Fundamentals of Machine Learning – Exercise 7

Dominique Cheray & Manuel Krämer

Proof – Ridge Regression Primal vs Dual

① $(X^T X + \tau \mathbb{1}_D)^{-1} X^T = X^T (X X^T + \tau \mathbb{1}_N)^{-1}$

$\Leftrightarrow (X^T X + \tau \mathbb{1}_D)(X^T X + \tau \mathbb{1}_D)^{-1} X^T (X X^T + \tau \mathbb{1}_N) = (X^T X + \tau \mathbb{1}_D) X^T (X X^T + \tau \mathbb{1}_N)^{-1} (X X^T + \tau \mathbb{1}_N)$

$\Leftrightarrow X^T (X X^T + \tau \mathbb{1}_N) = (X^T X + \tau \mathbb{1}_D) X^T$

$\Leftrightarrow X^T X X^T + X^T \tau \mathbb{1}_N = X^T X X^T + \tau \mathbb{1}_D X^T$

$\Leftrightarrow X^T \tau \mathbb{1}_N = \tau \mathbb{1}_D X^T$

$$\Leftrightarrow \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{D1} & \cdots & x_{DN} \end{pmatrix} \cdot \underbrace{\begin{pmatrix} \tau & & 0 \\ & \ddots & \\ 0 & & \tau \end{pmatrix}}_{N} \Bigg\} N = \underbrace{\begin{pmatrix} \tau & & 0 \\ & \ddots & \\ 0 & & \tau \end{pmatrix}}_{D} \Bigg\} D \cdot \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{D1} & \cdots & x_{DN} \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \tau x_{11} & \cdots & \tau x_{1N} \\ \vdots & \ddots & \vdots \\ \tau x_{D1} & \cdots & \tau x_{DN} \end{pmatrix} = \begin{pmatrix} \tau x_{11} & \cdots & \tau x_{1N} \\ \vdots & \ddots & \vdots \\ \tau x_{D1} & \cdots & \tau x_{DN} \end{pmatrix}$$

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