Data-driven statistical modelling with optimisation VT21
Hand-in exercise 1

Fanny Bergström

January 31 2021

Exercise 2.1

We will show that the smallest value of λ such that the regression coefficients estimated by the lasso are all equal to zero is given by

$$\lambda_{\max} = \max_j |\frac{1}{N}\langle \mathbf{x}_j, \mathbf{y}\rangle|.$$

Writing the Lasso optimization problem in the Lagrangian form, we get

$$\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \left\{ \frac{1}{2N} ||\mathbf{y} - \mathbf{X}\beta||_2^2 + \lambda ||\beta||_1 \right\}.$$

As this is a convex optimization problem the solution is given by

$$-\frac{1}{N}\langle \mathbf{x}_j, \mathbf{y} - \mathbf{X}\beta\rangle + \lambda s_j = 0 \qquad j = 1, ..., p, \tag{1}$$

where s_j is the subgradient for the absolute value function, which takes some value $[-1, 1]$ when $\beta_j = 0$ and else $\text{sign}(\beta_j)$. If $\beta_j = 0\ \forall j$, we can rewrite Eq. (1) as

$$\lambda s_j = \frac{1}{N}\langle \mathbf{x}_j, \mathbf{y}\rangle.$$

Now, taking the infinity norm of both sides we get

$$||\lambda s_j||_\infty = ||\frac{1}{N}\langle \mathbf{x}_j, \mathbf{y}\rangle||_\infty,$$

which is equal to

$$\lambda_{\max} = \max_j |\frac{1}{N}\langle \mathbf{x}_j, \mathbf{y}\rangle|,$$

and is what we wanted to show in this exercise.

Exercise 2.2

We will show that the soft-threshold estimator

$$S_\lambda(x) = \text{sign}(x)(|x| - \lambda)_+ \tag{2}$$

yields the solution to the single predictor lasso problem

$$\underset{\beta \in \mathbb{R}}{\text{minimize}} \left\{ \frac{1}{2N} ||\mathbf{y} - \mathbf{z}\beta||_2^2 + \lambda|\beta| \right\}. \tag{3}$$

We note that $\mathbf{z}$ is standardized, hence $\sum_i z_i^2 = N$. We can rewrite the expression in Eq. (3) as

$$\begin{aligned}
\frac{1}{2N} ||\mathbf{y} - \mathbf{z}\beta||_2^2 =& \frac{1}{2N}(||\mathbf{y}||_2^2 - 2\beta\langle \mathbf{z}, \mathbf{y}\rangle + \beta^2\langle \mathbf{z}, \mathbf{z}\rangle) + \lambda|\beta| \\
=& \frac{1}{2N}(||\mathbf{y}||_2^2 - 2\beta\langle \mathbf{z}, \mathbf{y}\rangle + \beta^2 N) + \lambda|\beta|.
\end{aligned}$$

Since this is a convex function, the solution of Eq. (3) is found by setting the derivative with respect to β of the expression above to 0. This gives us

$$0 = -\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle + \beta + \lambda \ \text{sign}(\beta).$$

If $\beta > 0$, then $\hat{\beta} = \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle - \lambda$; if $\beta < 0$, then $\hat{\beta} = \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle + \lambda$; and otherwise $\beta = 0$. We can summarize this as

$$\hat{\beta}(\lambda) = \begin{cases} \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle - \lambda & \text{if } \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle > 0 \text{ and } \lambda < \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle = |\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle| \\ \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle + \lambda & \text{if } \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle < 0 \text{ and } \lambda < -\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle = |\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle| \\ 0 & \text{if } \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle = 0 \text{ or } \lambda \geq \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle. \end{cases} \tag{4}$$

Since $\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle = 0$ is included in $\lambda \geq \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle$, we can write Eq. (4) as

$$\begin{aligned}
\hat{\beta}(\lambda) =& \begin{cases} \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle - \lambda & \text{if } \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle > 0 \text{ and } |\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle| > \lambda \\ \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle + \lambda & \text{if } \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle < 0 \text{ and } |\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle| > \lambda \\ 0 & \text{if } \lambda \geq \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle \end{cases} \\
=& \text{sign}\left(\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle\right)\left(|\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle| - \lambda\right)_+,
\end{aligned}$$

which is exactly the soft-threshold defined in as defined in Eq. (2) $S_\lambda(\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle)$ as we wanted to show in this exercise.