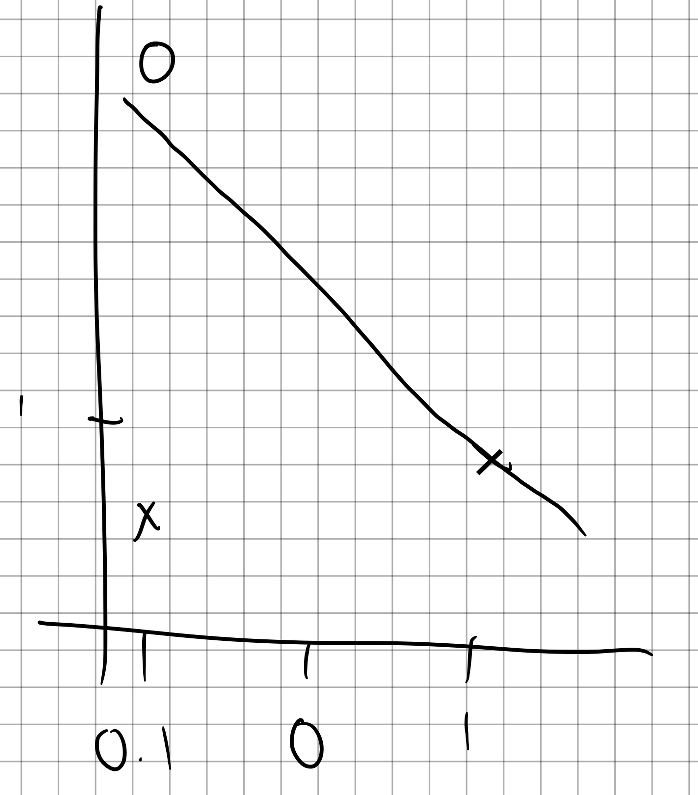


Empirical Risk =

Average loss on the training data

To train a model, you want to minimize the empirical risk.

We have no target function and don't know exact distribution of dataset

Target function =

What we want to reach, compare to actual function

$(1, \overset{\text{label}}{0.6}), (0.1, 0.1), (0.7, 0.4)$

$x \quad y$

$y = mx + b$

$L(y,v) = (y-v)^2$

P1.

$\hat{R}(w,b) = (w+b-0.6)^2 + (0.1w+b-0.1)^2 + (0.7w+b-0.4)^2$

P2.

w & b are learning parameters.

3. We minimize $\hat{R}(w,b)$

4.

$$\frac{\partial \hat{R}}{\partial w} = (w+b-0.6)^2 + (0.1w+b-0.1)^2 + (0.7w+b-0.4)^2$$

$$= 2(w+b-0.6)1 + 2(0.1w+b-0.1)0.1 + 2(0.7w+b-0.4)0.7$$

$$= 2(w+b-0.6) + 0.2(0.1w+b-0.1) + 1.4(0.7+b-0.4)$$

$$= 2w+2b-0.6+0.02w+0.2b-0.02+0.98+1.4b-0.4$$

$$= 2b+0.2b+1.4b$$

$$= 3.6b+3w-1.78$$

$$\frac{\partial \hat{R}}{\partial b} = 3.6w+6b-2.2$$

$$\begin{bmatrix} w^{t+1} \\ b^{t+1} \end{bmatrix} = \begin{bmatrix} w^t \\ b^t \end{bmatrix} - \eta \nabla \hat{R} = \begin{bmatrix} w^t \\ b^t \end{bmatrix} - \eta \begin{bmatrix} 3w^t + 3.6b^t - 1.76 \\ 3.6w^t + 6b^t - 2.2 \end{bmatrix}$$

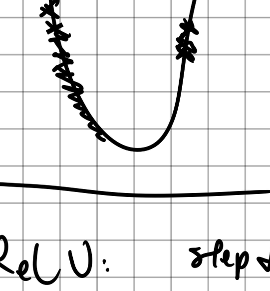
while $\hat{R}(w) \neq 0$

for $i=1 : I$ do

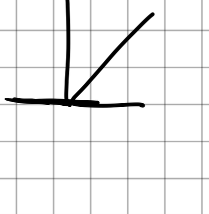
Determine $z_i := w^T x_i \leftarrow$ GD

if $s(z_i) \neq y_i$ then

$w \leftarrow w - \text{Sign}(z_i) x_i$



ReLU:



step function