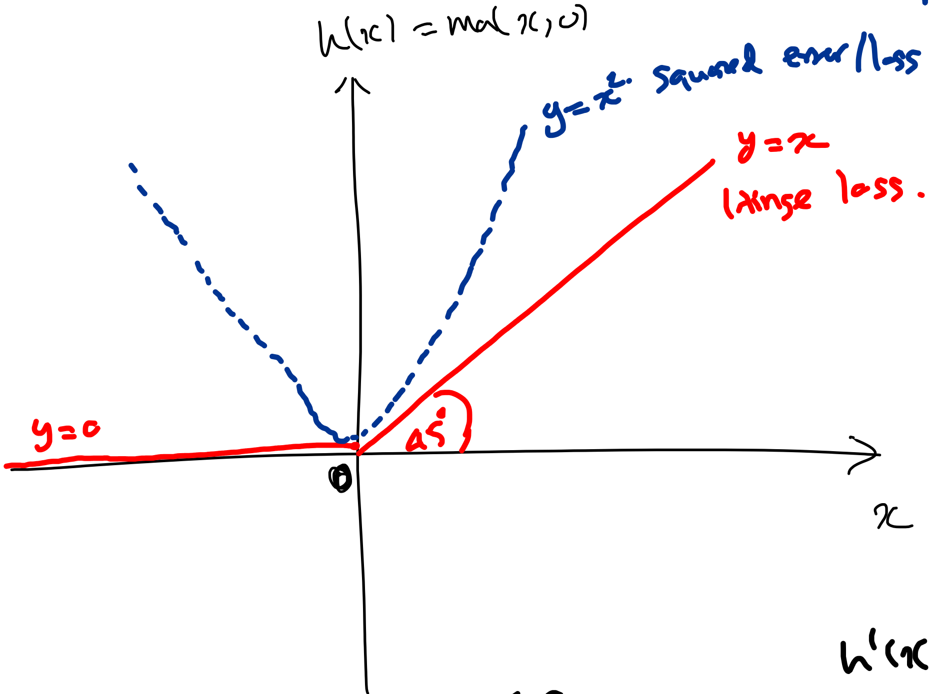
$\{(x_i, t_i)\}_{i=1}^{N} = \{(x_1, t_1), (x_2, t_2), \ldots, (x_N, t_N)\}.$

$h(x) = x^2$

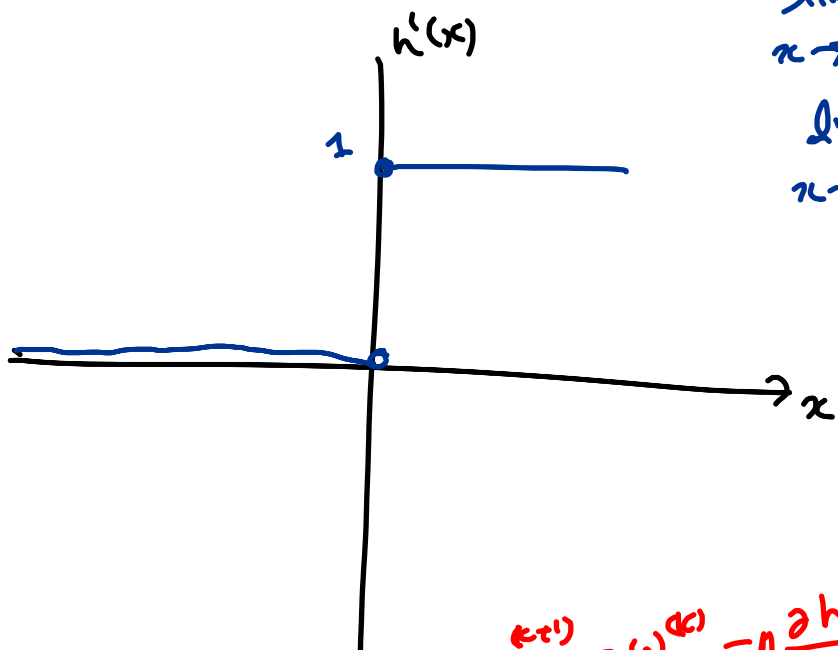


$$h(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$$

$$h'(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

$$\lim_{x \to 0^-} h'(x) = 0$$

$$\lim_{x \to 0^+} h'(x) = 1$$



$h(-t\, w^T x)$

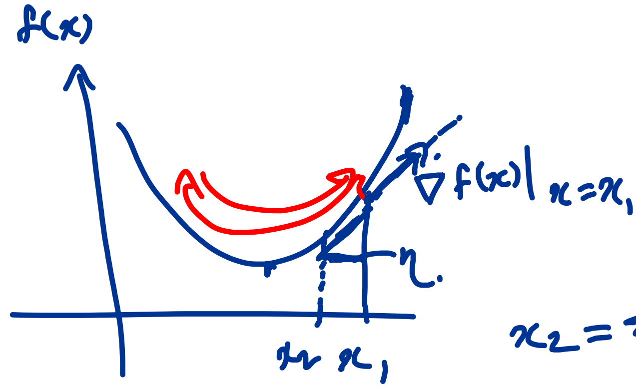
$$w^{(k+1)} = w^{(k)} - \eta \frac{\partial h(-t w^T x)}{\partial w}$$

$$= w^{(k)} - \eta\, h'(-t w^T x)(-t x)$$

$$= w^{(k)} + \eta\, h'(-t w^T x)\, t x.$$

$$= w^{(k)} + \eta\, t x.$$

SGD. SGA



ADAM
AdaDelta
AdaGrad.

$$x_2 = x_1 - \nabla f(x)\,\eta$$

$$\eta = \frac{\eta_0}{t^2} \quad 10^{-3}$$

$$w = (w_1, w_2, w_3)$$

$$x^T w$$

$$\|w\|_2^2 = w_1^2 + w_2^2 + w_3^2$$

l_2.

$$L(w, x) = h(-t x^T w) + \lambda \|w\|_2^2$$

$$\frac{\partial L}{\partial w} = h'(-t x^T w)\cdot(-t x) + 2\lambda w.$$

$$w^{(t+1)} = w^{(t)} - \eta \frac{\partial L}{\partial w}$$

$$= w^k - \eta\left[h'(-t x^T w)(-t x) + 2\lambda w^{(k)}\right]$$

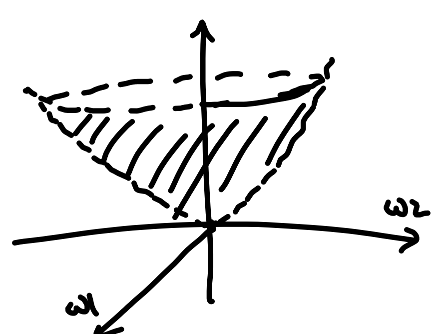
$$= w^k - \eta\left(-t x + 2\lambda w^{(t)}\right)$$

$$= w^k(1 - 2\eta\lambda) + \eta t x.$$

$$= w^{(t)}(1 - 2\lambda) + t x.$$

$$\eta = 1$$

$$w_2 = \sqrt{w_1^2 + w_2^2}$$



l_1 $\quad \|w\|_1 = |w_1| + |w_2| + \cdots + |w_d|$

l_1 $w = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

Sparse model

l_2 $w = \begin{pmatrix} 0.001 \\ 1.00 \\ 0.002 \end{pmatrix}$