

CS 378 Lecture 17: Transformer

Language Modeling

Announcements

- Custom FP proposals due today

Recap Attention
 v_1, \dots, v_n values

Keys k_1, \dots, k_n , query q

scores $s_i = k_i^T q$

attn probs = $\text{softmax}(\bar{s})$

result:

$$\sum \alpha_i \cdot v_i$$

Self-attention: embeddings e_1, \dots, e_n

E $\text{len} \times d$

"Basic" form: let $Q = E$
 $K = E$

len
queries
rather than
1

$len \times len$

$$s_{ij} = q_i \cdot k_j$$

$$S = QK^T$$

$(len \times d) \times (d \times len)$

$$A = \underset{\text{(row)}}{\text{softmax}}(S)$$

Result: AE alternately: AV

Example POS tagging

$[100]$ $[001]$
the fed

We want to tag fed
as N or V

$[010]$ $[001]$
i fed

the \rightarrow N
i \rightarrow V

query = $[100]$ "find the"

$$e=3$$

$$\begin{array}{l} \text{Scores:} \\ t \\ o \\ i \end{array} \begin{array}{l} r \\ f \\ o \\ f \end{array} \Rightarrow \begin{array}{l} \frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{2} \\ \frac{1}{2} \end{array}$$

result vectors: $\begin{bmatrix} 3/4 \\ 0 \\ 1/4 \end{bmatrix}$ ← found the

$\begin{bmatrix} 0 \\ 1/2 \\ 1/2 \end{bmatrix}$ ← did not find the

fed $[0 \ 0 \ 1] \Rightarrow$ query $[1 \ 0 \ 0]$

"fed looks for the" fed $-(W^Q)^T$

$$W^Q = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad [0 \ 0 \ 1] \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$W^K, W^V \text{ matrices} = [1 \ 0 \ 0]$$