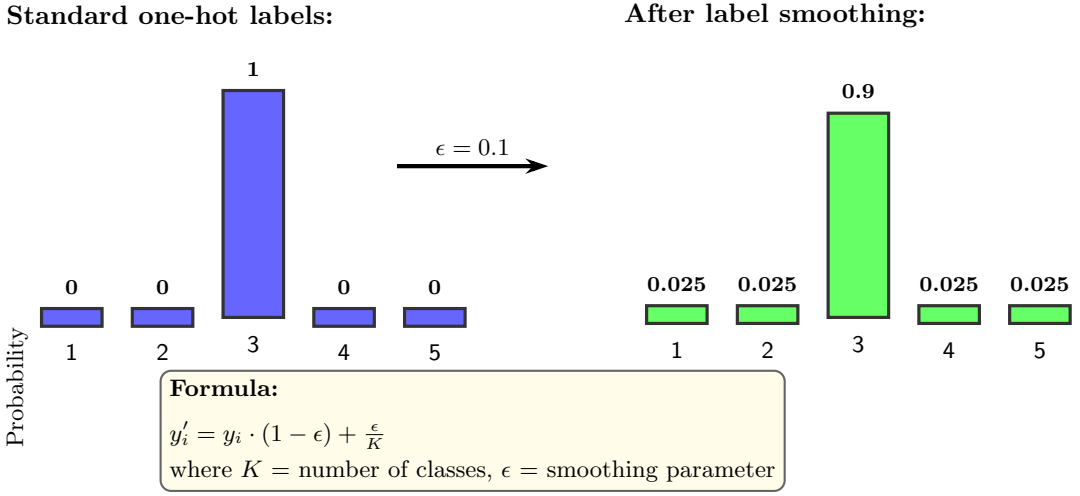
Label Smoothing



Formula:

$y'_i = y_i \cdot (1 - \epsilon) + \frac{\epsilon}{K}$

where K = number of classes, ϵ = smoothing parameter

Before:
- Hard labels (0 or 1)
- Model overfits
- Overconfident predictions

After:
- Soft labels
- Regularization effect
- Better calibration

Example calculation (for 5 classes, $\epsilon = 0.1$):

Correct class: $y'_3 = 1 \cdot (1 - 0.1) + \frac{0.1}{5} = 0.9 + 0.02 = 0.92$

Other classes: $y'_i = 0 \cdot (1 - 0.1) + \frac{0.1}{5} = 0 + 0.02 = 0.02$