2.2) 1

$$f(w_j) = \sum_{i=1}^{n} (w^T x_i - y_i)^2 + \lambda |w|$$

We will consider the case when $w_j > 0$

Differentiate

$$\frac{\partial f(w_j)}{\partial w_j} = \sum_{i=1}^{n} 2(w^T x_i - y_i)\, x_{ij} + \lambda .$$

$$= 2\sum_{i=1}^{n} \left[\sum_{k \neq j} 2\left(w_k^T x_{ik} - y_i + w_j x_{ij}\right) x_{ij} \right] + \lambda$$

$$= 2\sum_{i=1}^{n} \left[w_j x_{ij}^2 + \sum_{k \neq j} (w_k^T x_{ik} - y_i)\, x_{ij} \right] + \lambda$$

Substituting for a and we have

$$\frac{\partial f(w_j)}{\partial w_j} = a_j w_j - c_k + \lambda$$

For $w < 0$, $\frac{\partial |w|}{\partial w_j} = -1$.

$$\Rightarrow \cancel{\frac{\partial f(w_j)}{\partial w_j} = \begin{cases} a_k w_k - c_k - \lambda & w_k < 0 \\ a_k w_k - c_k + \lambda & w_k > 0 \end{cases}}$$

$$\frac{\partial f(w_j)}{\partial w_j} = \begin{cases} a_j w_j - c_j - \lambda & w_j < 0 \\ a_j w_j - c_j + \lambda & w_j > 0 \end{cases}$$

(2) f minimizes when $\frac{\partial f}{\partial w_j} = 0$.

For $w_j > 0$

~~$a_k w_k - c_k + \lambda = 0$~~

$a_j w_j - c_j + \lambda = 0$

$$\boxed{w_j = -\frac{1}{a_j}(\lambda - c_j)}$$

Here $w_j > 0$

$\Rightarrow \lambda - c_j < 0$

$$\boxed{c_j > \lambda}$$

For $w_j < 0$

$a_j w_j - c_j - \lambda = 0$

$$\boxed{w_j = \frac{1}{a_j}(\lambda + c)}$$

Here $w_j < 0$

$\Rightarrow \lambda + c < 0$

$$\boxed{c_j < -\lambda}$$

(3) $$f(w_j) = \sum_{i=1}^{n} \left[w_j x_{ij} + \sum_{k \neq j} w_k x_{ik} - y_i \right]^2 + \lambda |w_j| + \lambda \sum_{k \neq j} |w_k$$

$$f(0^+) = \lim_{h \to 0} \frac{f(0+h) - f(0)}{h}$$

$$f(0^+) = \frac{\sum_{i=1}^{n} \left[h x_{ij} + \sum_{k \neq j} w_k x_{ik} - y_i \right]^2 + \lambda h + \lambda \sum_{k \neq j} \cancel{|w_k|} - \sum_{i=1}^{n} \left[\sum_{k \neq j} w_k x_{ik} - y_i \right]^2 - \lambda \sum_{k \neq j} |w_k}{h}$$

$$f(0^+) = \frac{2\sum_{i=1} hx_{ij} \sum_{k\neq j} w_k x_{ik} - y_i + \lambda h}{h}$$

If it is minimizer at $w_j = 0$. $f(0^+) \geq 0$

$\cancel{f(0^+)}$ $c_j + \lambda \geq 0$

$$\boxed{c_j \geq -\lambda}$$

Similarly for $f(0^-)$, we get

$$f(0^-) = \frac{2\sum_{i=1}^{n} hx_{ij} \sum_{k\neq j} w_k x_{ik} - y_i - \lambda h}{h}$$

Minimizer at $w_j = 0 \Rightarrow f(0^-) > 0$

$$f(0^-) = c_j - \lambda \geq 0$$

$$\boxed{c_j \geq 0}$$

$\Rightarrow$ $c_j \in [-\lambda, \lambda]$ Minimizer $\Rightarrow w_j = 0 \Rightarrow c_j \in [-\lambda, \lambda]$

4) From equations of problem ① and ③, we have

$$w_j = \begin{cases} \frac{1}{a_j}(c_j - \lambda) & c_j > \lambda \\ 0 & c_j \in [-\lambda, \lambda] \\ \frac{1}{a_j}(c_j + \lambda) & c_j < -\lambda \end{cases}$$

Expression in ② $a_j = 2 \sum_{i=1}^{n} x_{ij}^2$

$$c_j = 2 \sum_{i=1}^{n} x_{ij} \left(y_i - w^T x_i + w_j x_{ij} \right)$$

a_j and c_j are same as what we had defined.

$$w_j = \text{Soft}\left(\frac{c_j}{a_j}, \frac{\lambda}{a_j} \right)$$

For $c_j > \lambda$

$$w_j = \text{sign}\left(\frac{c_j}{a_j} \right) \left(\left| \frac{c_j}{a_j} \right| - \frac{\lambda}{a_j} \right)$$

$$w_j = \frac{c_j - \lambda}{a_j}$$

For $c_j < -\lambda$

$$w_j = \frac{c_j + \lambda}{a_j}$$

For all other cases $w_j = 0$.

3.1) $L(w)$ can be written as

$$L(w) = \sum_{i=1}^{n} (w^T x_i - y_i)^2 + \lambda |w|$$

Now $w=0$, the function is not differentiable, we take one sided derivative

$$f'(x,v) = \lim_{h \downarrow 0} \frac{f(x+hv) - f(x)}{h}$$

Now

$$L'(0,v) = \lim_{h \downarrow 0} \frac{\sum_{i=1}^{n} ((hv)^T x_i - y_i)^2 + \lambda |hv| - \sum_{i=1}^{n} y_i^2}{h}$$

$$= \lim_{h \downarrow 0} \frac{\sum -2 (hv)^T x_i \cdot y_i + \cancel{y_i^2} + ((hv)^T x_i)^2 + \lambda |hv| - \cancel{\sum_{i=1}^{n} y_i^2}}{h}$$

This ~~term~~ should be greater than zero for all values of v since lasso objective is convex.

$$\Rightarrow \quad -2 \sum_{i=1}^{n} v^T x_i y_i + \lambda |v| \geq 0$$

$$\therefore \quad \lambda |v| \geq 2 \sum_{i=1}^{n} v^T x_i y_i$$

$$\lambda |v| \geq 2 v^T X^T y$$

~~For~~ Now we select that value of v (direction) which maximizes ~~$X^T y$~~ right side $\Rightarrow \lambda \geq 2 \|X^T y\|_\infty$

4.1) $\hat{R}_n(w) = \frac{1}{n}\sum_{i=1}^{n}(w^T x_i - y_i)^2$

$= \frac{1}{n}(Xw-y)^T(Xw-y)$

$= \frac{1}{n}\left[(X(X^TX)^{-1}X^Ty - y)^T(X(X^TX)^{-1}X^Ty - y)\right]$

$= \frac{1}{n}\left[(y^TX(X^TX)^{-1}X^T - y^T)(X(X^TX)^{-1}X^Ty - y)\right]$

$= \frac{1}{n}\left[y^TX(X^TX)^{-1}X^Ty - y^TX(X^TX)^{-1}X^Ty - y^TX\hat{w} + y^T\right.$

$= \frac{1}{n}\left[-y^TX\hat{w} + y^Ty\right]$

4.2) $\hat{R}_n(w) = \frac{1}{n}(Xw-y)^T(Xw-y)$

$= \frac{1}{n}\left[(w^TX^T - y^T)(Xw-y)\right]$

$= \frac{1}{n}\left[w^TX^TXw - w^TX^Ty - y^TXw - y^Ty\right]$ — 1

Given equation

$\hat{R}_n(w) = \frac{1}{n}(w-\hat{w})^TX^TX(w-\hat{w}) + \hat{R}_n(\hat{w})$

$= \frac{1}{n}\left[\left(w^T - ((X^TX)^{-1}X^Ty)^T\right)\left(X^TXw - X^TX(X^TX)^{-1}X^Ty\right)\right] + \hat{R}_n(\hat{w}$

$$= \frac{1}{n}\left[\left(w^T - y^T X (X^T X)^{-1}\right)\left(X^T X w - X^T y\right)\right] + \frac{1}{n}\left[-y^T X (X^T X)^{-1} X^T y + y^T y\right]$$

$$= \frac{1}{n}\left[w^T X^T X w - w^T X^T y - y^T X w + y^T X (X^T X)^{-1} X^T y - y^T X (X^T X)^{-1} X^T y + y^T y\right]$$

$$= \frac{1}{n}\left[w^T X^T X w - w^T X^T y - y^T X w - y^T y\right] \quad — (2)$$

Equation (1) and (2) are equal.

$$\Rightarrow \hat{R}_n(w) = \frac{1}{n}(w - \hat{w})^T X^T X (w - \hat{w}) + \hat{R}_n(\hat{w}) \quad — (3)$$

4.3) Note that from eq (3),

the first part of the term is either 0 or posi[tive] because $X^T X$ is positive semi definite. This means the lowest value $\hat{R}_n(w)$ can attain is $\hat{R}_n(\hat{w})$ and this is possible when first first term = 0.

$$\frac{1}{n}(w - \hat{w})^T X^T X (w - \hat{w}) = 0.$$

$$\Rightarrow (w - \hat{w}) = 0 \qquad w = \hat{w} \Rightarrow \hat{w} \text{ is empirical risk minimizer}$$

4.4) From (3)

$$\hat{R}_n(w) = \frac{1}{n}(w-\hat{w})^T x^T x (w-\hat{w}) + \hat{R}_n(\hat{w})$$

when $\hat{R}_n(w) \geq \hat{R}_n(\hat{w}) + c$.

$$\Rightarrow \quad \boxed{c \leq \frac{1}{n}(w-\hat{w})^T x^T x (w-\hat{w})} \quad - (1)$$

All w which satisfies this equation.

$\hat{R}_n(w)$ can be written as

$$\hat{R}_n(w) = \frac{1}{n}(x(w-\hat{w}))^T (x(w-\hat{w})) + \hat{R}_n(\hat{w})$$

At centre $c = 0$.

$$(1) \rightarrow \quad 0 = (w-\hat{w})^T x^T x (w-\hat{w})$$

$$(x(w-\hat{w}))^T (x(w-\hat{w})) = 0.$$

$$\Rightarrow \quad w - \hat{w} = 0$$

$$\boxed{w = \hat{w}}$$