①

a) $(A^TA+\lambda I)A^T = A^TAA^T+\lambda A^T = A^T(AA^T+\lambda I)$

$\therefore (A^TA+\lambda I)A^T = A^T(AA^T+\lambda I)$

$\therefore A^T(AA^T+\lambda I)^{-1} = (A^TA+\lambda I)^{-1}A^T$

b) The LHS of the formula computes more rapidly.

since $(A^TA+\lambda I)^{-1}$ is a psudo-inverse of A, it has dimension $A^{-1} \in \mathbb{R}^{100\times 8000}$

, where $A \in \mathbb{R}^{8000\times 100}$

$\therefore$ Computing $\mathbb{R}^{8000\times 100} \times \mathbb{R}^{100\times 8000}$ is faster than $\mathbb{R}^{8000\times 100} \times \mathbb{R}^{100\times 8000}$.

c(i) suppose $G^T = \begin{bmatrix} - g_1^T - \\ \vdots \end{bmatrix}$

$\therefore$ sol $= \min_{w} \| \text{sign}(G^Tw) - y \|_2^2$

It has a unique solution iff G is full-rank, which is unlikely to be the case for real data.

c(ii) $\min_{w} \| \text{sign}(Gw-y) - y \|_2^2 + \lambda \|w\|_2^2$

the LHS is more computationally efficient.

②

a) $w^{(k+1)} = \underset{w_i, i=1,\dots,M}{\text{argmin}} \sum_{i=1}^{M} (z_i^{(k)} - w_i)^2 + \lambda\tau w_i^2$

b) $w^{(k+1)} = \underset{w_i, i=1,\dots,M}{\text{arg min}} \sum_{i=1}^{M} (z_i^{(k)} - w_i)^2 + \lambda\tau |w_i|$.

③.