Causal Language Modeling

8

	I	work	in	the	city	of	New
	1	2	3	4	5	6	7
Queries	q_1	q_2	q_3	q_4	q_5	q_6	q_7
Keys	k_1	k_2	k_3	k_4	k_5	k_6	k_7
Values	v_1	v_2	v_3	v_4	v_5	v_6	v_7

Task: Compute a representation of the context to be able to predict the next word

Answer: York

Causal Language Modeling

	I	work	in	the	city	of	New	_____	Answer: York
	1	2	3	4	5	6	7		

Queries $\quad q_1 \quad q_2 \quad q_3 \quad q_4 \quad q_5 \quad q_6 \quad q_7$

Keys $\qquad k_1 \quad k_2 \quad k_3 \quad k_4 \quad k_5 \quad k_6 \quad k_7 \qquad \mathbb{R}^h$

Values $\quad v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \quad v_6 \quad v_7$

Task: Compute a representation of the context to be able to predict the next word

Attention Mechanism

- Representation of the context:

$$\sum_{i \leq 7} \qquad\qquad v_i$$