KMR: Estimate Multiple Kernel Effects **01/03/2019**

Problem Set 729

Author: Wenying Deng

Estimate Single Kernel Effect Let's start by estimating single kernel effect. The objective function is,

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - K_1\boldsymbol{\alpha}_1\|^2 + \lambda\boldsymbol{\alpha}_1^\top K_1\boldsymbol{\alpha}_1. \tag{1}$$

Differentiating (1) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_1$, we obtain,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{y} - K_1\hat{\boldsymbol{\alpha}}_1), \tag{2}$$

$$\hat{\boldsymbol{\alpha}}_1 = (K_1 + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{3}$$

Substituting (3) into (2), we get

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \{\mathbf{X}^\top[K_1(K_1+\lambda\mathbf{I})^{-1} - \mathbf{I}]\mathbf{X}\}^{-1}\mathbf{X}^\top[K_1(K_1+\lambda\mathbf{I})^{-1} - \mathbf{I}]\mathbf{y} \\
&= \{\mathbf{X}^\top(\mathbf{I} + \lambda^{-1}K_1)^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top(\mathbf{I} + \lambda^{-1}K_1)^{-1}\mathbf{y} && \text{(Woodbury matrix identity)} \\
&= \{\mathbf{X}^\top V_1^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_1^{-1}\mathbf{y}, && (4)
\end{aligned}$$

where we denote $V_1 = \mathbf{I} + \lambda^{-1}K_1$ and therefore $(\lambda V_1)^{-1} = (K_1 + \lambda\mathbf{I})^{-1}$.

Estimate Multiple Kernel Effects Now we move on to estimating multiple kernel effects. Here we consider two kernel effects, and the objective function becomes,

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - K_1\boldsymbol{\alpha}_1 - K_2\boldsymbol{\alpha}_2\|^2 + \lambda\boldsymbol{\alpha}_1^\top K_1\boldsymbol{\alpha}_1 + \lambda\boldsymbol{\alpha}_2^\top K_2\boldsymbol{\alpha}_2. \tag{5}$$

Similarly, differentiating (1) with respect to $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$, we obtain,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{y} - K_1\hat{\boldsymbol{\alpha}}_1 - K_2\hat{\boldsymbol{\alpha}}_2), \tag{6}$$

$$\hat{\boldsymbol{\alpha}}_1 = (K_1 + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - K_2\hat{\boldsymbol{\alpha}}_2) = (\lambda V_1)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - K_2\hat{\boldsymbol{\alpha}}_2), \tag{7}$$

$$\hat{\boldsymbol{\alpha}}_2 = (K_2 + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - K_1\hat{\boldsymbol{\alpha}}_1) = (\lambda V_2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - K_1\hat{\boldsymbol{\alpha}}_1). \tag{8}$$

Plugging (8) into (6), we obtain,

$$\hat{\boldsymbol{\beta}} = \{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}(\mathbf{y} - K_1\hat{\boldsymbol{\alpha}}_1). \tag{9}$$

Plugging (8) into (7), we obtain,

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_1 &= \lambda[\lambda^2\mathbf{I} - V_1^{-1}K_2V_2^{-1}K_1]^{-1}V_1^{-1}V_2^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&= \lambda W_1^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&= \lambda W_1^{-1}\{\mathbf{y} - \mathbf{X}\{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}(\mathbf{y} - K_1\hat{\boldsymbol{\alpha}}_1)\}, && \text{(plug in (9))}
\end{aligned}$$

$$\downarrow$$

$$\hat{\boldsymbol{\alpha}}_1 = \{\lambda^{-1}W_1 - \mathbf{X}\{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}K_1\}^{-1}\{\mathbf{I} - \mathbf{X}\{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}\}\mathbf{y}. \tag{10}$$

where we denote

$$W_1 = V_2V_1(\lambda^2\mathbf{I} - V_1^{-1}K_2V_2^{-1}K_1) = \lambda^2 V_2V_1 - V_2K_2V_2^{-1}K_1.$$

Similarly,

$$\hat{\boldsymbol{\alpha}}_2 = \{\lambda^{-1}W_2 - \mathbf{X}\{\mathbf{X}^\top V_1^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_1^{-1}K_2\}^{-1}\{\mathbf{I} - \mathbf{X}\{\mathbf{X}^\top V_1^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_1^{-1}\}\mathbf{y}. \tag{11}$$

Alternative Following the same calculation as constructing block matrices. Letting

$$\underset{n\times(p+n+n)}{K_3} = [\underset{n\times p}{X}, \underset{n\times n}{K_1}, \underset{n\times n}{K_2}], \qquad \underset{(p+n+n)\times(p+n+n)}{K_4} = \begin{bmatrix} \underset{p\times p}{\mathbf{0}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & K_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & K_2 \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \underset{p\times 1}{\boldsymbol{\beta}} \\ \underset{n\times 1}{\boldsymbol{\alpha}_1} \\ \underset{n\times 1}{\boldsymbol{\alpha}_2} \end{bmatrix}.$$

Then (5) becomes

$$\|\mathbf{y} - K_3\boldsymbol{\theta}\|^2 + \lambda\boldsymbol{\theta}^\top K_4\boldsymbol{\theta}. \tag{12}$$

Differentiating (12) with respect to $\boldsymbol{\theta}$, we obtain,

$$(\lambda K_4 + K_3^\top K_3)\boldsymbol{\theta} = K_3^\top\mathbf{y},$$
$$\boldsymbol{\theta} = (\lambda K_4 + K_3^\top K_3)^{-1}K_3^\top\mathbf{y}.$$

Specifically,

$$\begin{aligned}
&\left\{\lambda \begin{bmatrix} \underset{p\times p}{\mathbf{0}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & K_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & K_2 \end{bmatrix} + \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top K_1 & \mathbf{X}^\top K_2 \\ K_1^\top X & K_1^\top K_1 & K_1^\top K_2 \\ K_2^\top X & K_2^\top K_1 & K_2^\top K_2 \end{bmatrix}\right\} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top K_1 & \mathbf{X}^\top K_2 \\ K_1 X & K_1(K_1 + \lambda\mathbf{I}) & K_1K_2 \\ K_2X & K_2K_1 & K_2(K_2 + \lambda\mathbf{I}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{X}^\top\mathbf{y} \\ K_1\mathbf{y} \\ K_2\mathbf{y} \end{bmatrix}, \\
\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}}_1 \\ \hat{\boldsymbol{\alpha}}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top K_1 & \mathbf{X}^\top K_2 \\ K_1 X & K_1(K_1 + \lambda\mathbf{I}) & K_1K_2 \\ K_2X & K_2K_1 & K_2(K_2 + \lambda\mathbf{I}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^\top\mathbf{y} \\ K_1\mathbf{y} \\ K_2\mathbf{y} \end{bmatrix}.
\end{aligned}$$