Mathematical Formulation

Let:

- $\mathbf{z} \in \mathbb{R}^C$ be the model logits for one sample,
- t be the true class index,
- C be the number of classes,
- β be the SmoothMax temperature,
- α be the scaling factor.

1) SmoothMax

We approximate the maximum logit using:

$$\mathrm{SmoothMax}_\beta(\mathbf{z}) = \frac{1}{\beta} \log \left(\sum_{j=1}^{C} e^{\beta z_j} \right)$$

When β is large, this approaches the ordinary max:

$$\mathrm{SmoothMax}_\beta(\mathbf{z}) \approx \max_j z_j$$

2) True-class margin

We compute the difference between the true-class logit and the SmoothMax value:

$$m = z_t - \mathrm{SmoothMax}_\beta(\mathbf{z})$$

- If the true class is the largest logit, then $m \approx 0$.
- If another class is larger, then $m < 0$.

3) Smooth penalty

We transform this margin with a sigmoid:

$$p(\mathbf{z}) = 2 \cdot \sigma(-\alpha m) - 1$$

where

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

This maps the penalty into the range $[-1, 1]$:

- correct prediction $\to$ penalty near 0
- incorrect prediction $\to$ positive penalty

4) Final loss

The final loss is:

$$\mathcal{L}_{\text{FairLoss}} = \mathcal{L}_{\text{CE}}(\mathbf{z}, t) + \frac{1}{C}\, p(\mathbf{z})$$

where the base cross-entropy loss is:

$$\mathcal{L}_{\text{CE}}(\mathbf{z}, t) = -\log\left(\frac{e^{z_t}}{\sum_{j=1}^{C} e^{z_j}}\right)$$

Compact Form

Putting everything together:

$$\boxed{\mathcal{L}_{\text{FairLoss}} = -\log\left(\frac{e^{z_t}}{\sum_{j=1}^{C} e^{z_j}}\right) + \frac{1}{C}\left[2 \cdot \sigma(\alpha\left(\text{SmoothMax}_{\beta}(\mathbf{z}) - z_t\right)) - 1\right]}$$

with

$$\text{SmoothMax}_{\beta}(\mathbf{z}) = \frac{1}{\beta}\log\left(\sum_{j=1}^{C} e^{\beta z_j}\right)$$