1) Atenção input
- Todos os encoder states $h1, h2, ... hn$
- Um decoder state st

2) Score de atenção e^t
- $SCORE(st, hL)$, $L = 1 .. 3$
- P/ dot-product, $SCORE = st^T \cdot hL$

Considerando

$$e^{t1} = st_1^T.h_{11}, \ st_1^T.h_{12}, \ st_1^T.h_{13}$$
$$e^{t2} = st_2^T.h_{21}, \ st_2^T.h_{22}, \ st_2^T.h_{23}$$
$$e^{t2} = st_3^T.h_{31}, \ st_3^T.h_{32}, \ st_3^T.h_{33}$$

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \qquad st = \begin{bmatrix} st_1 \\ st_2 \\ st_3 \end{bmatrix}$$

$$e^t = H \cdot st = \begin{bmatrix} (h_{11} st_1 + h_{12} st_1 + h_{13} st_1) \\ (h_{21} st_2 + h_{22} st_2 + h_{23} st_2) \\ (h_{31} st_3 + h_{32} st_3 + h_{33} st_3) \end{bmatrix}_{3x1}$$

3) Distribuição de atenção

$$\alpha^t = Softmax(H \cdot st) = \begin{bmatrix} Softmax(h_{11} st_1 + h_{12} st_1 + h_{13} st_1) \\ Softmax(h_{21} st_2 + h_{22} st_2 + h_{23} st_2) \\ Softmax(h_{31} st_3 + h_{32} st_3 + h_{33} st_3) \end{bmatrix}_{3x1}$$

4) Vetor de contexto

$$at = \sum_{i=1}^{3} \alpha_i^t . h_i$$

$$at = \begin{bmatrix} Softmax(h_{11} st_1 + h_{12} st_1 + h_{13} st_1) h_{11} + \\ Softmax(h_{21} st_2 + h_{22} st_2 + h_{23} st_2) h_{12} + \\ Softmax(h_{31} st_3 + h_{32} st_3 + h_{33} st_3) h_{13} \\ \\ Softmax(h_{11} st_1 + h_{12} st_1 + h_{13} st_1) h_{21} + \\ Softmax(h_{21} st_2 + h_{22} st_2 + h_{23} st_2) h_{22} + \\ Softmax(h_{31} st_3 + h_{32} st_3 + h_{33} st_3) h_{23} \\ \\ Softmax(h_{11} st_1 + h_{12} st_1 + h_{13} st_1) h_{31} + \\ Softmax(h_{21} st_2 + h_{22} st_2 + h_{23} st_2) h_{32} + \\ Softmax(h_{31} st_3 + h_{32} st_3 + h_{33} st_3) h_{33} \end{bmatrix}_{3x1}$$