Modern Statistical Methods — Example Sheet 3

Lucas Riedstra

November 17, 2020

In all of the below, assume that any design matrices X are $n \times p$ and have their columns centred and then scaled to have ℓ^2-norm $\sqrt{n}$, and that any responses $Y \in \mathbb{R}^n$ are centred.

Question 1. *When proving the theorems on the prediction error of the Lasso, we started with the so-called basic inequality that*

$$\frac{1}{2n}\left\|X(\beta^0 - \hat{\beta})\right\|_2^2 \leq \frac{1}{n}\varepsilon^\top X(\hat{\beta} - \beta^0) + \lambda\left\|\beta^0\right\|_1 - \lambda\left\|\hat{\beta}\right\|_1.$$

Show that in fact we can improve this to

$$\frac{1}{n}\left\|X(\beta^0 - \hat{\beta})\right\|_2^2 \leq \frac{1}{n}\varepsilon^\top X(\hat{\beta} - \beta^0) + \lambda\left\|\beta^0\right\|_1 - \lambda\left\|\hat{\beta}\right\|_1.$$

Proof. By the KKT conditions, we there exists $\hat{\nu} \in \mathbb{R}^p$ with $\|\hat{\nu}\|_\infty \leq 1$ and $\langle \hat{\nu}, \hat{\beta} \rangle = \|\beta\|_1$ such that

$$\begin{aligned}
\frac{1}{n}X^\top(Y - X\hat{\beta}) &= \lambda\hat{\nu} \\
\frac{1}{n}X^\top(X(\beta^0 - \hat{\beta}) + \varepsilon) &= \lambda\hat{\nu} \\
\frac{1}{n}X^\top X(\beta^0 - \hat{\beta}) &= -\frac{1}{n}X^\top\varepsilon + \lambda\hat{\nu} \\
\frac{1}{n}(\beta^0 - \hat{\beta})^\top X^\top X(\beta^0 - \hat{\beta}) &= -\frac{1}{n}(\beta^0 - \hat{\beta})^\top X^\top\varepsilon + \lambda(\beta^0 - \hat{\beta})^\top\hat{\nu} \\
\frac{1}{n}\left\|X(\beta^0 - \hat{\beta})\right\|_2^2 &= \frac{1}{n}\varepsilon^\top X(\hat{\beta} - \beta^0) + \lambda\langle\hat{\nu}, \beta^0\rangle - \lambda\left\|\hat{\beta}\right\|_1,
\end{aligned}$$

and now plugging in $\langle \beta^0, \hat{\nu} \rangle \leq \left\|\beta^0\right\|_1 \|\hat{\nu}\|_\infty \leq \left\|\beta^0\right\|_1$ yields the result. □

Question 2. *Under the assumptions of Theorem 23 on the prediction and estimation properties of the Lasso under a compatibility condition, show that, with probability* $1 - 2p^{-(A^2/8-1)}$*, we have*

$$\frac{1}{n}\left\|X(\beta^0 - \hat{\beta})\right\|_2^2 \leq \frac{9A^2\log(p)}{4\varphi^2}\frac{\sigma^2 s}{n}.$$