

- ③ ② で計算した内積を softmax で確率的表現に変換する
(sometimes called attention weights)

$$\text{softmax} \begin{bmatrix} h_1 \cdot d_1 \\ h_2 \cdot d_2 \\ h_3 \cdot d_3 \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

w'_i が w_i に 2 次元の attention になる

$$\text{softmax} \begin{bmatrix} h_1 \cdot d_1 \\ h_2 \cdot d_2 \\ h_3 \cdot d_2 \end{bmatrix} = \begin{bmatrix} p'_1 \\ p'_2 \\ p'_3 \end{bmatrix}$$

$$\text{attention weight} = \begin{bmatrix} p_1, p'_1 \\ p_2, p'_2 \\ p_3, p'_3 \end{bmatrix}$$

- ④ h_3 の各要素を attention weight で掛け付け、足し合わせ 1 本のベクトルにする (called context vector).
- $$\begin{aligned} h_1 \cdot p_1 &= p_1 \text{ で } \text{掛け付け} + h'_1 \\ h_2 \cdot p_2 &= + \\ h_3 \cdot p_3 &= + \\ &\parallel \\ &\text{Context vector} \end{aligned}$$

- ⑤ コーシストベクトル d_i を結ぶ、
一本のベクトルにする

output of w'_1

output of w'_2

