Regul°:

3.1.1. Linear regression:

$$w^* = \underset{w}{\operatorname{argmin}} \mathcal{L}, \quad \mathcal{L} = \frac{1}{N}\sum_{n=1}^{N} \left(w^T x_n - t_n\right)^2$$

Extremum when $\nabla_{\vec{w}} \mathcal{L} = \vec{0}$

$$\nabla \mathcal{L} = 0 \iff \frac{1}{N}\sum_{n}^{N} 2\left(\vec{w}\vec{x}_n - t_n\right)\vec{x}_n = 0$$

$$\iff \underbrace{(\underbrace{XW}_{\text{shape}(N,D)(D,1)\to(N,1)} - T)^T}_{\text{shape}(N,1)^T = (1,N)} \underbrace{X}_{\text{shape}(N,D)} = 0$$

$$\iff (XW)^T X - T^T X = 0$$

$$W^T X^T X - T^T X = 0$$

$$W^T = T^T X \left(X^T X\right)^{-1}$$

$$W = \left(\left(X^T X\right)^{-1}\right)^T X^T T = \left(X^T X\right)^{-1} X^T T$$

So, we just need the inverse of $X^T X$, which is a $(D,N)(N,D) \to (D,D)$-shaped matrix. It's fine. Computing $X^T X$ is time $O(ND^2)$ $(O(D^3))$