CBOW Architecture: Predict Word from Context
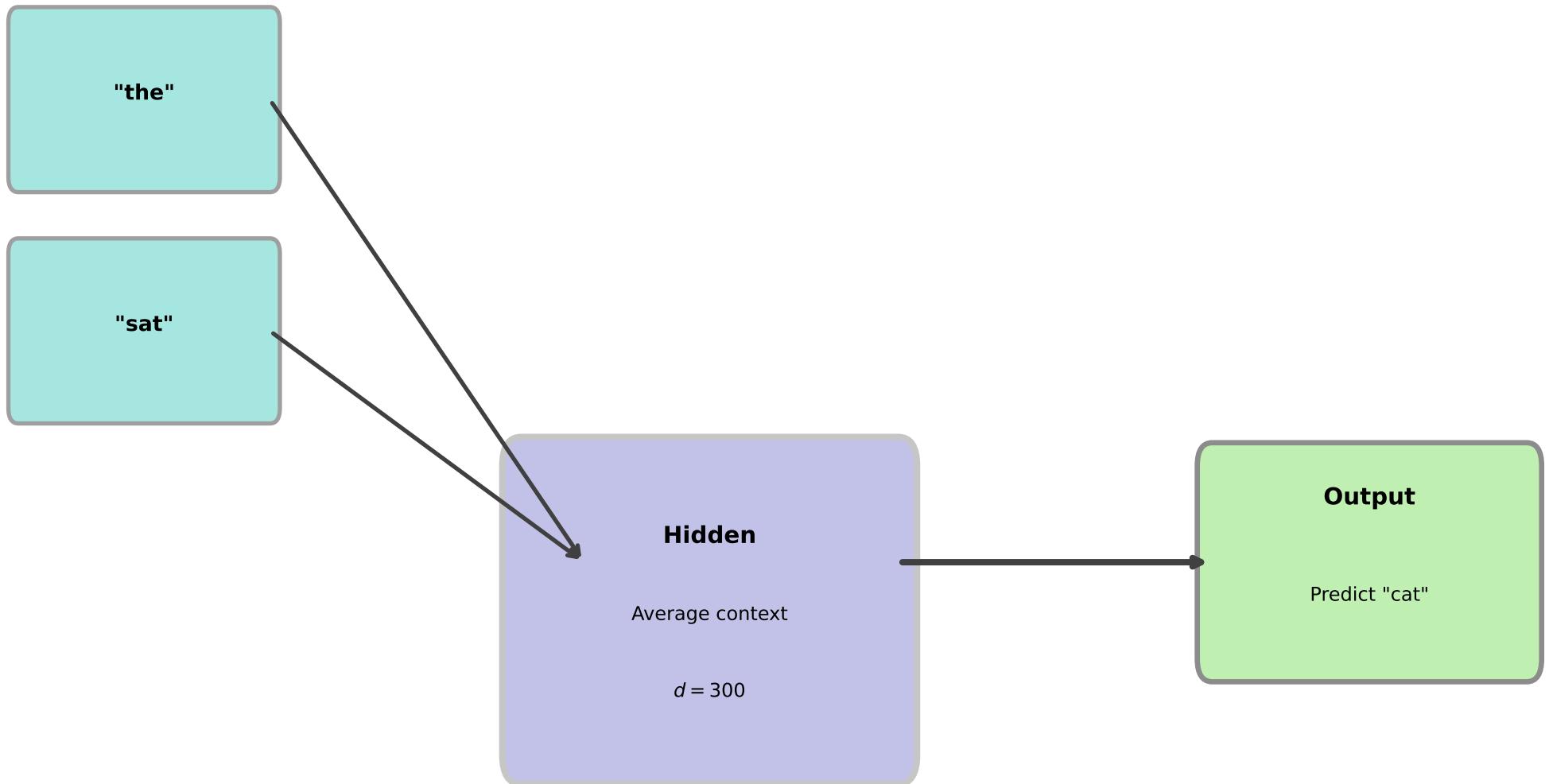
"the"

"sat"

Hidden

Average context

$d = 300$

Output

Predict "cat"

CBOW vs Skip-gram

CBOW: Context → Word (faster, frequent words)

Skip-gram: Word → Context (better rare words)

Both produce similar quality embeddings