Machine Learning over Networks

CA 1, 2 Group 6

CA1:

1.Define

$$X = [x_1| \dots |x_N], \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

So, we can write

$$\begin{aligned} f(w) &= \frac{1}{N}\sum_{i=1}^{N}\|w^T x_i - y_i\|_2^2 + \lambda\|w\|_2^2 \\ &= \frac{1}{N}\|X^T w - y\|_2^2 + \lambda\|w\|_2^2 \\ &= \frac{1}{N}(X^T w - y)^T(X^T w - y) + \lambda w^T w \end{aligned}$$

Hence,

$$\nabla f(w) = \left(\frac{1}{N}XX^T - \lambda I\right)w - \frac{1}{N}Xy$$

Therefore, the optimal solution w^* is obtained

$$\nabla f(w^*) = 0 \Longrightarrow w^* = \boxed{(XX^T - \lambda I)^{-1}Xy}$$

4 . We can use algorithms like Gradient Decent.

CA2:

3. For all the cases, the following variables were the same:

$$\lambda = 0.1, \quad \alpha = 0.1, \quad \#iterations = 1000,$$

	GD	**SGD**	**SVRG**	**SAG**
Time (ms)	43865	19.2	2002	59
$\boldsymbol{f}(w^*)$	0.356730	0.356732	0.356698	0.578958