

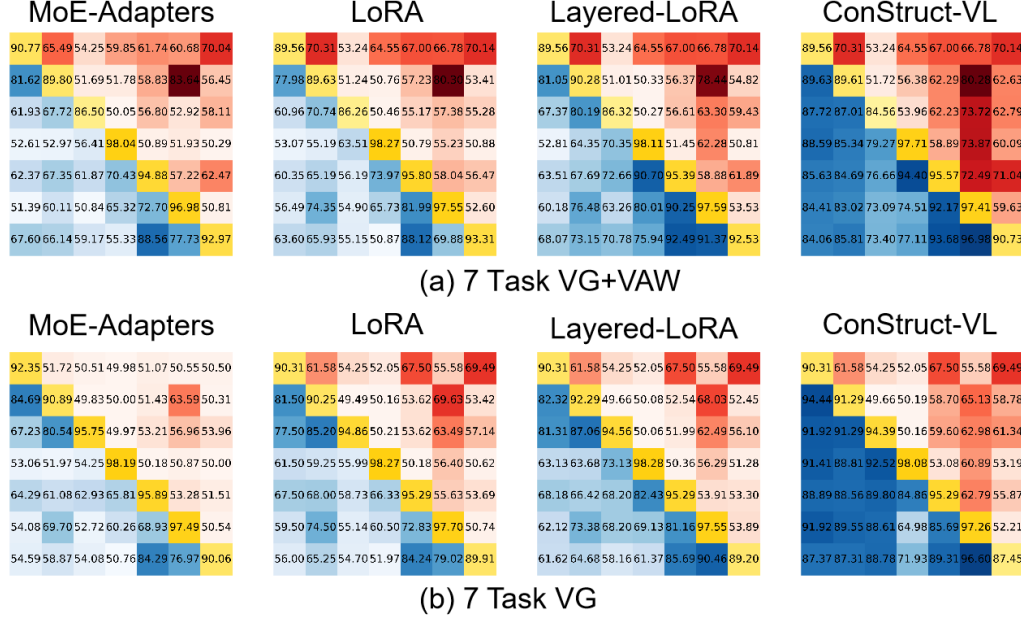


Figure R1: Empirical study of anti-forgetting, learning, and zero-shot performance in CL methods.

Table R1: The empirical and theoretical metrics for zero-shot stability and anti-forgetting ability, where  $|\mathcal{E}_k(\mathcal{M}^t) - \mathcal{E}_k(\mathcal{M}^s)| \leq \max\{\hat{\mathcal{E}}_{1:t}(\mathcal{M}^t) + \frac{1}{2t} \sum_{i=1}^t \text{Div}(\mathcal{T}_i, \mathcal{T}_k) + \sqrt{\frac{d[\ln(\tilde{N}/d)] + \ln(1/\delta)}{2N}}, \hat{\mathcal{E}}_{1:s}(\mathcal{M}^s) + \frac{1}{2t} \sum_{i=1}^s \text{Div}(\mathcal{T}_i, \mathcal{T}_k) + \sqrt{\frac{d[\ln(\tilde{N}_s/d)] + \ln(1/\delta)}{2N_s}}\}$  and  $\mathcal{E}_s(\mathcal{M}^t) - \mathcal{E}_s(\mathcal{M}^s) \leq \max\{\hat{\mathcal{E}}_{1:t}(\mathcal{M}^t) + \frac{1}{2t} \sum_{i=1}^t \text{Div}(\mathcal{T}_i, \mathcal{T}_s) + \sqrt{\frac{d[\ln(\tilde{N}/d)] + \ln(1/\delta)}{2N}}, \hat{\mathcal{E}}_{1:s}(\mathcal{M}^s) + \frac{1}{2t} \sum_{i=1}^s \text{Div}(\mathcal{T}_i, \mathcal{T}_s) + \sqrt{\frac{d[\ln(\tilde{N}_s/d)] + \ln(1/\delta)}{2N_s}}\}$ . Notation remains consistent with the original paper, including  $s \in \{1, \dots, t-1\}$  and  $k \in \{t+1, \dots, n\}$ .

|                     | Zero-shot stability (ZSS)                                                                         | Anti-forgetting ability (FFM)                                                     |
|---------------------|---------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Empirical metrics   | $\frac{1}{n-1} \sum_{t=2}^n  \text{Avg}(\sum_{i=1}^{t-1} A_{i,\cdot}) - \text{Avg}(A_{t,\cdot}) $ | $\frac{1}{n-1} \sum_{j=1}^{n-1} \max_{t \in \{1, \dots, n-1\}} (A_{tj} - A_{nj})$ |
| Theoretical metrics | $ \mathcal{E}_k(\mathcal{M}^t) - \mathcal{E}_k(\mathcal{M}^s) $                                   | $\mathcal{E}_s(\mathcal{M}^t) - \mathcal{E}_s(\mathcal{M}^s)$                     |