Transformers & Large Language Models (LLMs)

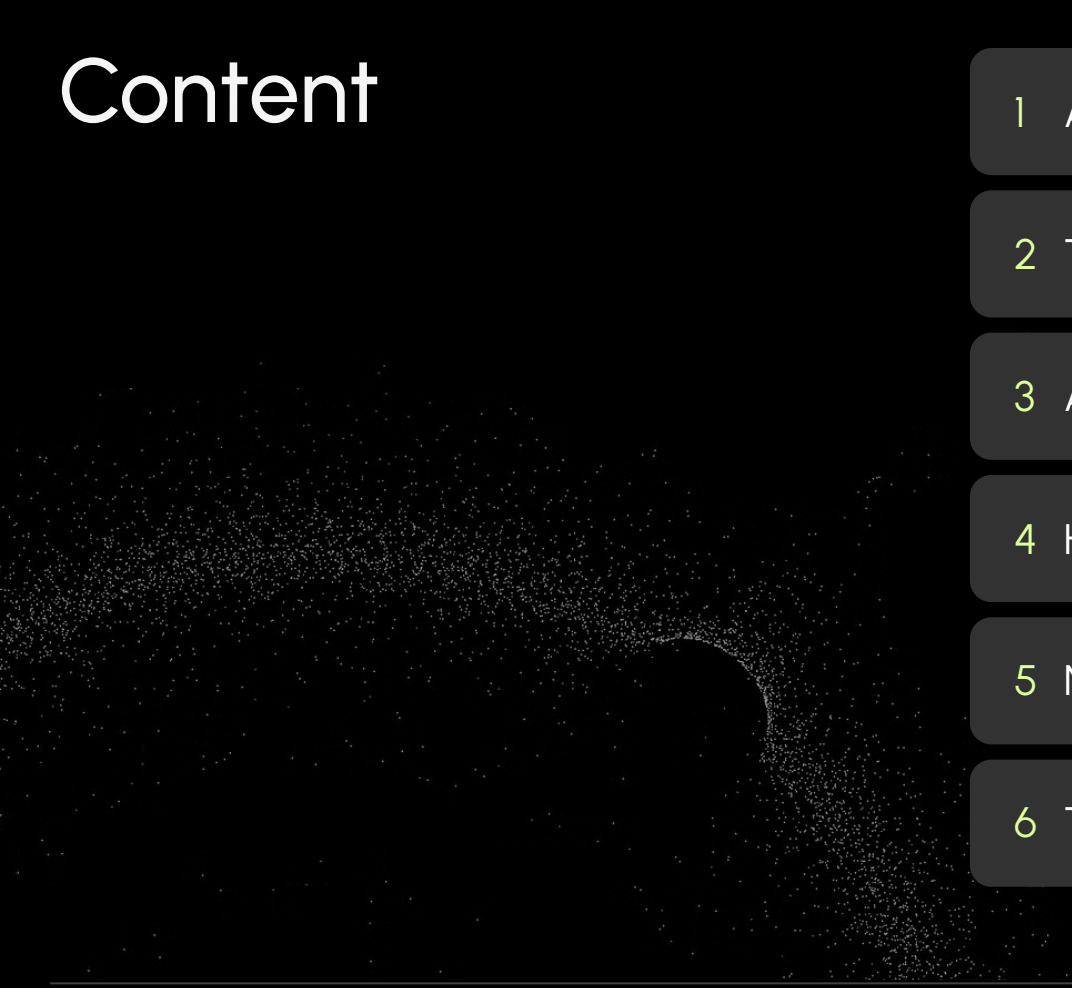
From Concepts to Practice - ML Course Day 4

Presented by

Mahamad Salah

Anton Popov

Content



1 Attention is All You Need!

2 The Anatomy of a Transformer

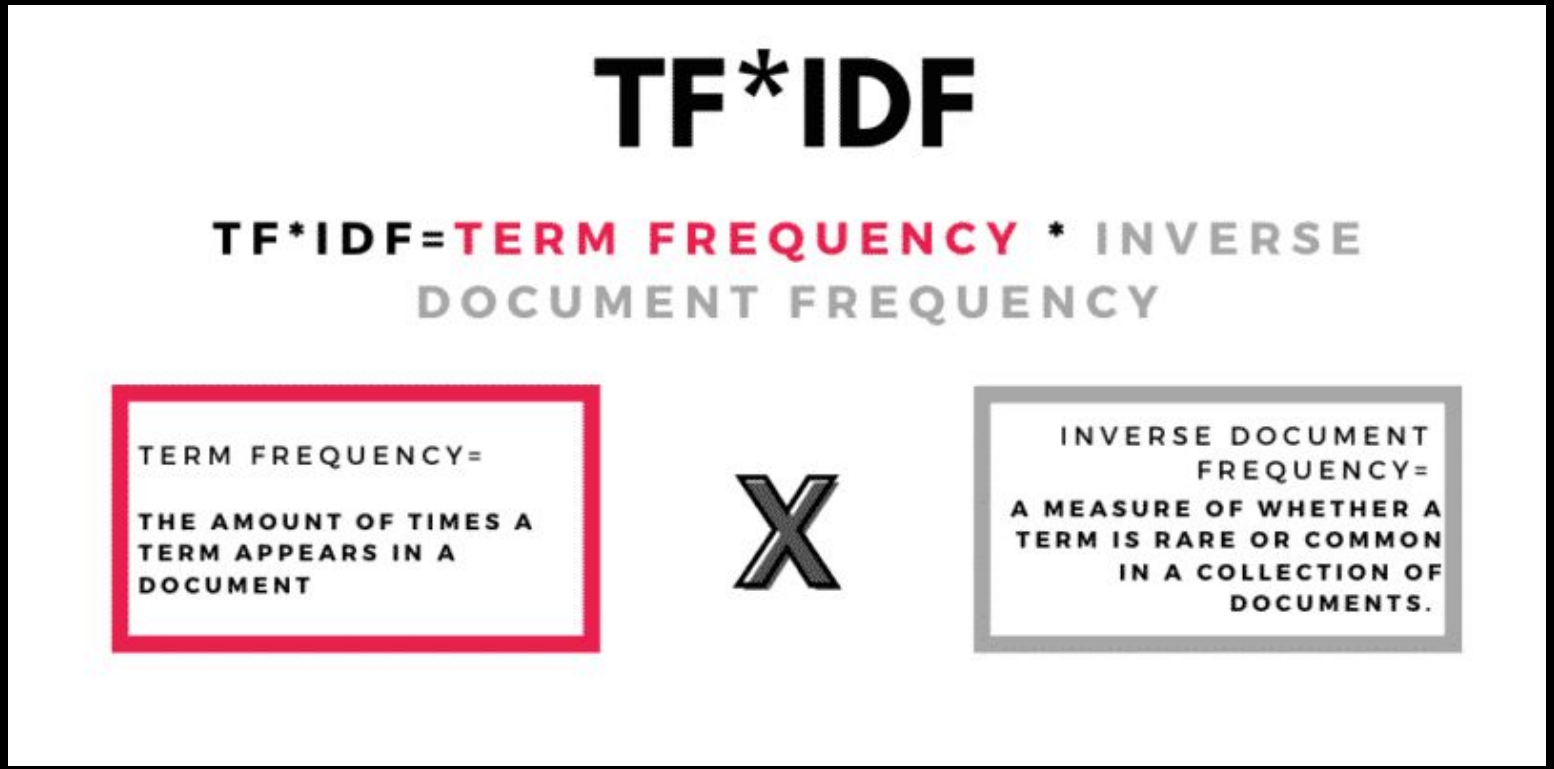
3 An Intro to LLMs

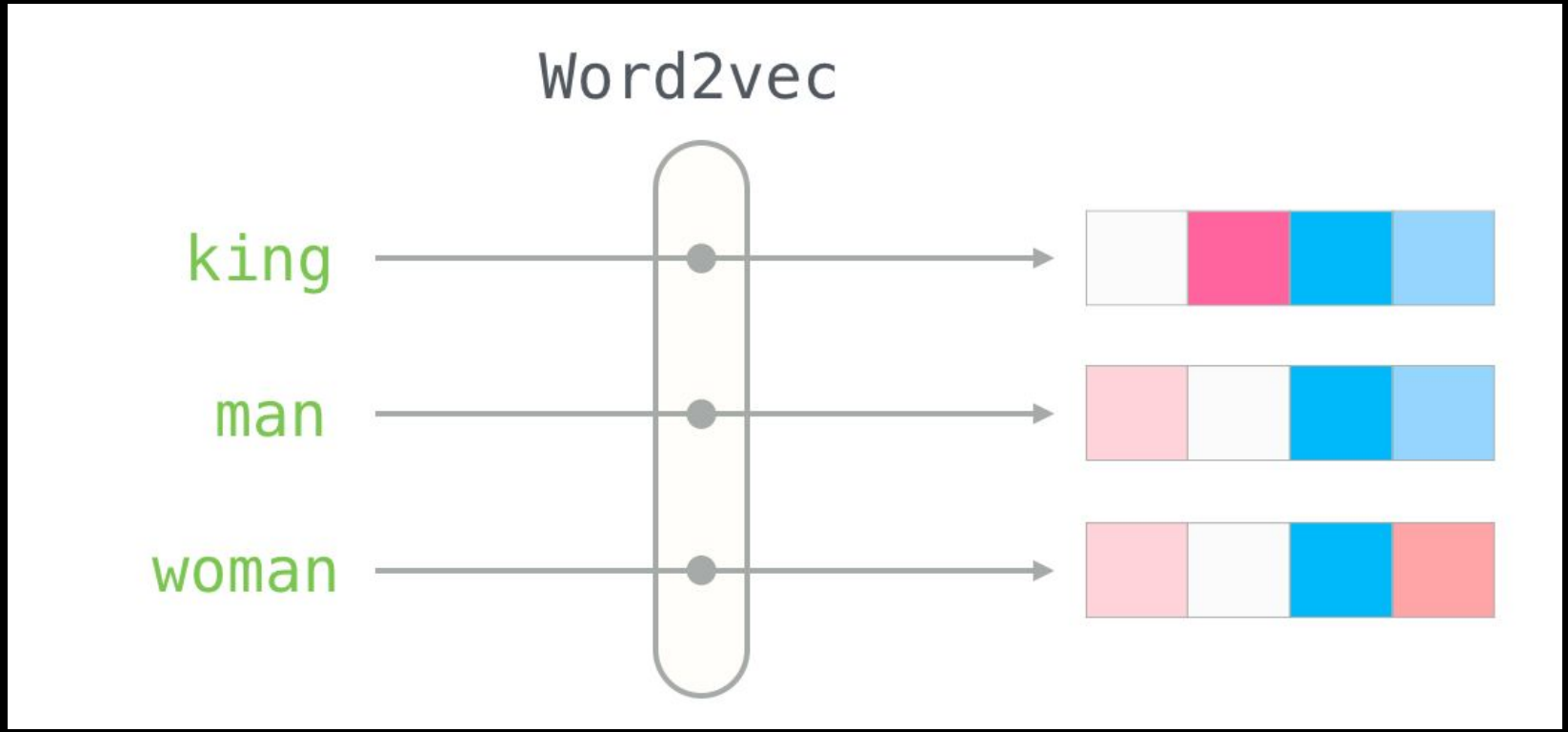
4 HuggingFace & Open Source Models

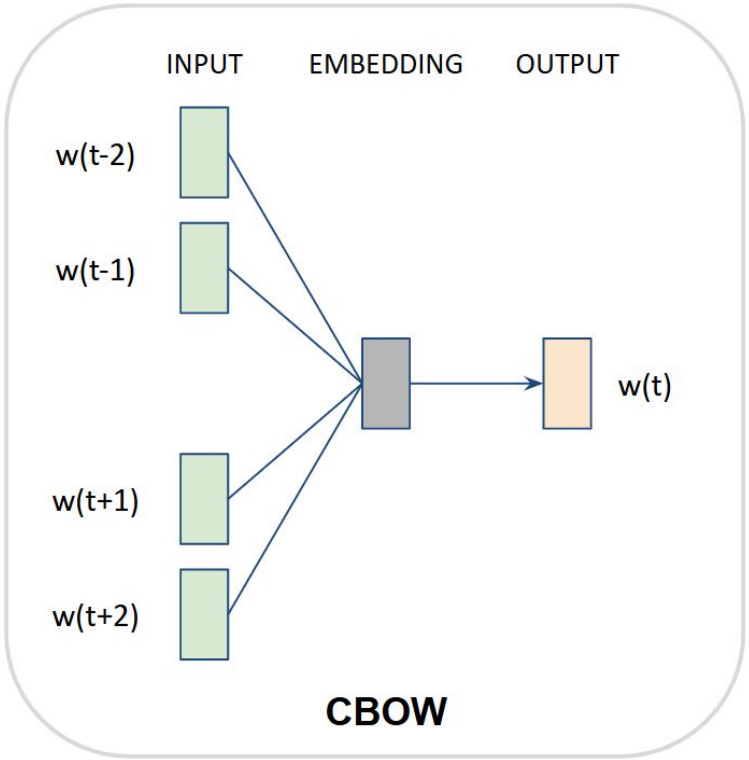
5 NLP Problems: A Demo

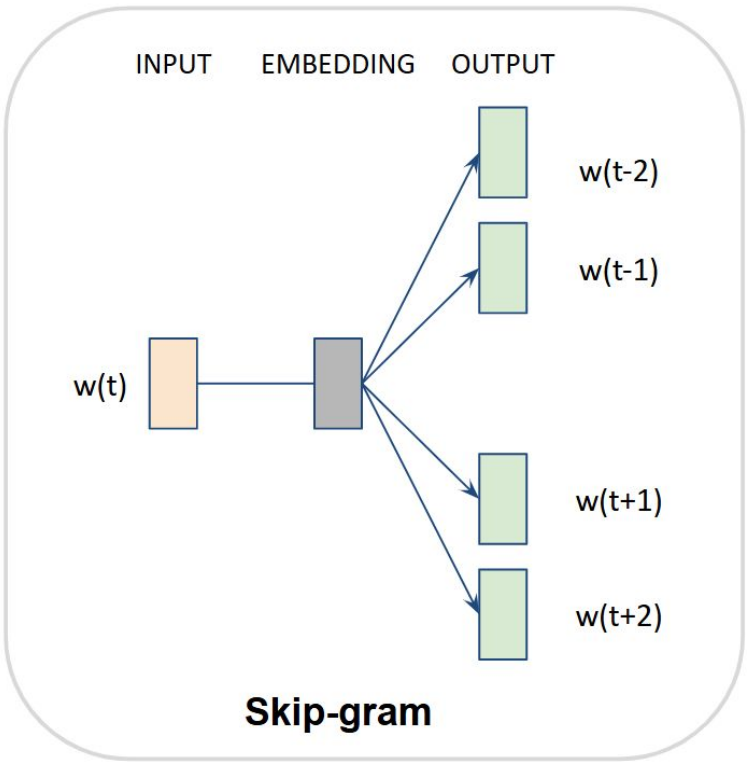
6 Transformers Beyond Text

Attention is All You Need


TF*IDF
TF*IDF=TERM FREQUENCY * INVERSE DOCUMENT FREQUENCY
TERM FREQUENCY=
THE AMOUNT OF TIMES A TERM APPEARS IN A DOCUMENT
X
INVERSE DOCUMENT FREQUENCY=
A MEASURE OF WHETHER A TERM IS RARE OR COMMON IN A COLLECTION OF DOCUMENTS.


Word2vec
king
man
woman

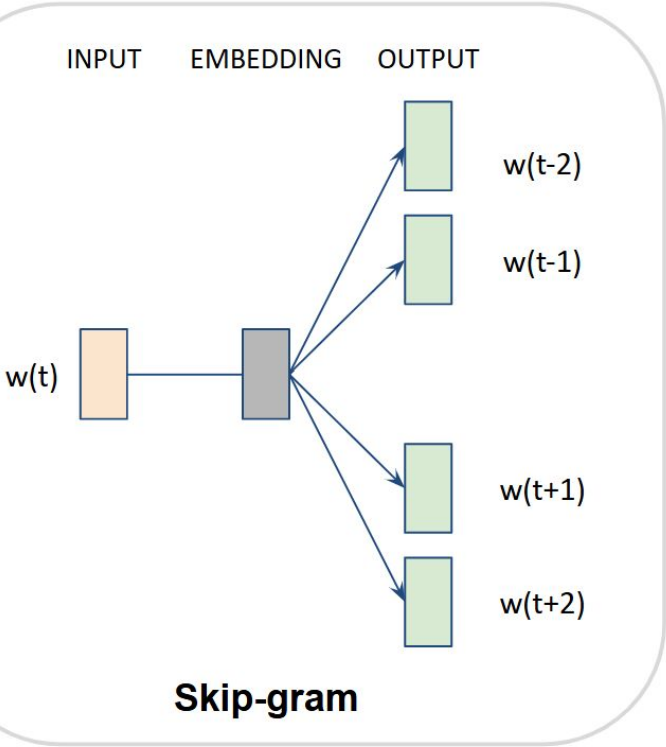

INPUT
EMBEDDING
OUTPUT
w(t-2)
w(t-1)
w(t+1)
w(t+2)
w(t)
CBOW


INPUT
EMBEDDING
OUTPUT
w(t)
w(t-2)
w(t-1)
w(t+1)
w(t+2)
Skip-gram

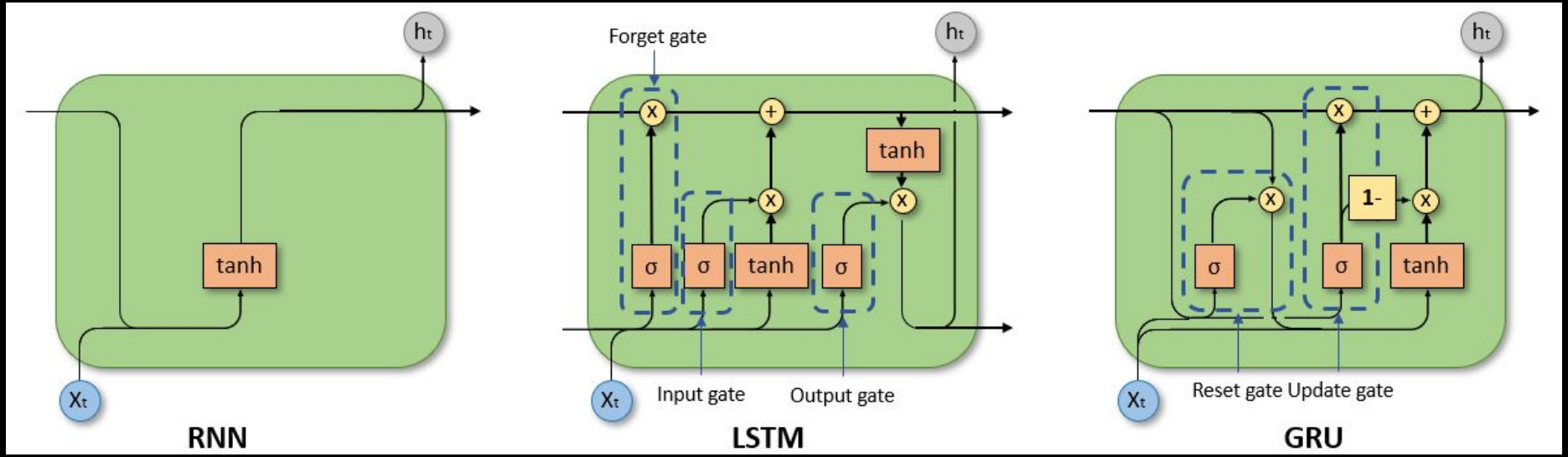
Cosine similarity:

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|}$$

EN2 EN1 Entrez:2020 Entrez:2020	$\cos\theta = 0.99$
Nausea Headache MESH:D009325 MESH:D006261	$\cos\theta = 0.976$
Cyclophosphamide Neutropenia MESH:D003520 MESH:D009503	$\cos\theta = 0.819$



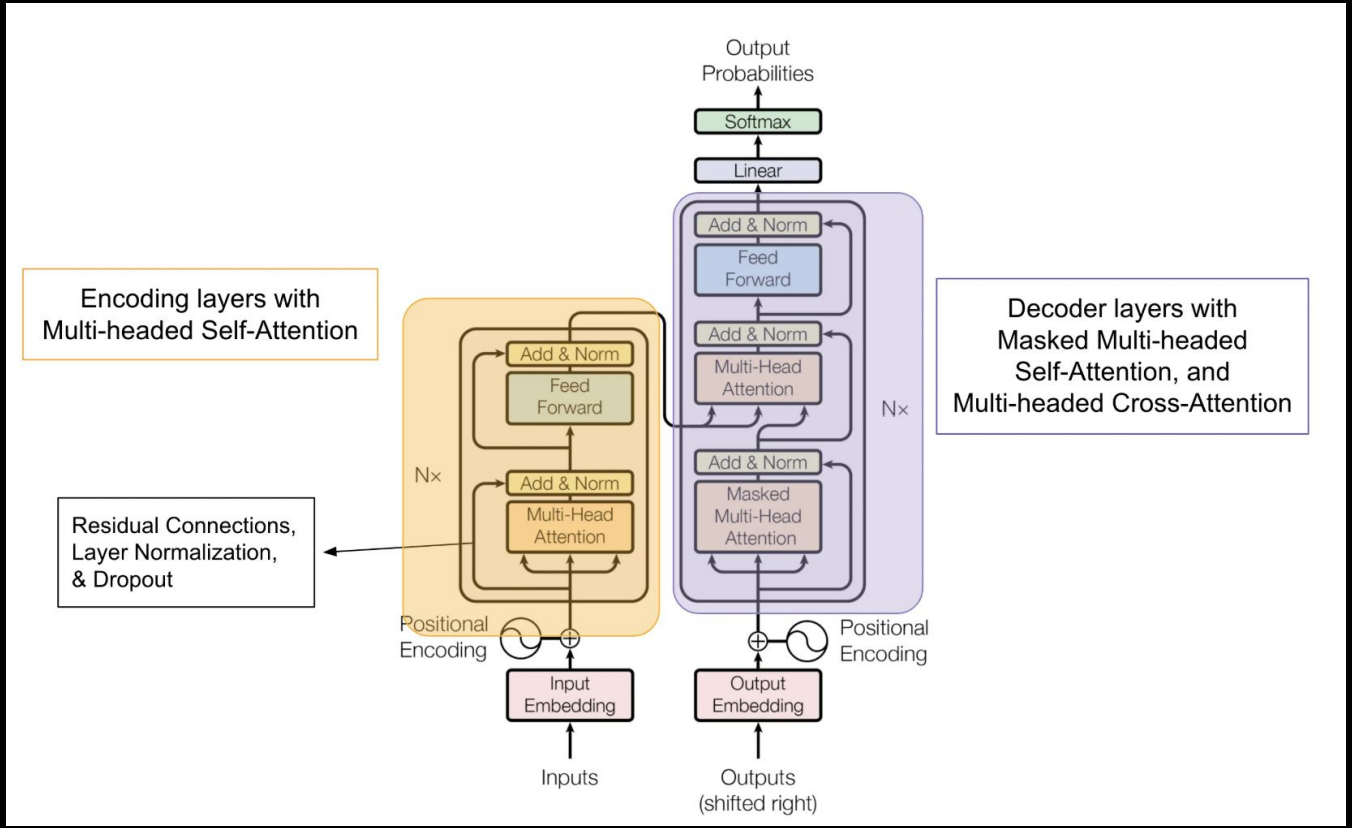
Attention is All You Need



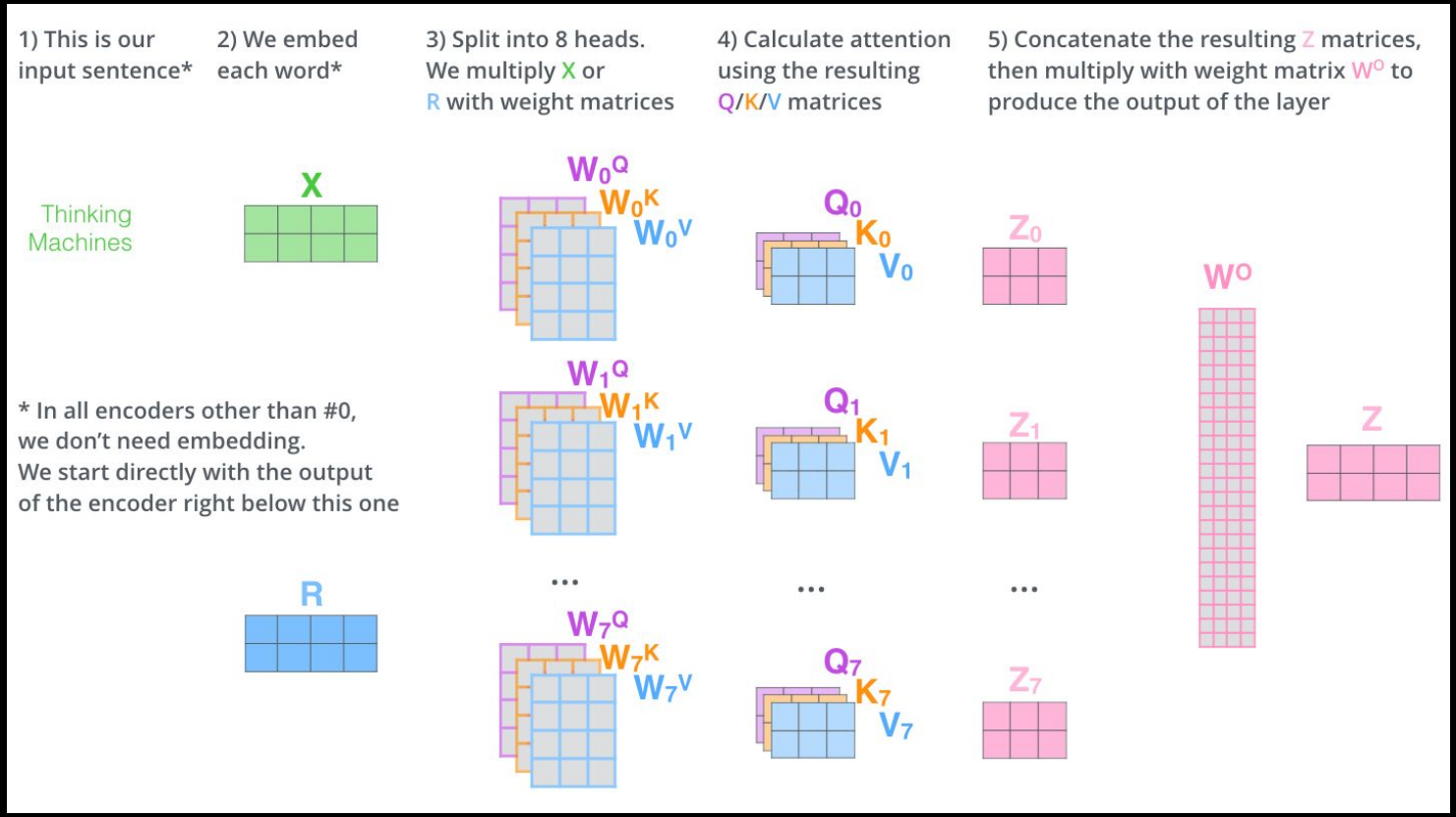
Major Challenges:

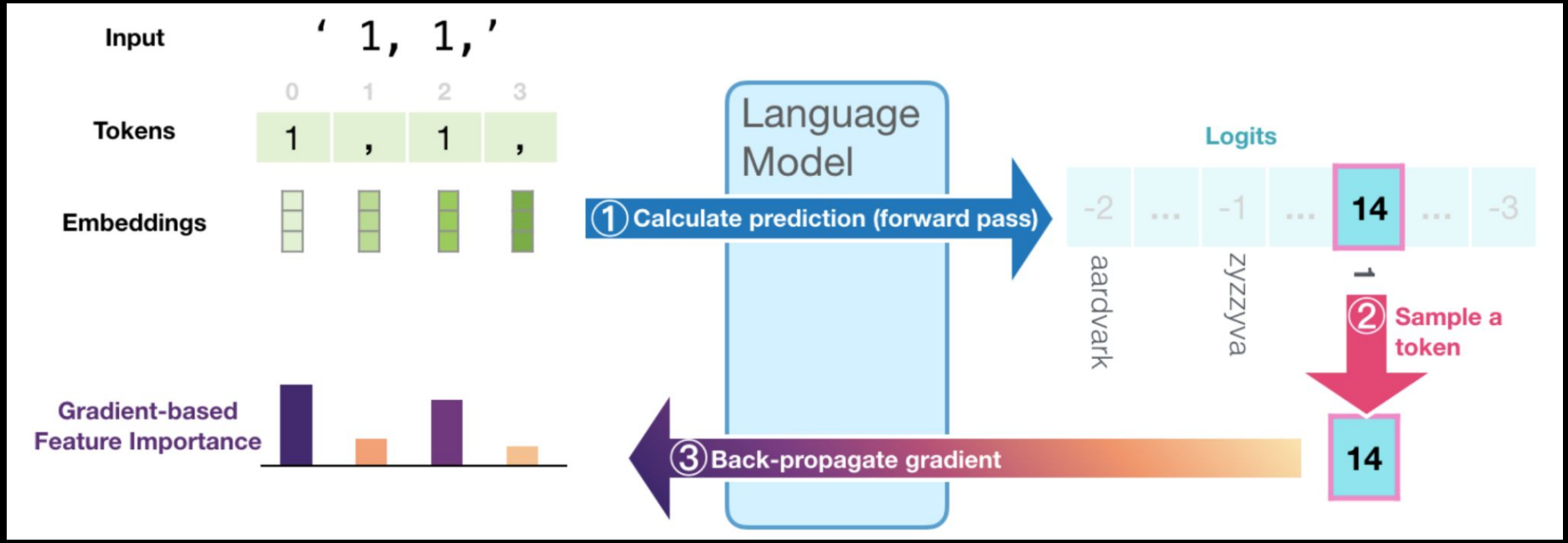
- Context Dependence
- Order Awareness
- Long-range dependencies
- Vanishing gradients

The Anatomy of a Transformer



The Anatomy of a Transformer




Input
' 1, 1,'
0
1
2
3
Tokens
1
,
1
,
Embeddings
Language Model
① Calculate prediction (forward pass)
Logits
-2
...
-1
...
14
...
-3
aardvark
zyzzyva
1
② Sample a token
14
③ Back-propagate gradient
Gradient-based Feature Importance

Hugging Face

NLP Problems: A Demo

- Sentence Classification
- Named Entity Recognition
- Text Generation
- Questions & Answers

Beyond Text: Vision Transformers