Temperature Calculation: Step-by-Step

Given: Logits = [2.0, 1.0, 0.5, 0.2]

Tokens = ["cat", "dog", "bird", "fish"]

Step 1: Scale by T=0.5 (MORE FOCUSED)

Scaled = [2.0/0.5, 1.0/0.5, 0.5/0.5, 0.2/0.5]

= [4.0, 2.0, 1.0, 0.4]

Softmax → [0.73, 0.19, 0.07, 0.04]

→ 73% on "cat" (VERY FOCUSED)

Step 2: Scale by T=1.0 (NORMAL)

Scaled = [2.0/1.0, 1.0/1.0, 0.5/1.0, 0.2/1.0]

= [2.0, 1.0, 0.5, 0.2]

Softmax → [0.53, 0.20, 0.12, 0.10]

→ 53% on "cat" (BALANCED)

Step 3: Scale by T=2.0 (FLATTER)

Scaled = [2.0/2.0, 1.0/2.0, 0.5/2.0, 0.2/2.0]

= [1.0, 0.5, 0.25, 0.1]

Softmax → [0.38, 0.23, 0.17, 0.15]

→ 38% on "cat" (MUCH FLATTER)

$$p_i = \frac{\exp(\text{logit}_i / T)}{\sum_j \exp(\text{logit}_j / T)}$$

Lower T → More confident (peaky)

Higher T → More random (flat)