Laboratorium 4

Aleksander Milach

22/05/2020

Exercise from the manuscript

Point i)

$$f\left(||A^Tb||_\infty/2\right) = \left(||A^Tb||_\infty/2\right)\left(1+\frac{||b||_2}{||A^Tb||_\infty}\right) - ||b||_2 = ||A^Tb||_\infty/2 + ||b||_2/2 - ||b||_2 =$$
$$= ||A^Tb||_\infty/2 - ||b||_2/2 \leq 0,$$

since $||A^Tb||_\infty = \max\{|\langle A_i^T, b\rangle|\} \leq \max\{||A_i^T||_2||b||_2\} \leq ||b||_2$, by Cauchy-Schwarz inequality. Simirarly

$$g\left(||A^Tb||_\infty/2\right) = -\frac{||A^Tb||_\infty||b||_2}{||A^Tb||_\infty/2} + ||b||_2 + ||A^Tb||_\infty = -||b||_2 + ||A^Tb||_\infty \leq 0.$$

Obviously, when values of both functions are smaller or equal to 0, then their values cannot be larger than $|A_j^Tb|$, thus in that case both rules do not discard the column j.

Point ii)

$$f\left(||A^Tb||_\infty\right) = \left(||A^Tb||_\infty\right)\left(1+\frac{||b||_2}{||A^Tb||_\infty}\right) - ||b||_2 = ||A^Tb||_\infty + ||b||_2 - ||b||_2 = ||A^Tb||_\infty,$$

and

$$g\left(||A^Tb||_\infty\right) = -\frac{||A^Tb||_\infty||b||_2}{||A^Tb||_\infty} + ||b||_2 + ||A^Tb||_\infty = -||b||_2 + ||b||_2 + ||A^Tb||_\infty = ||A^Tb||_\infty.$$

$$f(\lambda) = 0 \iff \lambda\left(1+\frac{||b||_2}{||A^Tb||_\infty}\right) - ||b||_2 = 0 \iff \frac{\lambda\left(||A^Tb||_\infty + ||b||_2\right)}{||A^Tb||_\infty} = ||b||_2 \iff$$
$$\iff \lambda\left(||A^Tb||_\infty + ||b||_2\right) = ||A^Tb||_\infty||b||_2 \iff -\left(||A^Tb||_\infty + ||b||_2\right) = -\frac{||A^Tb||_\infty||b||_2}{\lambda} \iff$$
$$\iff -\frac{||A^Tb||_\infty||b||_2}{\lambda} + ||b||_2 + ||A^Tb||_\infty = 0 \iff g(\lambda) = 0.$$

Point iii)

$$f(|A_j^Tb|) = |A_j^Tb|\left(1+\frac{||b||_2}{||A^Tb||_\infty}\right) - ||b||_2 = |A_j^Tb| + \frac{|A_j^Tb|||b||_2}{||A^Tb||_\infty} - ||b||_2 = |A_j^Tb| + ||b||_2\left(\frac{|A_j^Tb|}{||A^Tb||_\infty} - 1\right) =$$
$$= |A_j^Tb| - ||b||_2\left(1 - \frac{|A_j^Tb|}{||A^Tb||_\infty}\right) \leq |A_j^Tb|$$

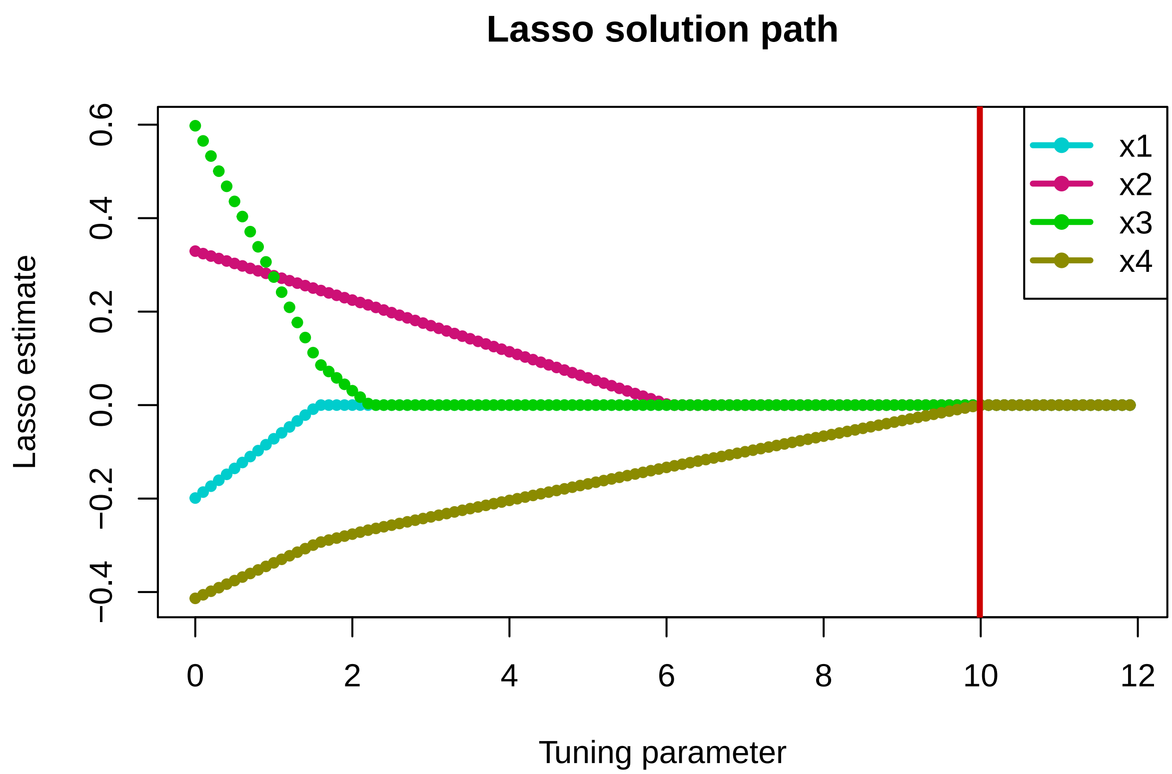
$$g(|A_j^Tb|) = -\frac{||A^Tb||_\infty||b||_2}{|A_j^Tb|} + ||b||_2 + ||A^Tb||_\infty = \frac{||A^Tb||_\infty|A_j^Tb| - ||A^Tb||_\infty||b||_2}{|A_j^Tb|} + ||b||_2 =$$
$$= \frac{||A^Tb||_\infty}{|A_j^Tb|}\left(|A_j^Tb| - ||b||_2\right) + ||b||_2 \leq \left(|A_j^Tb| - ||b||_2\right) + ||b||_2 \leq |A_j^Tb|.$$

In both subproofs we use the definition of infinity norm, in particular for every j: $|A_j^Tb| \leq ||A^Tb||_\infty$. From that reasoning and definitions of safe and DPP rules, when $\lambda \leq |A_j^T|$, both rules do not discard a column A_j.

Exercise 1

For values of λ higher or equal than those written in the following table, by safe and DPP rules $x_i(\lambda) = 0$.

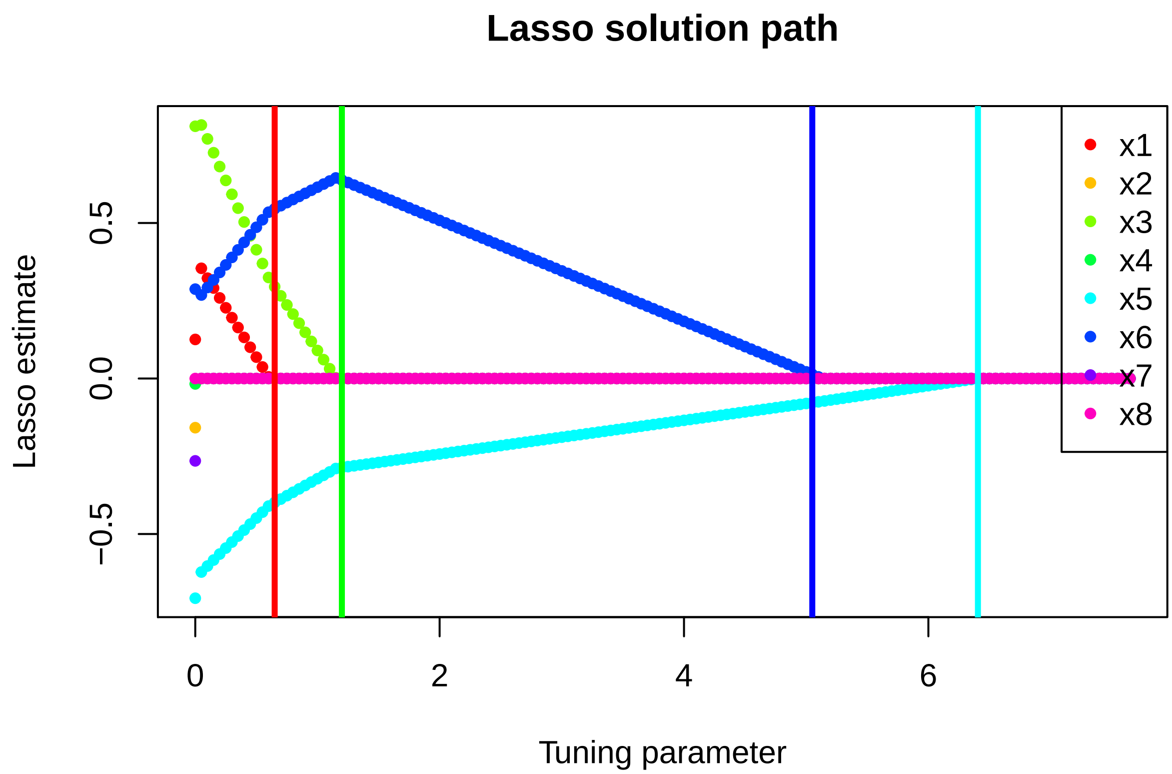
	x1	x2	x3	x4
Safe rule	8.022238	8.779657	6.788471	9.988666
DPP rule	7.964537	8.522724	6.519407	9.988666



Indeed, both rules provide an upper bound for value of λ, for which $x_i(\lambda) = 0$. For the column of matrix A, for which $|A_j^T b| = ||A_j^T b||_\infty$, which in this case is column 4, both rules provide the exact bound.

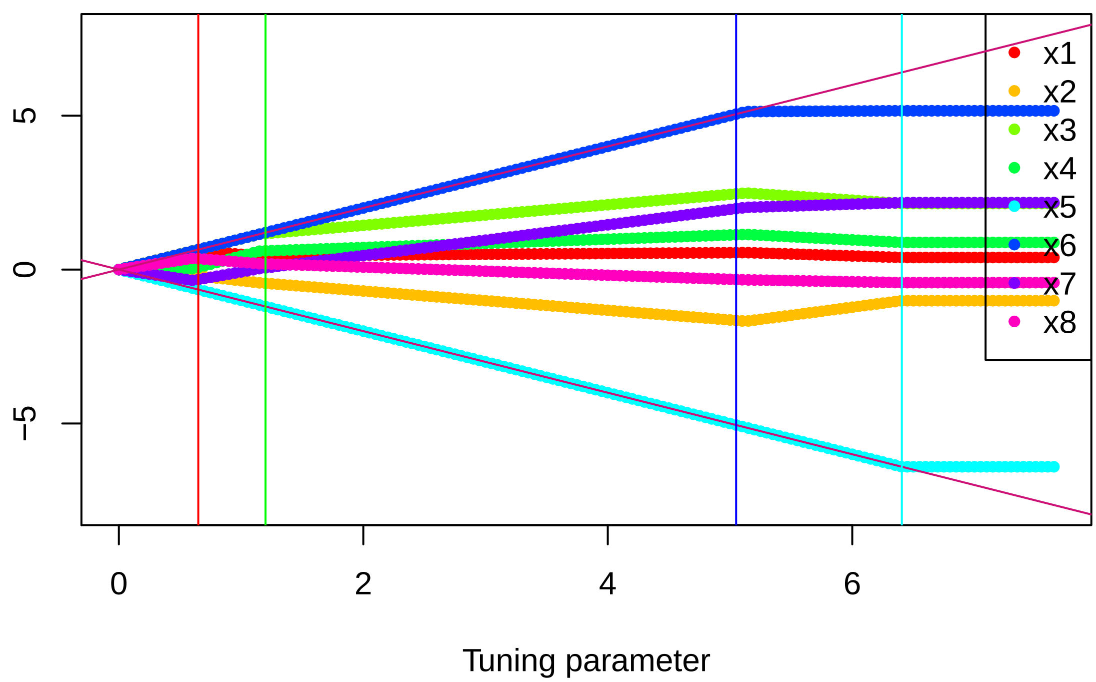
Exercise 2

Question 1



If, and only if the value of tuning parameter exceeds the value of λ_i corresponding to that column of matrix A, the value of LASSO estimate for corresponding variable equals. After the value of tuning parameter exceeds $||A^T b||_\infty$, then LASSO estimates for all variables equals zero (for convenience I deleted from the plot some lines corresponding to the values of λ_i very close to zero).

Question 2



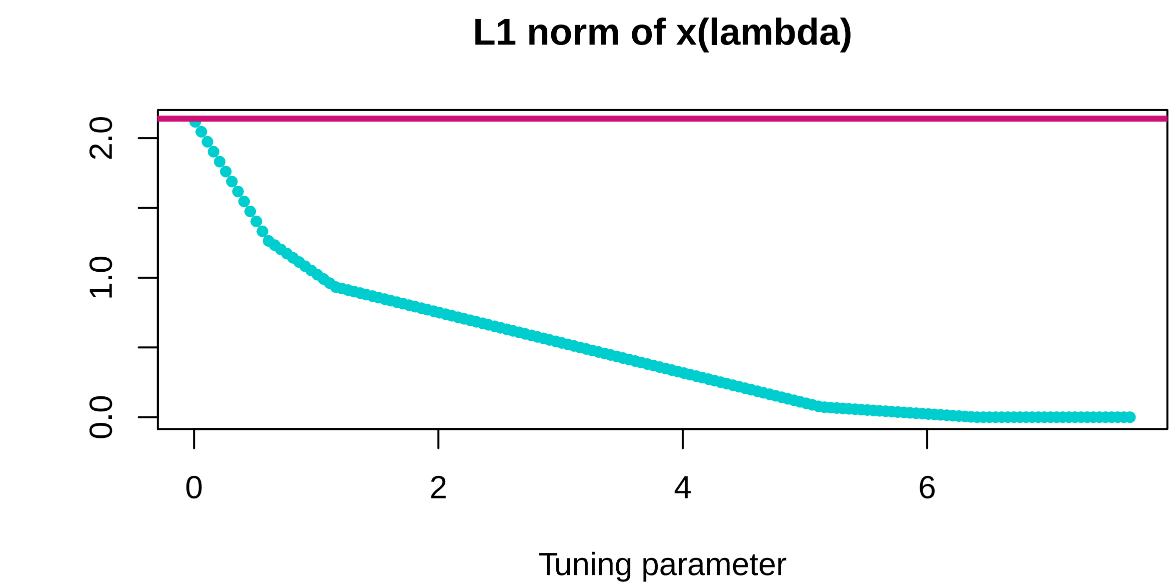
One can notice, that until the tuning parameter reaches the value λ_i, the value of $A_i^T(b - Ax(\lambda))$ equals to either λ or $-\lambda$. After it exceeds the value λ_i, all the values of $A_i^T(b - Ax(\lambda))$ lie between $\pm\lambda$. Finally, after the value of tuning parameter exceeds $||A^T b||_\infty$, all values of $A_i^T(b - Ax(\lambda))$ remain constant.

Question 3

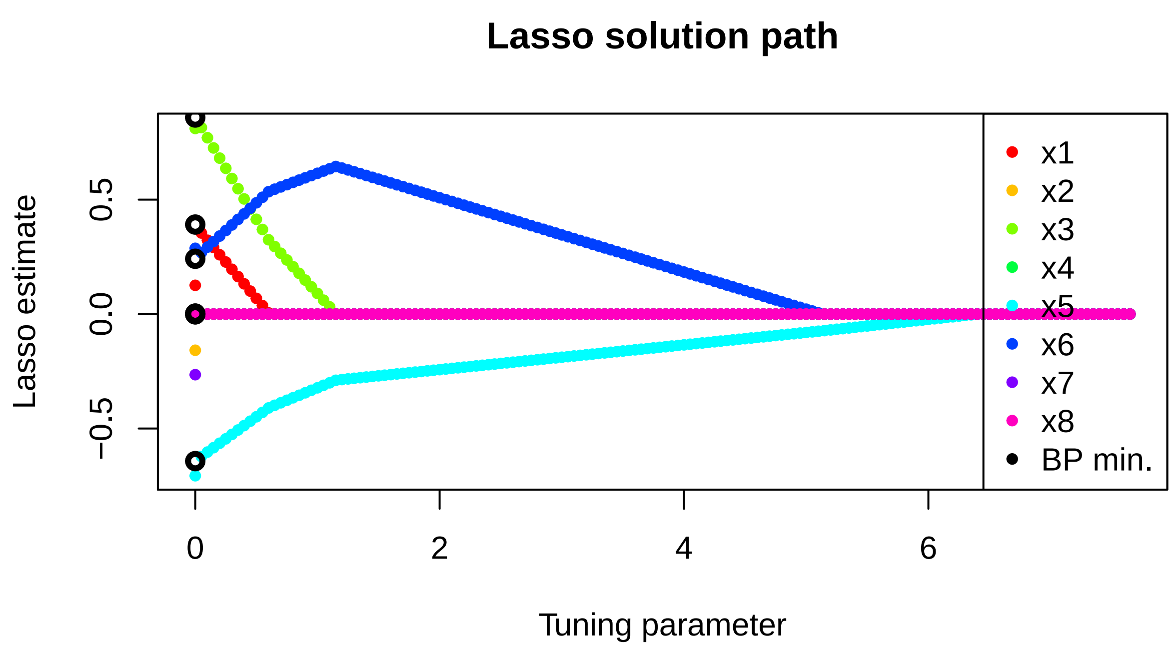
Let $x(\lambda)$ be a LASSO minimizer and x^* be a BP minimizer. Let us prove an inequality $||x(\lambda)||_1 \leq ||x^*||_1$.

$$\lambda||x(\lambda)||_1 \leq \lambda||x(\lambda)||_1 + \frac{1}{2}||Ax(\lambda) - b||_2 \leq \lambda||x^*||_1 + \frac{1}{2}||Ax^* - b||_2 = \lambda||x^*||_1,$$

which gives the desired inequality. In the second inequality we use definition of LASSO minimizer, thus we can replace $x(\lambda)$ with x^*. Finally we use the property of BP minimizers, such that $||Ax^* - b||_2 = 0$.



Question 4



The points obtained for $\lambda = 0$ obtained in Question 2 are points of discontinuity in curves following them. However if we would replace these points with points $(0, x_i^*)$, the curves of the functions $\lambda \to x_i(\lambda)$ will be continue.