

Possible formulation for Kernel SMM (updated on 04/12/2020)

Let $\mathbf{U} \in \mathbb{R}^{m \times r}$, $\mathbf{V} \in \mathbb{R}^{n \times r}$ denote the factor matrices of interest in the low-rank kernel.

- Dual problem

$$(D) \quad f(\mathbf{U}, \mathbf{V}) := \max_{\alpha \geq 0} \left\{ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{X}_i, \mathbf{X}_j) \right\}, \quad (1)$$

subject to $\sum_{i=1}^N y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, N.$

where the Kernel

$$K(\mathbf{X}_i, \mathbf{X}_j) = \exp \left(-\frac{\|\mathbf{P}_U \mathbf{X}_i \mathbf{P}_V - \mathbf{P}_U \mathbf{X}_j \mathbf{P}_V\|_2^2}{\sigma^2} \right),$$

implicitly depends on \mathbf{U} and \mathbf{V} .

- Primal problem:

$$(P) \quad g(\mathbf{U}, \mathbf{V}) := \min_{\xi \in \mathbb{R}^N, \mathbf{D} \in \mathbb{R}^{r \times r}, b} \left\{ \|\mathbf{D}\|_F^2 + C \sum_{i=1}^N \xi_i \right\}, \quad (2)$$

subject to $y_i (\langle \mathbf{D}, h(\mathbf{P}_U \mathbf{X}_i \mathbf{P}_V) \rangle + b) \geq 1 - \xi_i,$
 $\xi_i \geq 0, \quad i = 1, \dots, N.$

Remark 1. The duality gap between (1) and (2) is zero for all (\mathbf{U}, \mathbf{V}) . Therefore,

$$\min_{(\mathbf{U}, \mathbf{V}) \in \text{feasible domain}} g(\mathbf{U}, \mathbf{V}) = \min_{(\mathbf{U}, \mathbf{V}) \in \text{feasible domain}} f(\mathbf{U}, \mathbf{V}).$$

Recall that we have proposed the non-linear SMM problem as $\min_{\mathbf{U}, \mathbf{V}} g(\mathbf{U}, \mathbf{V})$. Equivalently, we seek the solution to the following (strong) dual problem:

$$(D') \quad \min_{\substack{\alpha \geq 0, \\ \mathbf{U} \in \mathbb{R}^{m \times r}, \mathbf{V} \in \mathbb{R}^{n \times r}}} \left\{ -\sum_{i=1}^N \alpha_i + \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{X}_i, \mathbf{X}_j) \right\},$$

subject to $\sum_{i=1}^N y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, N.$

Note that no orthogonality constraints are imposed on \mathbf{U} and \mathbf{V} . This is because only the column spaces of \mathbf{U} and \mathbf{V} are relevant in the dual problem, and their actual values are non-important.

Remark 2 (Implementation). For linear kernel $K(\mathbf{X}_i, \mathbf{X}_j) = \langle \mathbf{X}_i \mathbf{P}_V, \mathbf{P}_U \mathbf{X}_j \rangle$, the above optimization can be easily solved using alternating SVM (see earlier notes). How about non-linear kernel? In particular, how to update \mathbf{U}, \mathbf{V} ? Gradient descent?