Q: What is $\hat{\beta}^{Lasso}$ for $Y = X\beta + \varepsilon$ where X is ortonormal design matrix

see also example 6.4 in WNvW

Orthonormal design matrix:

$$\underset{N \times p}{X}$$

$$\underset{p \times N}{X^T}\underset{N \times p}{X} = \underset{p \times p}{I}$$

corr = 0 all pairs of x's

$$\min_{\beta} \left((Y - X\beta)^T (Y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

$$= \min_{\beta} \left(Y^TY - 2\beta^T \underbrace{X^TY}_{\hat{\beta}_{LS}} + \beta^T \underbrace{X^TX}_{I} \beta + \lambda \sum |\beta_j| \right)$$

$$\hat{\beta}_{LS} = (\underbrace{X^TX}_{I})^{-1} X^TY = X^TY$$

$$\propto \min_{\beta} \left(-2\beta^T \hat{\beta}_{LS} + \beta^T\beta + \lambda \sum |\beta_j| \right)$$

$$\beta_1 \cdot \hat{\beta}_{LS1} + \beta_2 \hat{\beta}_{LS2} + \dots \qquad \beta_1^2 + \beta_2^2 + \dots$$

$$= \min_{\beta} \left(-2 \sum_{j=1}^{p} \beta_j \cdot \hat{\beta}_{LS_j} + \sum_{j=1}^{p} \beta_j^2 + \lambda \cdot \sum_{j=1}^{p} |\beta_j| \right)$$

$$= \min_{\beta_1, \dots, \beta_p} \sum_{j=1}^{p} \left(-2\beta_j \hat{\beta}_{LSj} + \beta_j^2 + \lambda |\beta_j| \right)$$

$$= \sum_{j=1}^{p} \min_{\beta_j} \left(-2\beta_j \hat{\beta}_{LSj} + \beta_j^2 + \lambda |\beta_j| \right)$$

can handle each j separately!

Let : $\text{loss}(\beta_j) = -2\hat{\beta}_{LSj}\beta_j + \beta_j^2 + \lambda|\beta_j|$ ← minimize wrt β_j

Dropping j's and LS for ease of notation.

We want to find $\underline{\beta}$ such that $-2\hat{\beta}\beta + \beta^2 + \lambda|\beta|$ is minimal

$\underline{\beta \geq 0}$

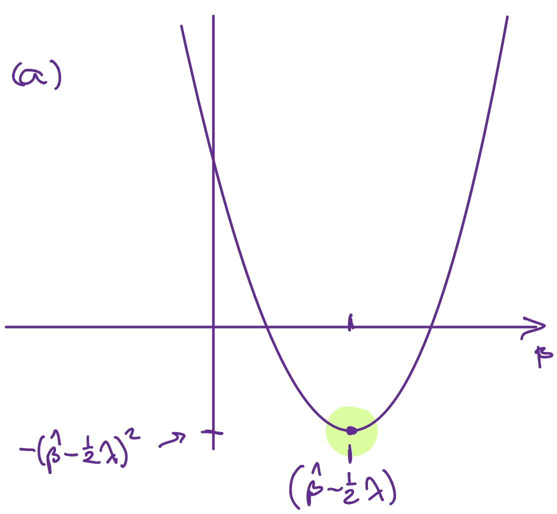
$$-2\hat{\beta}\beta + \beta^2 + \lambda\beta = \beta^2 - 2(\hat{\beta} - \tfrac{1}{2}\lambda)\beta + (\hat{\beta} - \tfrac{1}{2}\lambda)^2 - (\hat{\beta} - \tfrac{1}{2}\lambda)^2$$

completing sq.

$$= (\beta - (\hat{\beta} - \tfrac{1}{2}\lambda))^2 - (\hat{\beta} - \tfrac{1}{2}\lambda)^2 \qquad (\beta - a)^2 - a^2$$

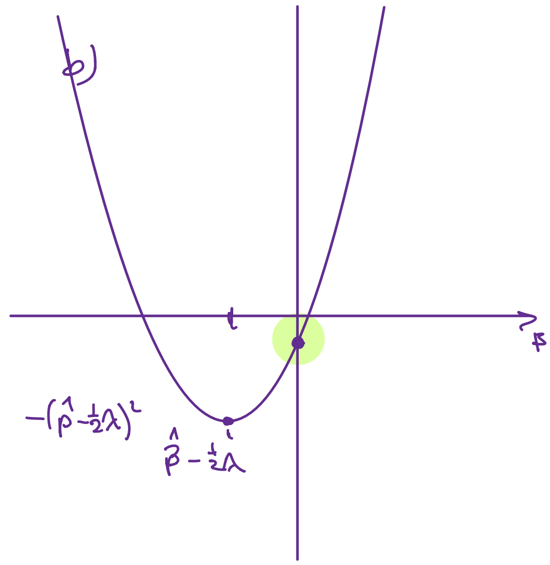
Look at $(\hat{\beta} - \tfrac{1}{2}\lambda) \geq 0$ (a) $\qquad (\hat{\beta} - \tfrac{1}{2}\lambda) < 0$ (b)

(a)



Minimum is for $\beta = \hat{\beta} - \tfrac{1}{2}\lambda$

b)



We required that $\beta \geq 0$, thus means the minimum is at $\beta = 0$

In conclusion: for $\beta \geq 0$

$\beta = \max(\hat{\beta} - \tfrac{1}{2}\lambda, 0)$ gives the minimum

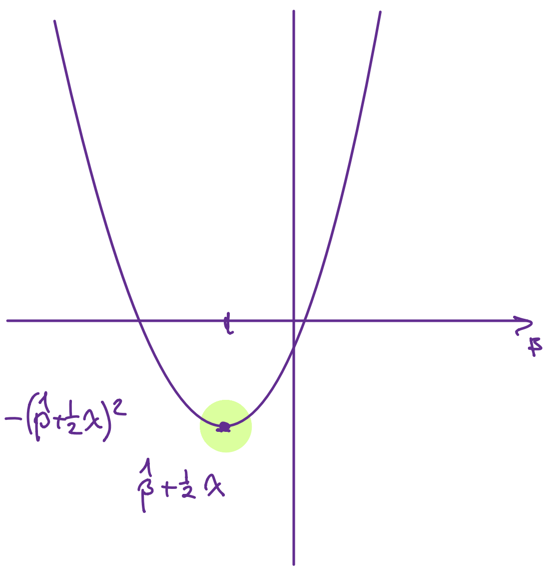
$\underline{\beta < 0}$

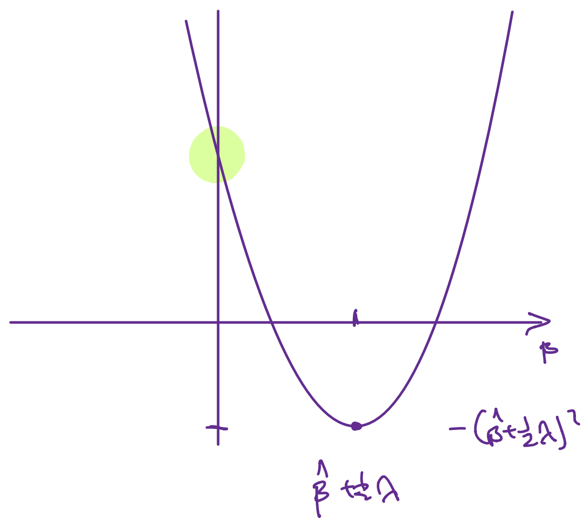
$$-2\hat{\beta}\beta + \beta^2 - \lambda\beta = \beta^2 - 2(\hat{\beta} + \tfrac{1}{2}\lambda)\beta + (\hat{\beta} + \tfrac{1}{2}\lambda)^2 - (\hat{\beta} + \tfrac{1}{2}\lambda)^2$$

$$= \left(\beta - (\hat{\beta} + \tfrac{1}{2}\lambda)\right)^2 - (\hat{\beta} + \tfrac{1}{2}\lambda)^2 \qquad (\beta - a)^2 - a^2, \quad a = \hat{\beta} + \tfrac{1}{2}\lambda$$

Same type of argument as above: Among $\beta \geq 0$ we find the minimum at $\beta = \min(\hat{\beta} + \frac{1}{2}\lambda, 0)$

(a) $\hat{\beta} + \frac{1}{2}\lambda \leq 0$ b) $\hat{\beta} + \frac{1}{2}\lambda > 0$





here $\beta < 0$

minimum at $\beta = (\hat{\beta} + \frac{1}{2}\lambda)$ minimum at $\beta = 0$

$$\beta = \min(\hat{\beta} + \tfrac{1}{2}\lambda, 0)$$

We need to combine the two results:

$$\beta = \max(\hat{\beta} - \tfrac{1}{2}\lambda, 0) \qquad \text{for } \beta \geq 0$$
$$\beta = \min(\hat{\beta} + \tfrac{1}{2}\lambda, 0) \qquad \text{for } \beta < 0$$
(could be $\beta \leq 0$ really)

And prefer to do that conditional on values of $\hat{\beta}$ and not of β (as so far)

$\hat{\beta} < -\frac{1}{2}\lambda$: $\quad \beta = \hat{\beta} + \frac{1}{2}$

i.e negative

$-\frac{1}{2}\lambda \leq \hat{\beta} < \frac{1}{2}\lambda$: $\quad \beta = 0$

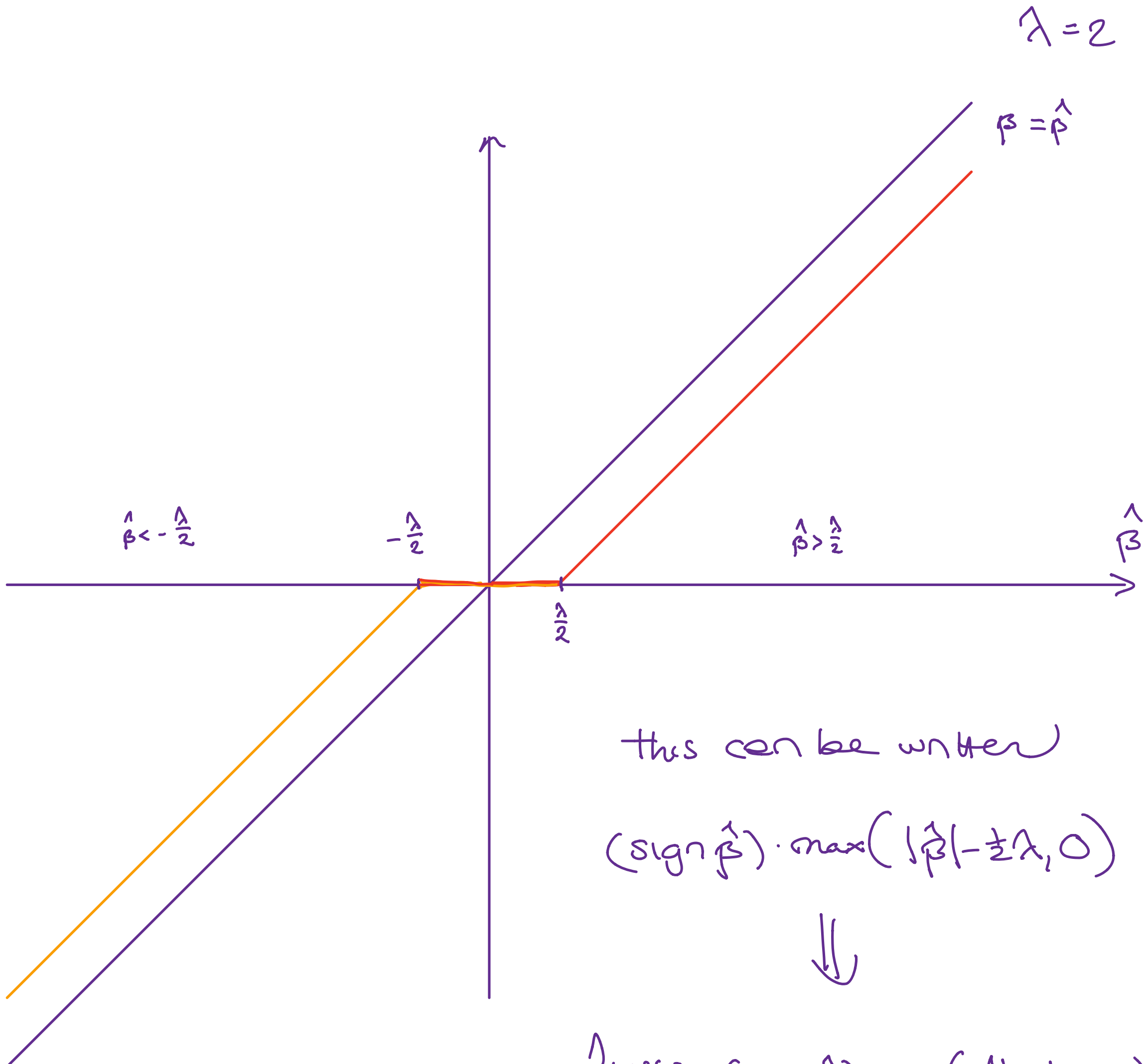
$\hat{\beta} \geq \frac{1}{2}\lambda$: $\quad \beta = \hat{\beta} - \frac{1}{2}\lambda$

i.e. positive

$\Rightarrow$ this is combined into

$$\operatorname{sgn}(\hat{\beta}) \cdot \max(|\hat{\beta}| - \tfrac{1}{2}\lambda, 0)$$

$$= \operatorname{sgn}(\hat{\beta})\left(|\hat{\beta}| - \tfrac{1}{2}\lambda\right)_+$$



this can be written

$$(\text{sign}\,\hat\beta)\cdot\max\left(|\hat\beta|-\tfrac{1}{2}\lambda, 0\right)$$

$$\Downarrow$$

$$\hat\beta^{LASSO} = (\text{sign}\,\hat\beta)\max\left(|\hat\beta|-\tfrac{1}{2}\lambda, 0\right)$$

$$= (\text{sign}\,\hat\beta)\left(|\hat\beta|-\tfrac{1}{2}\lambda\right)_+$$

NB: $\hat\beta = \hat\beta_{LS_j}$ and $\hat\beta^{LASSO} = \hat\beta_j^{LASSO}$