MLoN Computer Homework 1

Group 1

Consider

$$w^* = \underset{w \in \mathbb{R}^d}{minimize} \frac{1}{N} \sum_{I \in [N]} \left\| w^T x_i - y_i \right\|^2 + \lambda \left\| w \right\|_2^2 \tag{1}$$

for the dataset $\{(x_i, y_i)\}$.

1

Question
Find a closed-form solution for this problem.
Solution
Define:

$$f(w) = \frac{1}{N} \sum_{I \in [N]} \left\| w^T x_i - y_i \right\|^2 + \lambda \left\| w \right\|_2^2 \tag{2}$$

Then, we could write:

$$f(w) = \frac{1}{N} \left\| X^T w - y \right\|_2^2 + \lambda \left\| w \right\|_2^2 \tag{3}$$

Taking derivative w.r.t. w:

$$\bigtriangledown f(w) = \left(\frac{2}{N} X^T X + 2\lambda I \right) w - \frac{2}{N} X^T y \tag{4}$$

Set it to 0, we get the candidate solution:

$$w = (X^T X + \lambda N I)^{-1} X^T y \tag{5}$$

As the value function f is positively quadratic, this closed-form is the exact solution.

2

Question
Consider "Individual household electric power consumption" dataset, find the optimal linear regressor from the closed-form expression.
Solution
The closed-form solution is found as:

$$w = (X^T X + \lambda N I)^{-1} X^T y \tag{*}$$

The calculation of $(*)$ is composed of matrix multiplication and inversion, with ordinary algorithms, the complexities are of $\mathcal{O}\{d^2 N\}$ for multiplication and of $\mathcal{O}\{d^3\}$ for inversion. With "Individual household electric power consumption" dataset, the running result is given in table below

Solution	**regularization**	**clock-time/s**	**w**	**pure loss**
Our	$\lambda = 0.1$	0.0475	[36.82, 1.14, -0.00, -7.06, -0.41, -0.39]	31.03
Reference	$\alpha = 0.1$	0.1382	[40.55, 1.78, -0.19, -8.02, -0.40, -0.39]	2209.55

As shown, the solution we calculated is close to the solution given by the library function.

3

Question
Consider "Greenhouse gas observing network" dataset, observe the scalability issue of the closed-form expression:
Solution
Similar to last question, with "Greenhouse gas observing network" dataset, the running result is given in table below

Solution	**regularization**	**clock-time/s**	**w**	**pure loss**
Our	$\lambda = 0.1$	0.0475	[0.5946, -0.5934, ..., 0.0191, -0.0527]	0.028
Reference	$\alpha = 0.1$	0.1382	[0.5945, -0.5436, ..., 0.0238, -0.0585]	8674.28

As shown, the solution we calculated is close to the solution given by the library function with less clock-time and a bit larger pure loss. Comparing with last question, the clock-times have difference with approximately 2 orders of magnitude which aligns with the complexity. As result, the closed-form solution clock-time scales up when data size and dimension scale up following the complexity.

4

Question
Address bigger datasets.
Solution
Here are several approaches:

1. More efficient algorithms for multiplication and inversion.
2. Iterative methods for regression, e.g. stochastic gradient methods.
3. Use approximation methods, e.g. dimension reduction.