

ARF :  $\mathbf{w}^{\text{adv}} + \boldsymbol{\alpha}^{\text{adv}^\top} \mathbf{D}_{[15, :] }^{\text{adv}}$

RF :  $\mathbf{w}^{\text{adv}}$

