

Forget Gate: What to Erase?

Example: "The cat was hungry. The dog ..."

Inputs:

h_{t-1} : Previous output

x_t : Current word ("dog")

Forget Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Output: 0 to 1

Decision:

"cat" info



10%

Forget! (new subject)

"hungry" info



20%

Forget! (not relevant)

Lower values (close to 0) = FORGET
Higher values (close to 1) = KEEP

Intuition: When you see "dog", forget information about "cat"