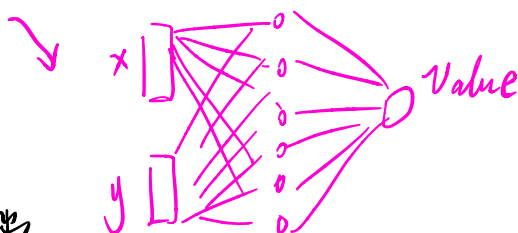


三种计算相关性

1. 内积

2. $X^T W Y$ 如果 X Y 维度不相同

3. 训练神经网络



↙ 对应LSTM输出
score (h_j, s_{t-1})

↑ 上一层隐含层

soft-Attention (所有 h_j)

hard-Attention (只有几个 h_j)

↑ 不可导 (强化学习方法)

求 LSTM 输出 $P(y_t | y_1, y_2, \dots, y_{t-1}, x)$
 $= f(y_{t-1}, s_t, c_t)$

$$s_t = f(s_{t-1}, y_{t-1}, c_t)$$

$\left\{ \begin{array}{l} \text{LSTM} \\ \text{GRU} \end{array} \right.$

↘ 输入进去依然做 embedding

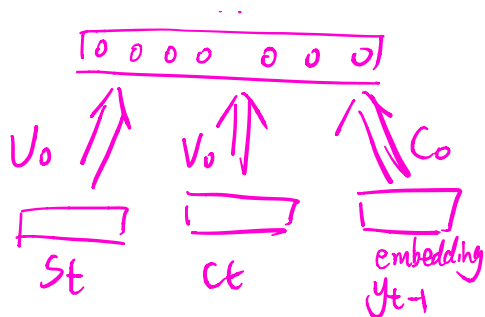
$$c_t = \sum_{i=1}^I \alpha_{t,i} h_i$$

$$\alpha_{t,i} = \text{score}(s_{t-1}, h_i) \Rightarrow \text{softmax}$$

字典 W_0 softmax

维 1000001

$W_0 \uparrow \uparrow$



解决对齐问题

我是中国人
 $\downarrow \quad \downarrow \quad \quad \downarrow \quad \downarrow$
 I am chinese

self-attention

$$y_t = f(s_{t-1}, c_t, y_{t-1})$$

$$1. \alpha_{t,i} = \text{score}(s_{t-1}, h_i)$$

$$2. \alpha_{t,i} = \text{score}(v, h_i)$$

\uparrow 句子向量