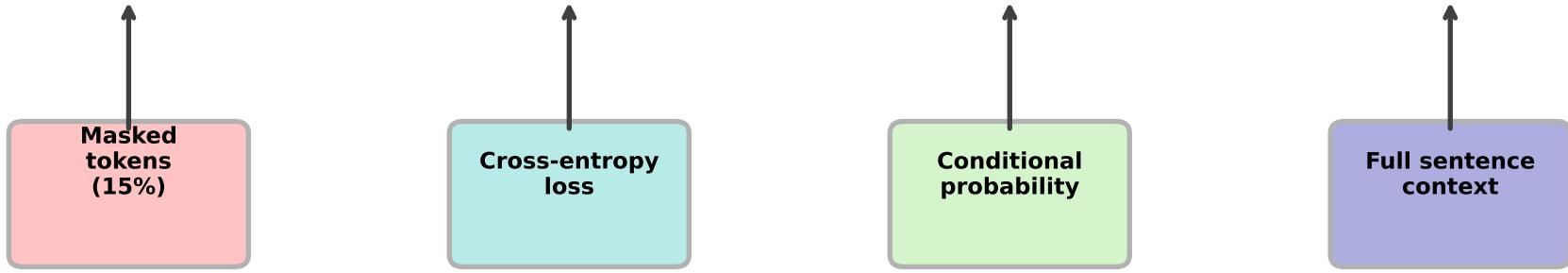


# BERT Training Objective: Mathematics

## Masked Language Modeling Objective

$$\mathcal{L}_{MLM} = - \sum_{i \in \text{masked}} \log P(w_i | \text{context})$$



### Example:

"The [MASK] sat" → P(cat) = 0.73

$$\text{Loss} = -\log(0.73) = 0.31$$