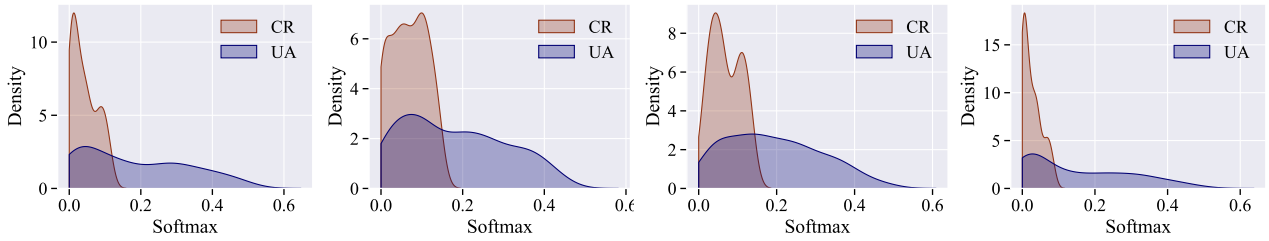
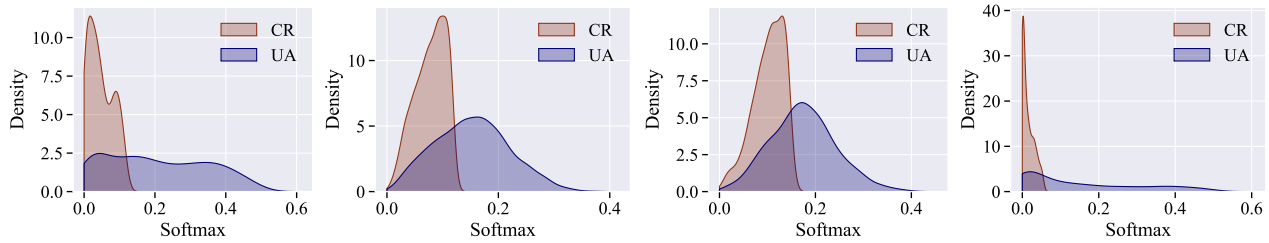


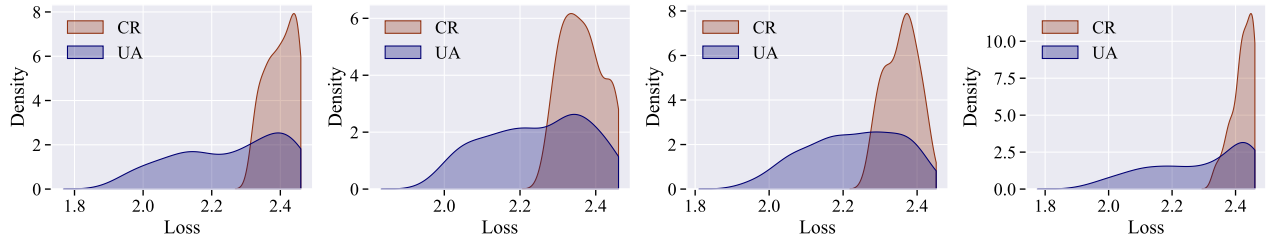
(a) FT (b) RL (c) Salun (d) SFRon

Figure 5: Softmax distribution in 10% **random data forgetting** scenario. We analyze the softmax distributions of ground truth labels for data identified as truly forgotten by CR and UA respectively. The distribution curves are fitted using KDE for clearer visualization. The results illustrate the softmax distributions of CR consistently closer to 0 when compared to UA, providing strong evidence that CR is better than UA in accurately capturing and measuring "real forgetting".



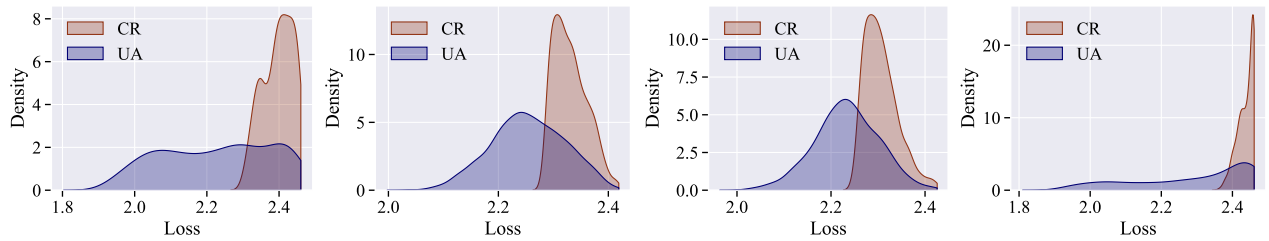
(a) FT (b) RL (c) Salun (d) SFRon

Figure 6: Softmax distribution in 50% **random data forgetting** scenario.



(a) FT (b) RL (c) Salun (d) SFRon

Figure 7: Loss distribution in 10% **random data forgetting** scenario. We analyze the cross entropy loss distributions of ground truth labels for data identified as truly forgotten by CR and UA respectively. Forgotten data identified by CR consistently show higher cross entropy loss than UA. Higher loss indicates better forgetting quality, which further validates that CR better captures "real forgetting".



(a) FT (b) RL (c) Salun (d) SFRon

Figure 8: Loss distribution in 50% **random data forgetting** scenario.