

$$y_t \rightarrow -\log\left(\frac{e^{y_i}}{\sum_j e^{y_j}}\right) \rightarrow L_t$$

cross entropy loss (softmax)

y_t
 \rightarrow
 $-\log\left(\frac{e^{y_i}}{\sum_j e^{y_j}}\right)$
 \rightarrow
 L_t

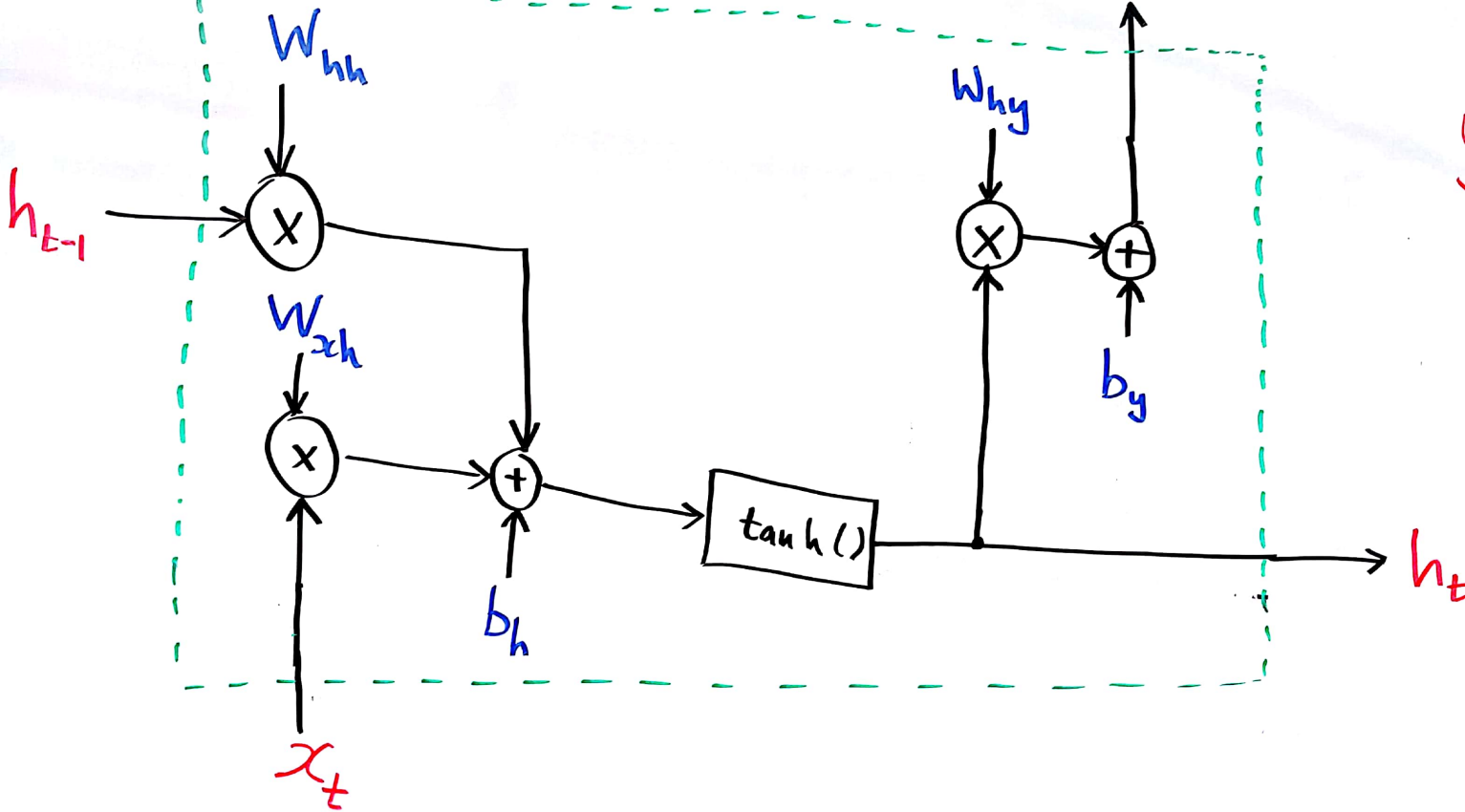
softmax
cross entropy loss

$$\frac{\partial L}{\partial y_i} = -\left(\frac{\sum_j e^{y_j}}{e^{y_i}}\right) \cdot \frac{\left(\sum_j e^{y_j} \cdot \cancel{e^{y_i}} - e^{y_i} \cdot \cancel{e^{y_i}}\right)}{\left(\sum_j e^{y_j}\right)^2} = \left(\frac{e^{y_i}}{\sum_j e^{y_j}} - 1\right)$$

$$\frac{\partial L}{\partial y_k} = -\left(\frac{\sum_j e^{y_j}}{e^{y_i}}\right) \cdot \cancel{e^{y_i}} \cdot \frac{-1}{\left(\sum_j e^{y_j}\right)^2} \cdot e^{y_k} = \frac{e^{y_k}}{\sum_j e^{y_j}}$$

$$\therefore \boxed{\frac{\partial L}{\partial y_n} = \frac{e^{y_n}}{\sum_j e^{y_j}} - 1 \quad (n=i)}$$

Vanilla RNN



$$y_t \rightarrow -\log\left(\frac{e^{y_i}}{\sum_j e^{y_j}}\right) \rightarrow L_t$$

cross entropy loss
(softmax)

$$\frac{\partial L}{\partial y_i} = - \left(\frac{\sum_j \cancel{e^{y_j}}}{\cancel{e^{y_i}}} \right) \cdot \frac{\left(\sum_j \cancel{e^{y_j}} \cdot \cancel{e^{y_i}} - \cancel{e^{y_i}} \cdot \cancel{e^{y_i}} \right)}{\left(\sum_j \cancel{e^{y_j}} \right)^2} = \left(\frac{e^{y_i}}{\sum_j e^{y_j}} - 1 \right)$$

$$\frac{\partial L}{\partial y_k} = - \left(\frac{\sum_j \cancel{e^{y_j}}}{\cancel{e^{y_i}}} \right) \cdot \cancel{e^{y_i}} \cdot \frac{-1}{\left(\sum_j \cancel{e^{y_j}} \right)^2} \cdot e^{y_k} = \frac{e^{y_k}}{\sum_j e^{y_j}}$$

$$\therefore \boxed{\frac{\partial L}{\partial y_n} = \frac{e^{y_n}}{\sum_j e^{y_j}} - 1 \quad (n=i)}$$