

Attention Mechanism

25 फरवरी 2025 22:22

Code Simplified Attention
Code Self Attention
Causal Self Attention
Multi Head Attention

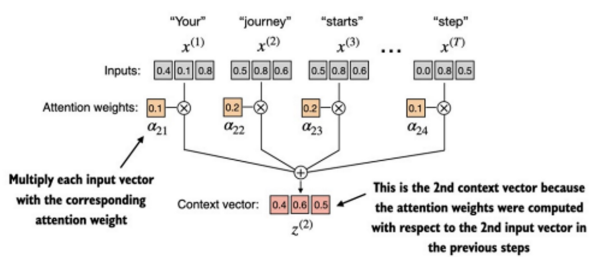
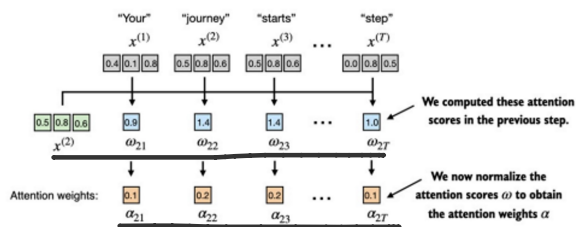
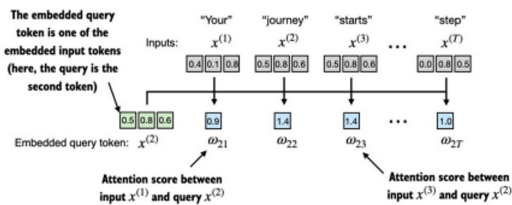
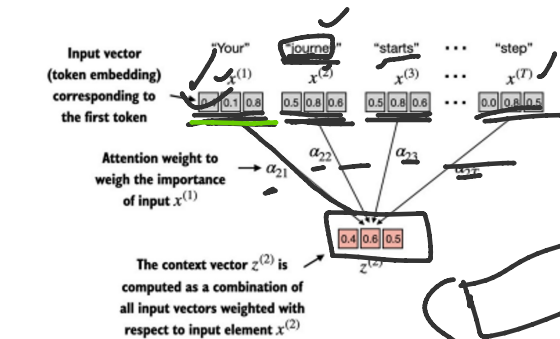


Figure 3.12

- 1) Compute attention scores
Compute the attention scores as dot products between the inputs
- 2) Compute attention weights
The attention weights are a normalized version of the attention scores
- 3) Compute context vectors
The context vectors are computed as a weighted sum over the inputs

Self Attention with Trainable parameters

