Prop 21

i) Lasso solutions exist

ii) $X\hat{\beta}^L_\lambda$ is unique

Proof

i) Provided $\lambda > 0$

$$\inf_{\beta : \lambda\|\beta\|_1 > \frac{1}{2n}\|Y\|_2^2} Q_\lambda(\beta) \geq \frac{1}{2n}\|Y\|_2^2 = Q_\lambda(0)$$

$$\geq \inf_{\beta : \lambda\|\beta\|_1 \leq \frac{1}{2n}\|Y\|_2^2} Q_\lambda(\beta) \qquad (*)$$

$$= \inf_\beta Q_\lambda(\beta)$$

But at $(*)$ we are minimising the cts f^n Q_λ over a closed bounded set, so a minimiser must exist.

ii) Fix $\lambda \geq 0$ and suppose $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(2)}$ are two Lasso solutions with

$$Q_\lambda(\hat{\beta}^{(1)}) = Q_\lambda(\hat{\beta}^{(2)}) = c^* .$$

By strict convexity $\|\cdot\|_2^2$,

$$\|\tfrac{1}{2}(Y - X\hat{\beta}^{(1)}) + \tfrac{1}{2}(Y - X\hat{\beta}^{(2)})\|_2^2 \leq \tfrac{1}{2}\|Y - X\hat{\beta}^{(1)}\|_2^2 + \tfrac{1}{2}\|Y - X\hat{\beta}^{(2)}\|_2^2$$

with equality iff $X\hat{\beta}^{(1)} = X\hat{\beta}^{(2)}$.

$$c^* \leq Q_\lambda(\tfrac{1}{2}\hat{\beta}^{(1)} + \tfrac{1}{2}\hat{\beta}^{(2)})$$

$$\leq \frac{1}{2n}\left(\tfrac{1}{2}\|Y - X\hat{\beta}^{(1)}\|_2^2 + \tfrac{1}{2}\|Y - X\hat{\beta}^{(2)}\|_2^2\right) + \lambda\|\tfrac{1}{2}\hat{\beta}^{(1)} + \tfrac{1}{2}\hat{\beta}^{(2)}\|_1 \qquad (*)$$

$$\leq \tfrac{1}{2}Q_\lambda(\hat{\beta}^{(1)}) + \tfrac{1}{2}Q_\lambda(\hat{\beta}^{(2)}) = c^*$$

Therefore, we must have equality at $(*)$, so $X\hat{\beta}^{(1)} = X\hat{\beta}^{(2)}$. □

Define the equicorrelation set $\hat{E}_\lambda$ to be the set of variables k s.t.

$$\frac{1}{n}|X_k^T(Y - X\hat{\beta}^L_\lambda)| = \lambda .$$

This is well-defined as it only depends on the (unique) fitted vals. By the KKT conditions, $\hat{E}_\lambda$ contains the set of non-zeros of all Lasso solutions (at λ).

If $\text{rank}(X_{\hat{E}_\lambda}) = |\hat{E}_\lambda|$, then the Lasso soln is unique:

$$X_{\hat{E}_\lambda}(\hat\beta^{(1)}_{\hat{E}_\lambda} - \hat\beta^{(2)}_{\hat{E}_\lambda}) \Rightarrow \hat\beta^{(1)}_{\hat{E}_\lambda} = \hat\beta^{(2)}_{\hat{E}_\lambda} \Rightarrow \hat\beta^{(1)} = \hat\beta^{(2)}.$$

2.2.5 Variable selection

Noiseless linear model $Y = X\beta^0$

$S = \{k : \beta^0_k \neq 0\} = \{1, \dots, s\}$

$N = \{1, \dots, p\} \setminus S$

Assume $\text{rank}(X_S) = s$.

Thm 14

Let $\lambda > 0$ and $\Delta = X_N^T X_S (X_S^T X_S)^{-1} \text{sgn}(\beta^0_S)$.

i) If $\|\Delta\|_\infty \leq 1$ $\quad \left(\max_{k \in N} |\text{sgn}(\beta^0_S)^T (X_S^T X_S)^{-1} X_S^T X_k| \leq 1 \right)$

$\cdot\ |\beta^0_k| > \lambda |\text{sgn}(\beta^0_S)^T \{(\frac{1}{n} X_S^T X_S)^{-1}\}_k| \quad (k \in S) \qquad (*)$

then $\exists$ a Lasso soln $\hat\beta^L_\lambda$ with $\text{sgn}(\hat\beta^L_\lambda) = \text{sgn}(\beta^0)$

ii) If $\exists$ a Lasso soln $\hat\beta^L_\lambda$ with $\text{sgn}(\hat\beta^L_\lambda) = \text{sgn}(\beta^0)$

then $\|\Delta\|_\infty \leq 1$.

<u>Proof</u>: Write $\hat\beta = \hat\beta_\lambda$, $\hat{S} = \{k : \hat\beta_k \neq 0\}$. KKT conditions:

$$\frac{1}{n} X^T X(\beta^0 - \hat\beta) = \lambda \hat\nu . \quad \|\hat\nu\|_\infty \leq 1 \text{ and } \hat\nu_{\hat S} = \text{sgn}(\hat\beta_{\hat S})$$

Expand this

$$\frac{1}{n}\begin{pmatrix} X_S^T X_S & X_S^T X_N \\ X_N^T X_S & X_N^T X_N \end{pmatrix} \begin{pmatrix} \beta^0_S - \hat\beta_S \\ -\hat\beta_N \end{pmatrix} = \lambda \begin{pmatrix} \hat\nu_S \\ \hat\nu_N \end{pmatrix} \qquad \begin{matrix} (1) \\ (2) \end{matrix}$$

Prove (ii) first

From (1) $\beta^0_S - \hat\beta_S = \lambda (\frac{1}{n} X_S^T X_S)^{-1} \text{sgn}(\beta^0_S)$

Substituting into (2)

$$\lambda \underbrace{\frac{1}{n} X_N^T X_S (\frac{1}{n} X_S^T X_S)^{-1} \text{sgn}(\beta^0_S)}_{\Delta} = \lambda \hat\nu_N$$

Then $\|\Delta\|_\infty \leq 1$ as $\|\hat\nu_N\|_\infty \leq 1$

Now (i). Try $(\hat\beta_S, \hat\beta_N) = (\beta^0_S - \lambda(\frac{1}{n}X_S^T X_S)^{-1}\,\mathrm{sgn}(\beta^0_S)\ ,\ 0)$

$(\hat\nu_S, \hat\nu_N) = (\mathrm{sgn}(\beta^0_S)\ ,\ \Delta)$

Only need to check $\mathrm{sgn}(\hat\beta_S) = \mathrm{sgn}(\beta^0_S)$, but this follows from $(*)$.

2.2.6 Prediction estimation

Now consider $Y = X\beta^0 + \varepsilon - \mathbf{1}\bar\varepsilon$ where ε_i are independent mean-zero sub-Gaussian with parameter σ.

Let S, s, N be defined as before.

Def^n 6

Define
$$\phi^2 = \inf_{\substack{\beta\in\mathbb{R}^p : \|\beta_N\|_1 \le 3\|\beta_S\|_1 \\ \|\beta_S\|_1 \neq 0}} \frac{\frac{1}{n}\|X\beta\|_2^2}{\frac{1}{s}\|\beta_S\|_1^2}$$

where compatibility factor $\phi \geq 0$. The compatibility condition is that $\phi > 0$.