DATASCI W207
Applied Machine Learning

Week 13
Live Session Slides

Outline

- Introduction to Recurrent Neural Networks (RNNs)
    - Architecture of RNNs
    - Working of RNNs
    - Limitations of RNNs
- Introduction to Transformers
    - Architecture of Transformers
    - Working of Transformers
    - Types of Transformers

The cat sat on the _____.

a. mat
b. roof
c. cushion
d. chair

Introduction to Recurrent Neural Networks (RNNs)

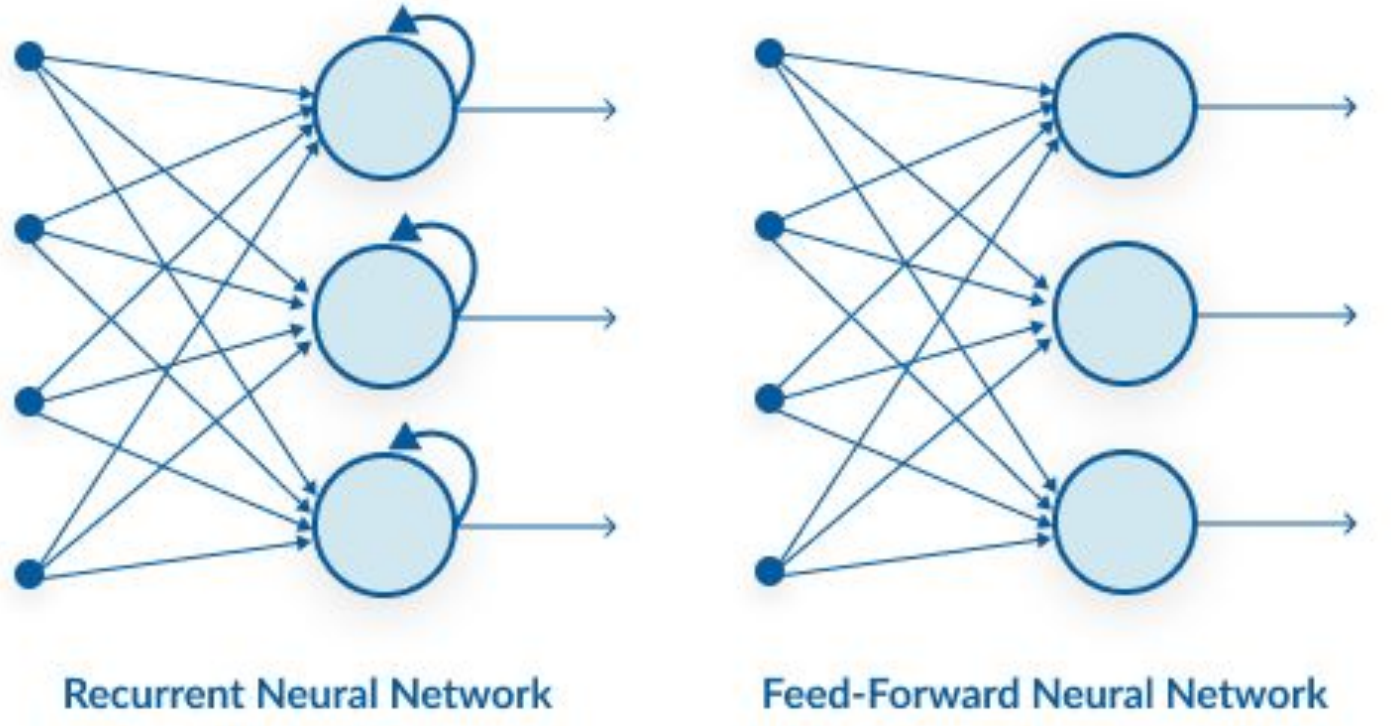
In the quiet evening, the family shared stories in their cozy living room. Outside, raindrops tapped gently on the windows, creating a soothing melody. Seeking shelter, the cat leapt onto the roof, finding comfort in the gentle rain. The cat sat on the _____ , a content observer of the world below.

a. mat
b. roof
c. cushion
d. chair

Introduction to Recurrent Neural Networks (RNNs) : Applications

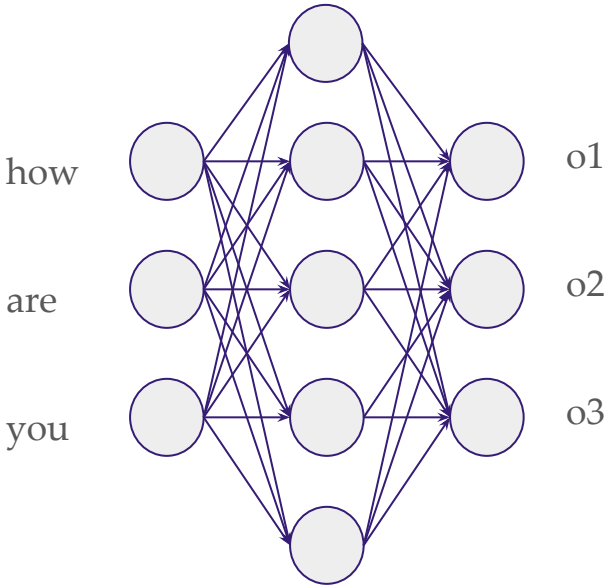
- Prediction Problems
  - Language Modelling and Generating Text
  - Music Composition
- Machine Translation
  - Speech Recognition
- Generating Image Descriptions
  - Named Entity Recognition
- Sentiment Analysis
- Image Recognition

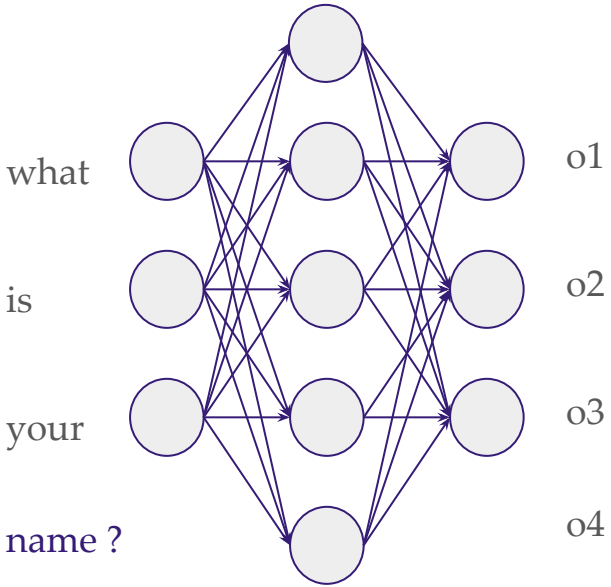
Architecture of RNNs



Architecture of RNNs: Why RNNs

- No fixed size of neurons in a layer
- Too much computation
- Parameters are not shared





Architecture of RNNs: Why RNNs

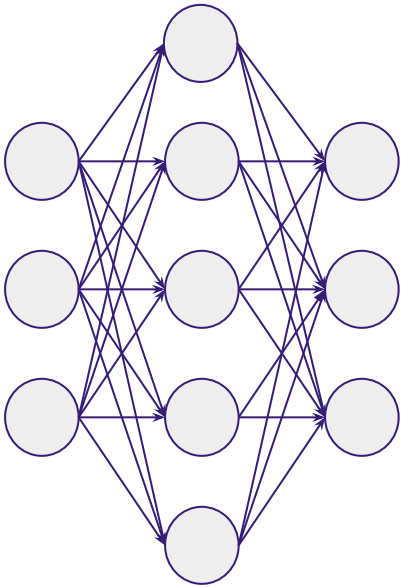
- No fixed size of neurons in a layer
- Too much computation
- Parameters are not shared

25000 words in a vocabulary

how [0, 0, 0, .., 1, 0, 0,...,0]

are [0, 1, 0, 0, 0 ,..,0, 0,..0]

you [0, 0, 0, 0, 0 ,..,0, 1,..0]



42000 words in a vocabulary

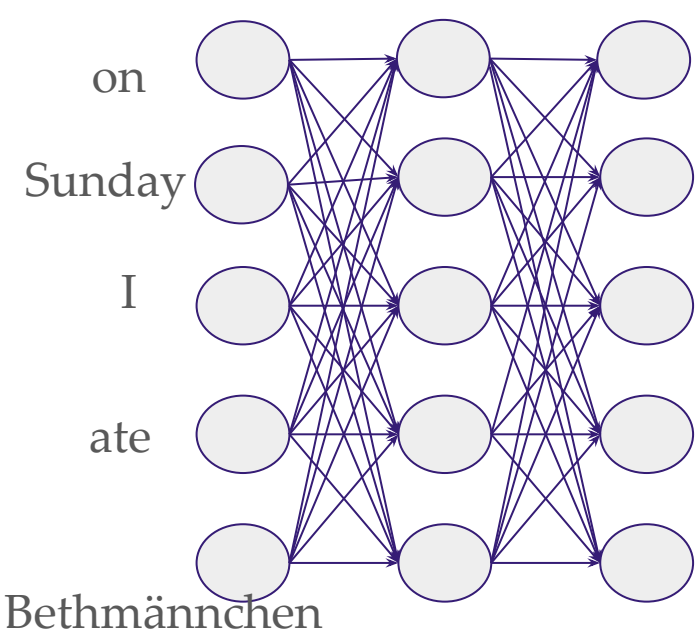
[0, 0, 0, .., 1, 0, 0,...,0]

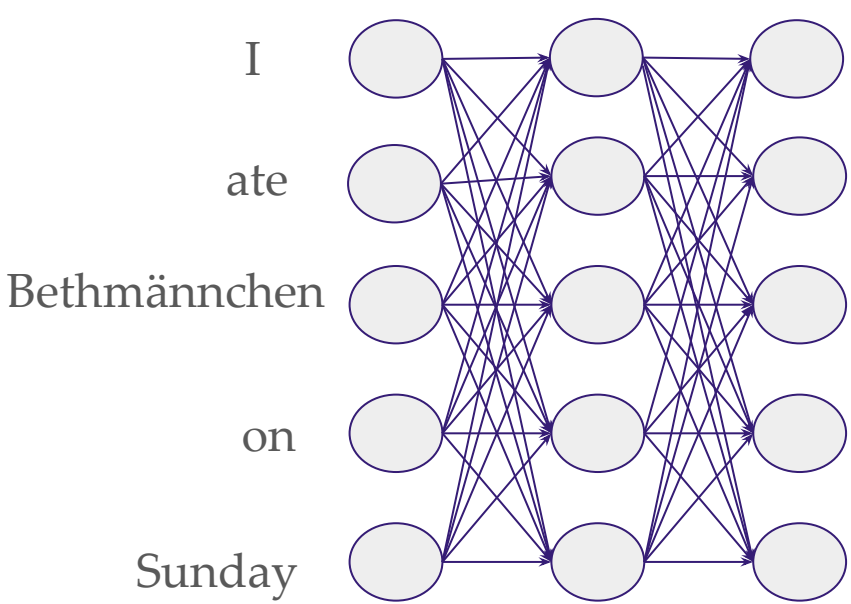
[0, 0, 0, .., 1, 0, 0,...,0]

[0, 0, 0, .., 1, 0, 0,...,0]

Architecture of RNNs: Why RNNs

- No fixed size of neurons in a layer
- Too much computation
- Parameters are not shared





Working of RNNs: Why RNNs

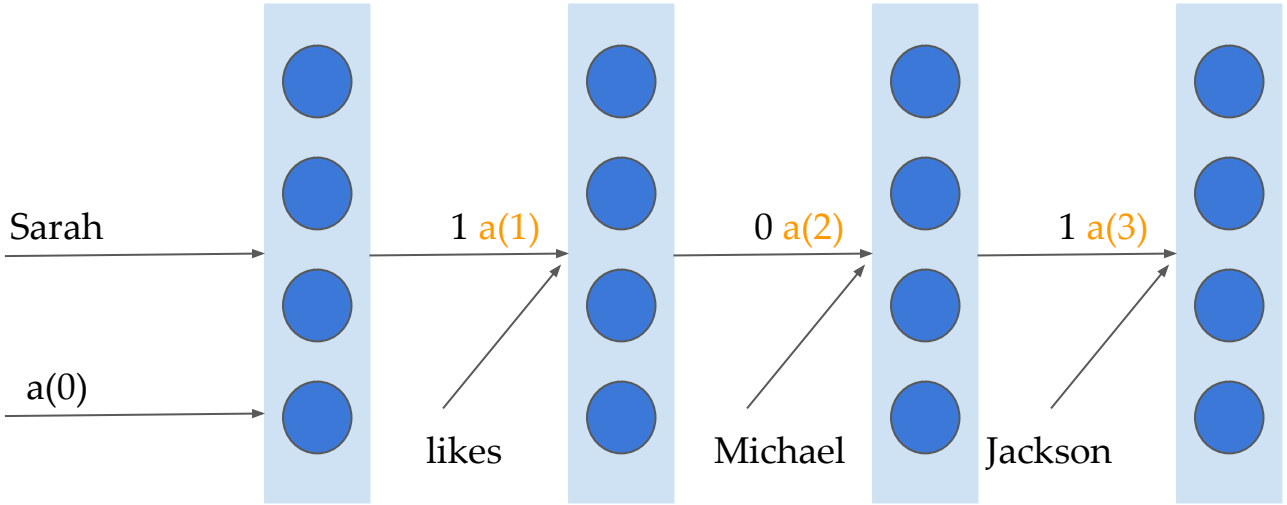
Sarah likes Michael Jackson

person person

Sarah likes Michael Jackson

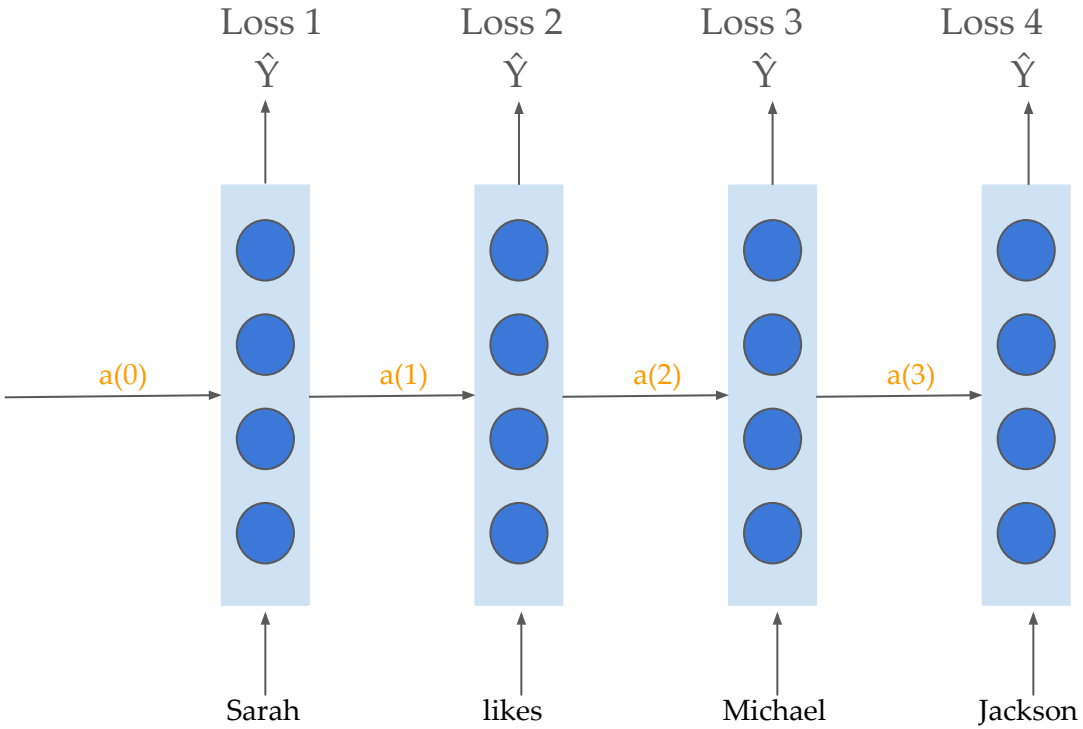
1 0 1 1

Working of RNNs



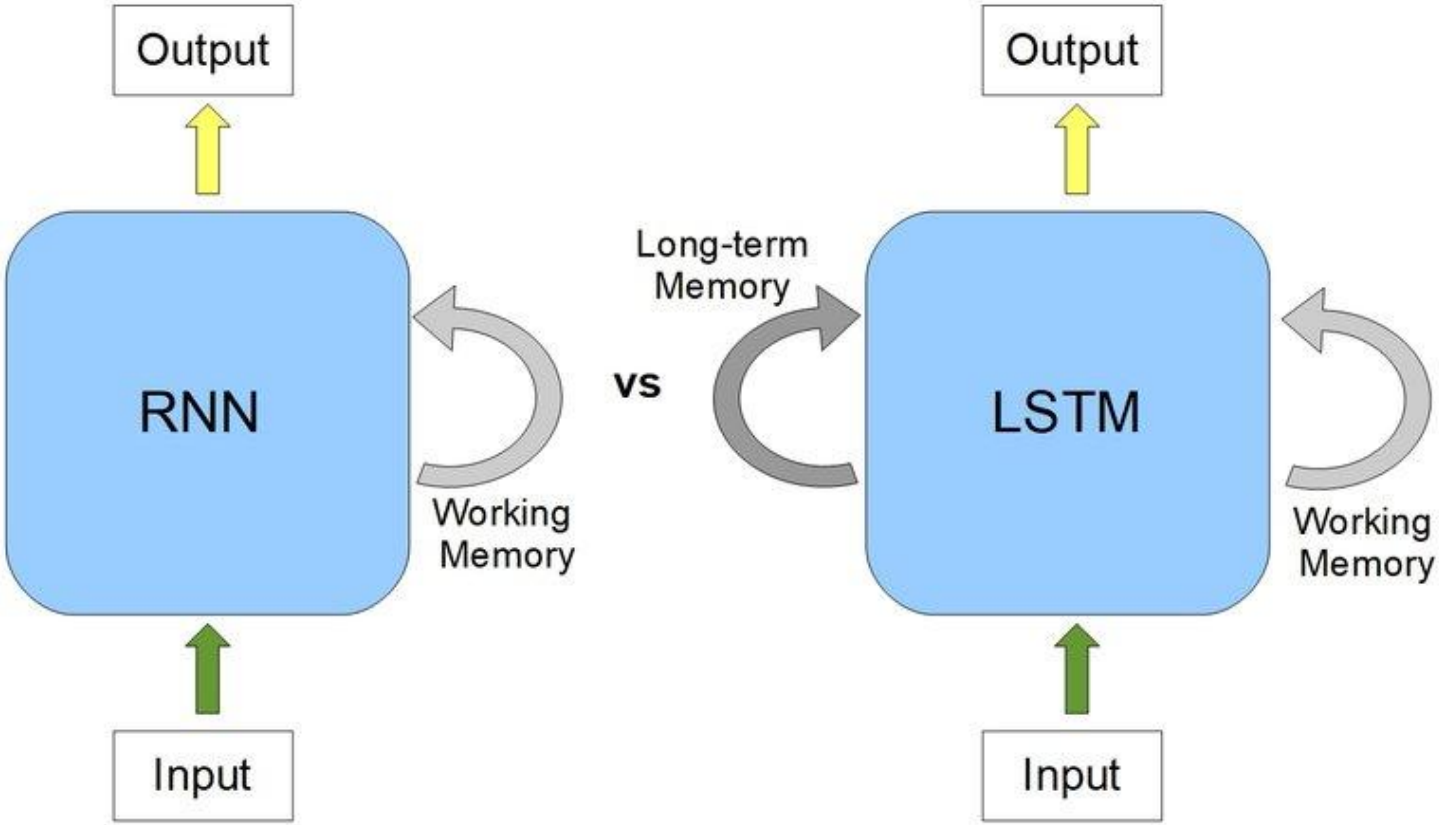
Working of RNNs

Sarah likes Michael Jackson **1 0 1 1**

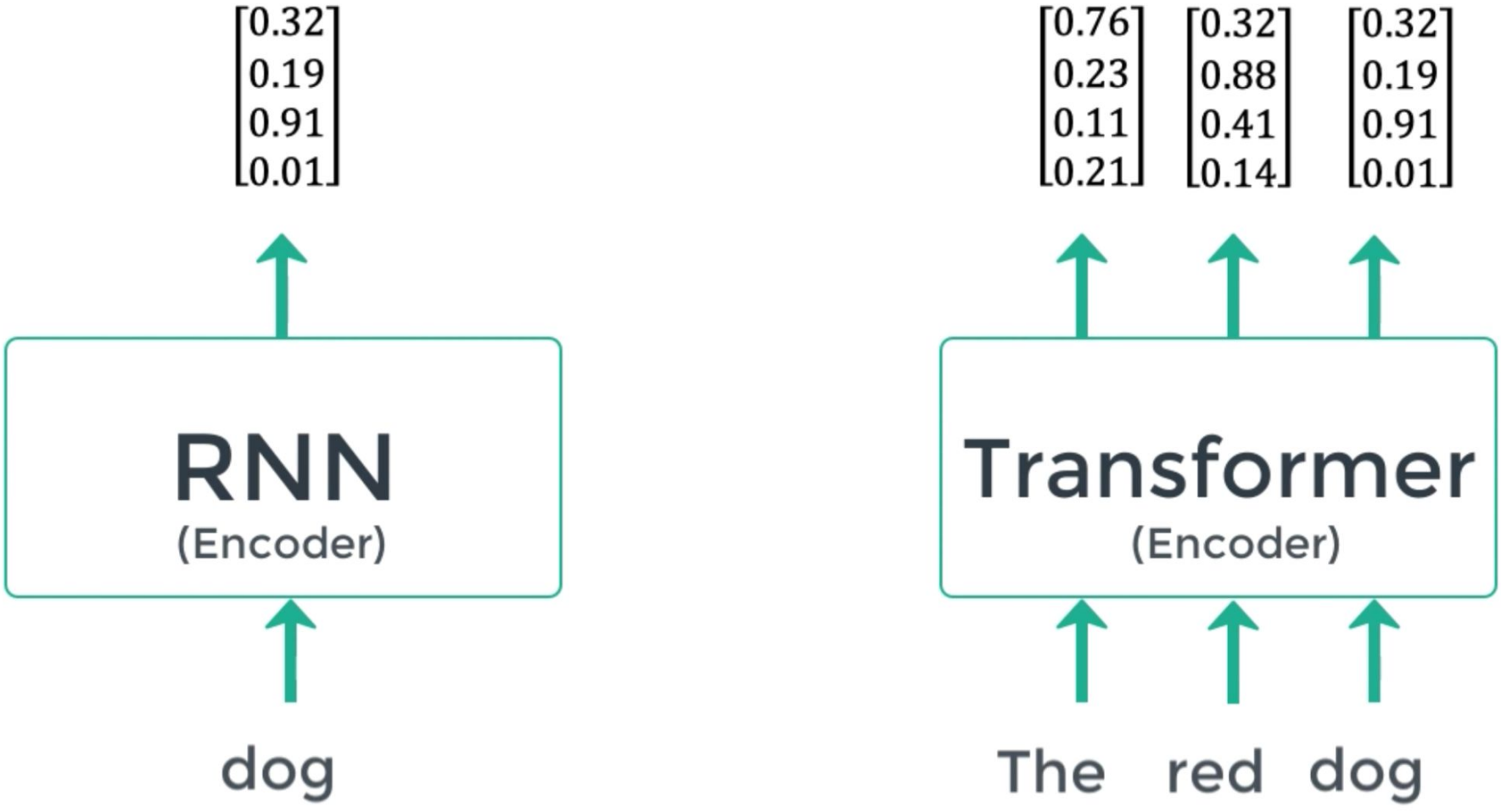


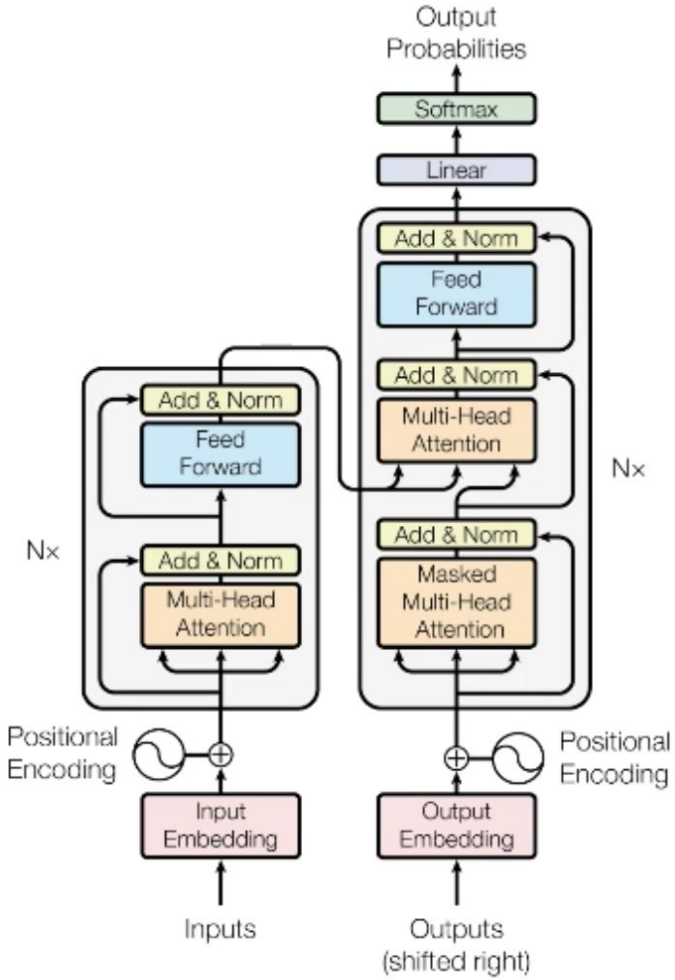
Total Loss = Loss 1 + Loss 2 + Loss 3 + Loss 4

Limitations RNNs



Transformers


0.32
0.19
0.91
0.01
RNN
(Encoder)
dog
0.76
0.23
0.11
0.21
0.32
0.88
0.41
0.14
0.32
0.19
0.91
0.01
Transformer
(Encoder)
The
red
dog


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)

Transformer Components

Input Embedding

dog →

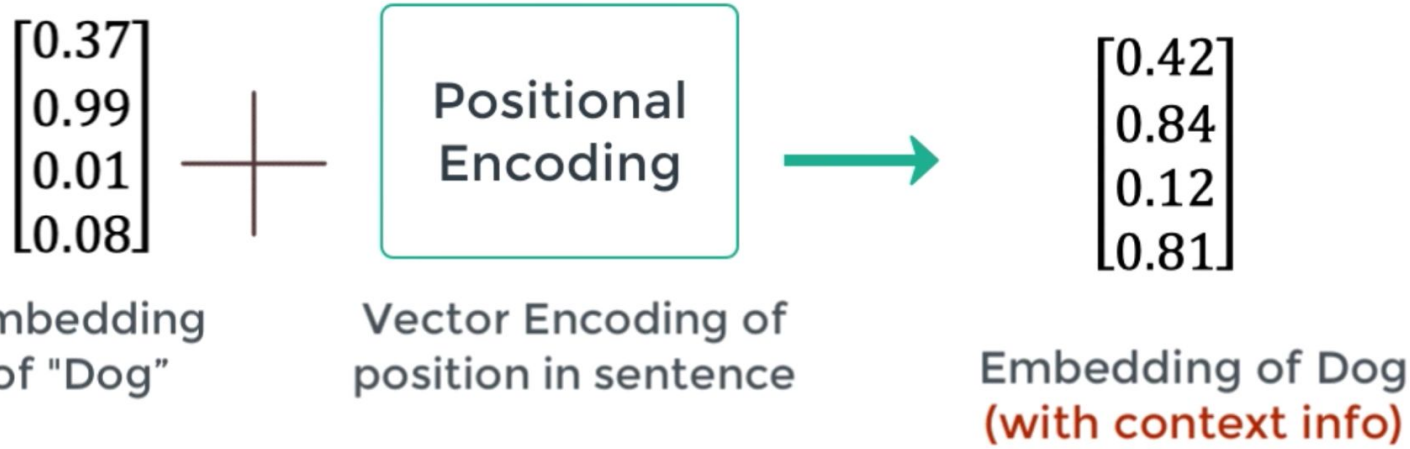
→ $\begin{bmatrix} 0.37 \\ 0.99 \\ 0.01 \\ 0.08 \end{bmatrix}$

AJ's dog is a cutie

AJ looks like a dog

Transformer Components

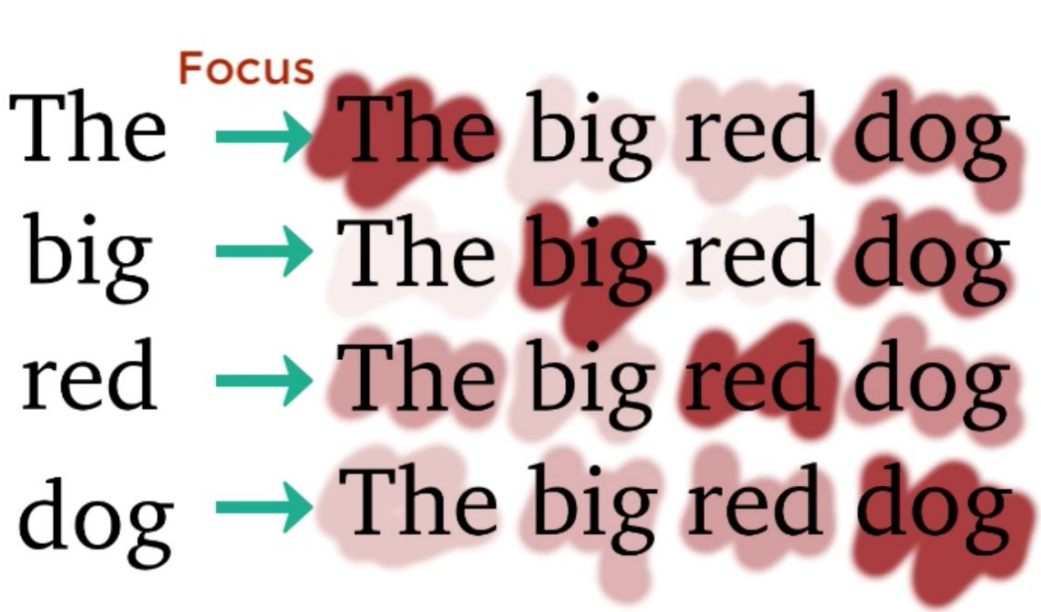
Positional Encoder :vector that gives context based on position of word in sentence



$$PE_{(pos,2i)} = sin(pos/10000^{2i/d_{\text{model}}})$$
$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d_{\text{model}}})$$

Transformer Components

Attention : What part of the input should we focus?



Attention Vectors

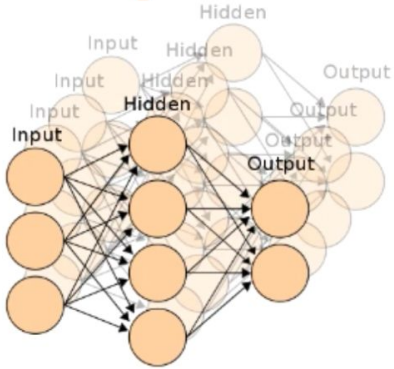
$[0.71 \quad 0.04 \quad 0.07 \quad 0.18]^T$

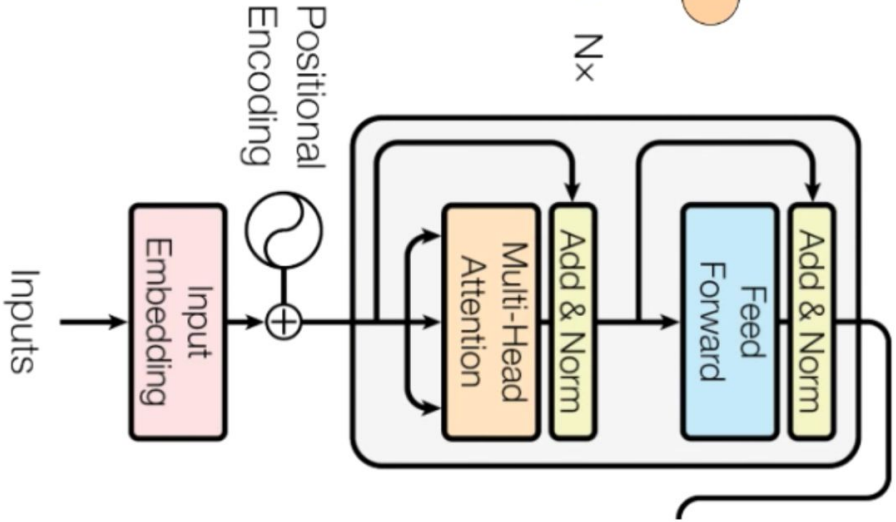
$[0.01 \quad 0.84 \quad 0.02 \quad 0.13]^T$

$[0.09 \quad 0.05 \quad 0.62 \quad 0.24]^T$

$[0.03 \quad 0.03 \quad 0.03 \quad 0.91]^T$

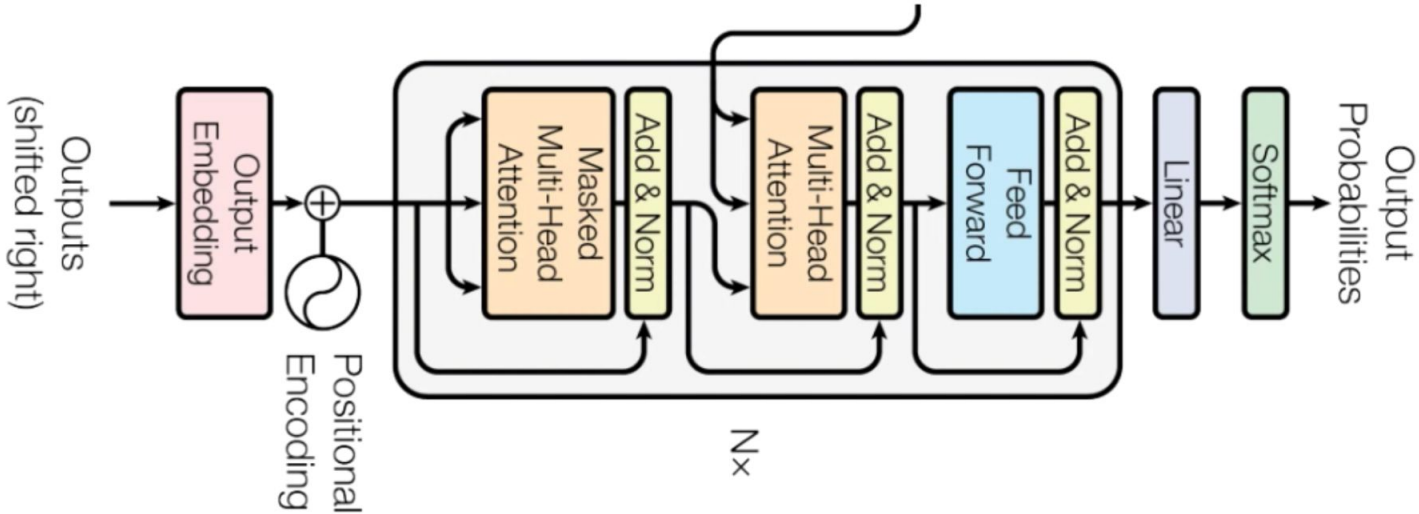
Transformer Components





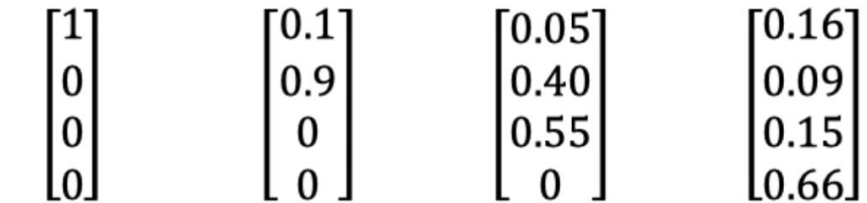
Transformer Components

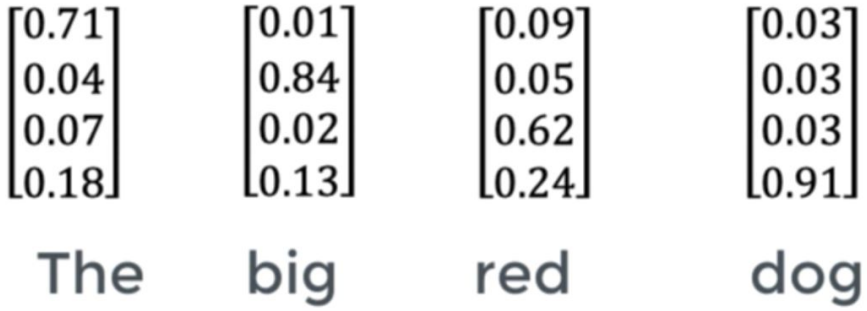
Decoder



Transformer Components

Decoder



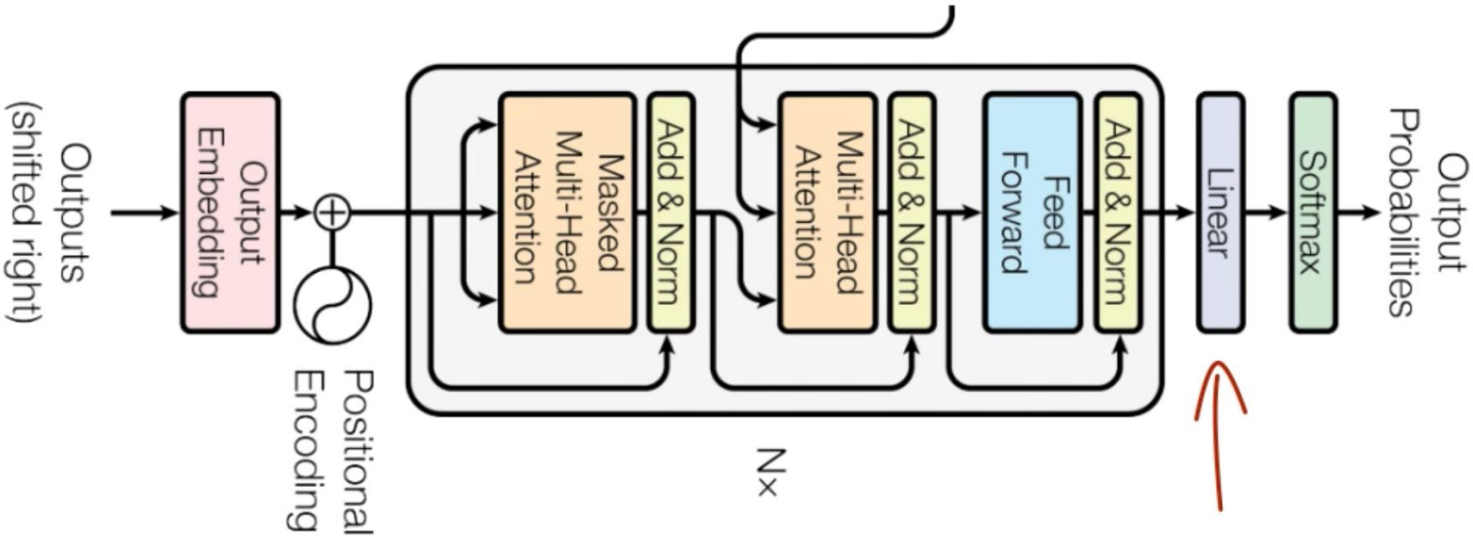


Encapsulates English-French Interactions

Encoder-Decoder Attention

Transformer Components

Le



Feed Forward Layer

Neurons =
words in French