

# COMPUTATIONAL GRAPH

Why?

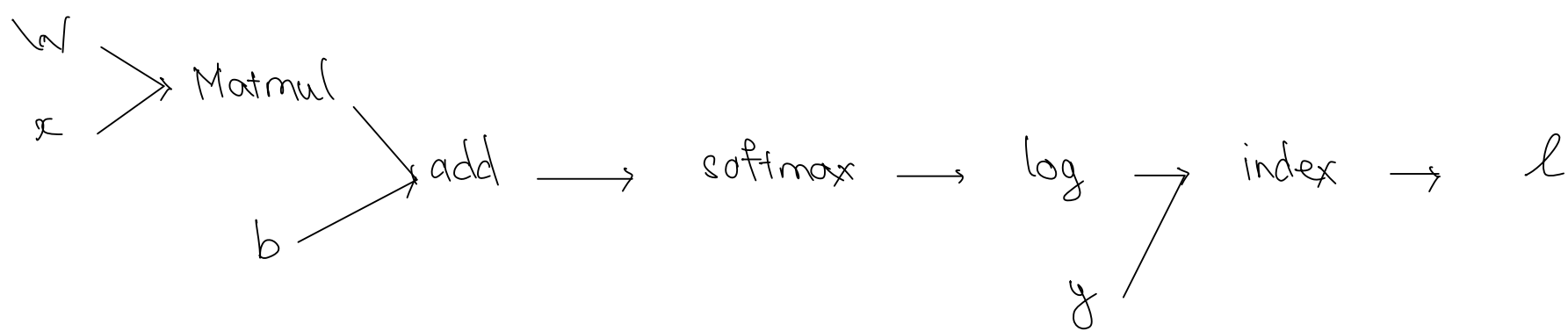
To calculate gradient easier with backpropagation

Example

Calculate gradient of this loss function:

$$\begin{aligned} \ell(\theta | x, y) &= \log(\text{softmax}(Wx + b))_y \\ &= \text{index}(\log(\text{softmax}(\text{add}(\text{matmul}(W, x), b))), y) \end{aligned}$$

Build computational graph:



Calculate the gradient manually:

$$\nabla_{\theta} \ell(\theta | x, y) \in \mathbb{R}$$

$$= \nabla_{\text{index}}(\dots) \nabla_{\log}(\dots) \nabla_{\text{softmax}}(\dots) \nabla_{\text{add}}(\dots) \left[ \nabla_{\text{matmul}}(W, x) \nabla_{\theta} W + \nabla_{\theta} b \right]$$

$1 \times n \quad n \times m \quad m \times l \quad l \times k \quad k \times \dots \quad k \times \dots$

*number of parameters*

Question: Should we calculate forward (matmul  $\rightarrow$  index) or backward (index  $\rightarrow$  matmul)?

Answer: Calculate backward because the operations require will be lesser (vector-matrix multiplication) than forward (matrix-matrix multiplication)

Back propagation Computational Graph

