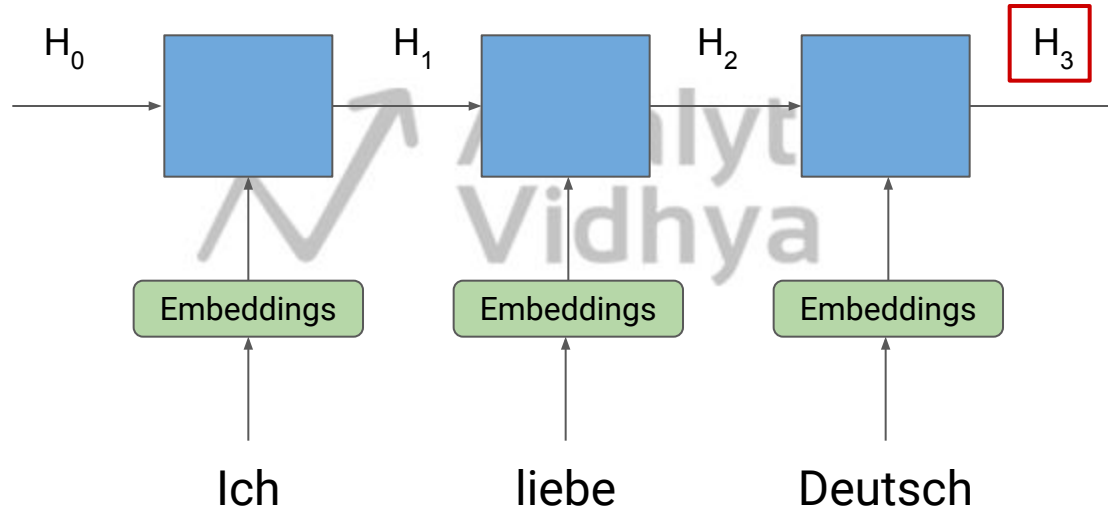


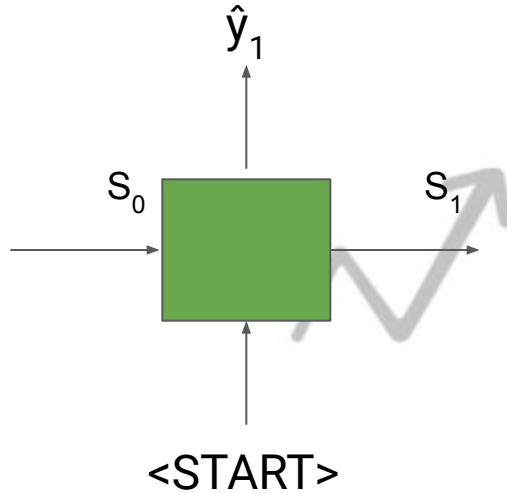
Understanding Attention Mechanism

Encoder during Training Phase



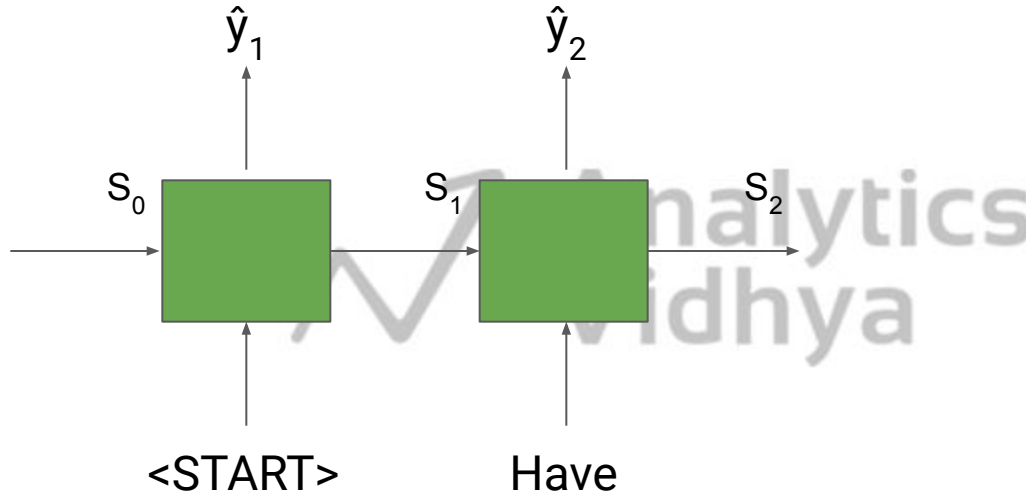
Decoder during Training Phase

Target: Have your cake



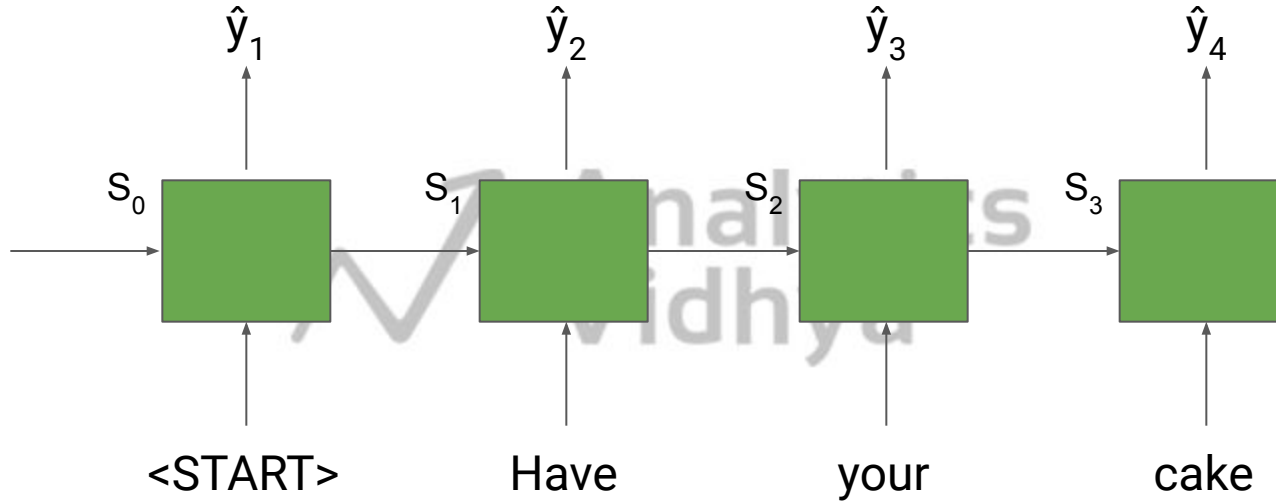
Decoder during Training Phase

Target: Have your cake



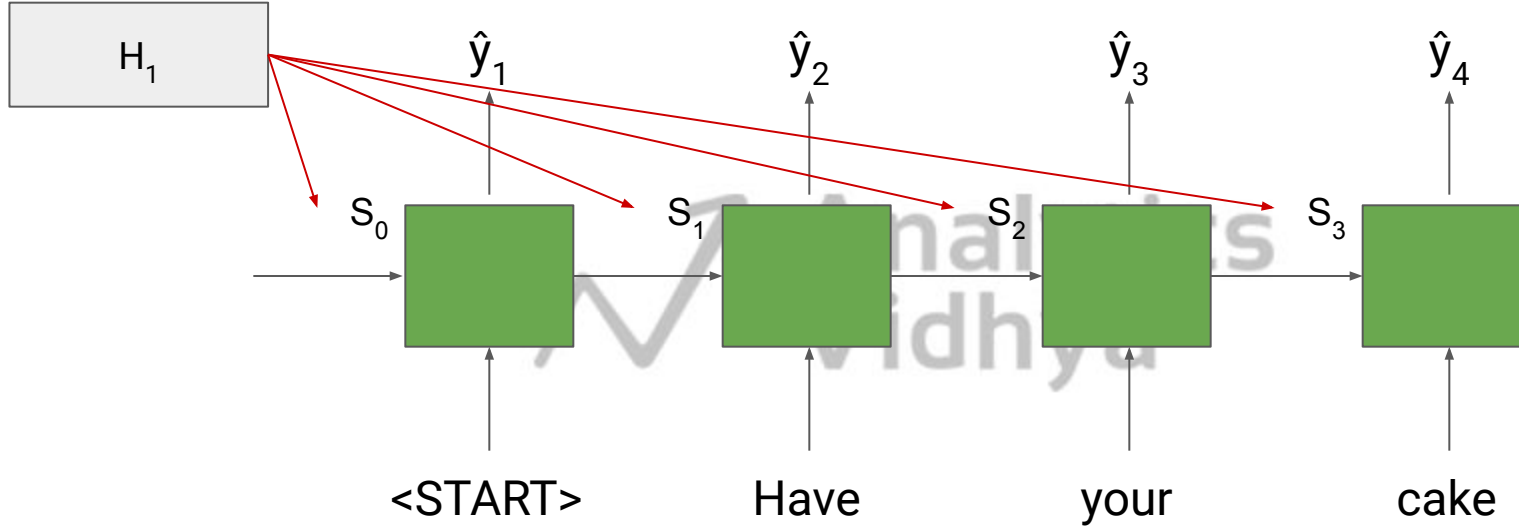
Decoder during Training Phase

Target: Have your cake



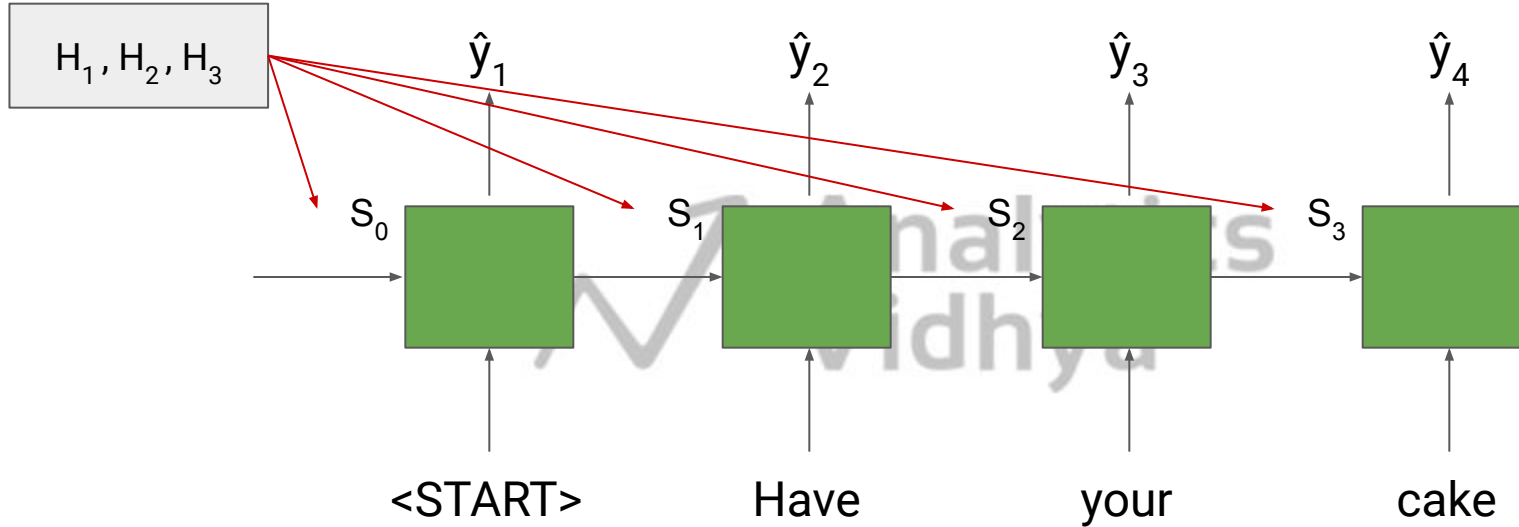
Decoder during Training Phase

Target: Have your cake



Decoder during Training Phase

Target: Have your cake



Attention Mechanism

Encoder Hidden
States

H_1

H_2

H_3



Attention Mechanism

Encoder Hidden
States

H_1

H_2

H_3

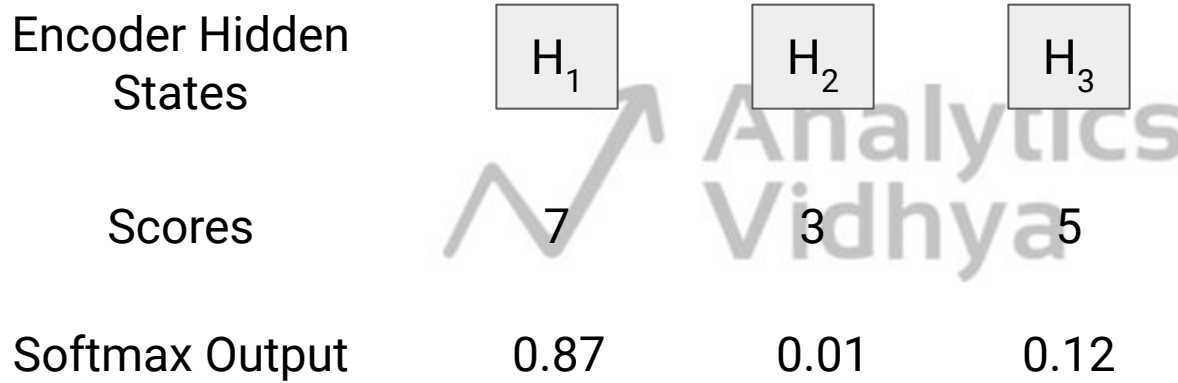
Scores

7

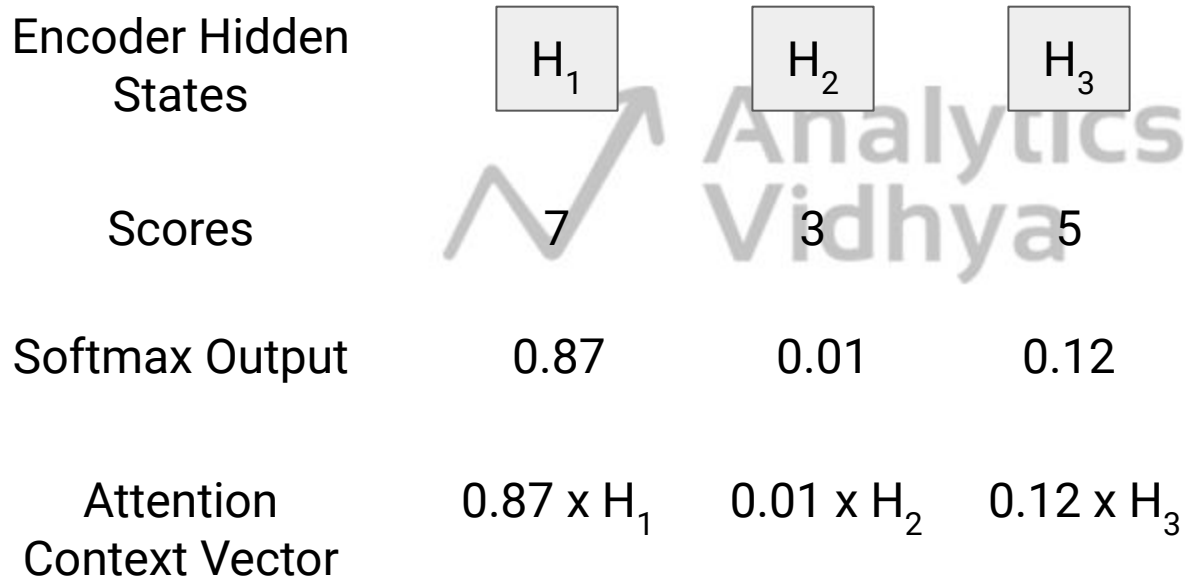
3

5

Attention Mechanism



Attention Mechanism



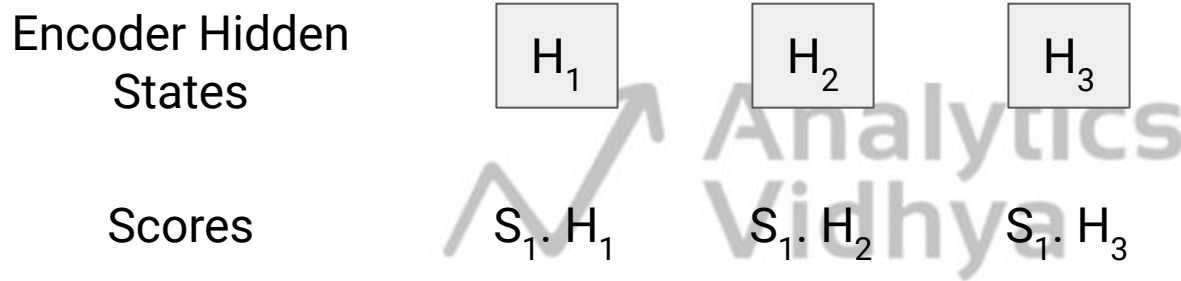
The diagram illustrates the Attention Mechanism process. It shows three encoder hidden states, H_1 , H_2 , and H_3 , each in a box. Below each box is a score: 7 for H_1 , 3 for H_2 , and 5 for H_3 . A large, faint watermark 'Analytics Vidhya' is visible in the background. A grey line graph with an upward arrow is also present. Below the scores are the softmax outputs: 0.87 for H_1 , 0.01 for H_2 , and 0.12 for H_3 . At the bottom, the attention context vector is calculated as the weighted sum of the hidden states: $0.87 \times H_1$, $0.01 \times H_2$, and $0.12 \times H_3$.

Encoder Hidden States	H_1	H_2	H_3
Scores	7	3	5
Softmax Output	0.87	0.01	0.12
Attention Context Vector	$0.87 \times H_1$	$0.01 \times H_2$	$0.12 \times H_3$

Attention Mechanism

Encoder Hidden States	H_1	H_2	H_3
Scores	7	3	5
Softmax Output	0.87	0.01	0.12
Attention Context Vector	$0.87 \times H_1$	$0.01 \times H_2$	$0.12 \times H_3$

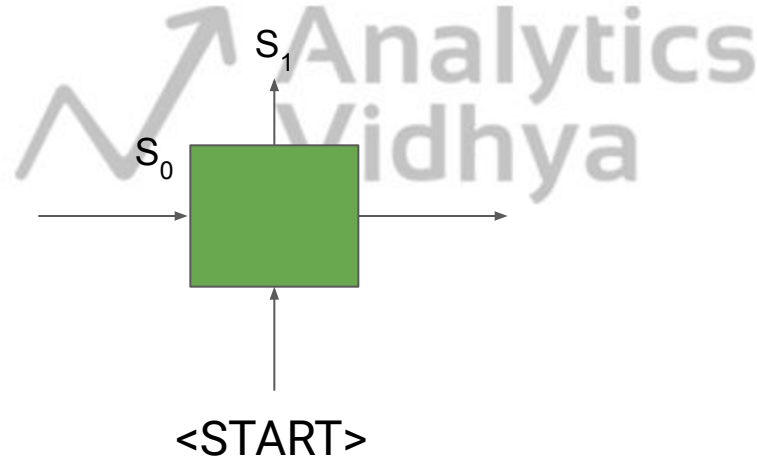
Attention Score Calculation



Decoder

Attention Scores = $f(S_1, H_t)$

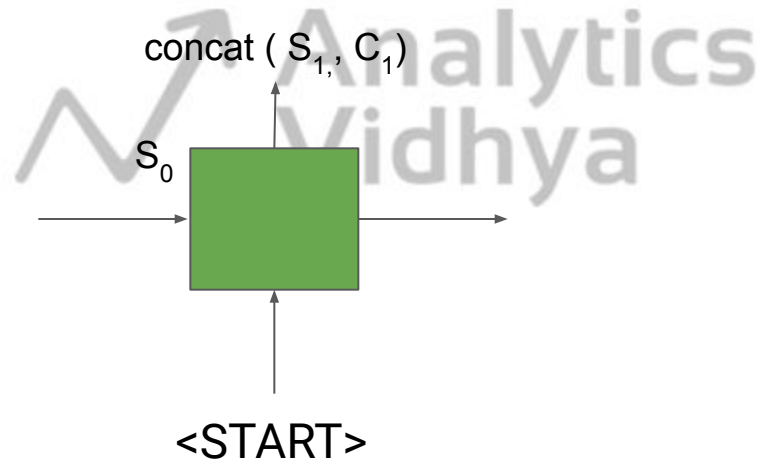
Context Vector = C_1



Decoder

Attention Scores = $f(S_1, H_t)$

Context Vector = C_1

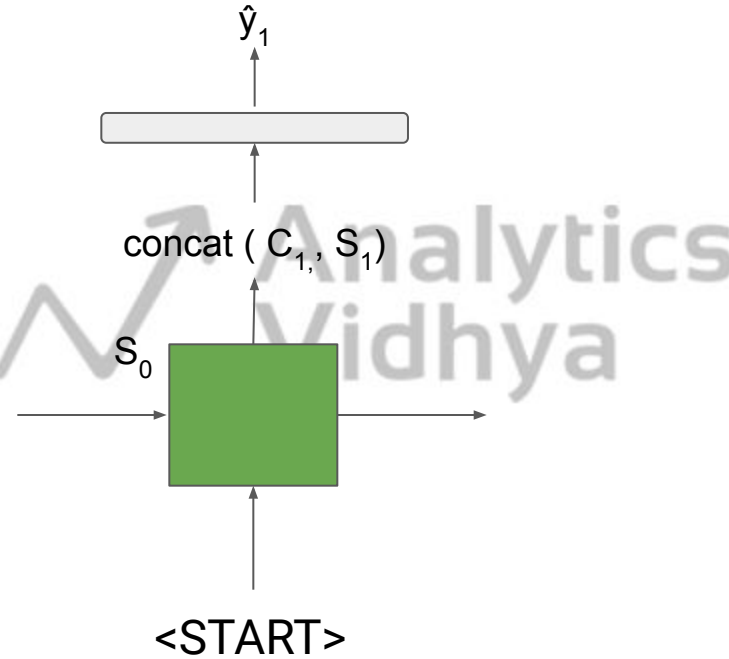


Decoder

Attention Scores = $f(S_1, H_t)$

Context Vector = C_1

$\hat{y}_1 = \text{Softmax}(\tanh([C_1 : S_1]))$

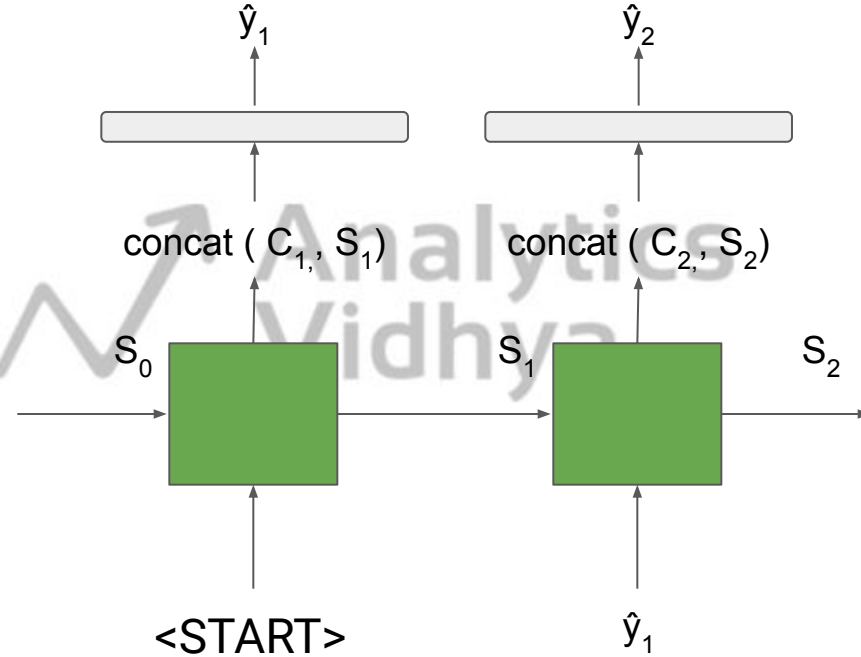


Decoder

Attention Scores = $f(S_2, H_t)$

Context Vector = C_2

$\hat{y}_2 = \text{Softmax}(\tanh([C_2 : S_2]))$





Thank You