Exercise 7

Eugen Dizer
Luis Clemens
Robert Freund

2 - Proof - Ridge Regression - Primal vs. Dual

To show: $\hat{\beta} = X^T \hat{\alpha}$.

Proof:

As written in the excercise, we know already

$$\hat{\beta} = (X^T X + \tau \mathbb{1}_D)^{-1} X^T \vec{y},$$
$$\hat{\alpha} = (X X^T + \tau \mathbb{1}_N)^{-1} \vec{y}.$$

So what's left to show is that $(X^T X + \tau \mathbb{1}_D)^{-1} X^T = X^T (X X^T + \tau \mathbb{1}_N)^{-1}$

because then:
$$\begin{aligned} \hat{\beta} &= (X^T X + \tau \mathbb{1}_D)^{-1} X^T \vec{y} \\ &= X^T (X X^T + \tau \mathbb{1}_N)^{-1} \vec{y} \\ &= X^T \hat{\alpha}. \end{aligned}$$

But this is obviously true if the inverse exists: and from the given form of $\hat{\alpha}, \hat{\beta}$ we know they do...

$$\begin{aligned} & X^T X X^T + \tau X^T = X^T X X^T + \tau X^T \\ \Leftrightarrow\quad & X^T (X X^T + \tau \mathbb{1}_N) = (X^T X + \tau \mathbb{1}_D) X^T \\ \Leftrightarrow\quad & (X^T X + \tau \mathbb{1}_D)^{-1} X^T = X^T (X X^T + \tau \mathbb{1}_N)^{-1}. \end{aligned}$$

$\square$