SMM Kernel Method

Chanwoo Lee, April 22, 2020

1 Kernel method

I am suggesting new kernel method which makes optimization easier. Define feature mapping $h : \mathbb{R}^{m\times n} \to \mathbb{R}^{m'\times n}$ where $m < m'$. Kernels for matrix case can be define as

$$K(X, X') = h(X)^T h(X') \in \mathbb{R}^{n\times n}.$$

Our objective primal problem is

$$\begin{aligned}
&\min_{U\in\mathbb{R}^{m'\times r}, V\in\mathbb{R}^{n\times r}, \boldsymbol{\xi}} \frac{1}{2}\|UV^T\|^2 + c\sum_{i=1}^N \xi_i \\
&\text{subject to } y_i(\langle UV^T, h(X_i)\rangle + b) \leq 1 - \xi_i \\
&\xi_i \geq 0, \quad i = 1, \ldots, N.
\end{aligned} \tag{1}$$

First, fix V and solve (1) with respect to U. We have the following dual problem.

$$\begin{aligned}
&\min_{\boldsymbol{\alpha}} -\sum_{i=1}^N \alpha_i + \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \alpha_i\alpha_j y_i y_j \langle h(X_i)H_V, h(X_j)H_V\rangle \\
&\text{subject to } \sum_{i=1}^N y_i\alpha_i = 0 \\
&0 \leq \alpha_i \leq C, \quad i = 1, \ldots, N.
\end{aligned}$$

Notice $\langle h(X_i)H_V, h(X_j)H_V\rangle = \operatorname{tr}\left(H_V h(X_i)^T h(X_j)\right) = \operatorname{tr}\left(H_V K(X_i, X_j)\right)$. Therefore, we can update U as

$$U = \sum_{i=1}^N \alpha_i y_i h(X_i) V (V^T V)^{-1} \tag{2}$$

where h function is not known. We are going to borrow this formula to update V in the next step. Now, fix U and solve (1) with respect to V. We have the following dual problem.

$$\begin{aligned}
&\min_{\boldsymbol{\beta}} -\sum_{i=1}^N \beta_i + \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \beta_i\beta_j y_i y_j \langle H_U h(X_i), H_U h(X_j)\rangle \\
&\text{subject to } \sum_{i=1}^N y_i\beta_i = 0 \\
&0 \leq \beta_i \leq C, \quad i = 1, \ldots, N.
\end{aligned} \tag{3}$$

To get an optimal $\boldsymbol{\beta}$ in (3), we need the information of $\langle H_U h(X_i), H_U h(X_j)\rangle$. Notice

$$\begin{aligned}
\langle H_U h(X_i), H_U h(X_j)\rangle &= \operatorname{tr}\left(H_U h(X_j) h(X_i)^T\right) = \operatorname{tr}\left(U(U^TU)^{-1}U^T h(X_j) h(X_i)^T\right) \\
&= \operatorname{tr}\left((U^TU)^{-1}U^T h(X_j) h(X_i)^T U\right)
\end{aligned} \tag{4}$$

Using the (2), we have the following expression of the component in (4).

$$
\begin{aligned}
U^T U &= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (V^T V)^{-1} V^T h(X_i)^T h(X_j) V (V^T V)^{-1} \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (V^T V)^{-1} V^T K(X_i, X_j) V (V^T V)^{-1},
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
U^T h(X_j) &= \sum_{l=1}^{N} \alpha_l y_l (V^T V)^{-1} V^T h(X_l)^T h(X_j) \\
&= \sum_{l=1}^{N} \alpha_l y_l (V^T V)^{-1} V^T K(X_l, X_j).
\end{aligned}
$$

Therefore, we can get an optimal $\boldsymbol{\beta}$ in (3) with (5). Finally, we update V with the help of Equation (5) as

$$
V = \sum_{i=1}^{N} \alpha_i y_i (U^T U)^{-1} U^T h(X_i).
$$

Remark 1. We can get explicit update formula V while U cannot be expressed as explicit value. This does make sense because in feature mapping $h : \mathbb{R}^{m \times n} \to \mathbb{R}^{m' \times n}$, dimension m' can be increased to arbitrary dimension while n being fixed.

Remark 2. There are some kernel functions that might be used often.

$$
\begin{aligned}
\text{Linear:} \quad & K(X, X') = X^T X' \\
\text{Polynomial:} \quad & K(X, X') = (X^T X' + I_n)^d \\
\text{Radial:} \quad & K(X, X') = \exp\left((X - X')^T (X - X')/\sigma\right),
\end{aligned}
$$

where $\exp(A) = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$. Notice that when $X \in \mathbb{R}^{m \times 1}$ i.e. X is a vector, all those definitions are reduced to SVM case. From this way, we can generalize linear SMM method to Kernel SMM with tractable algorithm.