \cdot \frac{\partial u_i}{\partial w_{ij}} \rightarrow -(t_i - u_i) \cdot \frac{\partial u_i}{\partial w_{ij}}$$

We know that $u_i = f(P_i) = f\left[\sum_j^M w_{ij} x_j \right]$ We see dependency of u_i on f (activation F)

$$-(t_i - u_i) \cdot \frac{\partial u_i}{\partial P_i} \cdot \frac{\partial P_i}{\partial w_{ij}} \rightarrow -(t_i - u_i) \cdot f'(P_i) \cdot \frac{\partial P_i}{\partial w_{ij}}$$

$u' = f'(P_i)$ $f \neq$ Heaviside Function.

if the activation function is linear $f' = 1$, if not it depends on the action potential P, slope at certain P.

$$-(t_i - u_i) \cdot f'(P_i) \cdot \frac{\partial \left(\sum_{j=1}^M w_{ij} x_j \right)}{\partial w_{ij}} \rightarrow \boxed{-(t_i - u_i) \cdot f'(P_i) \cdot x_j}$$

So we might finally conclude that

$$\Delta w_{ij} \approx \left(\frac{\partial E}{\partial w_{ij}} \right)$$

$$\Delta w_{ij} = \eta \cdot \sum_{K=1}^{R} \left(t_i^{(K)} - u_i^{(K)} \right) \cdot f'(P_i) \cdot x_j^{(K)}$$