P_θ : model output $\in \mathbb{R}^N$

g : one-hot ground truth $\in \mathbb{R}^N$

u : a prior distribution $\in \mathbb{R}^N$

- smoothed target ($\varepsilon \in [0, 1]$)

$$t = (1-\varepsilon)\, g + \varepsilon u \in \mathbb{R}^N$$

- loss (constant terms are ignored)

$$L = -t \cdot \log P_\theta$$ ← dot product

$$= (1-\varepsilon) \underbrace{\left(-g \cdot \log P_\theta\right)}_{\text{cross entropy}} - \varepsilon \underbrace{\left(u \cdot \log P_\theta\right)}_{(*)}$$

(*) : if u is the uniform distribution,

$$u = \left[\frac{1}{N} \ \frac{1}{N} \ \cdots\cdots \right]$$

$$\left(u \cdot \log P_\theta\right) = \text{mean}\left(\log P_\theta\right)$$

Note : we can use whatever u