Lasso regul°

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

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

$$= \text{————} + \lambda \sum_{d=1}^{D} \text{sign}(w_d)$$

Here we defined a <u>sub-gradient</u>:

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

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

$\Rightarrow$ <u>There is no simple</u> mathematical form (to frame it as an algebra problem)

- The GD update step is:

$$\vec{w} \to \vec{w} - \eta \vec{\nabla}_w \mathcal{L} = \vec{w} - \eta\lambda \sum_{d=1}^{D} \text{sign}(w_d) + (\text{usual MSE term})$$

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

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

$$w_d \mapsto w_d - \eta\lambda(+1) + \vec{\nabla}(MSE)$$

shrinks w_d, independently of its magnitude

$\to$ go code it!