Lasso regul°

- $\mathcal{L} = \frac{1}{N} \sum_{n=1}^{N} (w x - t_n)^2 + \lambda \|w\|_1$

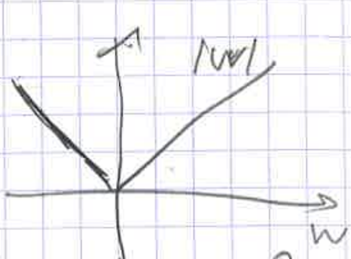
$$\nabla \mathcal{L} = \frac{2}{N} X^T (X^T w - T) + \lambda \vec{\nabla}_{\vec{w}} \sum_{d=1}^{D} |w_d|$$

$$= \text{\_\_\_\_\_\_\_\_} + \lambda \sum_{d=1}^{D} \text{sign}(w_d) \; e\_d$$

Where e_d is the canonical base vector associated to dimension d: i.e. (0,0,0,0,...,0,1,0,...0) (1 at the d-th position)

Here we defined a sub-gradient:

$$\nabla_w |w| = \begin{cases} +1 & \text{if } w_d > 0 \\ 0 & \text{if } w_d = 0 \\ -1 & \text{if } w_d < 0 \end{cases}$$



We may define $\vec{\text{sign}}(\vec{w})$, but it won't factor with $\vec{w}$.

$\Longrightarrow$ There is no simple mathematical form (to frame it as an algebra problem)

- The GD update step is:

$$\vec{w} \to \vec{w} - \eta \vec{\nabla}_w \mathcal{L} = \vec{w} - \eta \lambda \sum_{d=1}^{D} \text{sign}(w_d) \; e\_d$$

$$+ (\text{usual MSE term})$$

$$= \vec{w} - \eta \lambda \, \vec{\text{sign}}(\vec{w}) \, e\_d + \vec{\nabla}(MSE)$$

For a given dim°, say, where $w_d > 0$, we have

$$w_d \mapsto \underbrace{w_d - \eta \lambda (+1)}_{\text{shrinks } w_d \text{, independently of its magnitude}} + \vec{\nabla}(MSE)$$

$\longrightarrow$ go code it!