

Loss

$$\text{Loss} = \alpha * \text{KL}(\text{教師分布} \parallel \text{生徒分布}) + (1-\alpha) * \text{CE}(\text{正解ラベル}, \text{生徒予測})$$

1. KL項: 教師モデルの確率分布と生徒モデルの確率分布の差異を測定

例: 教師が [0.7, 0.2, 0.1] と予測する時、生徒も同様の分布を目指す

2. CE項: 通常の教師あり学習の損失 (正解ラベルとの差)

3. α : 両者のバランスを調整するパラメータ (通常0.5前後)