Supplementary Material for GKEAL: Gaussian Kernel Embedded Analytic Learning for Few-shot Class Incremental Task

Huiping Zhuang[1*], **Zhenyu Weng**[2], **Run He**[1], **Zhiping Lin**[2], **Ziqian Zeng**[1]
[1]Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, China
[2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
*corresponding: hpzhuang@scut.edu.cn

1. Proof

Proof. The problem is formulated as

$$\underset{\boldsymbol{W}_{\text{FCN}}}{\text{argmin}} \quad \left\| \boldsymbol{Y}_{0:k}^{\text{train}} - \boldsymbol{X}_{0:k}^{\text{train}} \boldsymbol{W}_{\text{FCN}}^{(k)} \right\|_F^2 + \gamma \left\| \boldsymbol{W}_{\text{FCN}}^{(k)} \right\|_F^2$$

with a solution

$$\hat{\boldsymbol{W}}_{\text{FCN}}^{(k)} = (\boldsymbol{X}_{0:k}^{(\text{ke})\text{T}} \boldsymbol{X}_{0:k}^{(\text{ke})} + \gamma \boldsymbol{I})^{-1} \boldsymbol{X}_{0:k}^{(\text{ke})\text{T}} \boldsymbol{Y}_{0:k}^{\text{train}}. \tag{a}$$

We break (a) into

$$\begin{aligned}
\hat{\boldsymbol{W}}_{\text{FCN}}^{(k)} &= (\boldsymbol{X}_{0:k-1}^{(\text{ke})\text{T}} \boldsymbol{X}_{0:k-1}^{(\text{ke})} + \gamma \boldsymbol{I} + \boldsymbol{X}_{k}^{(\text{ke})\text{T}} \boldsymbol{X}_{k}^{(\text{ke})})^{-1} \begin{bmatrix} \boldsymbol{X}_{0:k-1}^{(\text{ke})\text{T}} & \boldsymbol{X}_{k}^{(\text{ke})\text{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_{0:k-1}^{\text{train}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}_{k}^{\text{train}} \end{bmatrix} \\
&= (\boldsymbol{X}_{0:k-1}^{(\text{ke})\text{T}} \boldsymbol{X}_{0:k-1}^{(\text{ke})} + \gamma \boldsymbol{I} + \boldsymbol{X}_{k}^{(\text{ke})\text{T}} \boldsymbol{X}_{k}^{(\text{ke})})^{-1} \begin{bmatrix} \boldsymbol{X}_{0:k-1}^{(\text{ke})\text{T}} \boldsymbol{Y}_{0:k-1}^{\text{train}} & \boldsymbol{X}_{k}^{(\text{ke})\text{T}} \boldsymbol{Y}_{k}^{\text{train}} \end{bmatrix}.
\end{aligned} \tag{b}$$

According to Woodbury matrix identity, we have

$$\boldsymbol{R}_k = \boldsymbol{R}_{k-1} - \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}} (\boldsymbol{I} + \boldsymbol{X}_{k}^{(\text{fe})} \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}})^{-1} \boldsymbol{X}_{k}^{(\text{fe})} \boldsymbol{R}_{k-1}$$

and (b) can be written as

$$\hat{\boldsymbol{W}}_{\text{FCN}}^{(k)} = \left[\underbrace{\boldsymbol{R}_k \boldsymbol{X}_{0:k-1}^{(\text{ke})\text{T}} \boldsymbol{Y}_{0:k-1}^{\text{train}}}_{\boldsymbol{G}_k} \quad \boldsymbol{R}_k \boldsymbol{X}_{k}^{(\text{ke})\text{T}} \boldsymbol{Y}_{k}^{\text{train}} \right]. \tag{c}$$

This obviously violates the incremental setting as $\hat{\boldsymbol{W}}_{\text{FCN}}^{(k)}$ involves historical data $\mathcal{D}_{0:k-1}^{\text{train}}$. To avoid the violation, we connect historical data using the trained weight $\hat{\boldsymbol{W}}_{\text{FCN}}^{(k-1)}$. which rewrites $\boldsymbol{G}_k$ as

$$\begin{aligned}
\boldsymbol{G}_k &= \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1} - \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}} (\boldsymbol{I} + \boldsymbol{X}_{k}^{(\text{fe})} \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}})^{-1} \boldsymbol{X}_{k}^{(\text{fe})} \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1} \\
&= \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1} - \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}} \left(\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{X}_{k}^{(\text{fe})} \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}})^{-1} \boldsymbol{X}_{k}^{(\text{fe})} \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}} \right) \boldsymbol{X}_{k}^{(\text{fe})} \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1} \\
&= \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1} - \left(\boldsymbol{R}_{k-1} - \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}} (\boldsymbol{I} + \boldsymbol{X}_{k}^{(\text{fe})} \boldsymbol{R}_{k-1} \boldsymbol{X}_{k}^{(\text{fe})\text{T}})^{-1} \boldsymbol{X}_{k}^{(\text{fe})} \boldsymbol{R}_{k-1} \right) \boldsymbol{X}_{k}^{(\text{fe})\text{T}} \boldsymbol{X}_{k}^{(\text{fe})} \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1} \\
&= \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1} - \boldsymbol{R}_{k} \boldsymbol{X}_{k}^{(\text{fe})\text{T}} \boldsymbol{X}_{k}^{(\text{fe})} \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1}.
\end{aligned} \tag{d}$$

Hence, (c) can be written as

$$\hat{\boldsymbol{W}}_{\text{FCN}}^{(k)} = \left[\underbrace{\hat{\boldsymbol{W}}_{\text{FCN}}^{k-1} - \boldsymbol{R}_{k} \boldsymbol{X}_{k}^{(\text{fe})\text{T}} \boldsymbol{X}_{k}^{(\text{fe})} \hat{\boldsymbol{W}}_{\text{FCN}}^{k-1}}_{\text{old tasks}} \quad \underbrace{\boldsymbol{R}_k \boldsymbol{X}_{k}^{(\text{ke})\text{T}} \boldsymbol{Y}_{k}^{\text{train}}}_{\text{new tasks}} \right]. \tag{e}$$

□