Appendix E: The proof of Theorem 1

Proof. Let us consider Problem (6) directly, rewrite it as following

$$\min_{W,b,\zeta} \frac{1}{2}\text{tr}(W^TW) + C \sum_{i=1}^{n} \zeta_i + \tau\|W\|_* \quad (19)$$

s.t. $$y_i[\text{tr}(W^TX_i) + b] \geq 1 - \zeta_i, \quad i = 1, \ldots, n$$
$$\zeta_i \geq 0, \quad i = 1, \ldots, n.$$

We construct the following Lagrangian function for (19)

$$Q(W,b,\zeta,\beta,\mu) = \frac{1}{2}\text{tr}(W^TW) + C \sum_{i=1}^{n} \zeta_i + \tau\|W\|_*$$
$$-\sum_{i=1}^{n} \beta_i\{y_i[\text{tr}(W^TX_i) + b] - 1 + \zeta_i\} - \sum_{i=1}^{n} \mu_i\zeta_i. \quad (20)$$

Letting the first-order derivative of $Q$ with respect to $\zeta_i$ be 0, we have

$$\mu_i = C - \beta_i, \quad i = 1, \ldots, n. \quad (21)$$

Substituting (21) in (20) to eliminate $r_i$ and $\xi_i$, we obtain

$$Q(W,\alpha) = \frac{1}{2}\text{tr}(W^TW) + C \sum_{i=1}^{n} \zeta_i + \tau\|W\|_*$$
$$-\sum_{i=1}^{n} \alpha_i\{y_i[\text{tr}(W^TX_i) - 1]\}. \quad (22)$$

Since $\mu_i \geq 0$, $\beta_i$ should satisfy the following condition

$$0 \leq \beta_i \leq C. \quad (23)$$

By KKT condition (Borwein & Lewis, 2010), we also have $0$ should be in the subgradient of $Q$ when the optimal solution is taken:

$$0 \in \partial Q(\tilde{W},\tilde{\beta})\bigg|_{W=\tilde{W}}. \quad (24)$$

Based on Eqn. (1), we have

$$\tilde{W} = D_\tau\left(\sum_{i=1}^{n} \tilde{\beta}_i y_iX_i\right),$$

where $\tilde{\beta}$ is the corresponding value of Lagrangian multiplier when $\tilde{W}$ is the optimal solution.

$$S^* = \arg\max_S E(S) \quad (25)$$

References