KMR: Estimate Multiple Kernel Effects **01/06/2019**

Problem Set 0

Author: Wenying Deng

0.1 Single Kernel

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

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

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

$$(\mathbf{I}+\lambda^{-1}K)^{-1}=\mathbf{I}-K(K+\lambda\mathbf{I})^{-1}$$

And finally,

$$\hat{\boldsymbol{\alpha}}_1=\{\lambda V_1-\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top K_1\}^{-1}\{\mathbf{I}-\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\}\mathbf{y}. \tag{4}$$

0.2 Two Kernels

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 (5) 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,

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_1 &= \{\lambda^2 V_2 V_1 + K_1 - \lambda V_2 K_1\}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= W_{21}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= W_{21}^{-1}\{\mathbf{y} - \mathbf{X}\{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}(\mathbf{y} - K_1\hat{\boldsymbol{\alpha}}_1)\}, && \text{(plug in (9))}\end{aligned}$$

$$\downarrow$$

$$\hat{\boldsymbol{\alpha}}_1 = \Big\{W_{21} - \mathbf{X}\{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}K_1\Big\}^{-1}\Big\{\mathbf{I} - \mathbf{X}\{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}\Big\}\mathbf{y}. \tag{10}$$

where we denote

$$W_{21} = \lambda^2 V_2 V_1 + K_1 - \lambda V_2 K_1.$$

0.3 Three Kernels

We consider three kernel effects, and the objective function becomes,

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - K_1\boldsymbol{\alpha}_1 - K_2\boldsymbol{\alpha}_2 - K_3\boldsymbol{\alpha}_3\|^2 + \lambda\boldsymbol{\alpha}_1^\top K_1\boldsymbol{\alpha}_1 + \lambda\boldsymbol{\alpha}_2^\top K_2\boldsymbol{\alpha}_2 + \lambda\boldsymbol{\alpha}_3^\top K_3\boldsymbol{\alpha}_3. \tag{11}$$

Similarly, differentiating (11) with respect to $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$, we obtain,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{y} - K_1\hat{\boldsymbol{\alpha}}_1 - K_2\hat{\boldsymbol{\alpha}}_2 - K_3\hat{\boldsymbol{\alpha}}_3), \tag{12}$$

$$\hat{\boldsymbol{\alpha}}_1 = (\lambda V_1)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - K_2\hat{\boldsymbol{\alpha}}_2 - K_3\hat{\boldsymbol{\alpha}}_3), \tag{13}$$

$$\hat{\boldsymbol{\alpha}}_2 = (\lambda V_2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - K_1\hat{\boldsymbol{\alpha}}_1 - K_3\hat{\boldsymbol{\alpha}}_3), \tag{14}$$

$$\hat{\boldsymbol{\alpha}}_3 = (\lambda V_3)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - K_1\hat{\boldsymbol{\alpha}}_1 - K_2\hat{\boldsymbol{\alpha}}_2). \tag{15}$$

Plugging (14) into (15), we obtain,

$$\hat{\boldsymbol{\alpha}}_3 = W_{23}^{-1}\big\{\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - K_1\hat{\boldsymbol{\alpha}}_1\big\}, \tag{16}$$

where we denote

$$W_{23} = \lambda^2 V_2 V_3 + K_3 - \lambda V_2 K_3.$$

Plugging (14) into (12), we obtain,

$$\mathbf{X}^\top(\lambda V_2)^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\top(\lambda V_2)^{-1}[\mathbf{y} - K_1\hat{\boldsymbol{\alpha}}_1 - K_3\hat{\boldsymbol{\alpha}}_3]. \tag{17}$$

Plugging (16) into (17), we obtain,

$$\hat{\boldsymbol{\beta}} = \{\mathbf{X}^\top(\lambda V_2)^{-1}[\mathbf{I} - K_3 W_{23}^{-1}]\mathbf{X}\}^{-1}\mathbf{X}^\top(\lambda V_2)^{-1}[\mathbf{I} - K_3 W_{23}^{-1}][\mathbf{y} - K_1\hat{\boldsymbol{\alpha}}_1]. \tag{18}$$

Plugging (14), (16) and (18) subsequently into (13), we obtain,

$$\begin{aligned}&\Big\{[W_{21} - K_3 W_{21}^{-1}K_1] - \gamma\mathbf{X}\{\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\mathbf{X}\}^{-1}\mathbf{X}^\top(\lambda V_2)^{-1}\gamma K_1\Big\}\hat{\boldsymbol{\alpha}}_1 \\ =&\gamma\{\mathbf{I} - \mathbf{X}\{\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\mathbf{X}\}^{-1}\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\}\mathbf{y},\end{aligned}$$

where we denote

$$\gamma = [\mathbf{I} - K_3 W_{23}^{-1}].$$

$$\downarrow$$

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_1 =&\Big\{[W_{21} - K_3 W_{21}^{-1}K_1] - \gamma\mathbf{X}\{\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\mathbf{X}\}^{-1}\mathbf{X}^\top(\lambda V_2)^{-1}\gamma K_1\Big\}^{-1} \\ &\cdot\gamma\{\mathbf{I} - \mathbf{X}\{\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\mathbf{X}\}^{-1}\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\}\mathbf{y}.\end{aligned} \tag{19}$$

0.4 Summary

Finally, if we put all those $\hat{\boldsymbol{\alpha}}_1$s together,

$$\hat{\boldsymbol{\alpha}}_1 = \{\lambda V_1 - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top K_1\}^{-1}\{\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\}\mathbf{y}, \qquad \text{(single kernel)}$$
$$\hat{\boldsymbol{\alpha}}_1 = \big\{W_{21} - \mathbf{X}\{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}K_1\big\}^{-1}\big\{\mathbf{I} - \mathbf{X}\{\mathbf{X}^\top V_2^{-1}\mathbf{X}\}^{-1}\mathbf{X}^\top V_2^{-1}\big\}\mathbf{y}, \qquad \text{(two kernels)}$$
$$\begin{aligned}\hat{\boldsymbol{\alpha}}_1 = &\Big\{[W_{21} - K_3W_{21}^{-1}K_1] - \gamma\mathbf{X}\{\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\mathbf{X}\}^{-1}\mathbf{X}^\top(\lambda V_2)^{-1}\gamma K_1\Big\}^{-1}\\ &\cdot\gamma\{\mathbf{I} - \mathbf{X}\{\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\mathbf{X}\}^{-1}\mathbf{X}^\top(\lambda V_2)^{-1}\gamma\}\mathbf{y}. \qquad \text{(three kernels)}\end{aligned}$$

where,

$$\begin{aligned}V_1 &= \mathbf{I} + \lambda^{-1}K_1,\\ V_2 &= \mathbf{I} + \lambda^{-1}K_2,\\ W_{21} &= \lambda^2V_2V_1 + K_1 - \lambda V_2K_1,\\ W_{23} &= \lambda^2V_2V_3 + K_3 - \lambda V_2K_3,\\ \gamma &= [\mathbf{I} - K_3W_{23}^{-1}].\end{aligned}$$

0.5 Alternative for Two Kernels

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

$$\|\mathbf{y} - K_3\boldsymbol{\theta}\|^2 + \lambda\boldsymbol{\theta}^\top K_4\boldsymbol{\theta}. \tag{20}$$

Differentiating (20) with respect to $\boldsymbol{\theta}$, we obtain,

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

Specifically,

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