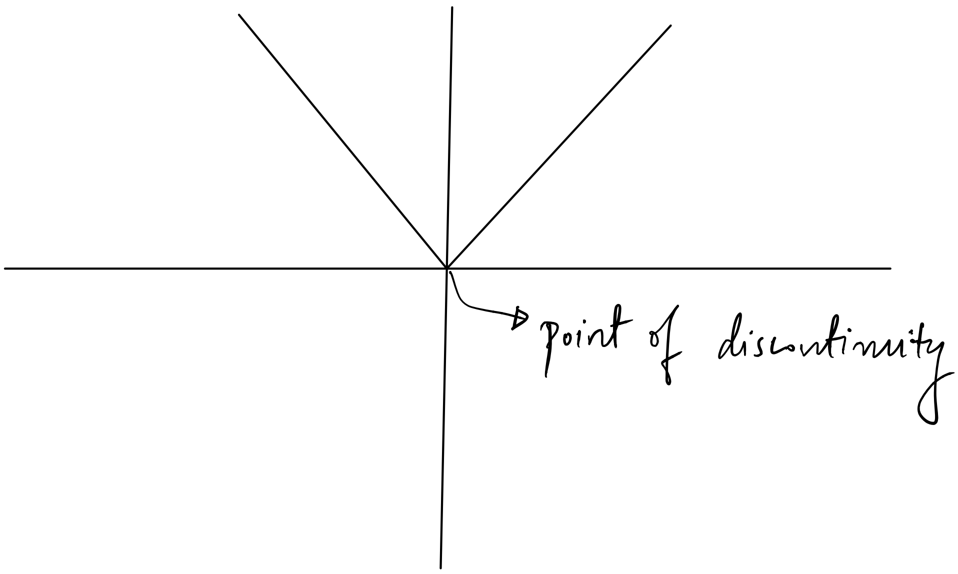
③

Answer : $L(w) = \frac{1}{2} \| x - w \|_2^2 + \lambda \| w \|_1$

This function is continous on both positive and negative sides except at zero. This is shown below.



Because of this we cannot differentiate $L(w)$ directly. We need to split $L(w)$ into ranges (+ve & -ve)

$$\| w \|_1 = \begin{cases} +w & \text{if } w > 0 \\ -w & \text{if } w < 0 \\ 0 & \text{otherwise} \end{cases}$$

$$L(w) = \frac{1}{2}\|x-w\|_2^2 + \lambda\|w\|_1$$

Differentiate $L(w)$ in the positive range of w

$$L(w) = \frac{1}{2}\|x-w\|_2^2 + \lambda \sum w_i$$

where all w_i are positive

This $L(w)$ can be differentiated.

$$\frac{dL(w)}{dw} = \frac{1}{2} \cdot 2 \cdot (x-w)(-1) + \lambda$$

$$0 = -(x-w) + \lambda$$

$$x - w = \lambda$$

$$\boxed{w = x - \lambda}$$

Here, $w_i > 0 \quad \therefore (x-\lambda) > 0 \rightarrow \boxed{x > \lambda}$

|||ly Differentiate $L(w)$ in the negative range of w

$$L(w) = \frac{1}{2}\|x-w\|_2^2 - \lambda \sum w_i$$

$$\frac{dL(w)}{dw} = (-1)(x-w) - \lambda$$

$$(w-x) - \lambda = 0$$

$$w - x = \lambda$$

$$\boxed{w = \lambda + x}$$

Here, $w < 0$

$$\lambda + x < 0 \rightarrow \lambda < -x \rightarrow \boxed{x < -\lambda}$$

So, the optimal values of w are

$$w_i^* = \begin{cases} x - \lambda & x > \lambda \\ x + \lambda & x < -\lambda \\ 0 & \text{otherwise} \end{cases}$$

(1) Answer :

a) All parameters w are free to be determined

$R(w)$ = Elastic Net

This allows all parameters w to be determined. L_1 and L_2 can be controlled using λ_1 and λ_2. Therefore all values of lasso or ridge are possible.

b) w should be sparse

$R(w)$ = L_1 Norm

The reason for using L_1 Norm is its shape. It has spikes that occur at sparse points. If this used to get a solution it is very likely to touch a point on the spike tip. Thereby giving a sparse solution

c) The coefficients of w should be small in magnitude on average

$R(w) =$ L2 Norm. L2 Norm minimizes all the coefficients without making them zero and minimization of the coefficients will be lesser than that of L1.

d) For most indices j, w_j should be equal to w_{j-1}

$$R(w) = \sum_{i=1}^{d-1} (w_i - w_{i+1})^2$$

e) w should have no negative-valued coefficients

$R(w) =$ L1 Norm