⊛ Softmax derivative note

$$y = i \rightarrow (soft_i - soft_i^2) \frac{\partial L}{\partial m_i}$$

$$y \neq i \quad -\sum_n soft_i \cdot soft_n \cdot \frac{\partial L}{\partial m_n}$$

- You must comm this before summing $soft_n$
- Because you can't indiv. remove $soft_n$ ith $\frac{\partial L}{\partial m_n}$ → or else deriv. won't be zero!

→ sum of rows the dw times $soft_i$

→ output w/ dx to $soft_i$