Normalize Scores
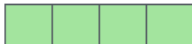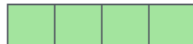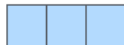
Input	**Thinking**	**Machines**
Embedding	x_1 ▢▢▢▢	x_2 ▢▢▢▢
Values	v_1 ▢▢▢	v_2 ▢▢▢
Score	$q_1 \cdot k_1 = 112$	$q_1 \cdot k_2 = 96$
Divide by 8 ($\sqrt{d_k}$)	14	12
Softmax	0.88	0.12