# Temperature Scaling

Softmax with Temperature

$$p_i = \exp(z_i/T) / \Sigma \exp(z_j/T)$$

T = Temperature Parameter


## T = 1

Standard Softmax

[0.05, 0.10, 0.80, 0.05]

Sharp Distribution

## T = 3~5

Optimal for Distillation

[0.15, 0.20, 0.45, 0.20]

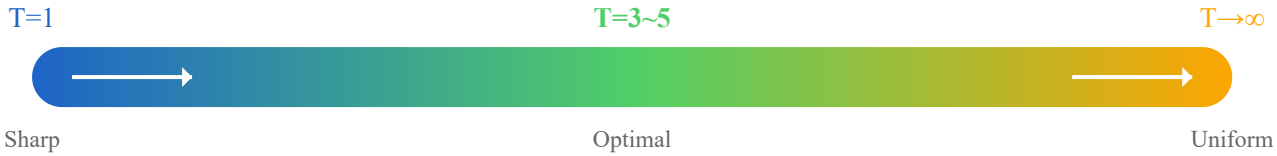
Soft Distribution

## T → ∞

Extreme Case

[0.25, 0.25, 0.25, 0.25]

Uniform Distribution




**Practical Tip:** T=3~4 is effective for medical image classification
Higher T → Richer inter-class relationship information → Improved Student learning