

Confidence Calibration



What is Calibration?

If model says 80% confidence, it should be correct 80% of the time

Temperature Scaling

Adjust logits with temperature T
 $p' = \text{softmax}(\text{logits} / T)$

Platt Scaling

Logistic regression on outputs
 $p' = 1 / (1 + \exp(A \cdot p + B))$

Isotonic Regression

Non-parametric calibration
Monotonic function fitting



Calibration Metrics

ECE (Expected Calibration Error): $|\text{confidence} - \text{accuracy}|$

MCE (Maximum Calibration Error): $\max|\text{confidence} - \text{accuracy}|$

Brier Score: Mean squared error of probabilities