$$\frac{\partial E}{\partial w} = \underbrace{\frac{1}{m}\sum^{1}\left(w^{(k)}x_i - t_i\right)x_i}_{\text{average error of sample}}$$

To find average of like 7×10^8 is unrealistic

So we need to find the average of a sample

$$w^{(k+1)} \leftarrow w^{(k)} - \eta \left.\frac{\partial E}{\partial w}\right|_{w=w^{(k)}} \quad [500 \text{ samples}]$$

$$(x_{i_1}, t_1), (x_{i_2}, t_{i_2}) \cdots (x_{i_k}, t_{i_k}) \quad k \leq m$$

$$\frac{1}{k}\sum_{i=1}^{k}\underbrace{(wx_{i_l} - t_{i_l})(x_{il})}_{\forall\, l \text{ in range } 0,(\text{len}(x))}$$

$$w^{k+1} \leftarrow w^{(k+1)} - \eta \left.\frac{\partial E}{\partial w}\right|_{w=w^{(k+1)}} \quad [500 \text{ other samples}]$$

This is (mini) batch gradient descent

for stochastic gradient descent, sample size = 1,