Mila

12-06-2025

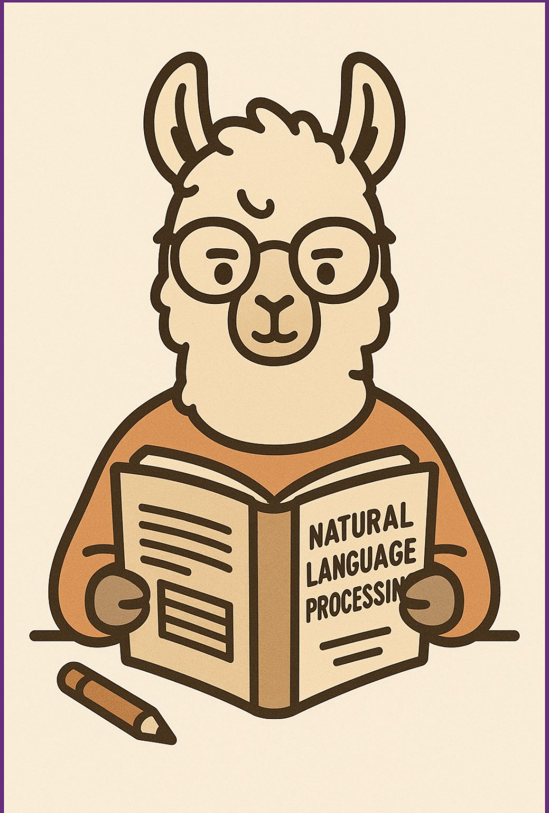


Image generated using Sora

Introduction to Natural Language Processing

Part 2

Prakhar Ganesh

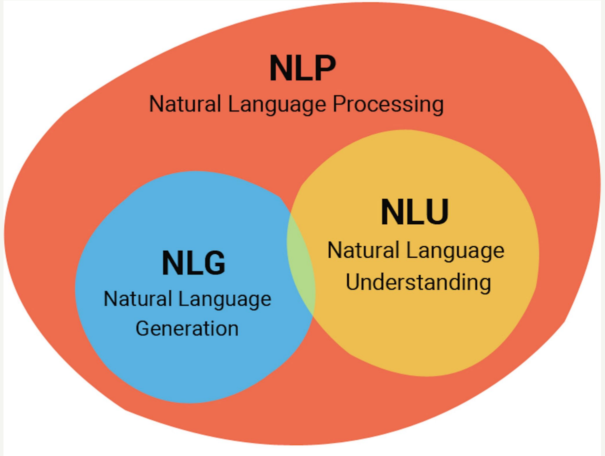
A quick recap …

A quick recap ...

- NLP, NLU and NLG

A quick recap ...

- NLP, NLU and NLG

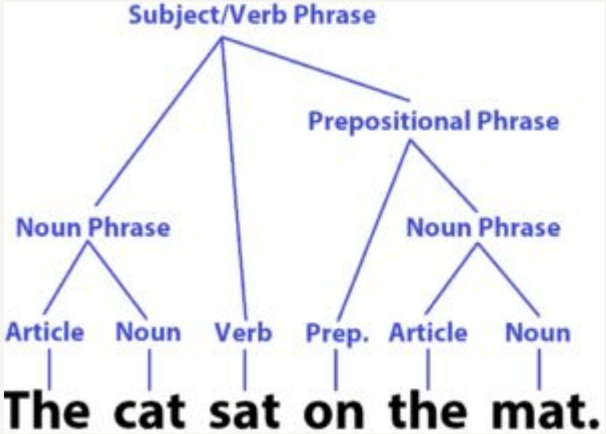


A quick recap ...

- NLP, NLU and NLG
- Syntax and Parsing

A quick recap …

- NLP, NLU and NLG
- Syntax and Parsing



Mila

A quick recap …

- NLP, NLU and NLG
- Syntax and Parsing
- Semantics and Pragmatics

A quick recap ...

- NLP, NLU and NLG
- Syntax and Parsing
- Semantics and Pragmatics

Word	Semantic
pen	a writing tool
pen	a livestock's enclosure
pen	a portable enclosure for a baby
pen	a correctional institution
pen	a female swan

A quick recap ...

- NLP, NLU and NLG
- Syntax and Parsing
- Semantics and Pragmatics
- Morphology and Tokenization

A quick recap ...

- NLP, NLU and NLG
- Syntax and Parsing
- Semantics and Pragmatics
- Morphology and Tokenization

A quick recap ...

- NLP, NLU and NLG
- Syntax and Parsing
- Semantics and Pragmatics
- Morphology and Tokenization
- Types of NLP Applications

A quick recap ...

- NLP, NLU and NLG
- Syntax and Parsing
- Semantics and Pragmatics
- Morphology and Tokenization
- Types of NLP Applications

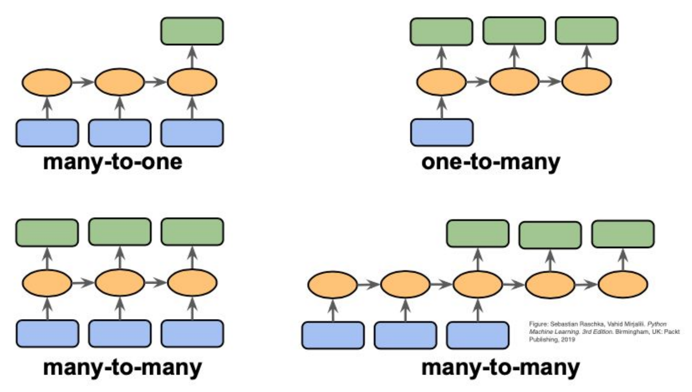


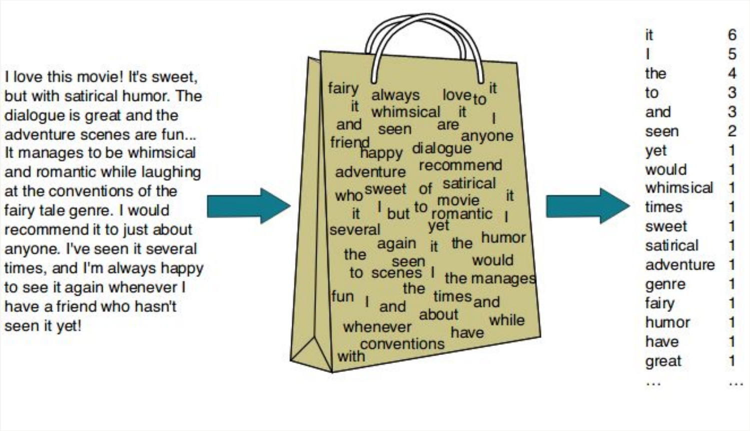
Figure: Sebastian Raschka, Vahid Mirjalili. *Python Machine Learning. 3rd Edition*. Birmingham, UK: Packt Publishing, 2019

A quick recap …

- Bag of Words

A quick recap …

- Bag of Words



Mila

A quick recap ...

- Bag of Words
- Bag of n-grams

A quick recap ...

- Bag of Words
- Bag of n-grams

The cat sat on the mat.

Bag of Words

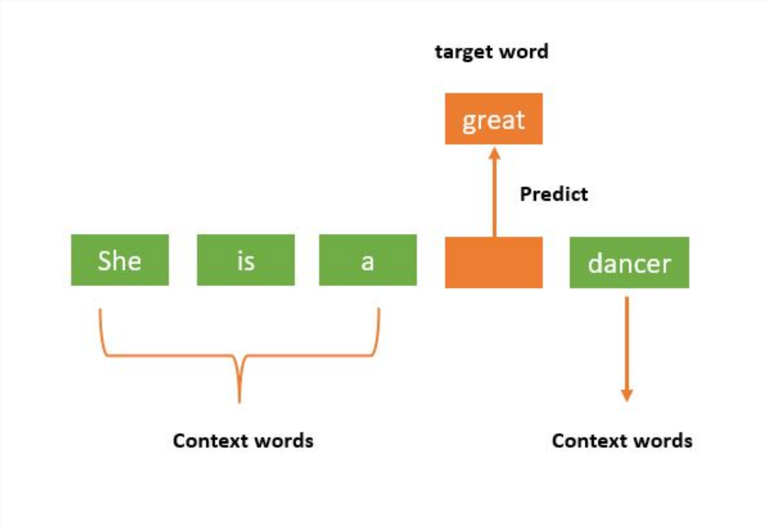
Bag of 2-grams

A quick recap …

- Bag of Words
- Bag of n-grams
- Continuous Bag of Words

A quick recap …

- Bag of Words
- Bag of n-grams
- Continuous Bag of Words

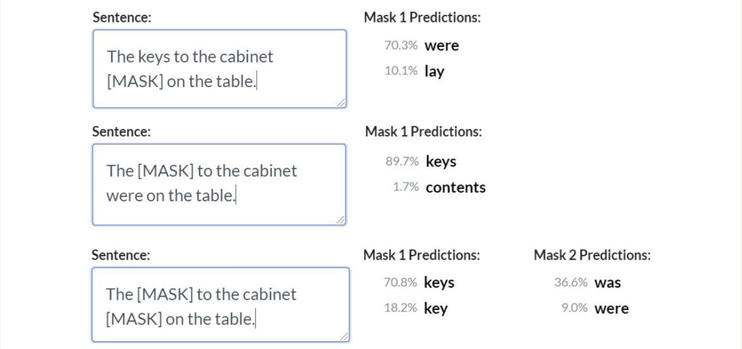


A quick recap …

- Bag of Words
- Bag of n-grams
- Continuous Bag of Words
- Masked Language Modeling

A quick recap ...

- Bag of Words
- Bag of n-grams
- Continuous Bag of Words
- Masked Language Modeling

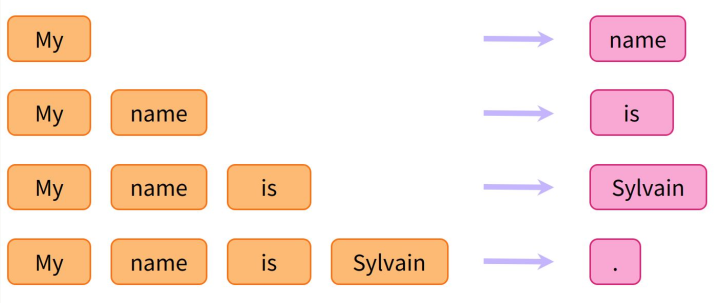


A quick recap ...

- Bag of Words
- Bag of n-grams
- Continuous Bag of Words
- Masked Language Modeling
- Causal Language Modeling

A quick recap ...

- Bag of Words
- Bag of n-grams
- Continuous Bag of Words
- Masked Language Modeling
- Causal Language Modeling



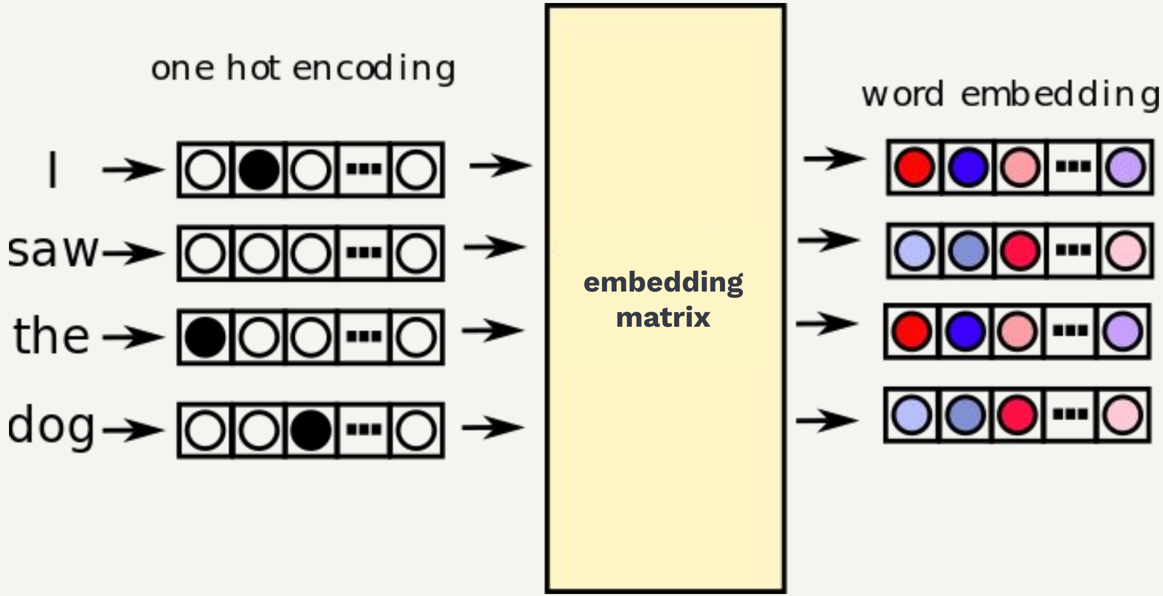
Mila

A quick recap …

- Embeddings

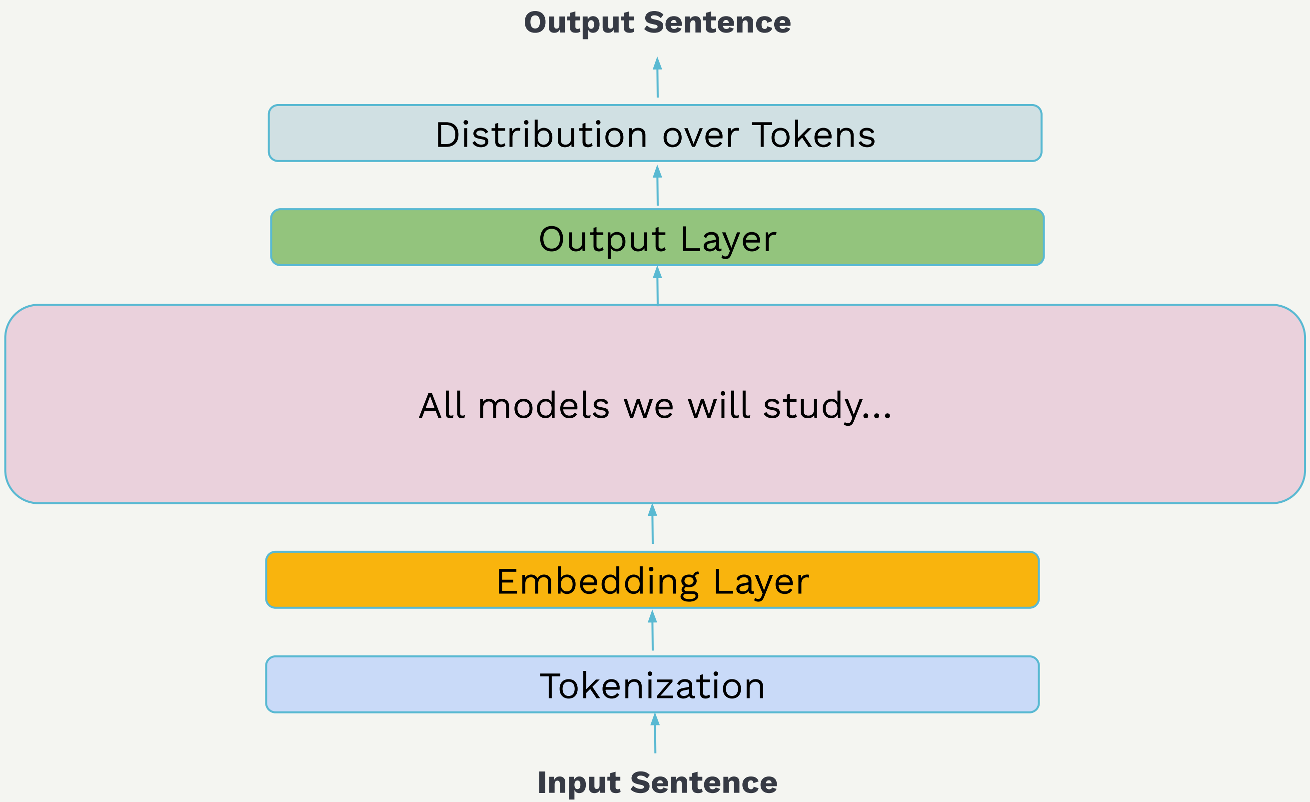
A quick recap ...

- Embeddings



Mila

NLP Pipeline



Any questions from previous sessions?

Goals today...

- Convolutional Neural Networks (CNNs)
- Recurrent Neural Networks (RNNs)
- Long Short-term Memory Networks (LSTMs)
- Attention
- Self-Attention and Transformers

Positional Equivariance and Positional Awareness

Positional Equivariance

Positional Equivariance

“This is awful.”

Positional Equivariance

"This is awful."

"The customer service was really a challenge to deal with. Honestly, this is awful."

Positional Equivariance

“This is awful.”

“The customer service was really a challenge to deal with. Honestly, this is awful.”

“This is awful. The customer service was really a challenge to deal with.”

Positional Equivariance

“This is awful.”

“The customer service was really a challenge to deal with. Honestly, this is awful.”

“This is awful. The customer service was really a challenge to deal with.”

“I was promised a delivery yesterday, but nothing arrived. I called customer support, got transferred four times, and each person gave me a different explanation. Now I’ve wasted my entire afternoon trying to fix something that wasn’t my fault — this is awful.”

Positional Equivariance

*“**This is awful.**”*

*“The customer service was really a challenge to deal with. Honestly, **this is awful.**”*

*“**This is awful**. The customer service was really a challenge to deal with.”*

*“I was promised a delivery yesterday, but nothing arrived. I called customer support, got transferred four times, and each person gave me a different explanation. Now I’ve wasted my entire afternoon trying to fix something that wasn’t my fault — **this is awful**.”*

Processing Time Series Data

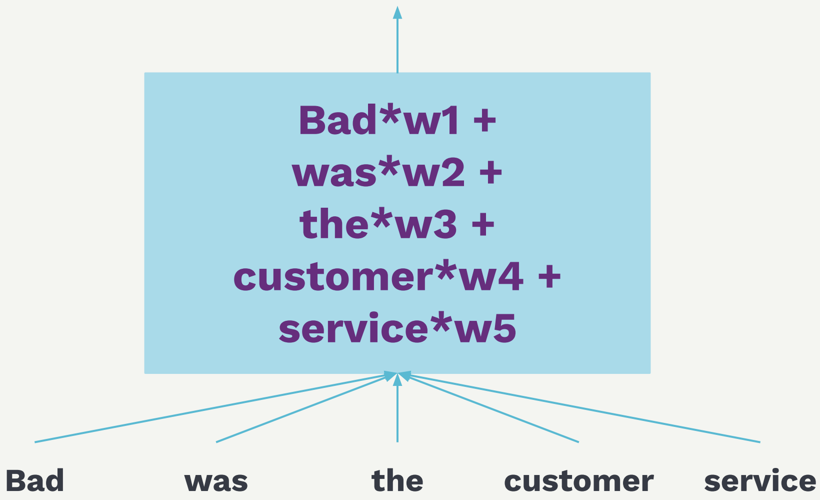
*To deal with language, we need **local** positional equivariance, i.e., they apply **the same function** regardless of position, but **global** positional awareness!*

Using MLPs?

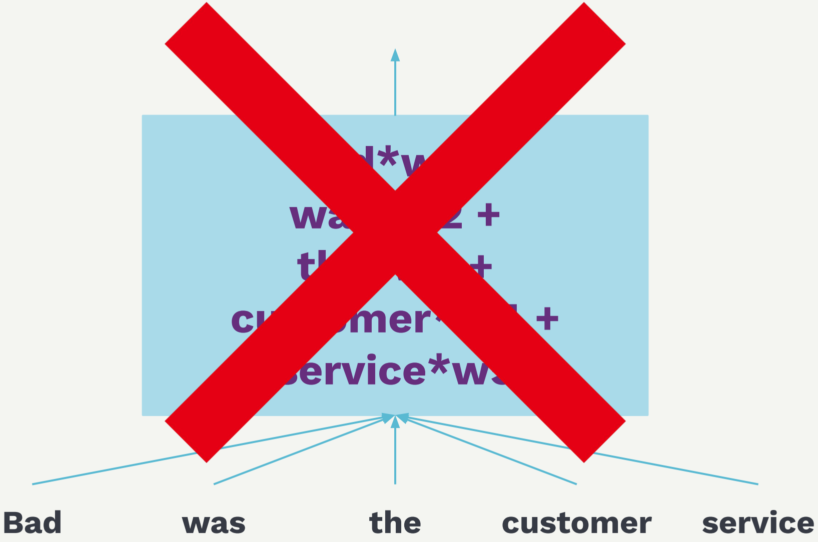
Using MLPs?


The*w1 +
customer*w2 +
service*w3 + was*w4
+
bad*w5
The
customer
service
was
bad

Using MLPs?


Bad*w1 +
was*w2 +
the*w3 +
customer*w4 +
service*w5
Bad
was
the
customer
service

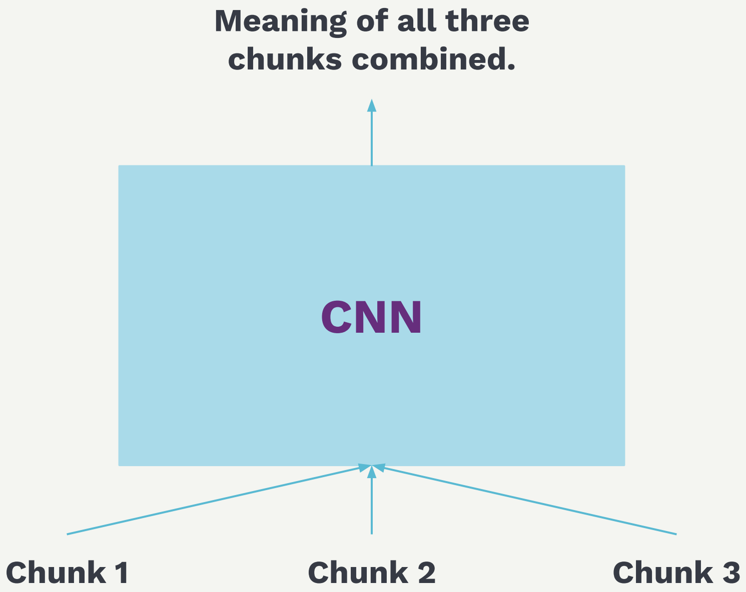
Using MLPs?



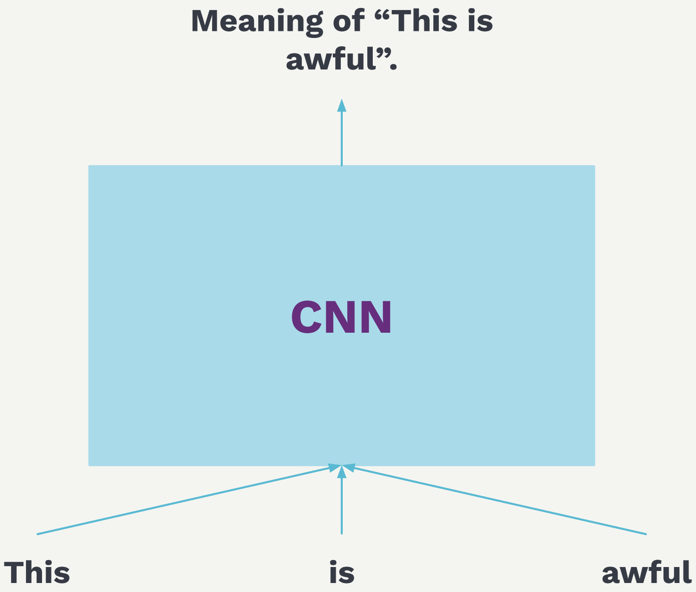
Mila

Using CNNs

Using CNNs

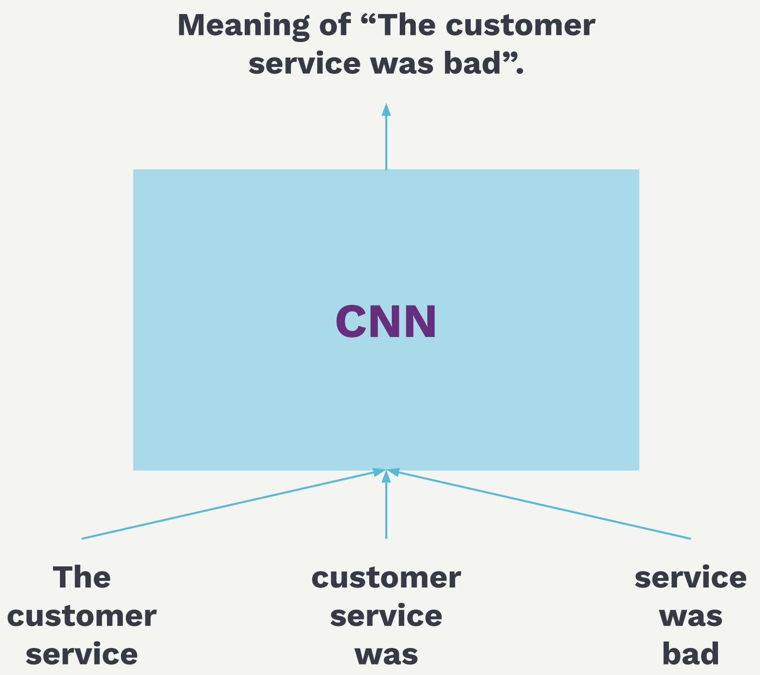

Meaning of all three chunks combined.
CNN
Chunk 1
Chunk 2
Chunk 3

Using CNNs



Mila

Using CNNs


Meaning of "The customer service was bad".
CNN
The customer service
customer service was
service was bad

Using CNNs

The customer service was really a challenge to deal with. Honestly, this is awful.

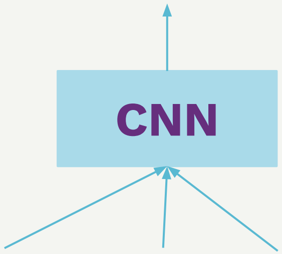
Using CNNs

The customer service was really a challenge to deal with. Honestly, this is awful.

Using CNNs

Meaning of “The customer service”

Meaning of “customer service was”



The customer service was really a challenge to deal with. Honestly, this is awful.

Using CNNs

Meaning of
“The customer
service”

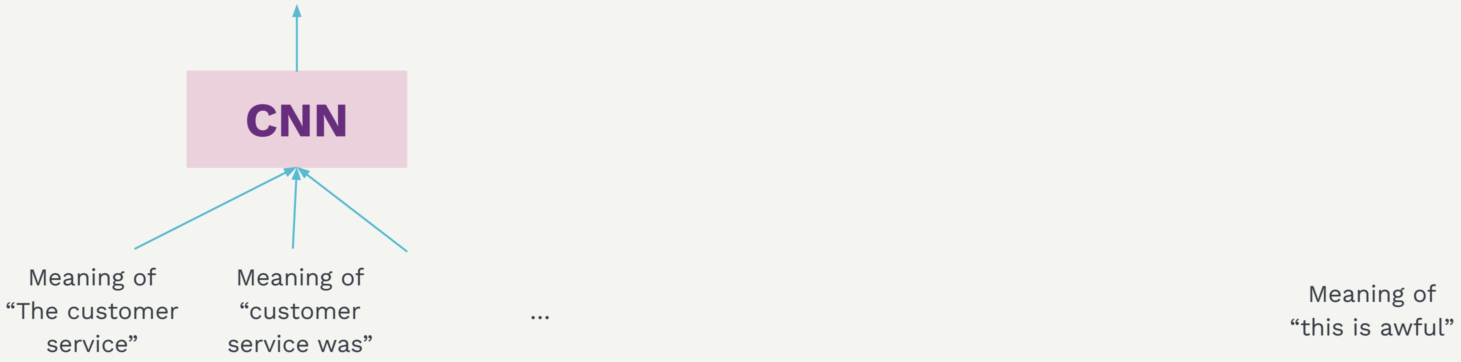
Meaning of
“customer
service was”

...

Meaning of
“this is awful”

The customer service was really a challenge to deal with. Honestly, this is awful.

Using CNNs

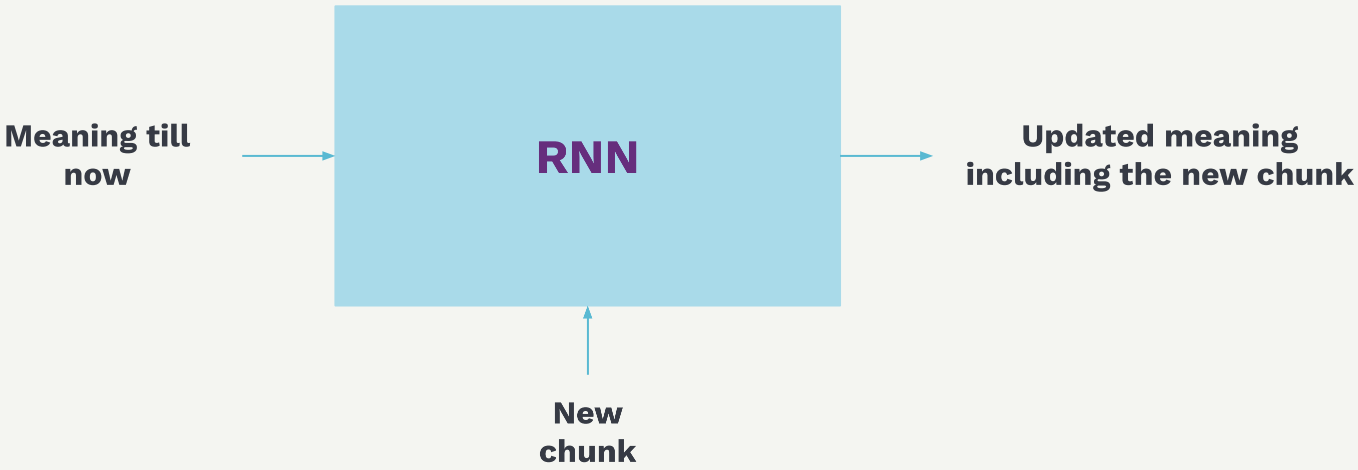


The customer service was really a challenge to deal with. Honestly, this is awful.

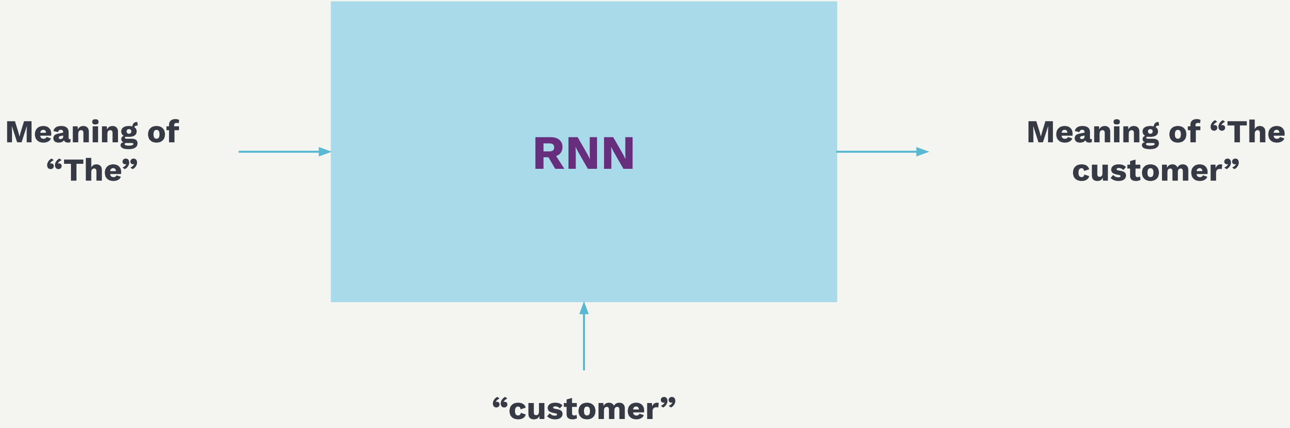
Mila

Using RNNs

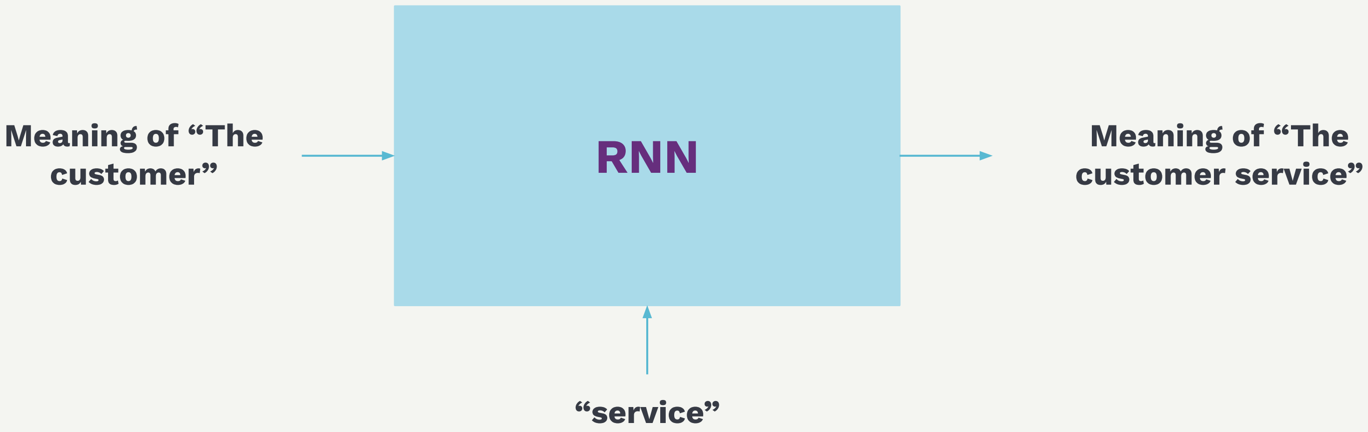
Using RNNs


Meaning till now
RNN
Updated meaning including the new chunk
New chunk

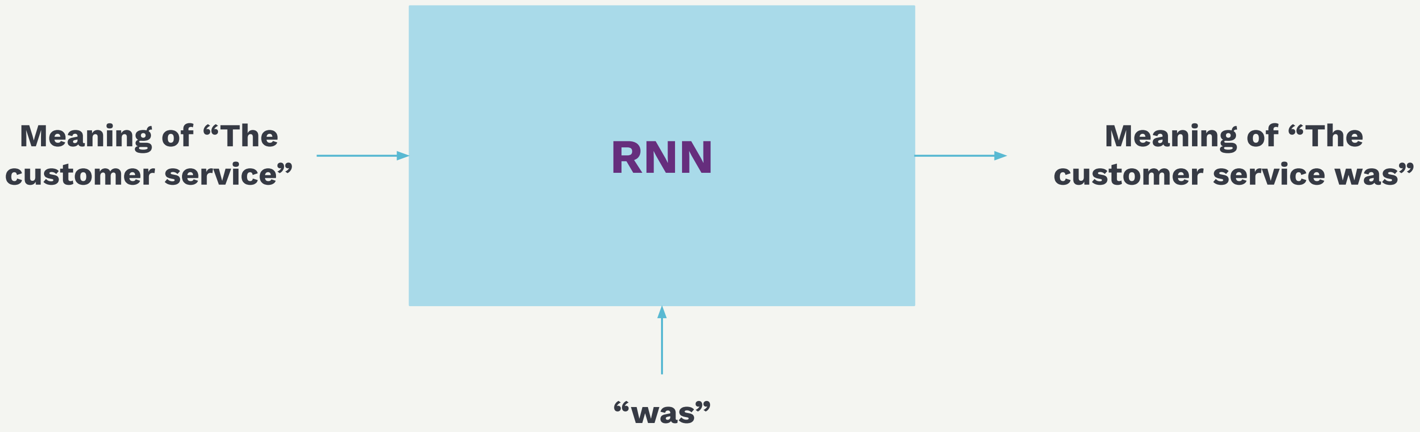
Using RNNs



Using RNNs


Meaning of “The customer”
RNN
Meaning of “The customer service”
“service”

Using RNNs

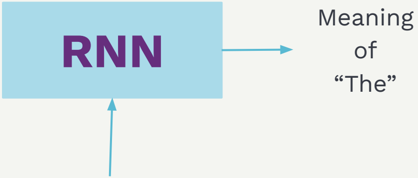


Meaning of “The customer service”
RNN
Meaning of “The customer service was”
“was”

Using RNNs

The customer service was really a challenge to deal with. Honestly, this is awful.

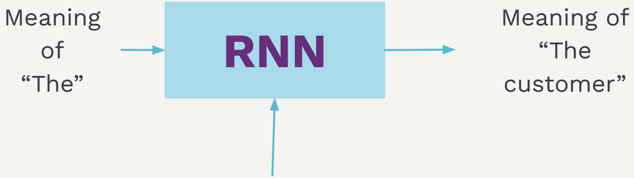
Using RNNs



The customer service was really a challenge to deal with. Honestly, this is awful.

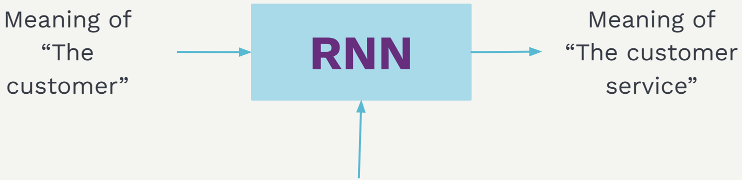
Mila

Using RNNs



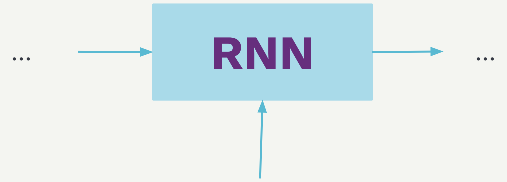
The customer service was really a challenge to deal with. Honestly, this is awful.

Using RNNs



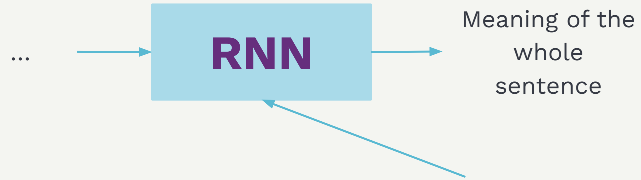
The customer service was really a challenge to deal with. Honestly, this is awful.

Using RNNs



The customer service was really a challenge to deal with. Honestly, this is awful.

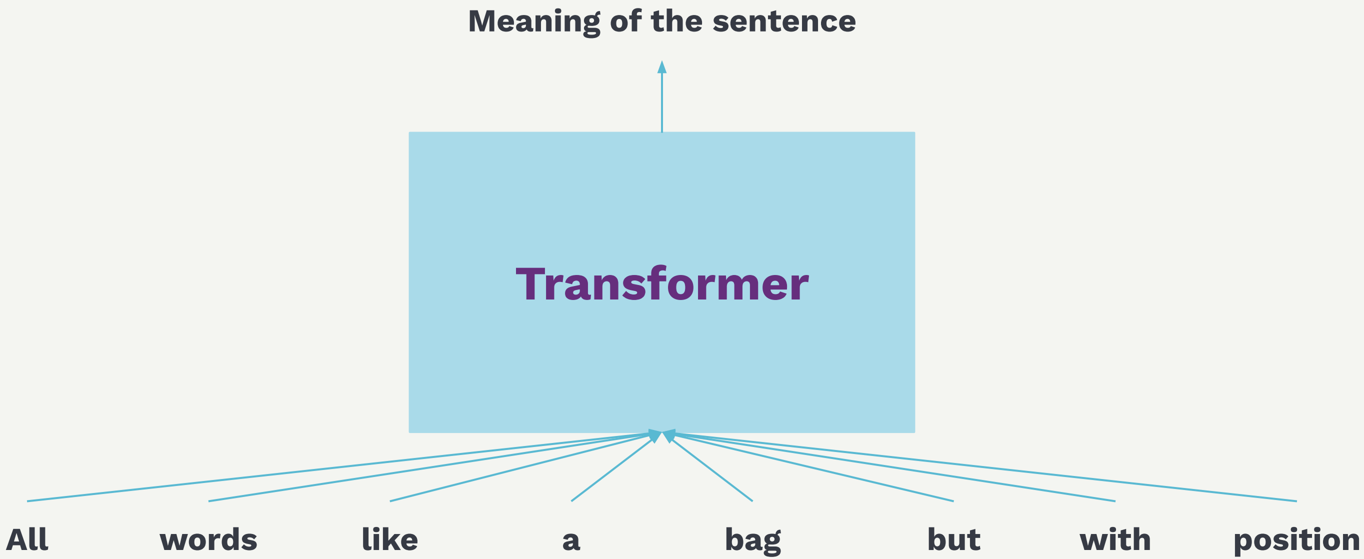
Using RNNs



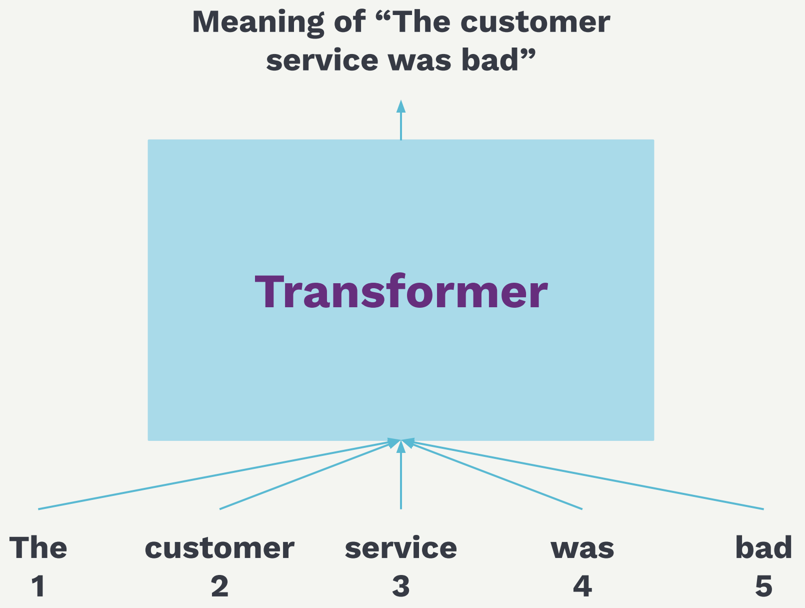
The customer service was really a challenge to deal with. Honestly, this is awful.

Using Transformers

Using Transformers



Using Transformers



Mila

Using Transformers

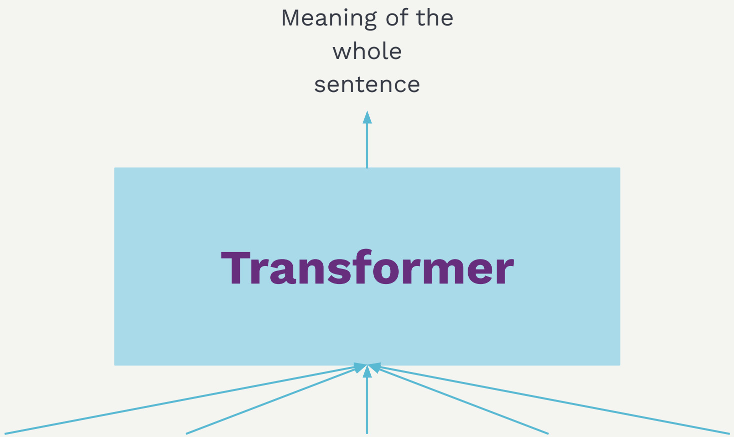
The customer service was really a challenge to deal with. Honestly, this is awful.

Using Transformers

The customer service was really a challenge to deal with. Honestly, this is awful.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

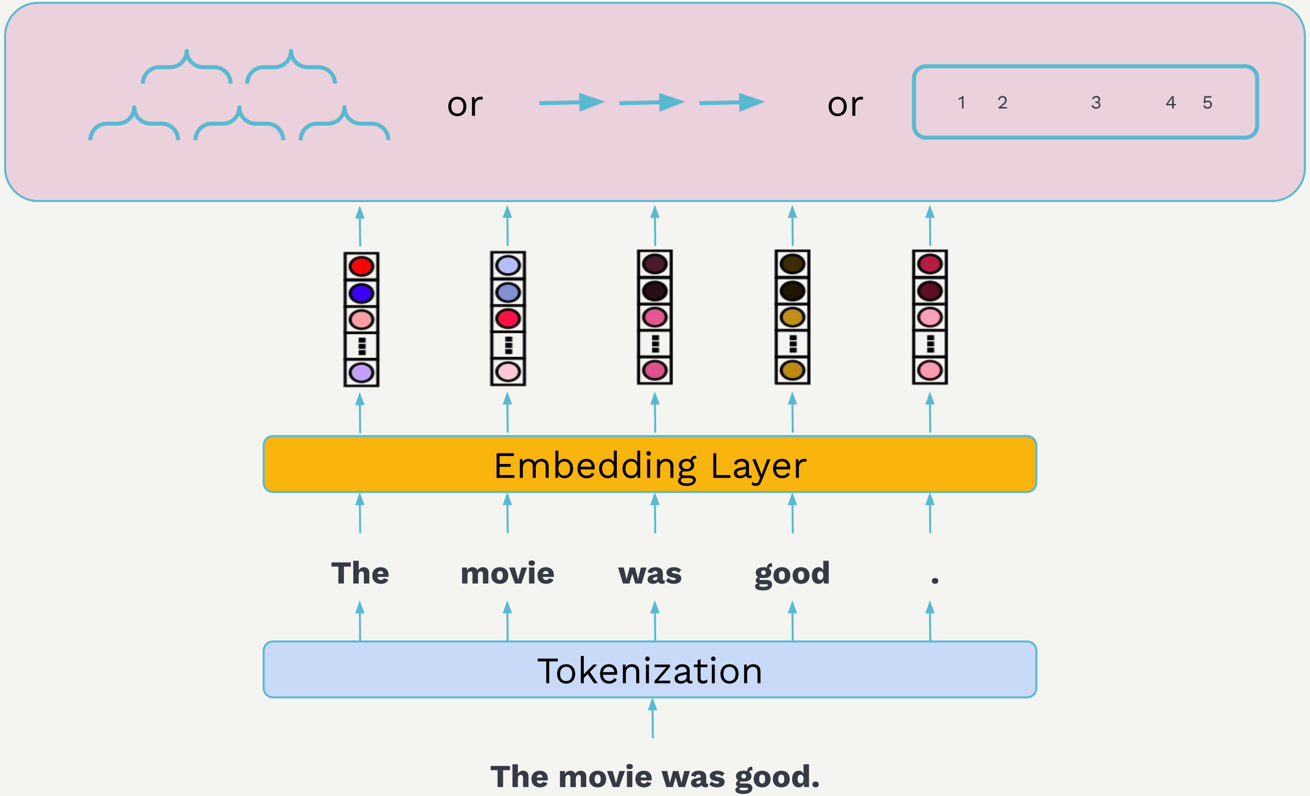
Using Transformers



The customer service was really a challenge to deal with. Honestly, this is awful.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

NLU Pipeline



Convolutional Neural Networks (CNNs)

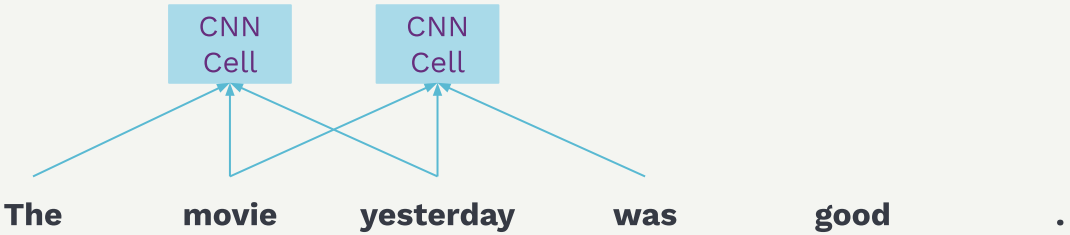
Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs)

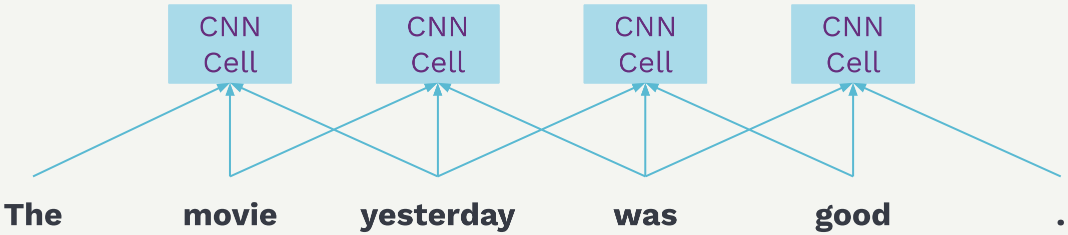
The movie yesterday was good .

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs)

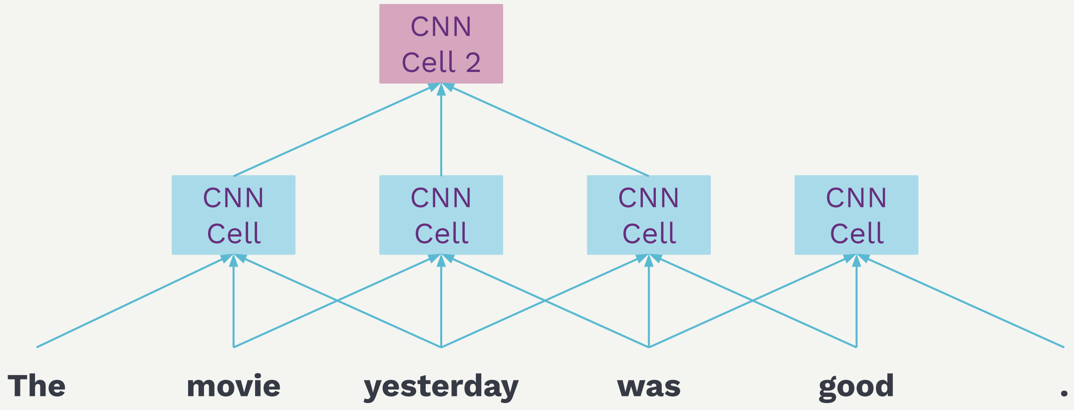


Convolutional Neural Networks (CNNs)

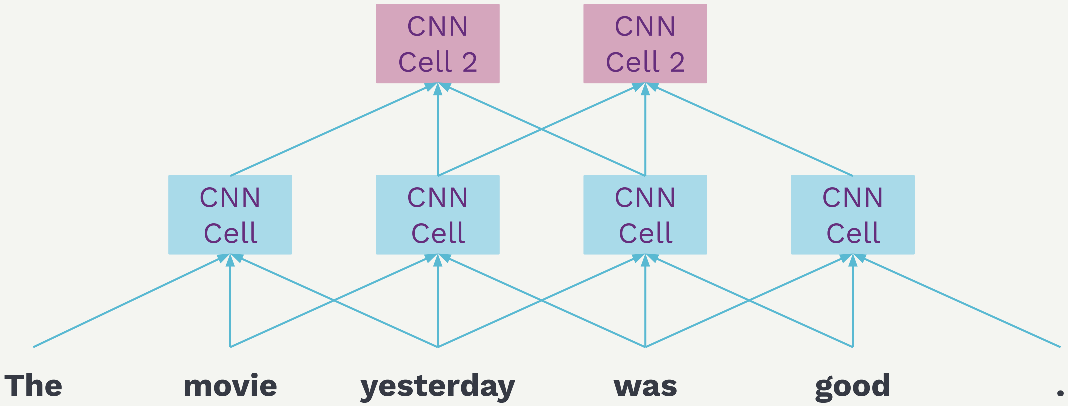


Mila

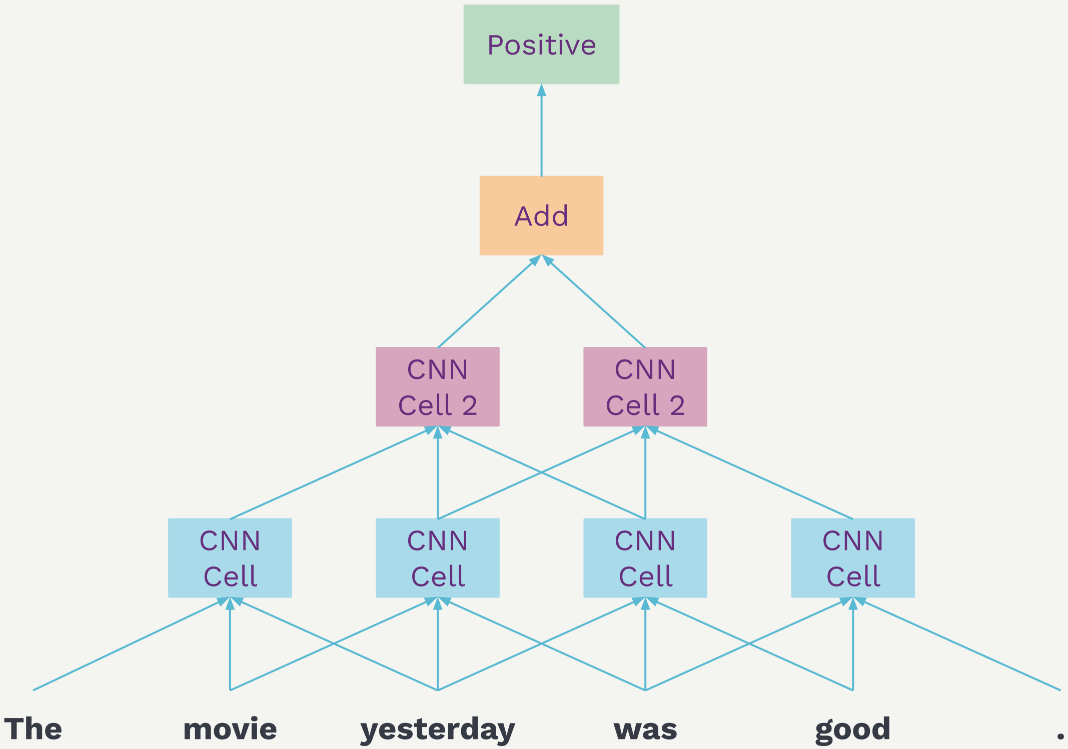
Convolutional Neural Networks (CNNs)



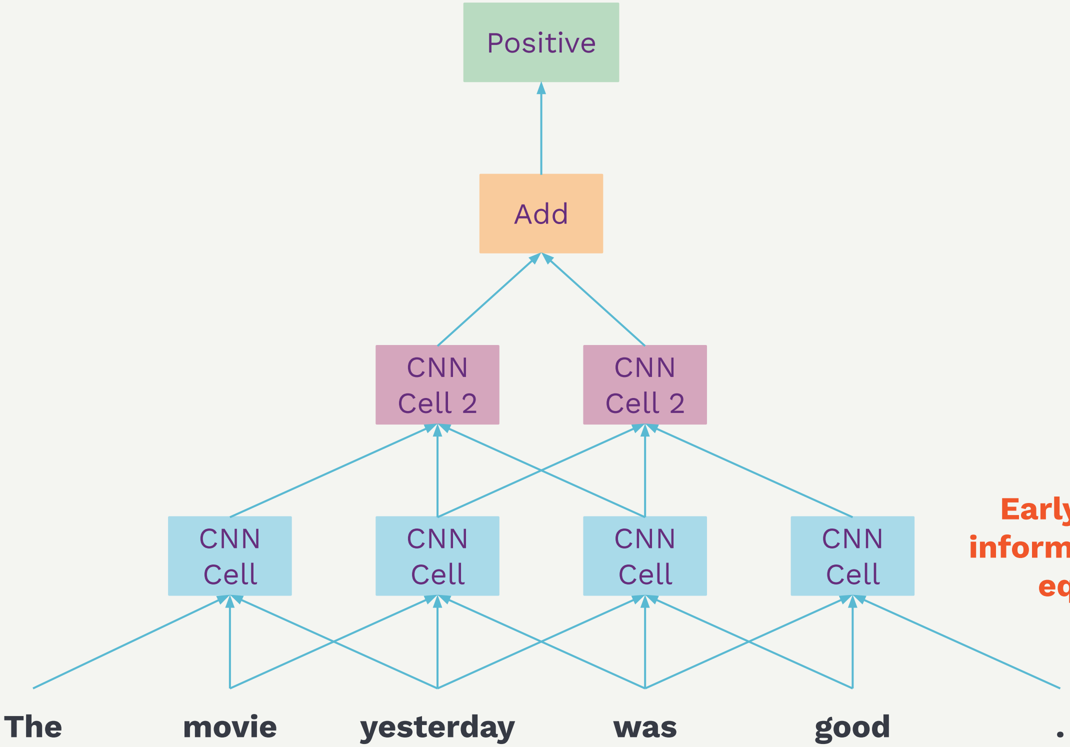
Convolutional Neural Networks (CNNs)



Convolutional Neural Networks (CNNs)

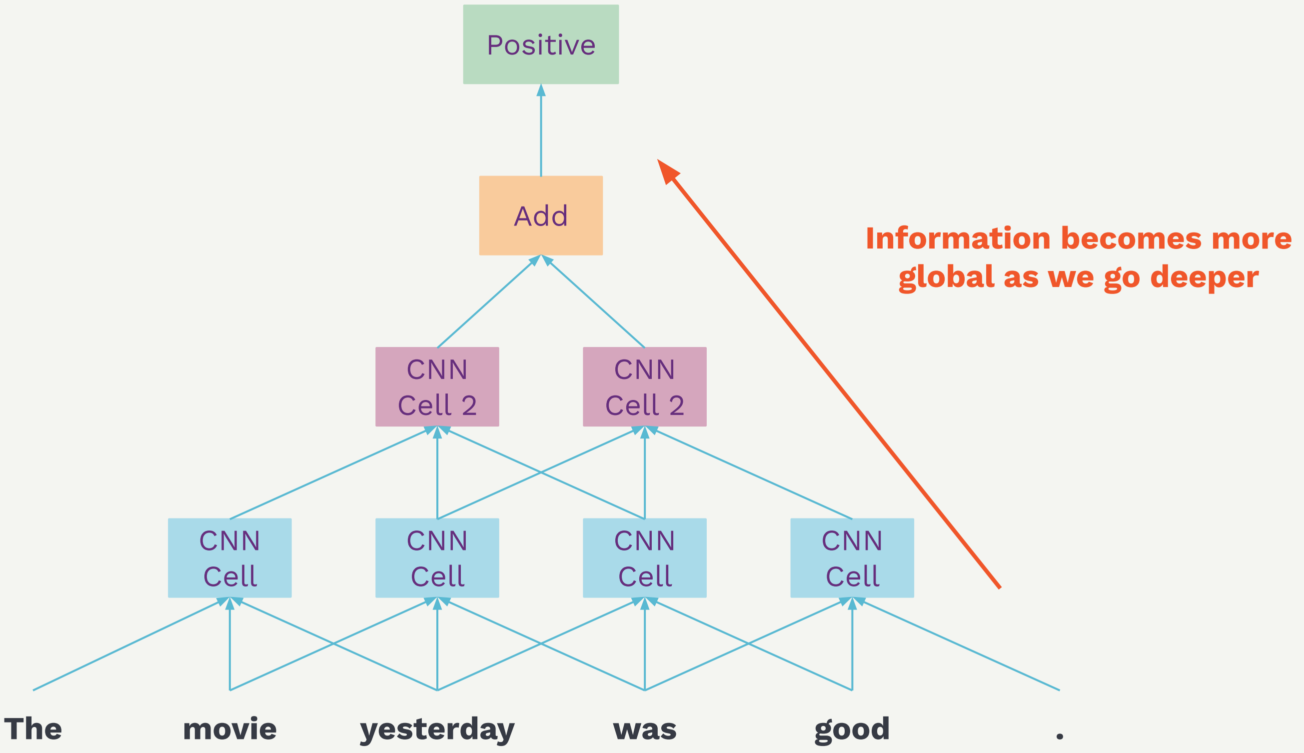


Convolutional Neural Networks (CNNs)



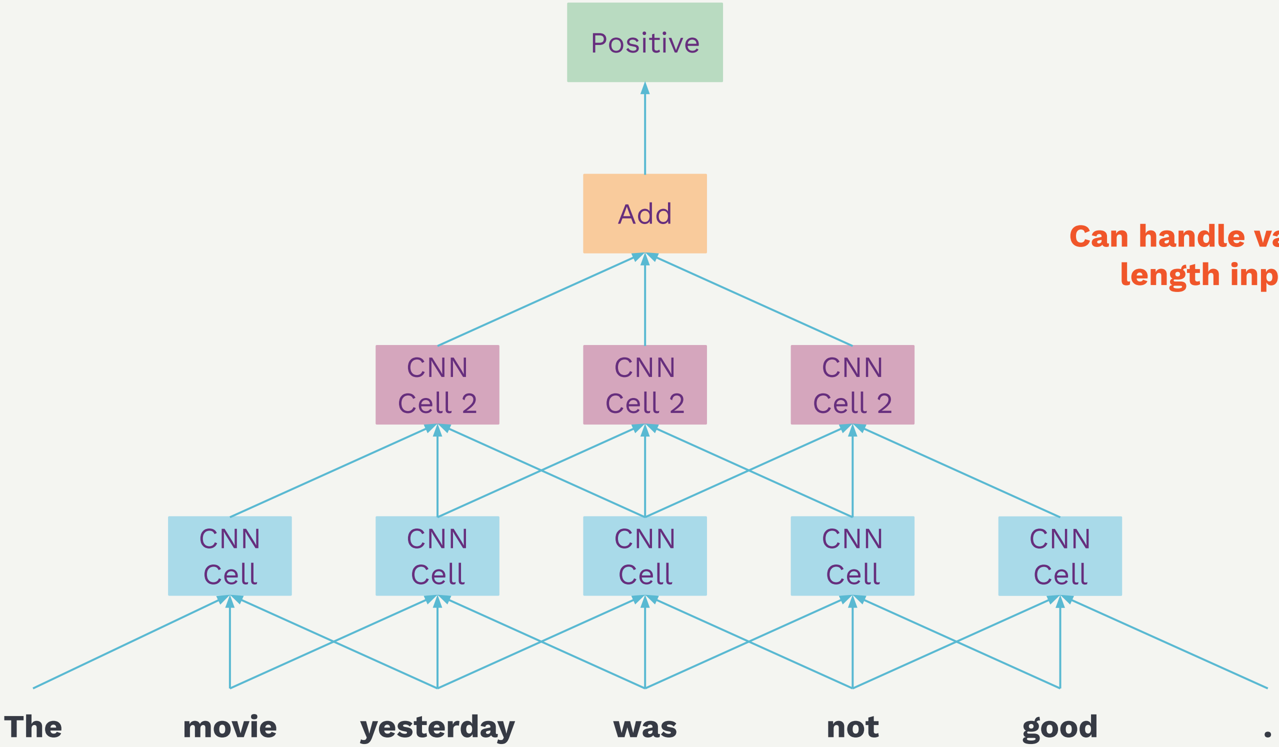
Early layers capture local information, in a 'positionally equivariant' manner.

Convolutional Neural Networks (CNNs)



Mila

Convolutional Neural Networks (CNNs)



Can handle variable length inputs

Recurrent Neural Networks (RNNs)

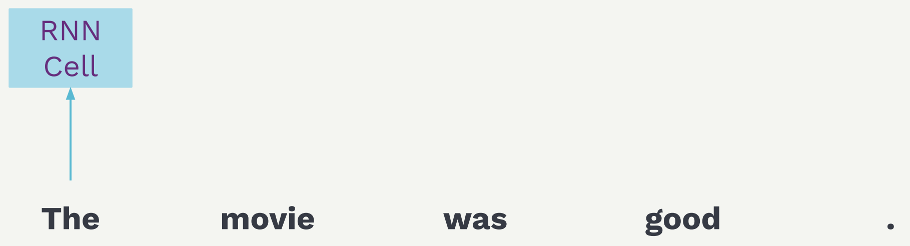
Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs)

The movie was good .

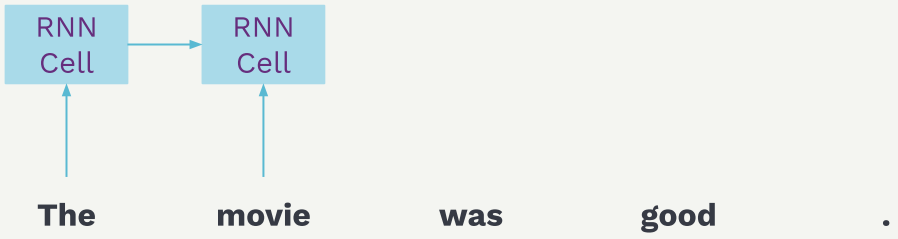
Mila

Recurrent Neural Networks (RNNs)

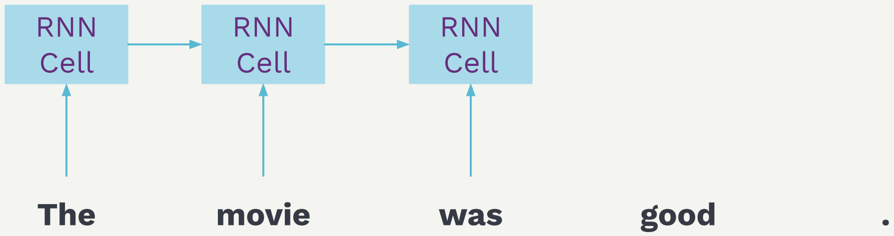


Mila

Recurrent Neural Networks (RNNs)

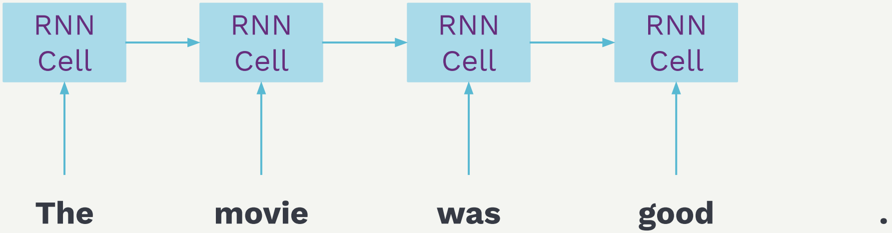


Recurrent Neural Networks (RNNs)

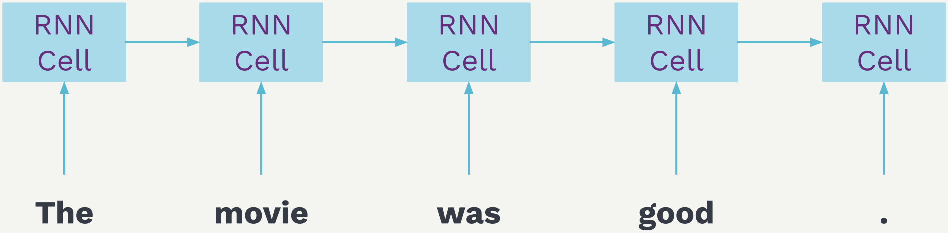


Mila

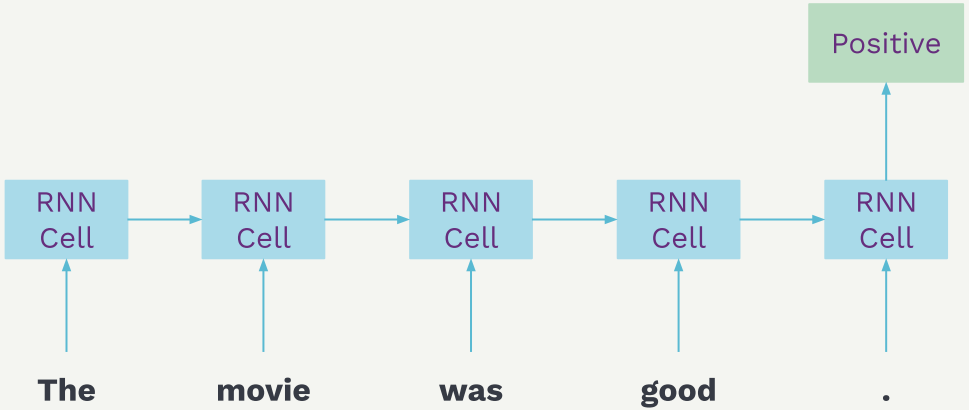
Recurrent Neural Networks (RNNs)



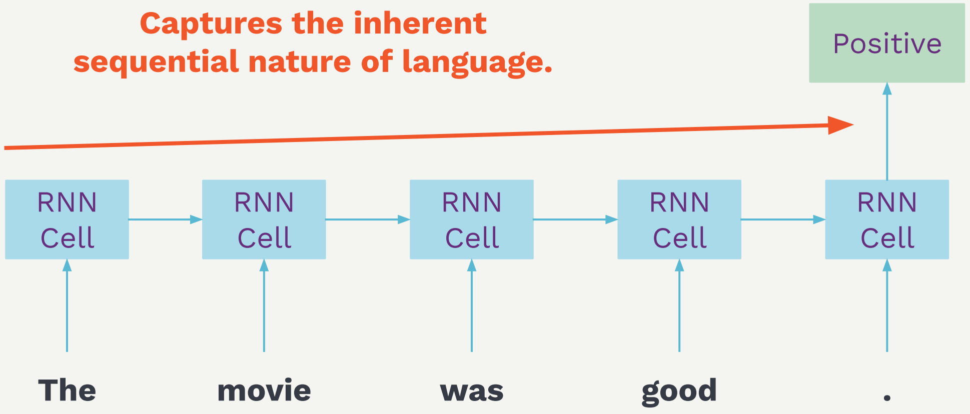
Recurrent Neural Networks (RNNs)



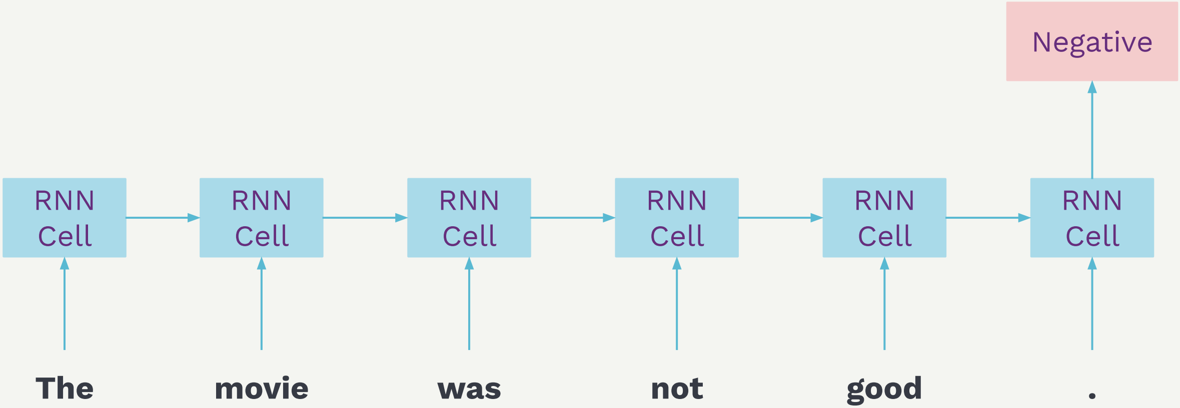
Recurrent Neural Networks (RNNs)



Recurrent Neural Networks (RNNs)

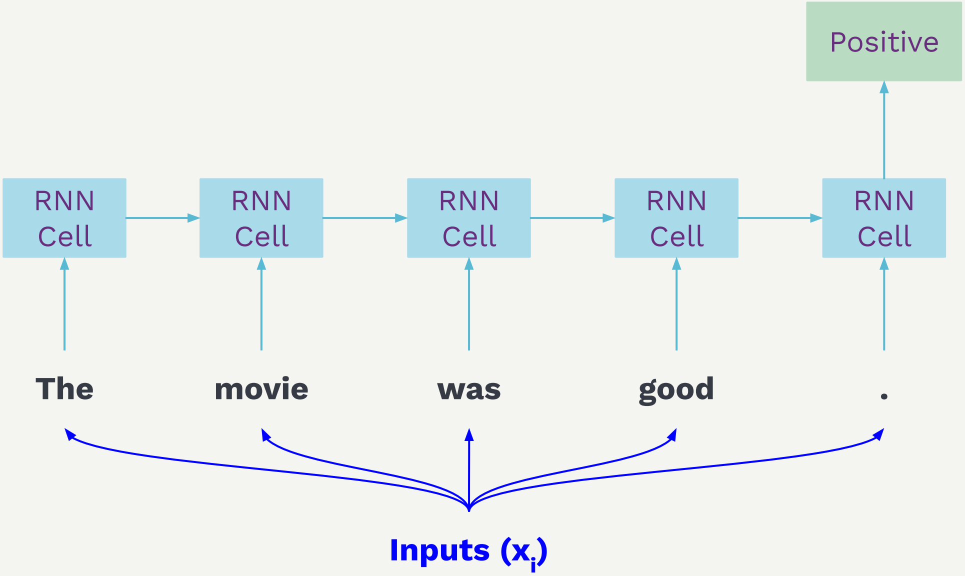


Recurrent Neural Networks (RNNs)

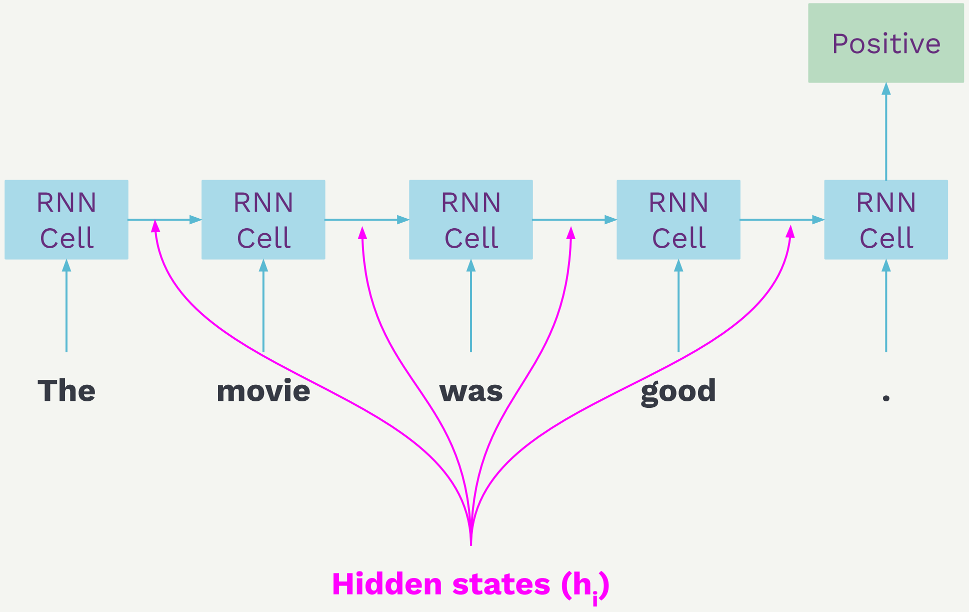


Can handle variable length inputs

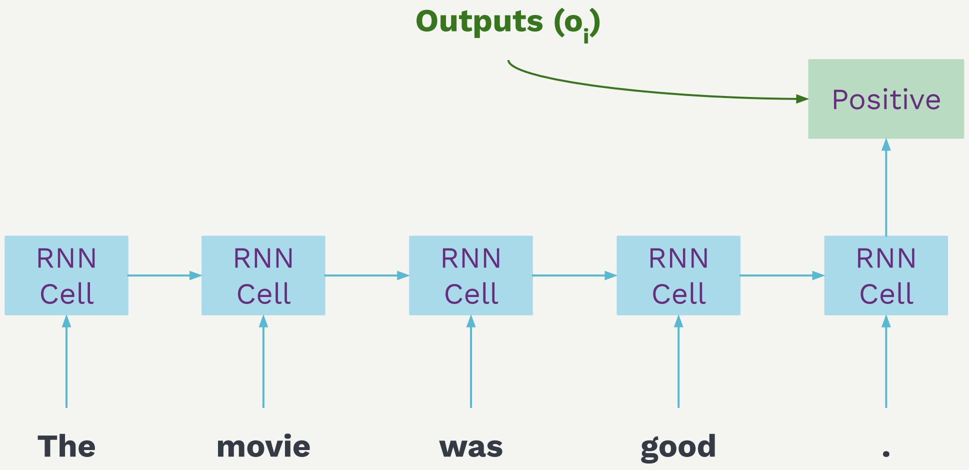
Recurrent Neural Networks (RNNs)



Recurrent Neural Networks (RNNs)

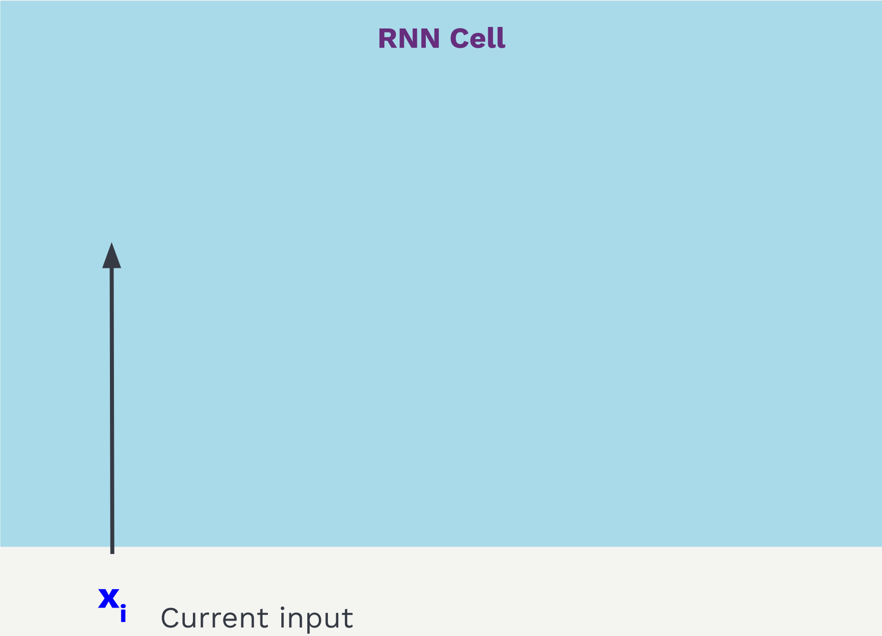


Recurrent Neural Networks (RNNs)

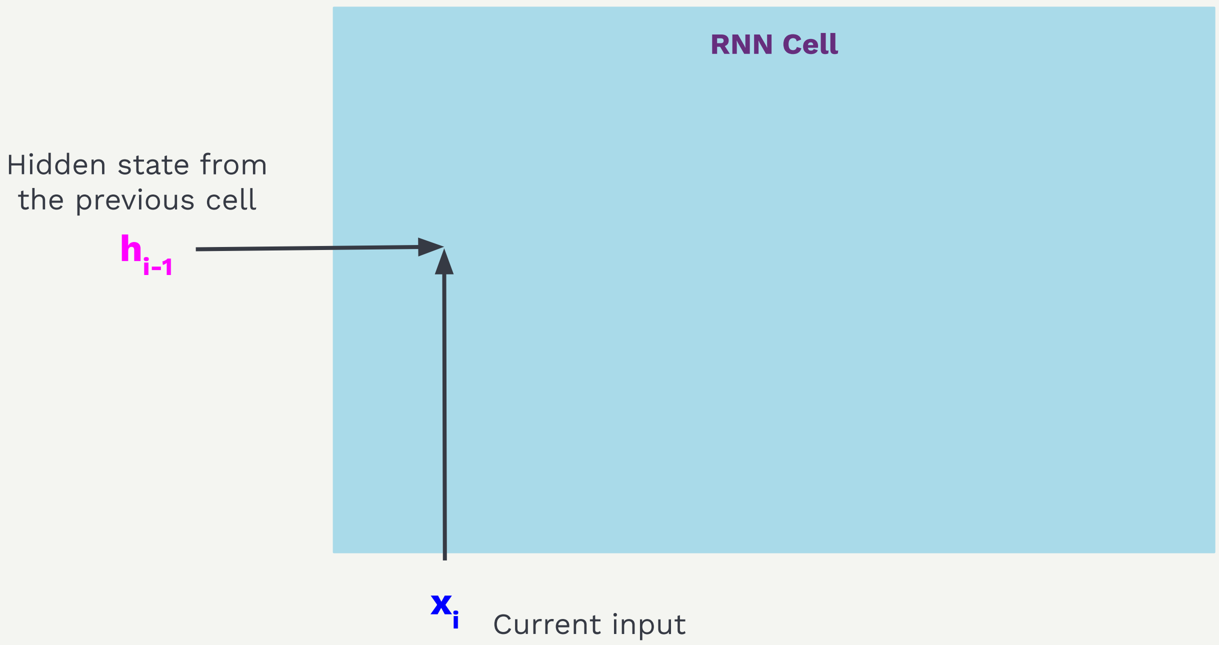


Mila

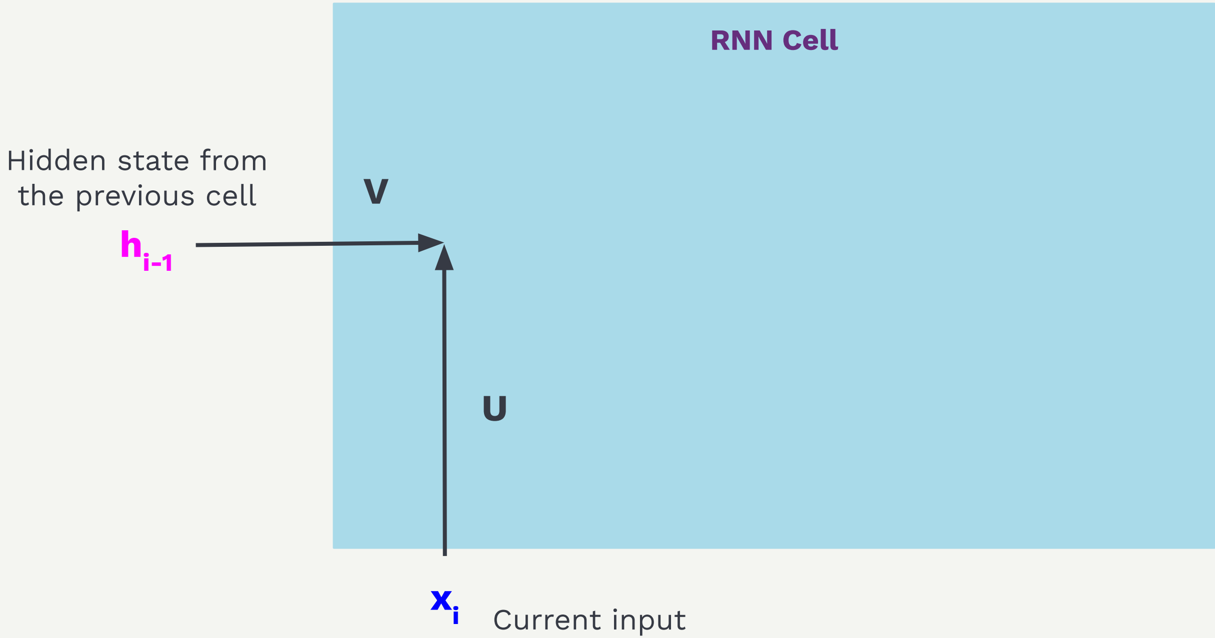
Recurrent Neural Networks (RNNs)



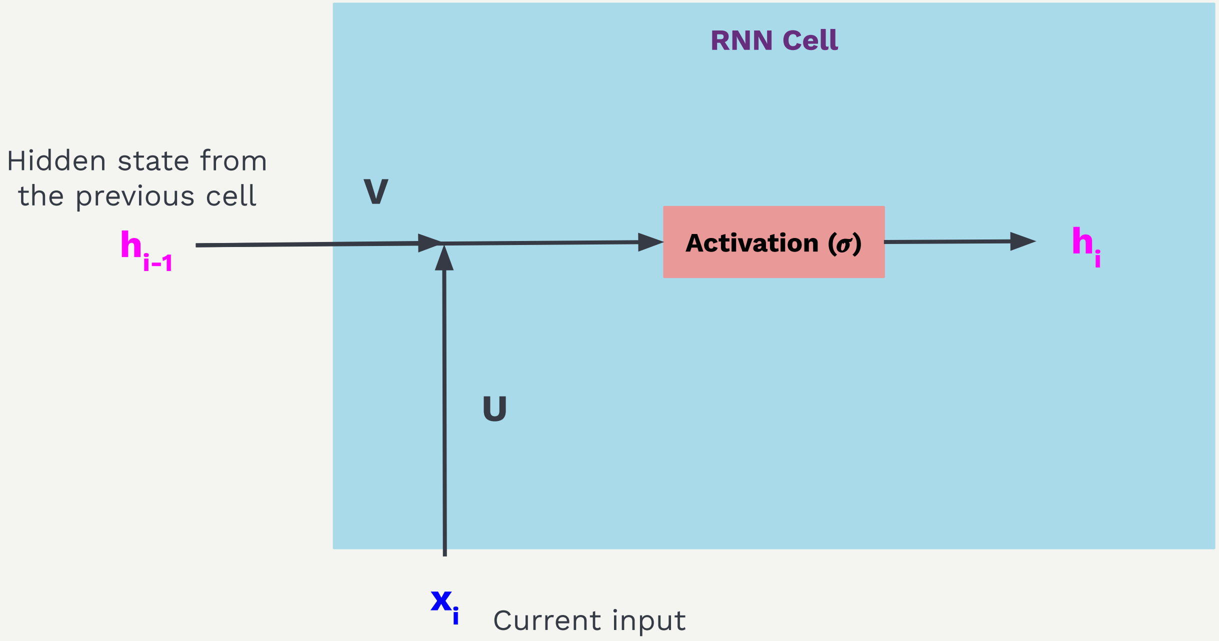
Recurrent Neural Networks (RNNs)



Recurrent Neural Networks (RNNs)

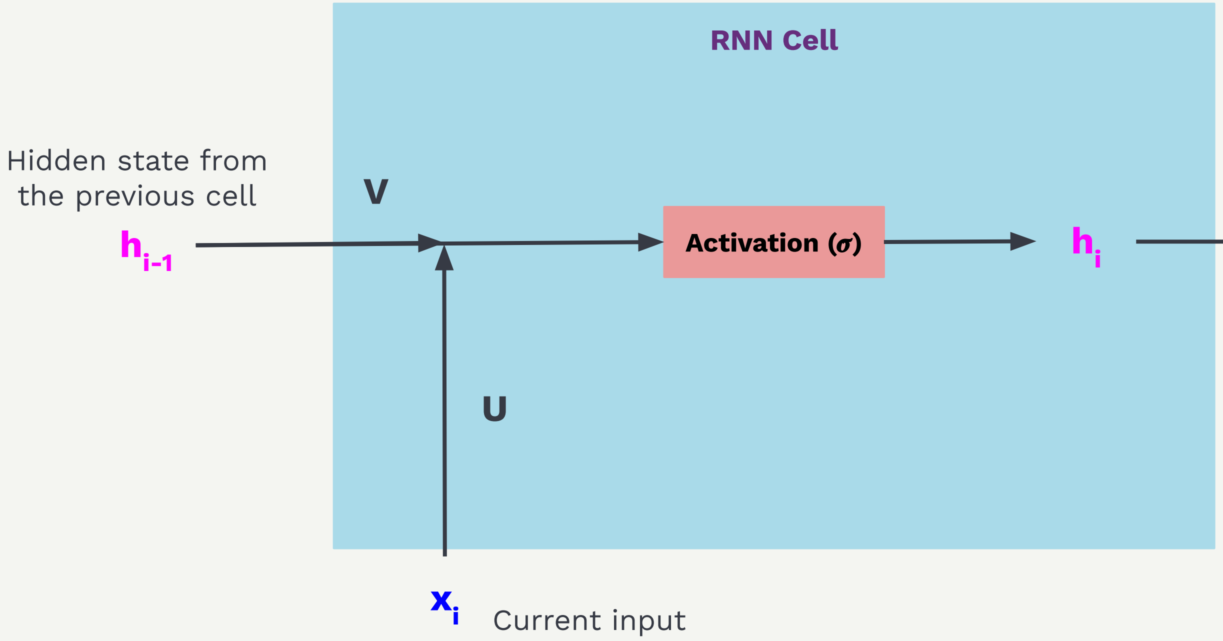


Recurrent Neural Networks (RNNs)



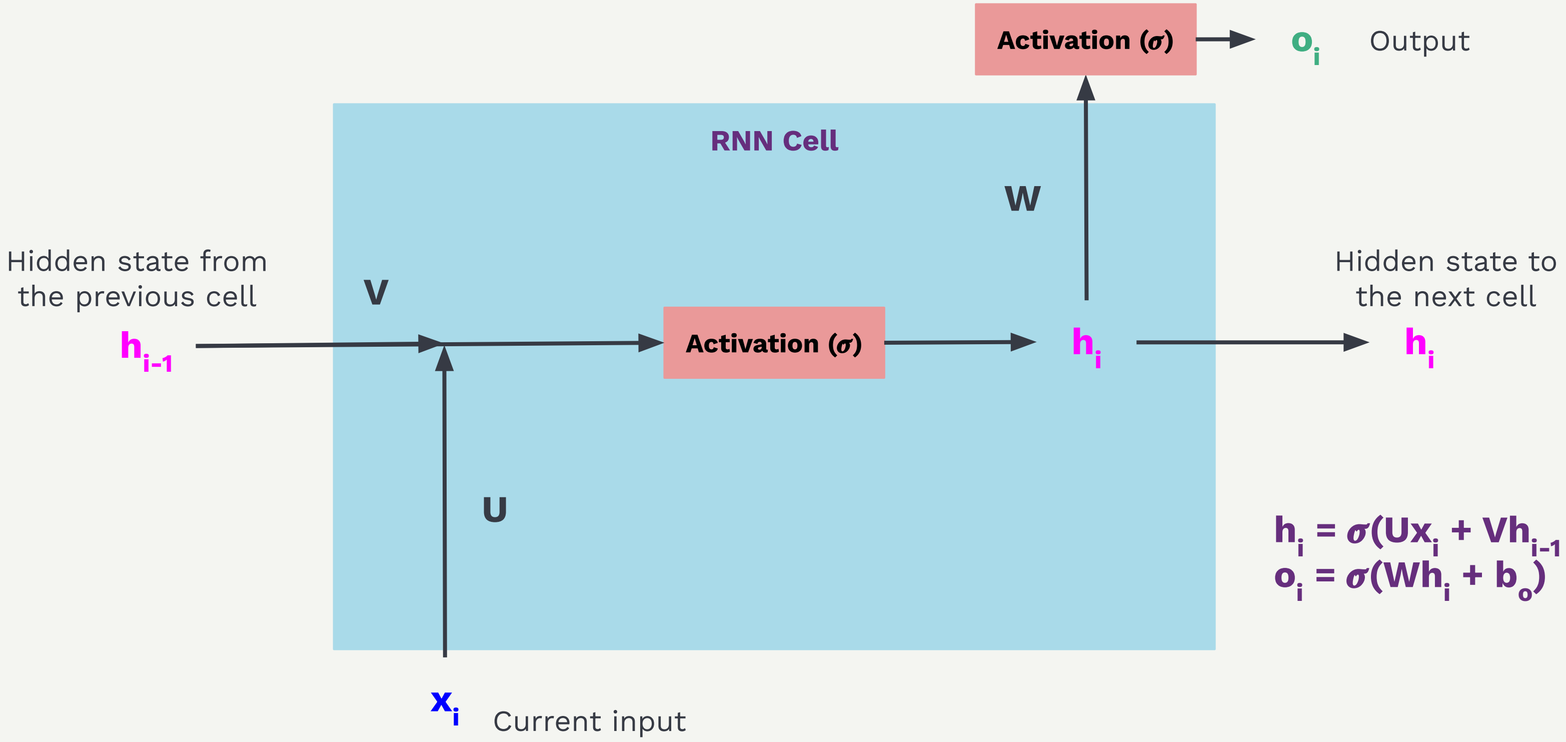
$$h_i = \sigma(Ux_i + Vh_{i-1} + b_h)$$

Recurrent Neural Networks (RNNs)



$$h_i = \sigma(Ux_i + Vh_{i-1} + b_h)$$

Recurrent Neural Networks (RNNs)



$$h_i = \sigma(Ux_i + Vh_{i-1} + b_h)$$
$$o_i = \sigma(Wh_i + b_o)$$

Example: RNN for Sentiment Classification

RNN for Text Classification

Objective: Given a sentence **s**, predict whether it contains positive or negative sentiments.

RNN for Text Classification

Objective: Given a sentence **s**, predict whether it contains positive or negative sentiments.

Eg: That movie was awful. → Negative

RNN for Text Classification

Step 1: Collect Data

Sentence	Prediction
This movie is great.	Positive
That movie was good.	Positive
This movie is awful.	Negative
That movie was bad.	Negative

RNN for Text Classification

Step 2: Tokenize Data and Create a Vocabulary

Sentence	Tokens
This movie is great.	“This” “movie” “is” “great” “.”
That movie was good.	“That” “movie” “was” “good” “.”
This movie is awful.	“This” “movie” “is” “awful” “.”
That movie was bad.	“That” “movie” “was” “bad” “.”
Vocabulary	“This” “That” “movie” “is” “was” “great” “good” “awful” “bad” “.”

RNN for Text Classification

Step 3: Encode Sentences

	This	That	movie	is	was	great	good	awful	bad	.
This	1	0	0	0	0	0	0	0	0	0
movie	0	0	1	0	0	0	0	0	0	0
is	0	0	0	1	0	0	0	0	0	0
great	0	0	0	0	0	1	0	0	0	0
.	0	0	0	0	0	0	0	0	0	1

RNN for Text Classification

Step 4: Initialize All Weights

Embedding Matrix (E)

e_11	...	e_1k
e_21	...	e_2k
...	...	...
e_v1	...	e_vk

v → vocabulary size
k → embedding size

Weight Matrix (U)

u_11	...	u_1k
u_21	...	u_2k
...	...	...
u_k1	...	u_kk

k → embedding size

Weight Matrix (V)

v_11	...	v_1k
v_21	...	v_2k
...	...	...
v_k1	...	v_kk

k → embedding size

Weight Matrix (W)

w_11
w_21
...
w_k1

k → embedding size

Biases (b_h)

b_11	...	b_1k

k → embedding size

Biases (b_o)

bo_11

RNN for Text Classification

Step 5: Forward Pass

1	0	0	0	0	0	0	0	0	0

This

RNN for Text Classification

Step 5: Forward Pass

$$\mathbf{x_1} = \mathbf{E} \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This

RNN for Text Classification

Step 5: Forward Pass

0	0	0	0	0	0	0	0	0	0

h_0

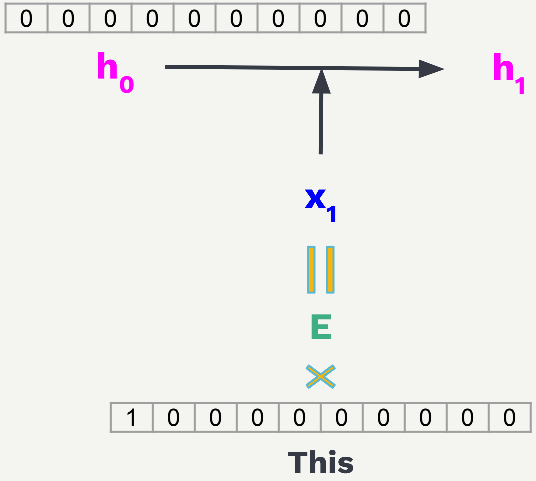
$x_1 = E \times$

1	0	0	0	0	0	0	0	0	0

This

RNN for Text Classification

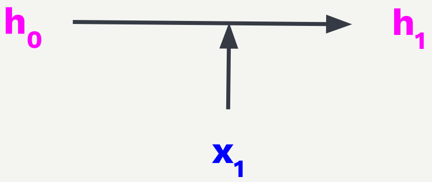
Step 5: Forward Pass



$$h_1 = \sigma(Ux_1 + Vh_0 + b_h)$$

RNN for Text Classification

Step 5: Forward Pass

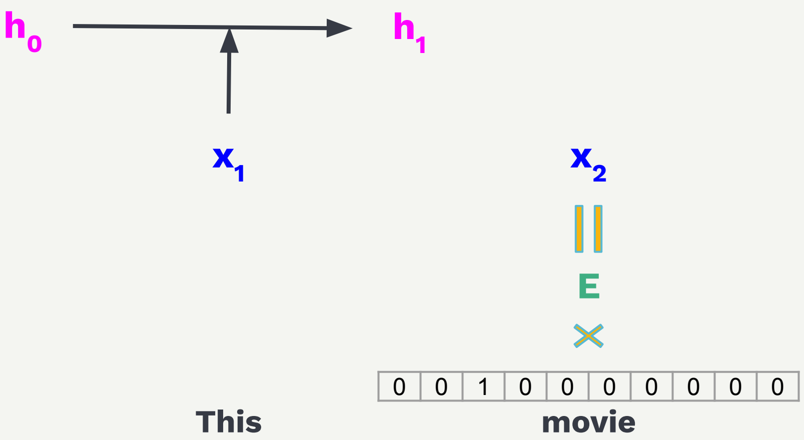


This movie

Mila

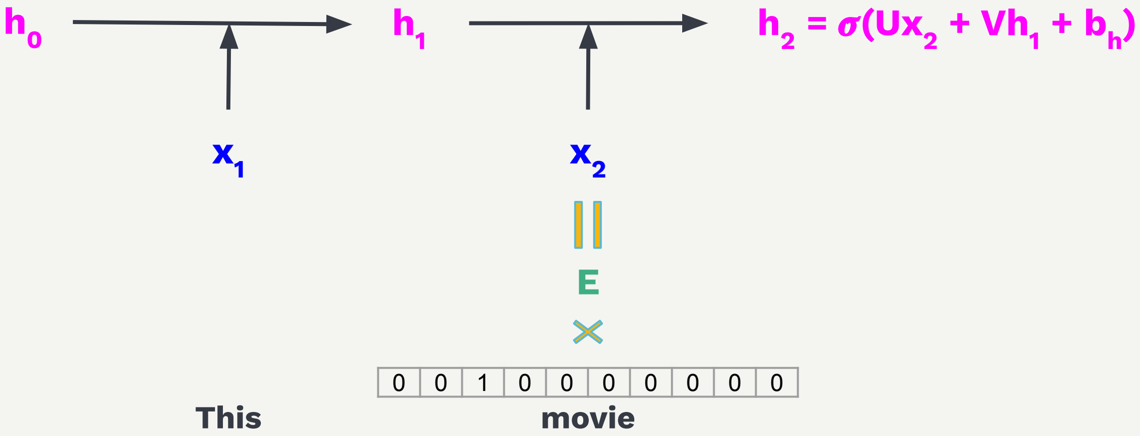
RNN for Text Classification

Step 5: Forward Pass



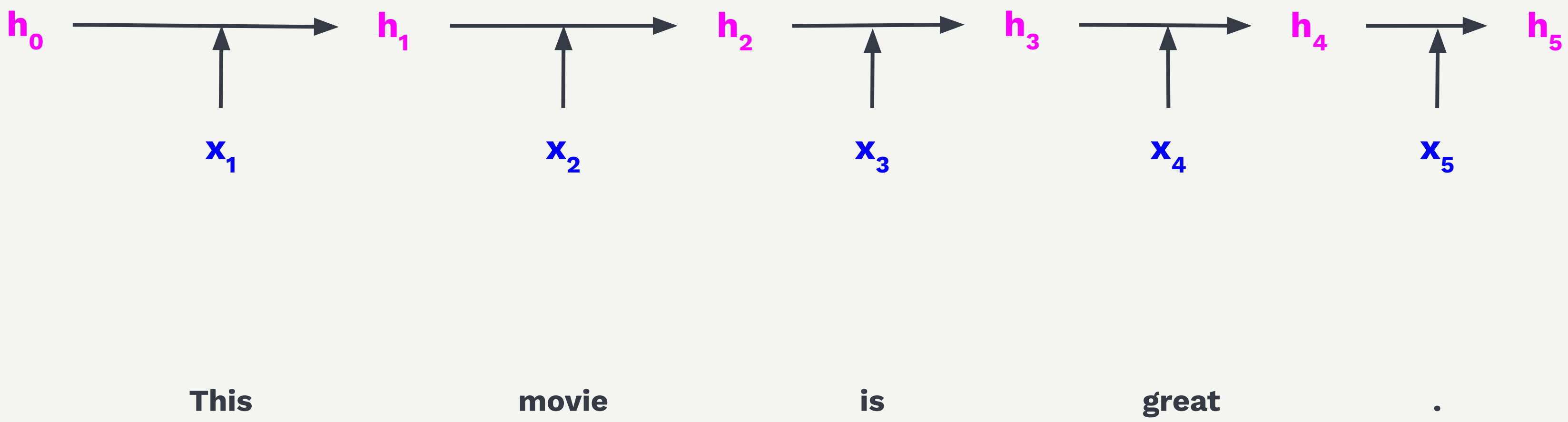
RNN for Text Classification

Step 5: Forward Pass



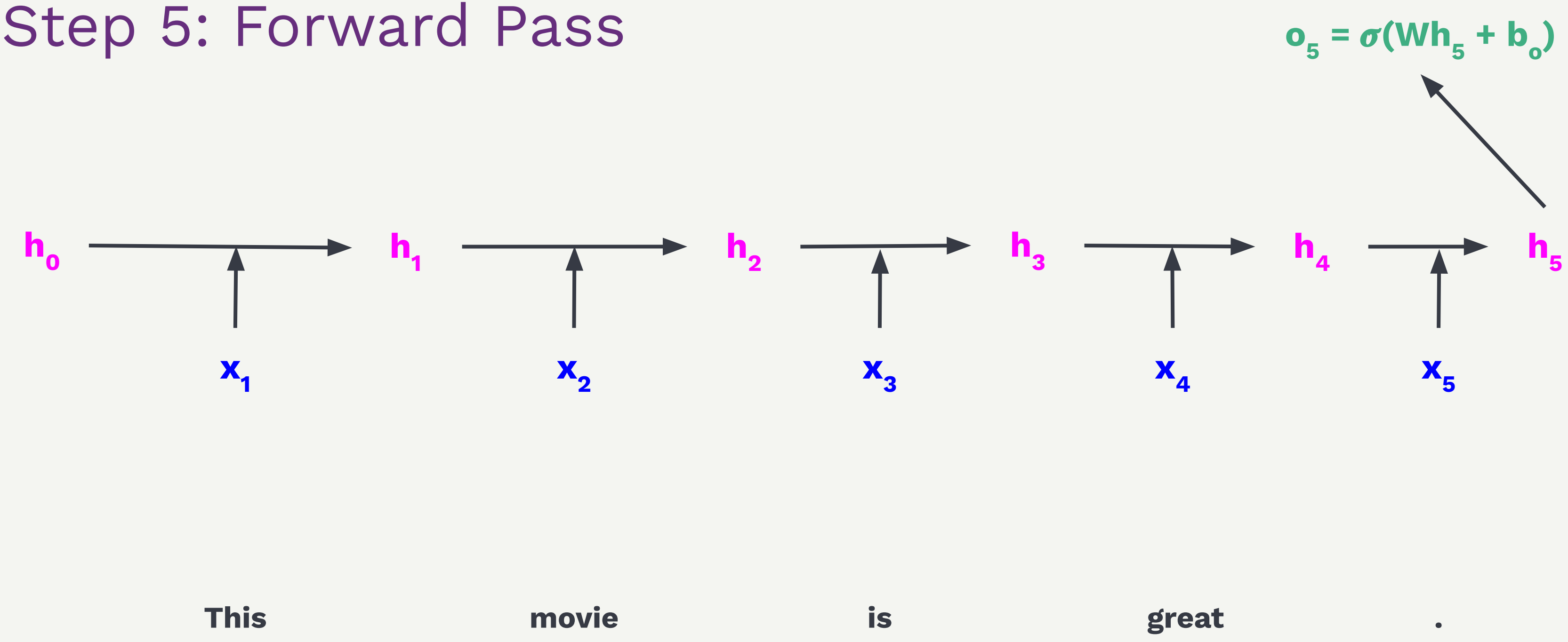
RNN for Text Classification

Step 5: Forward Pass



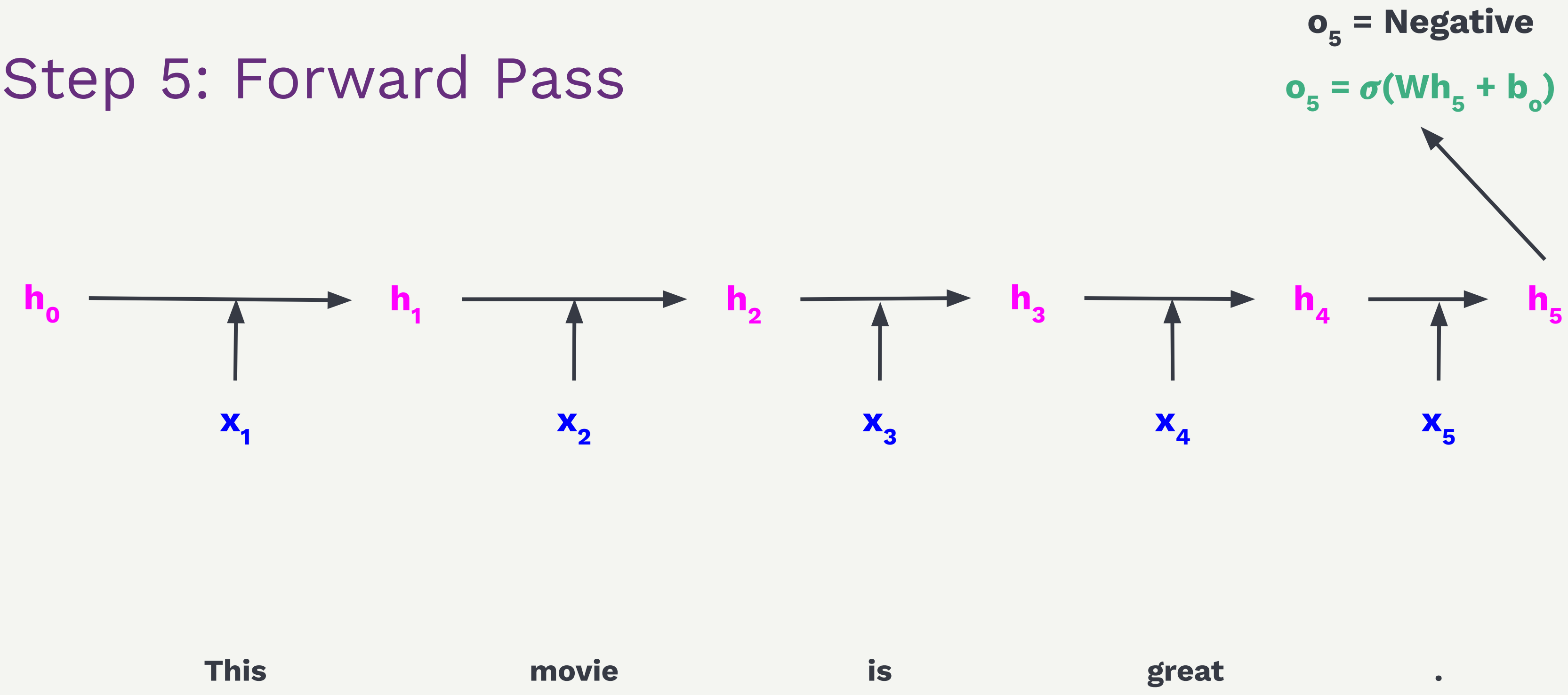
RNN for Text Classification

Step 5: Forward Pass



RNN for Text Classification

Step 5: Forward Pass



Mila

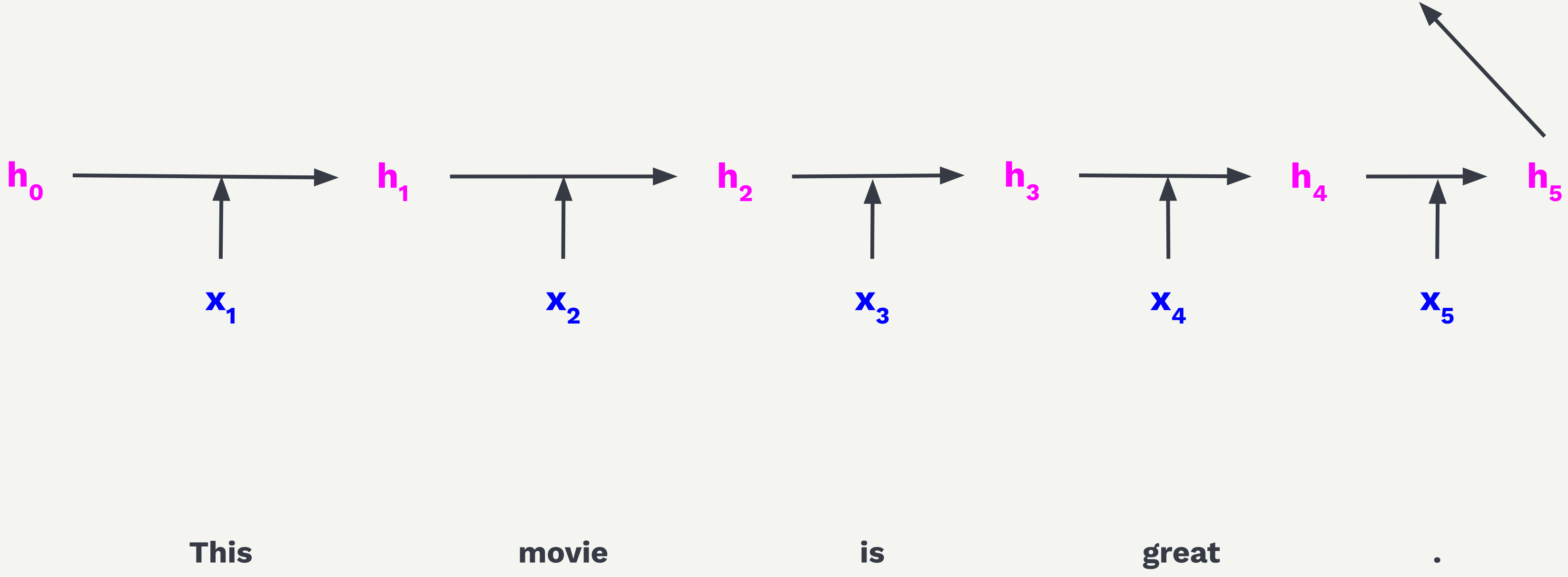
RNN for Text Classification

Step 6: Calculate Loss

Actual Label = Positive

o_5 = Negative

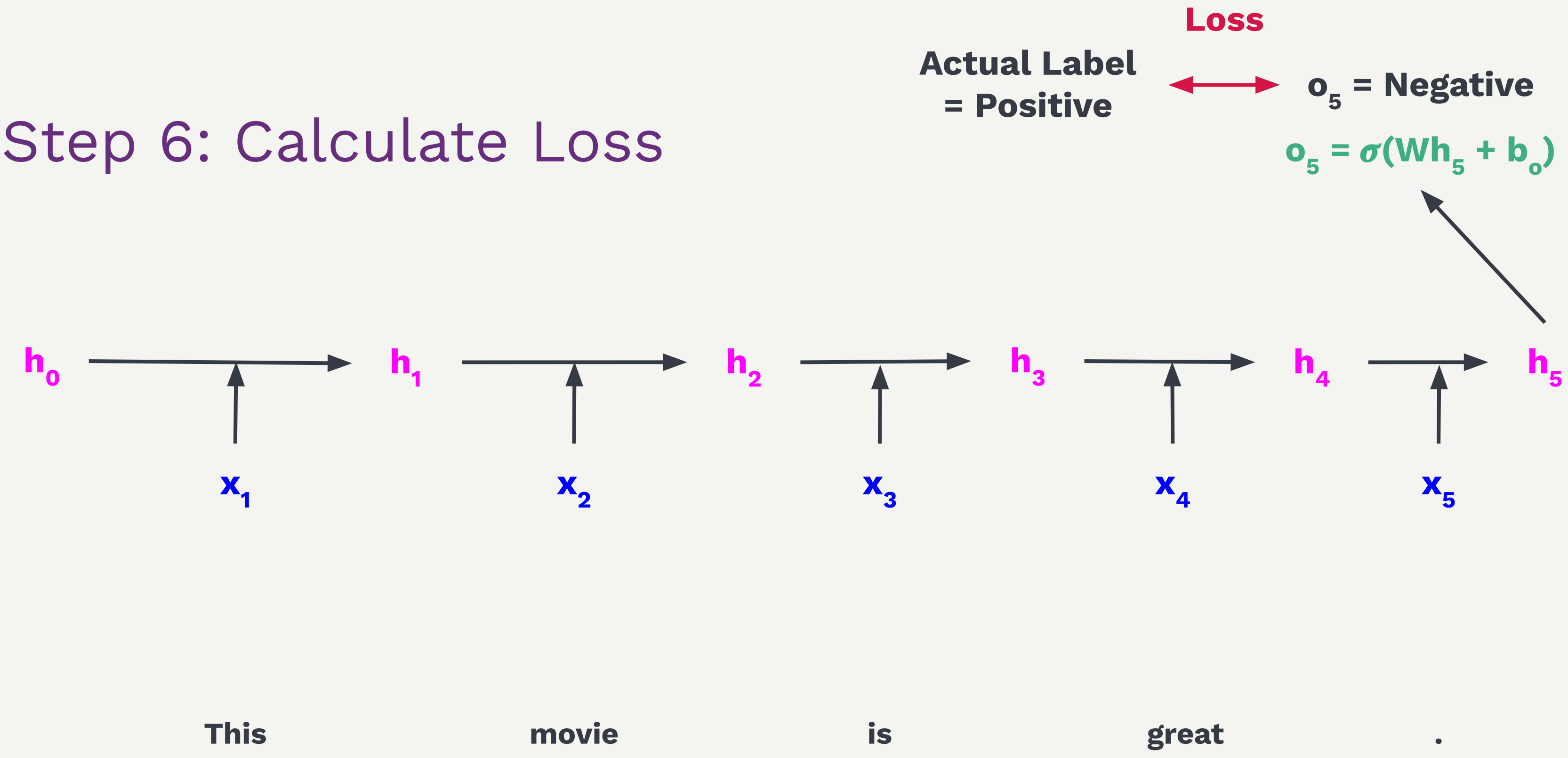
$o_5 = \sigma(Wh_5 + b_o)$



Mila

RNN for Text Classification

Step 6: Calculate Loss

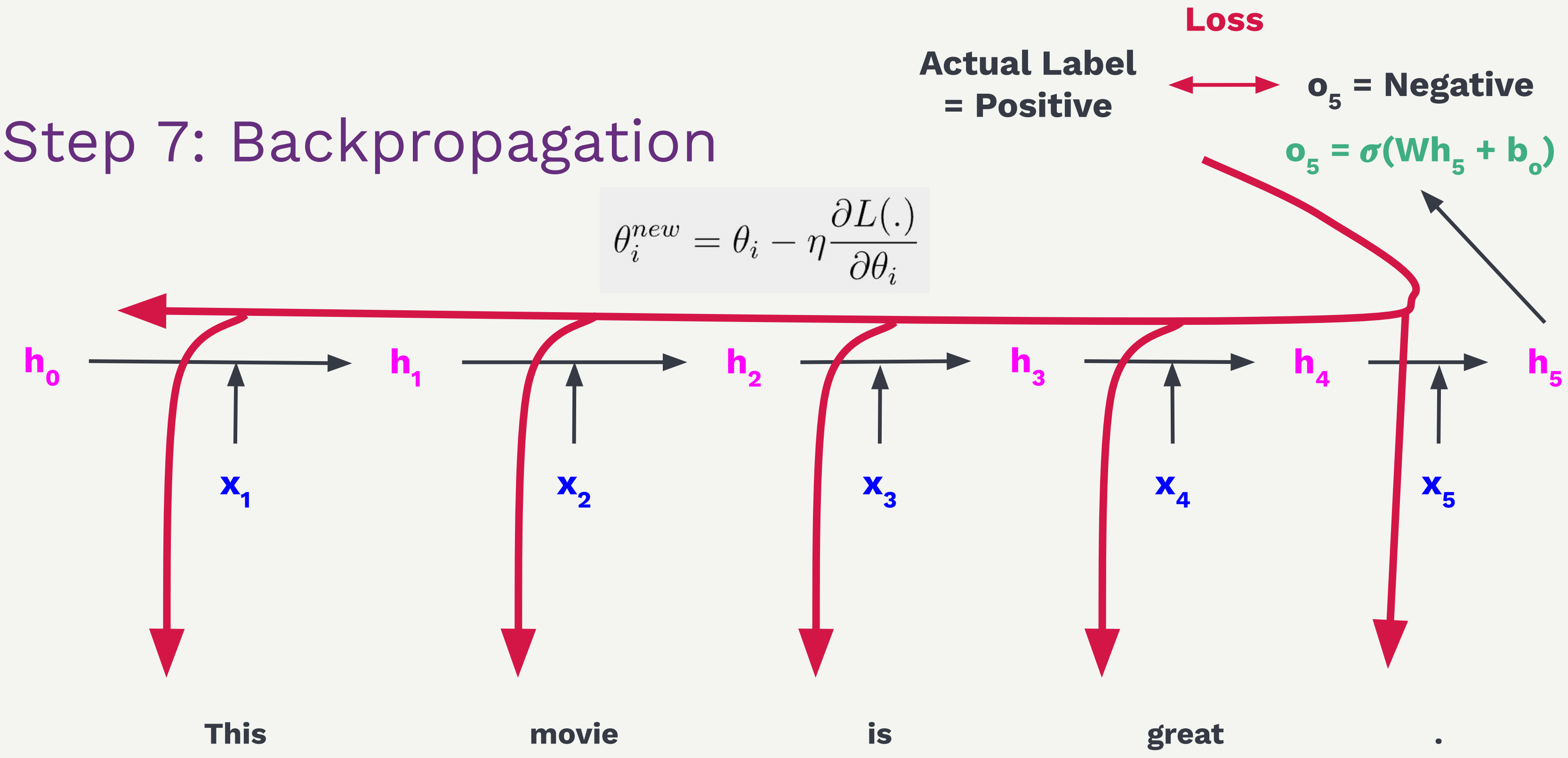


Mila

RNN for Text Classification

Step 7: Backpropagation

$$\theta_i^{new} = \theta_i - \eta \frac{\partial L(.)}{\partial \theta_i}$$



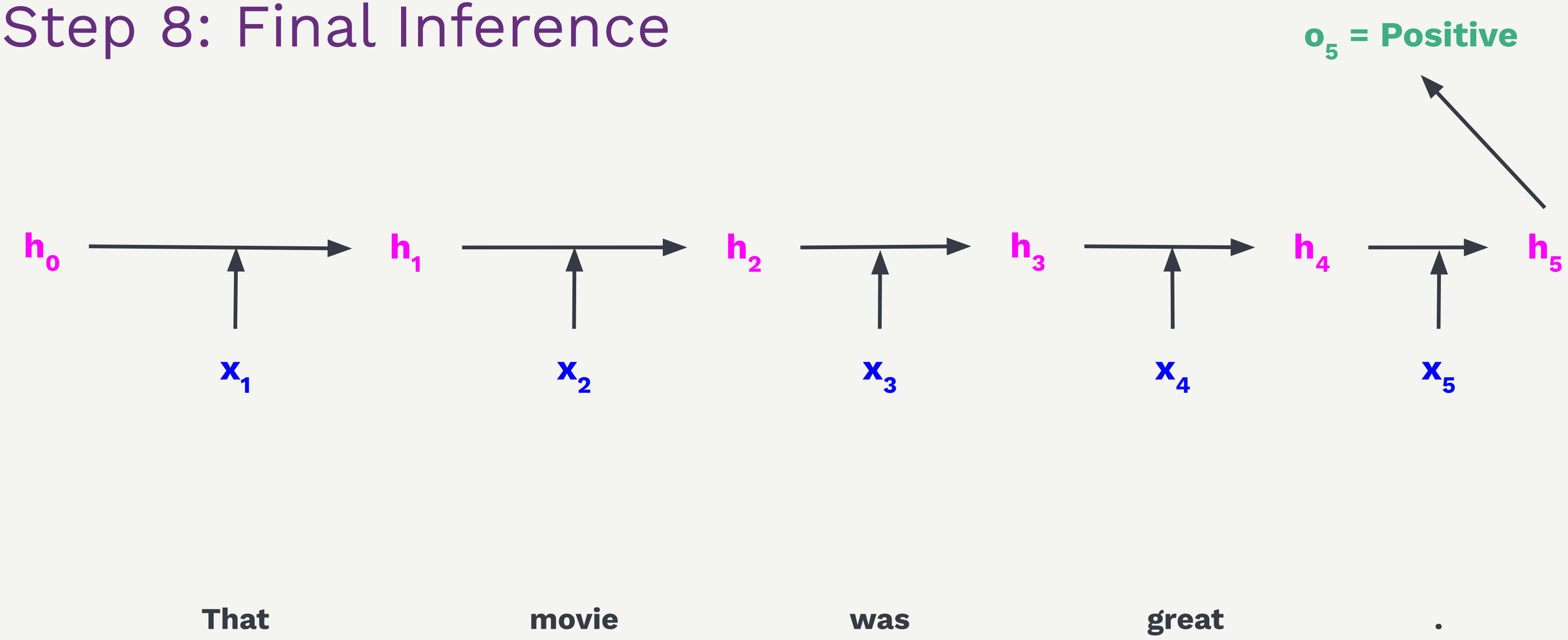
Mila

RNN for Text Classification

Gradient Descent: Repeat steps 5-7

RNN for Text Classification

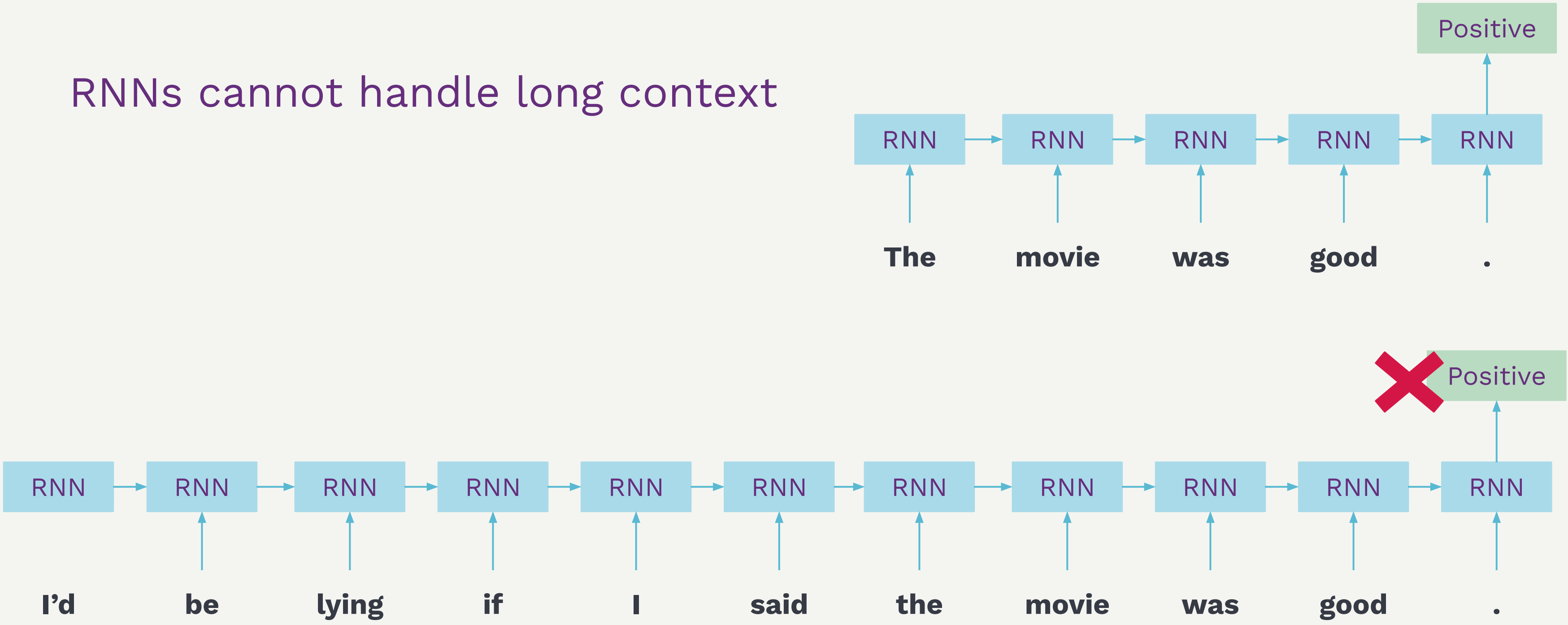
Step 8: Final Inference



Advanced RNNs: LSTMs and Attention

Long Short-Term Memory (LSTMs)

RNNs cannot handle long context



Mila

Long Short-Term Memory (LSTMs)

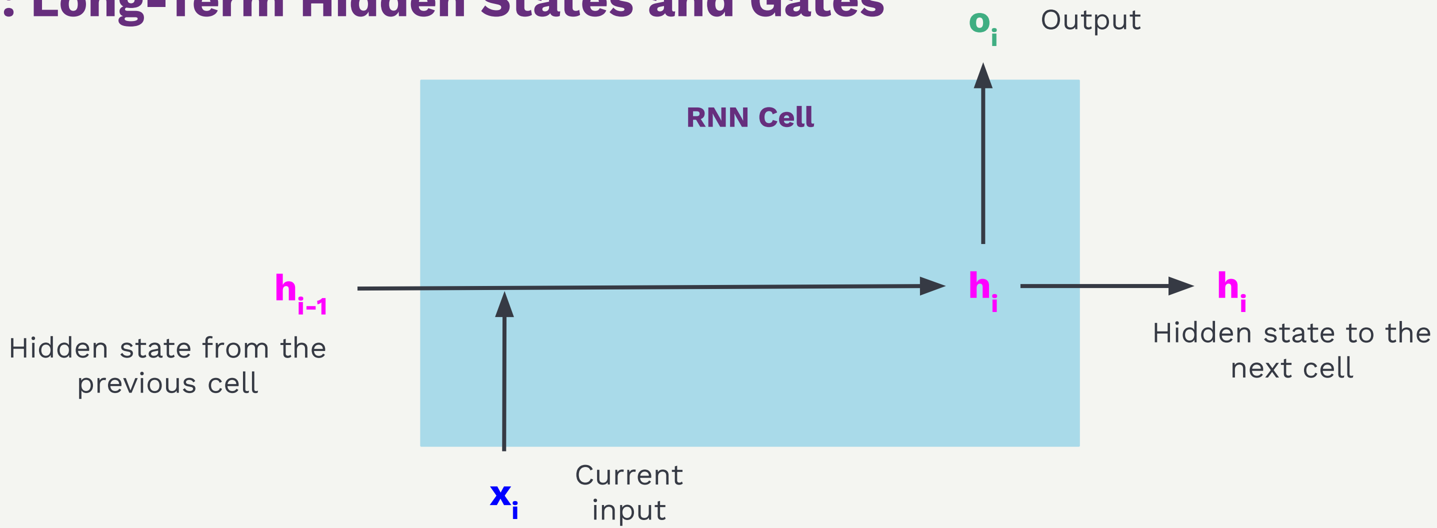
RNNs cannot handle long context

Solution: Long-Term Hidden States and Gates

Long Short-Term Memory (LSTMs)

RNNs cannot handle long context

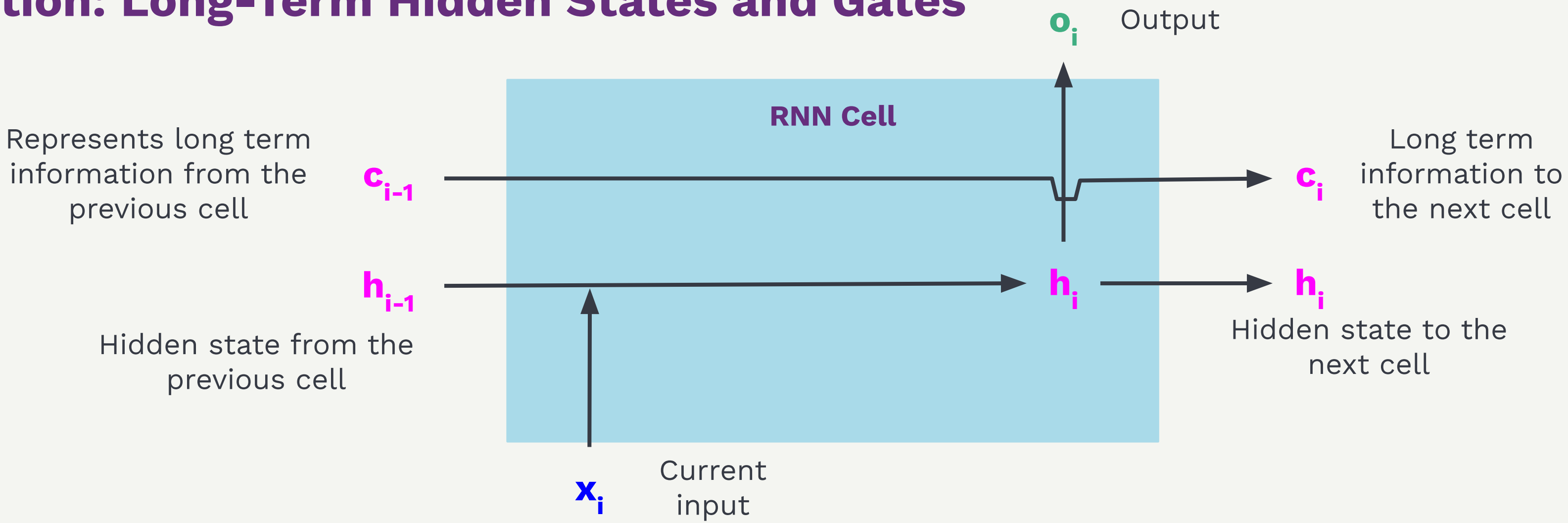
Solution: Long-Term Hidden States and Gates



Long Short-Term Memory (LSTMs)

RNNs cannot handle long context

Solution: Long-Term Hidden States and Gates

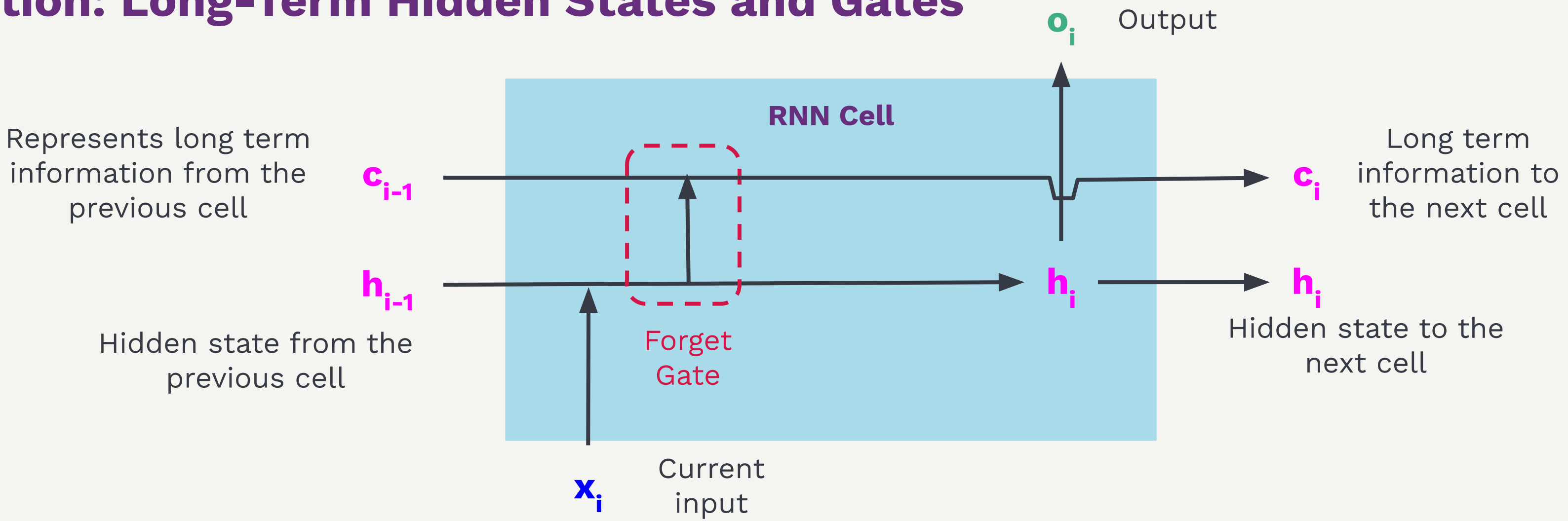


Mila

Long Short-Term Memory (LSTMs)

RNNs cannot handle long context

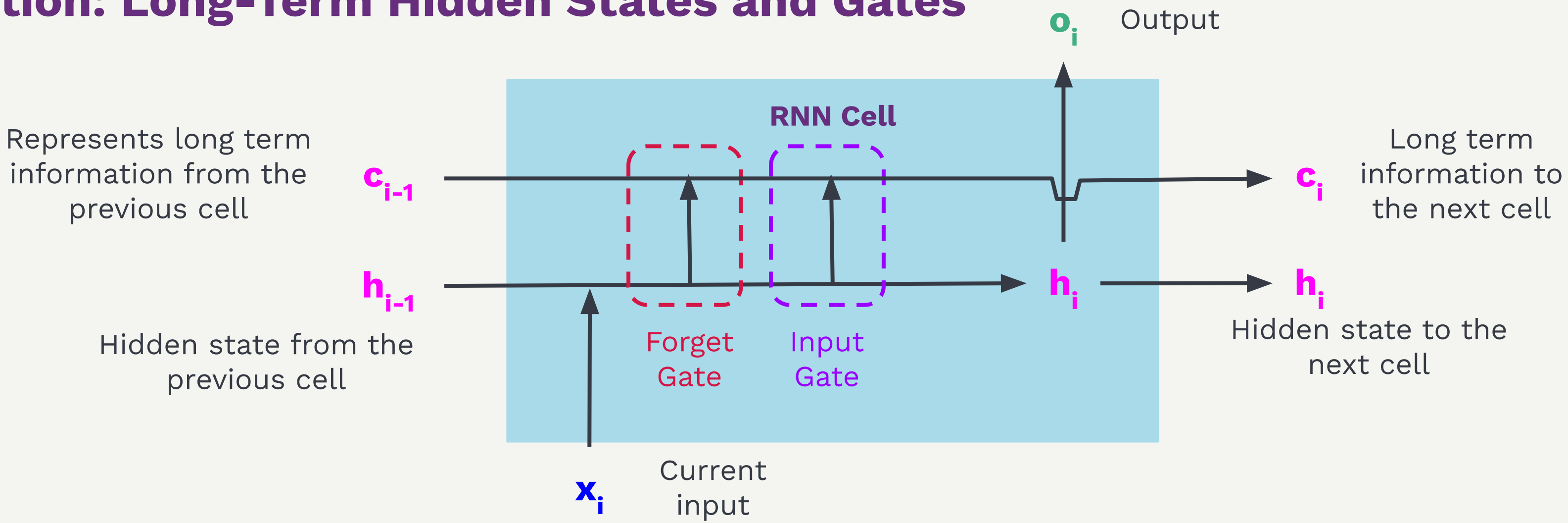
Solution: Long-Term Hidden States and Gates



Long Short-Term Memory (LSTMs)

RNNs cannot handle long context

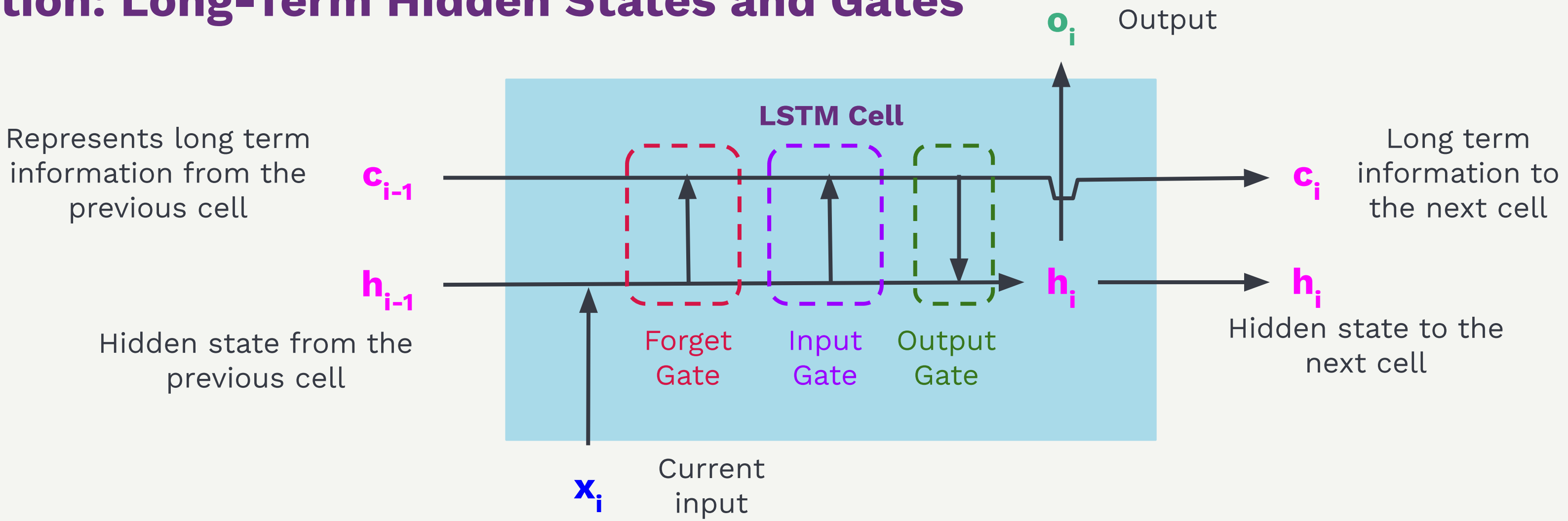
Solution: Long-Term Hidden States and Gates



Long Short-Term Memory (LSTMs)

RNNs cannot handle long context

Solution: Long-Term Hidden States and Gates



Attention

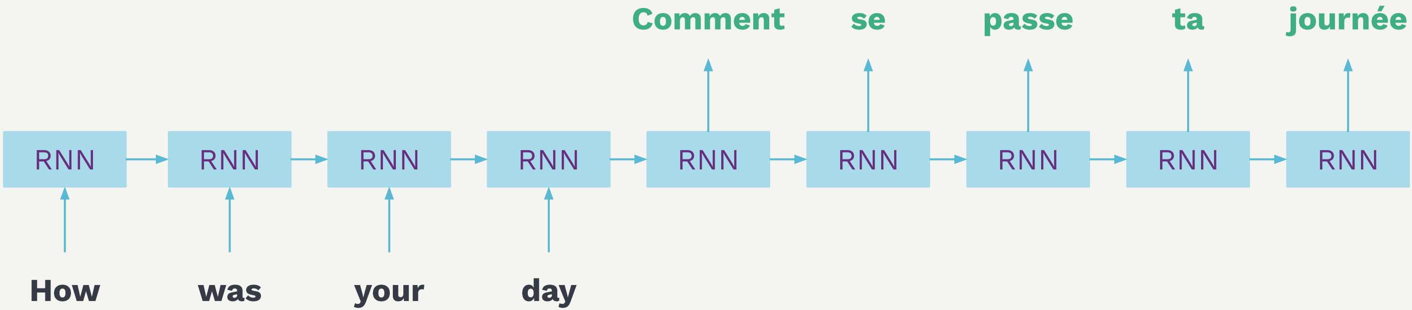
RNNs and LSTMs use the information about the complete sentence at all times.

Attention

RNNs and LSTMs use the information about the complete sentence at all times.

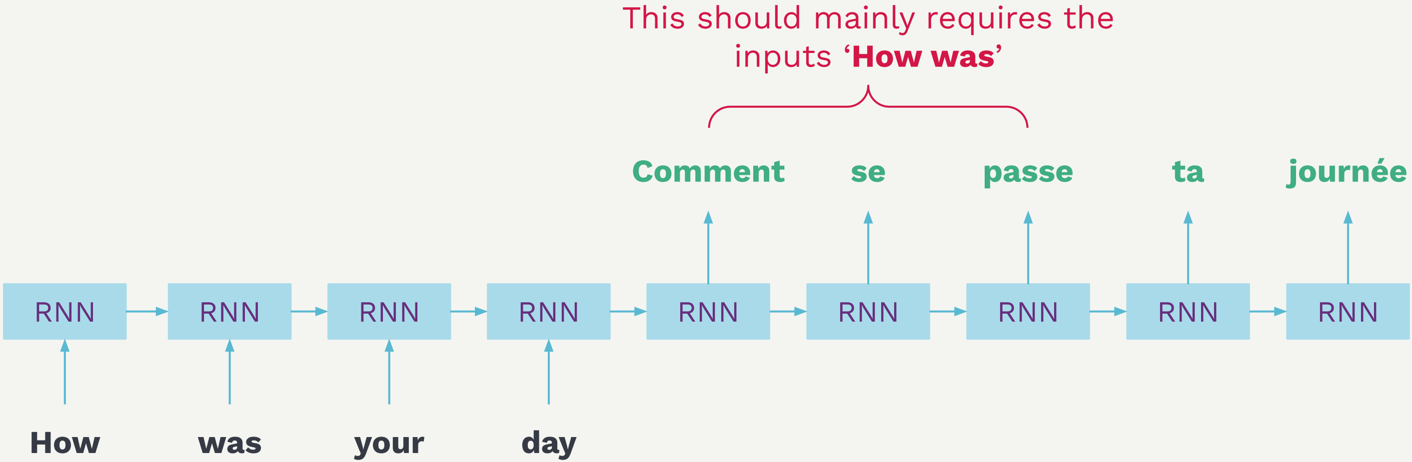
But is that really necessary?

Attention

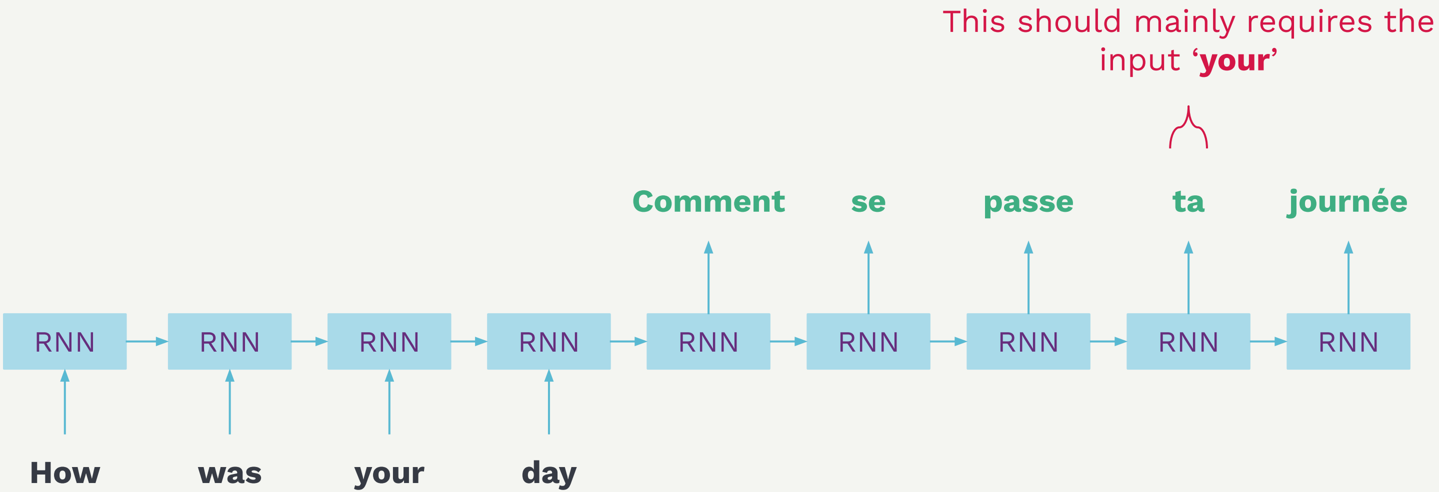


Comment
se
passe
ta
journée
RNN
RNN
RNN
RNN
RNN
RNN
RNN
RNN
RNN
How
was
your
day

Attention

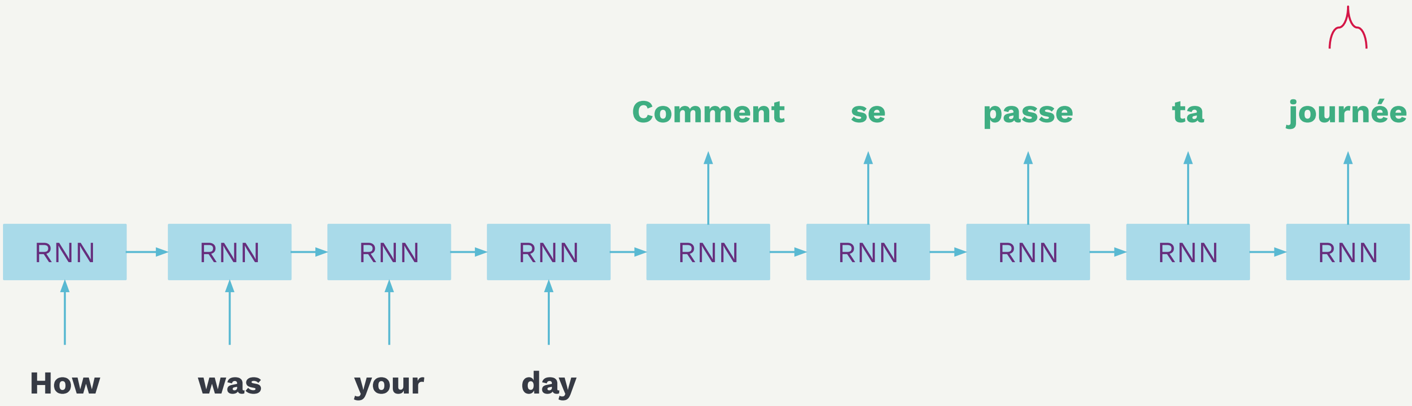

This should mainly requires the inputs ‘**How was**’
Comment
se
passe
ta
journée
RNN
RNN
RNN
RNN
RNN
RNN
RNN
RNN
RNN
How
was
your
day

Attention


This should mainly requires the input 'your'
Comment
se
passe
ta
journée
RNN
RNN
RNN
RNN
RNN
RNN
RNN
RNN
RNN
How
was
your
day

Attention

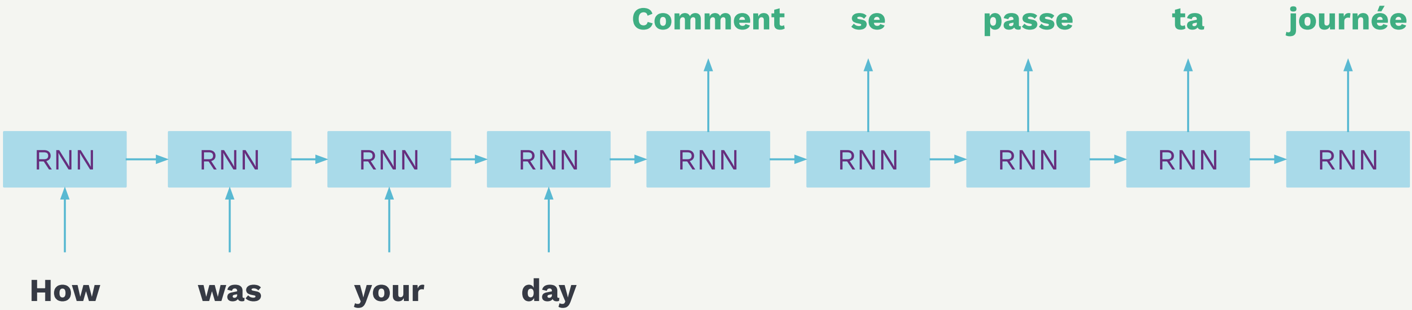
This should mainly requires the input ‘**day**’



Mila

Attention

Not all information is always needed, and ‘focusing’/‘attending’ on certain information more can help the language model

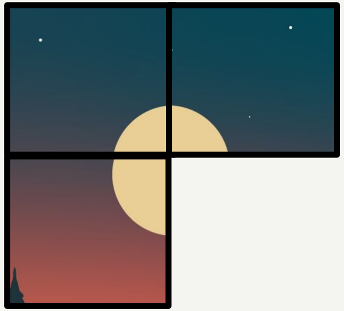


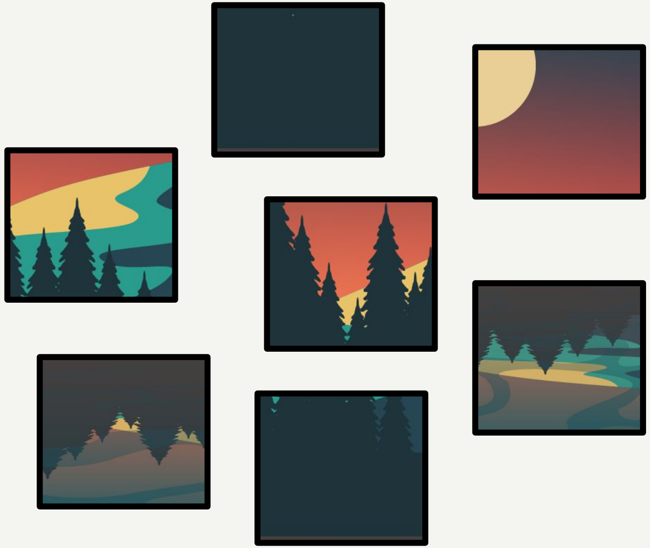
Mila

Attention

A mechanism to allow neural networks to dynamically focus on various parts of the input based on the current task.

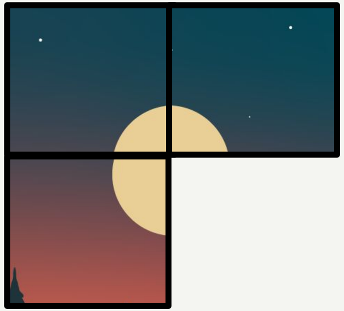
Attention





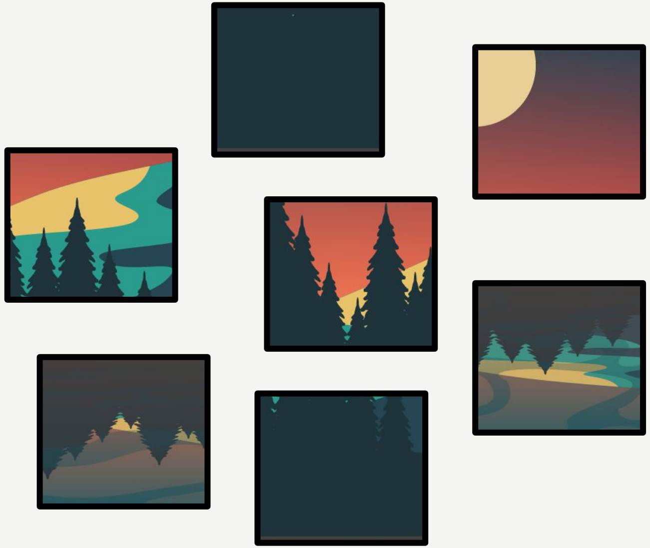
Mila

Attention

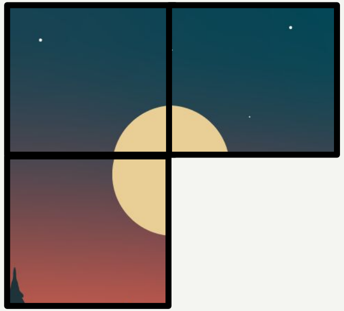


Query

I want a piece with yellow color

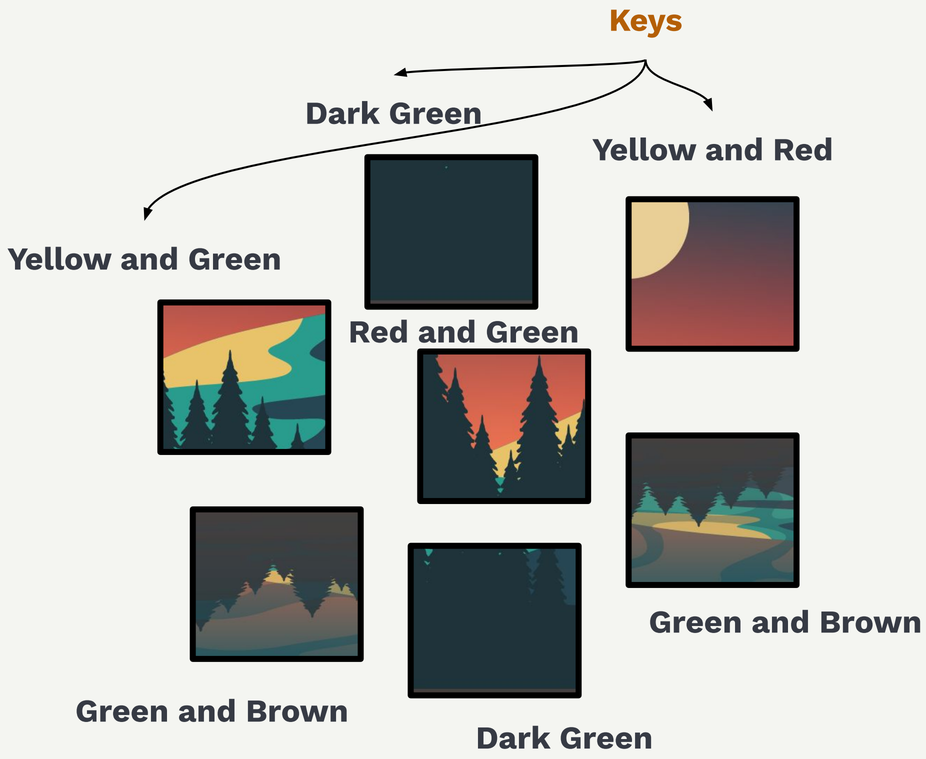


Attention

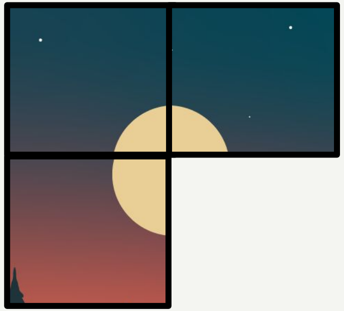


Query

I want a piece with yellow color

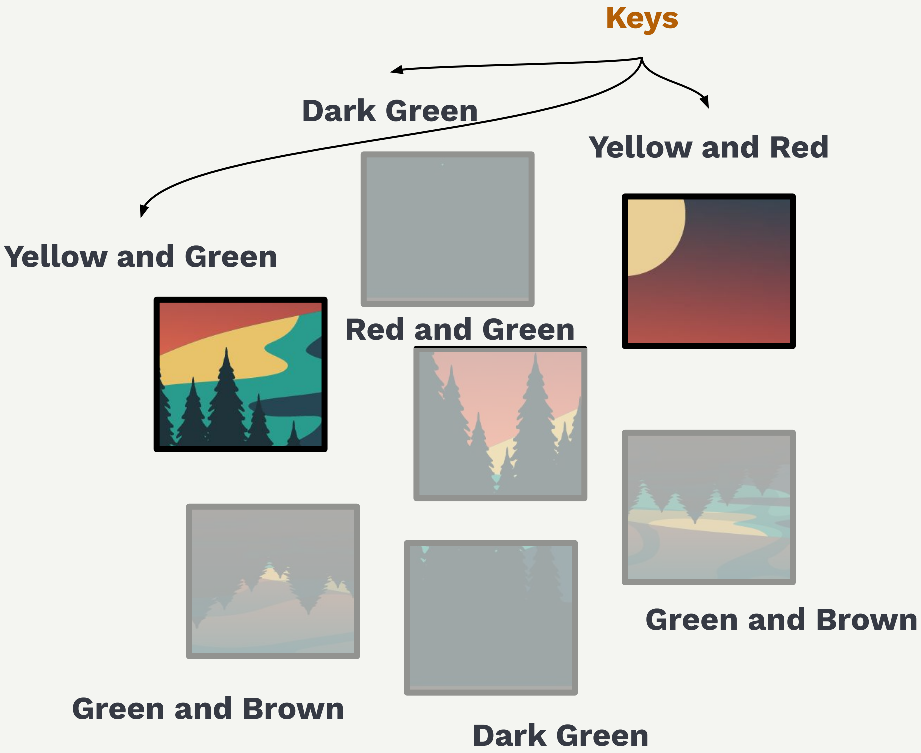


Attention



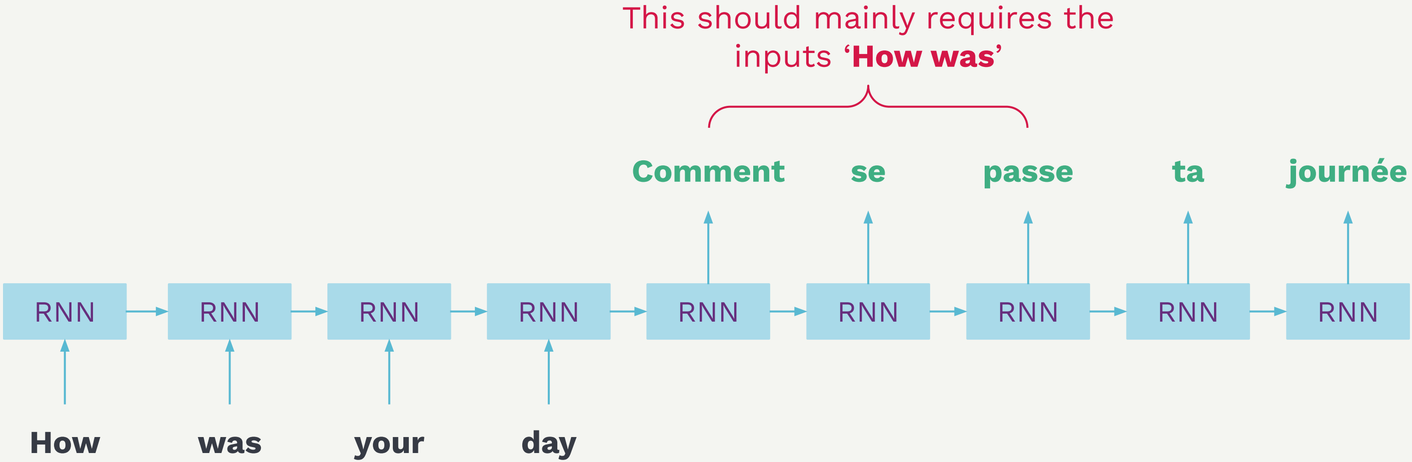
Query

I want a piece with yellow color



Attention

Attention



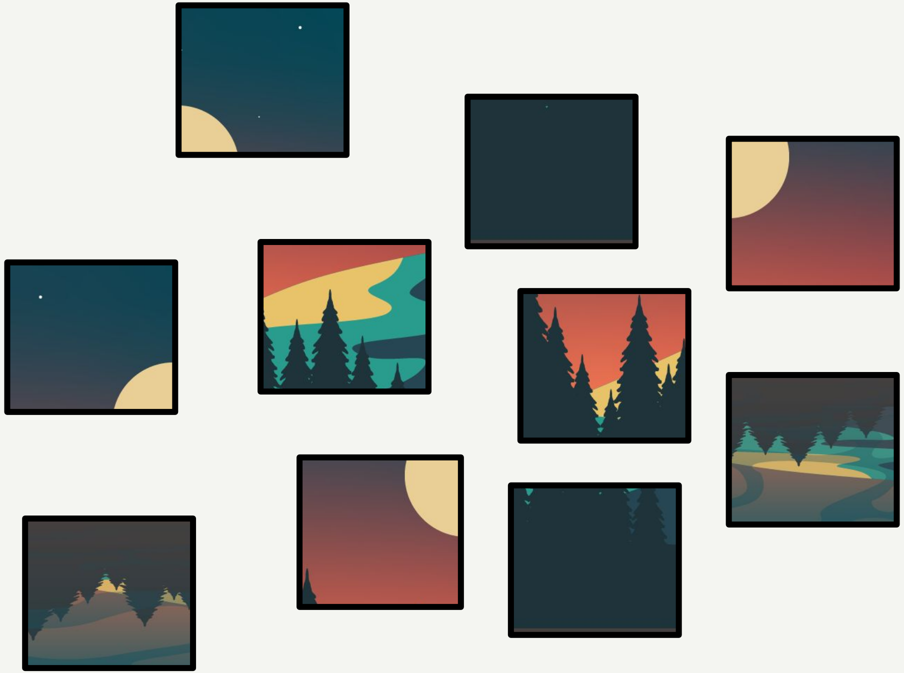
This should mainly requires the inputs '**How was**'
Comment
se
passe
ta
journée
RNN
RNN
RNN
RNN
RNN
RNN
RNN
RNN
RNN
How
was
your
day

Self-Attention and Transformers

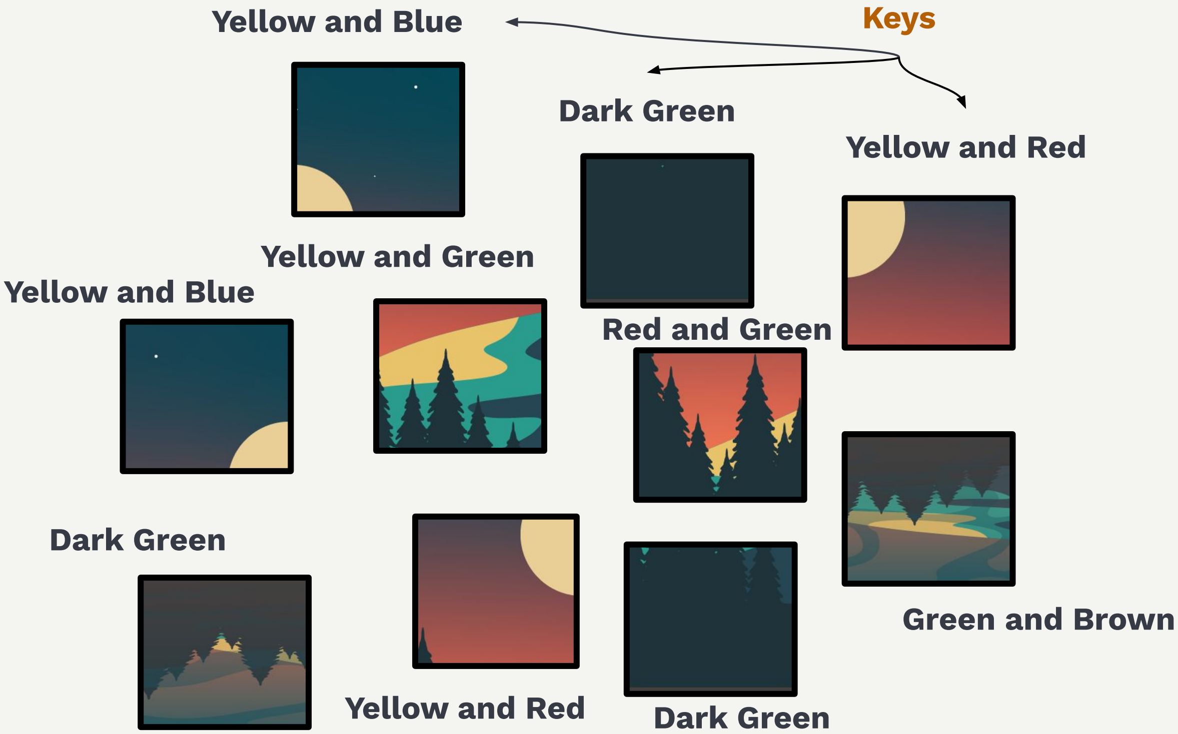
Self-Attention

Self-attention is assigning importance to various words in context of other words in the same sentence, capturing dependencies between different words in the sentence.

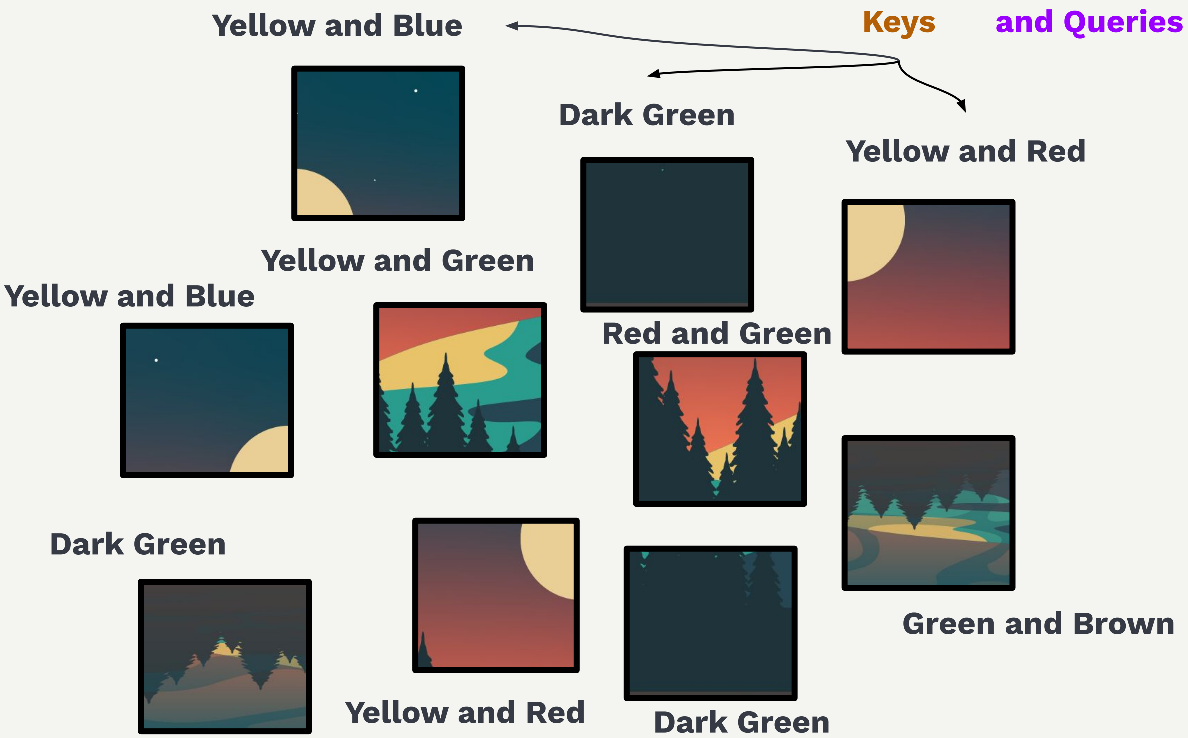
Self-Attention



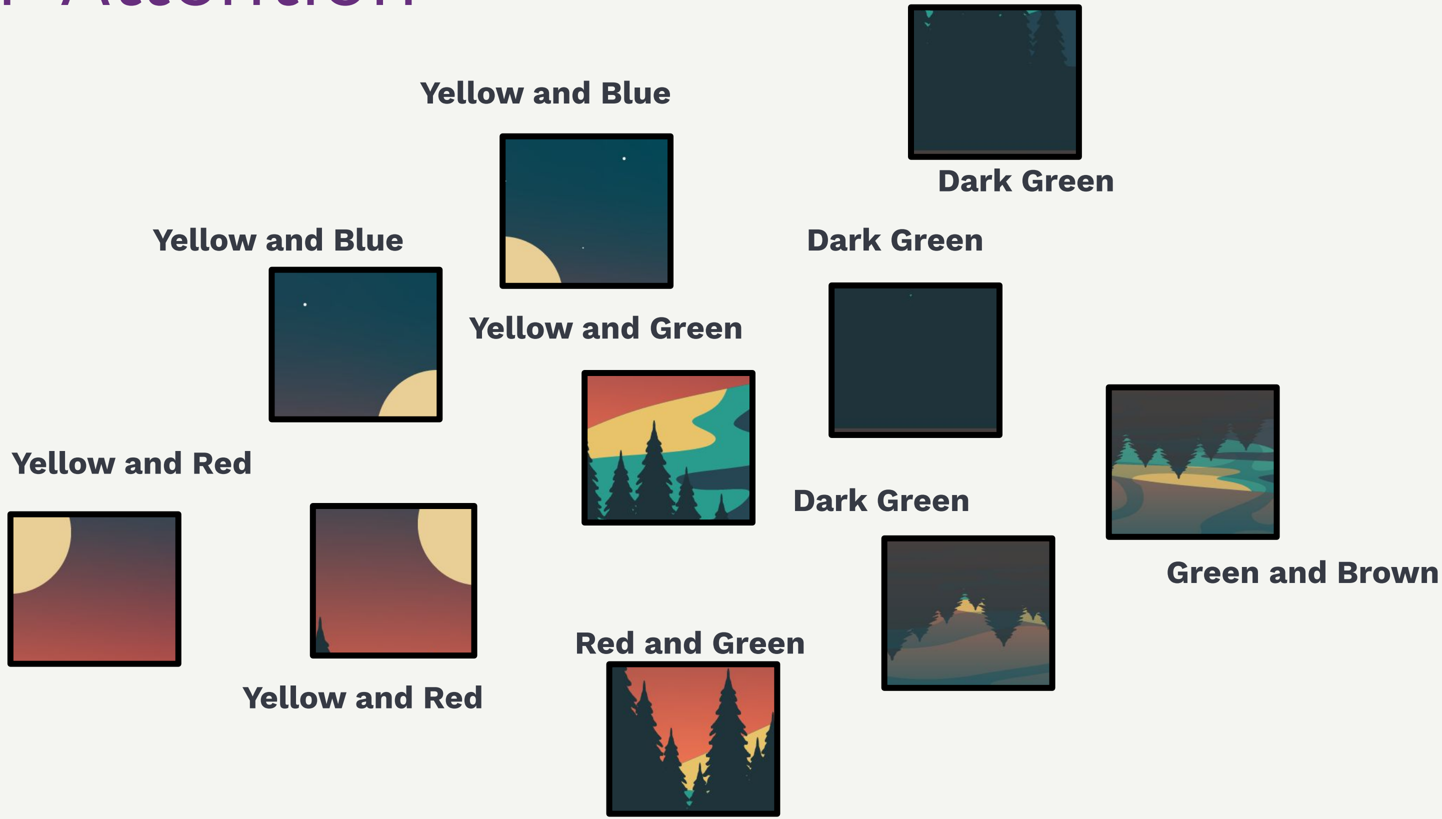
Self-Attention



Self-Attention

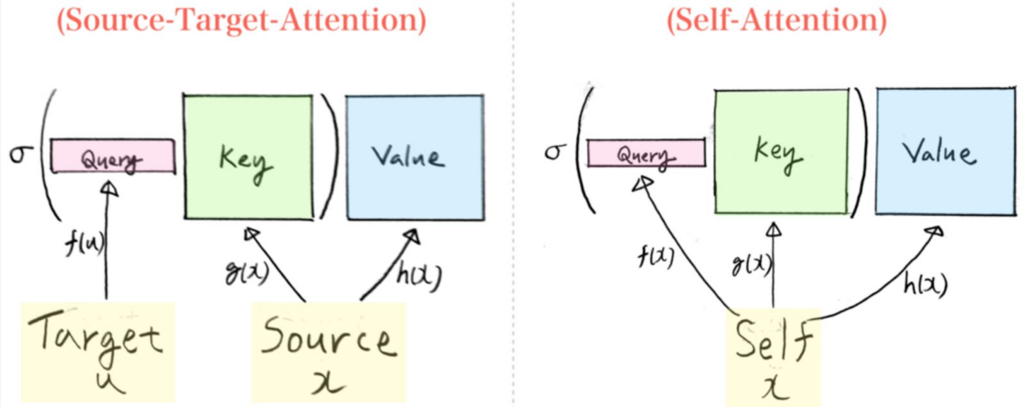


Self-Attention

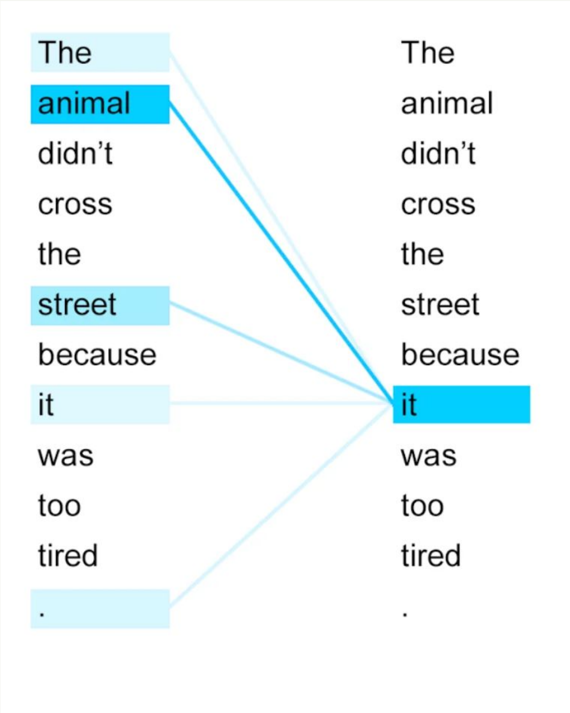


Self-Attention

Self-attention is assigning importance to various words in context of other words in the same sentence, capturing dependencies between different words in the sentence.



Self-Attention



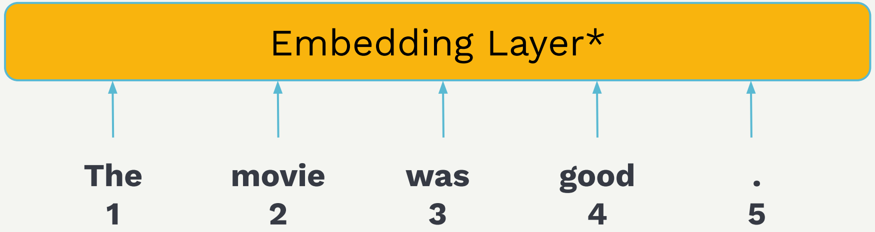
Transformers

The movie was good .

Transformers

The	movie	was	good	.
1	2	3	4	5

Transformers

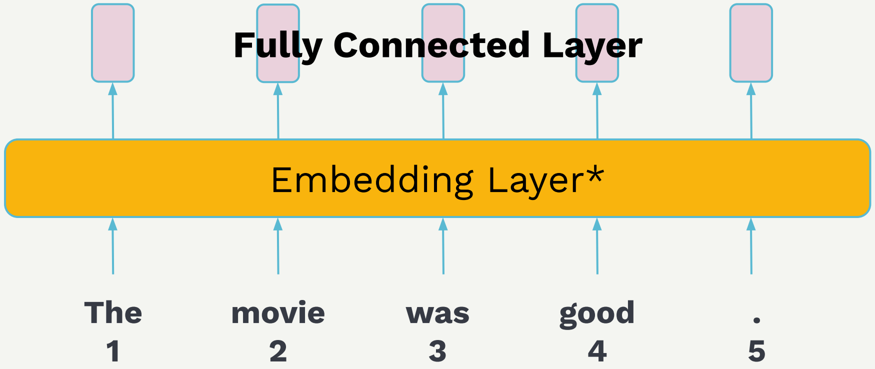


***Word embedding AND Position Embedding**

Mila

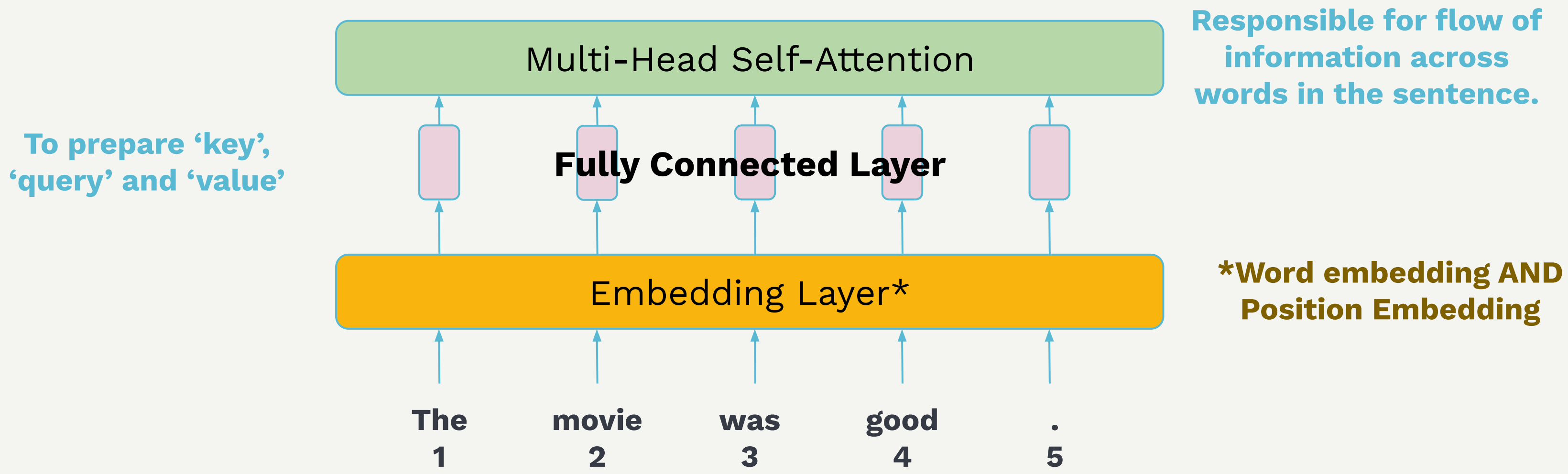
Transformers

To prepare 'key', 'query' and 'value'



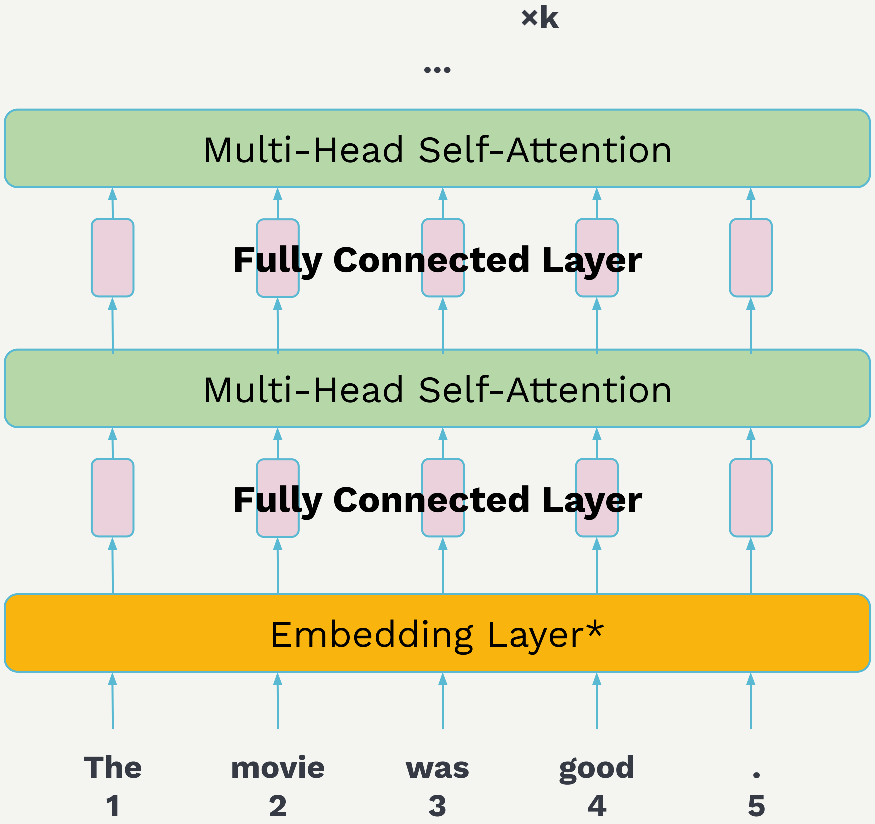
*Word embedding AND Position Embedding

Transformers



Mila

Transformers



*Word embedding AND Position Embedding

Sneak Peek

- **Large Language Models (LLMs) - ChatGPT, Claude, etc.**
- **Responsible NLP**

In the next class