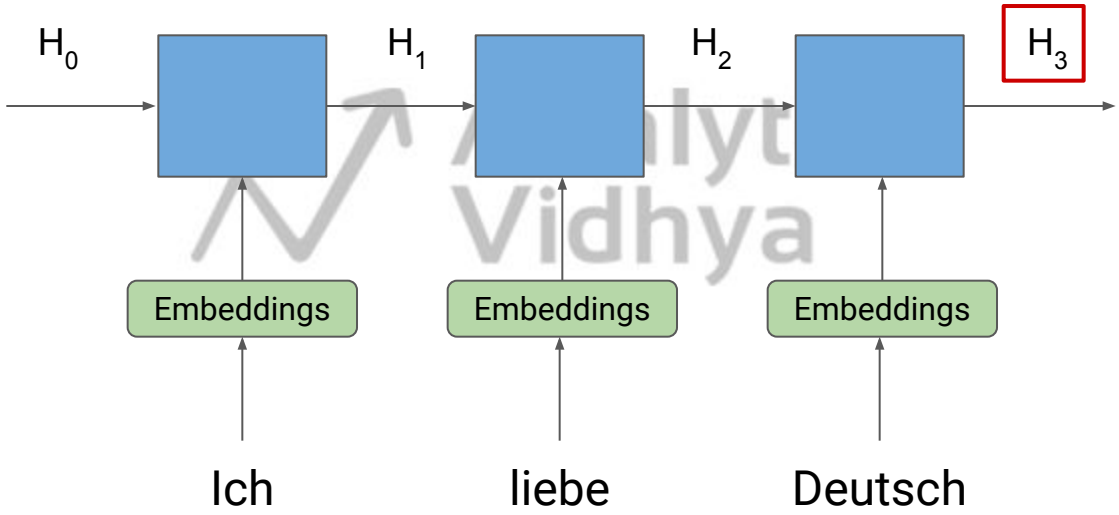
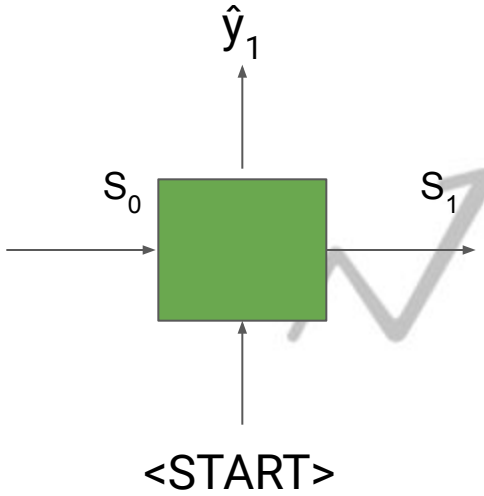
Understanding Attention Mechanism

Encoder during Training Phase



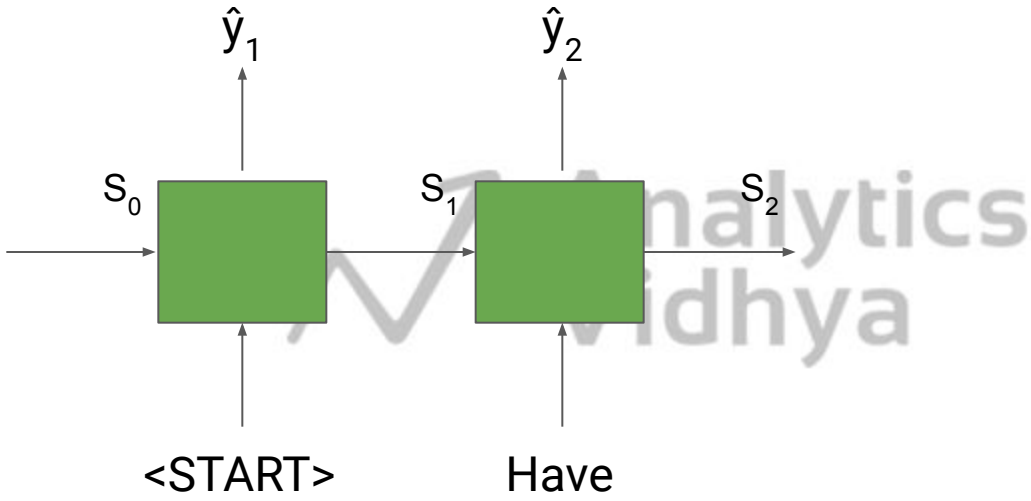
Decoder during Training Phase

Target: Have your cake



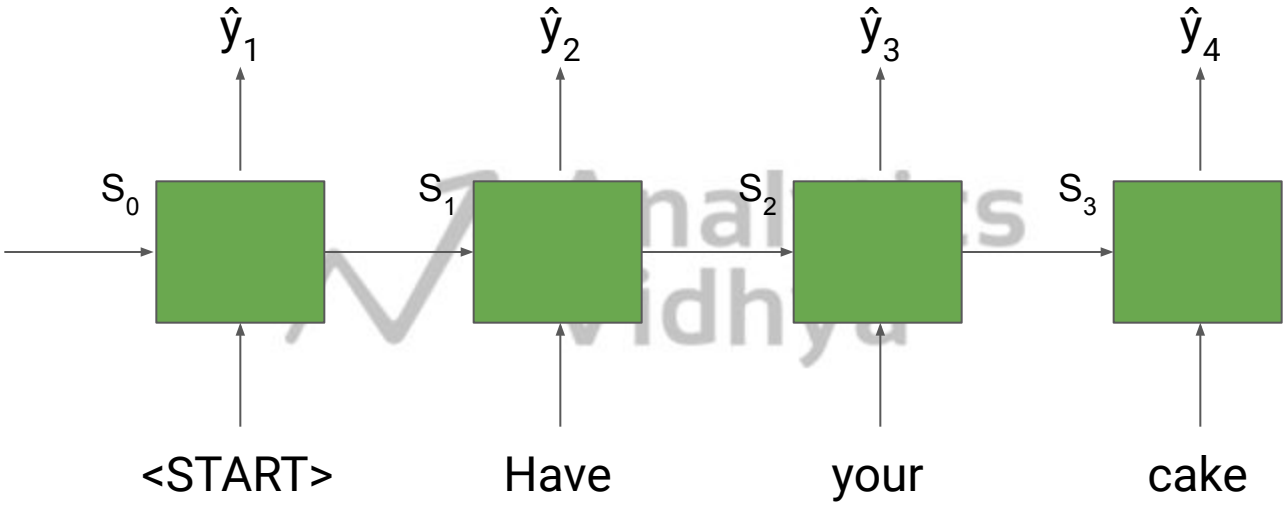
Decoder during Training Phase

Target: Have your cake

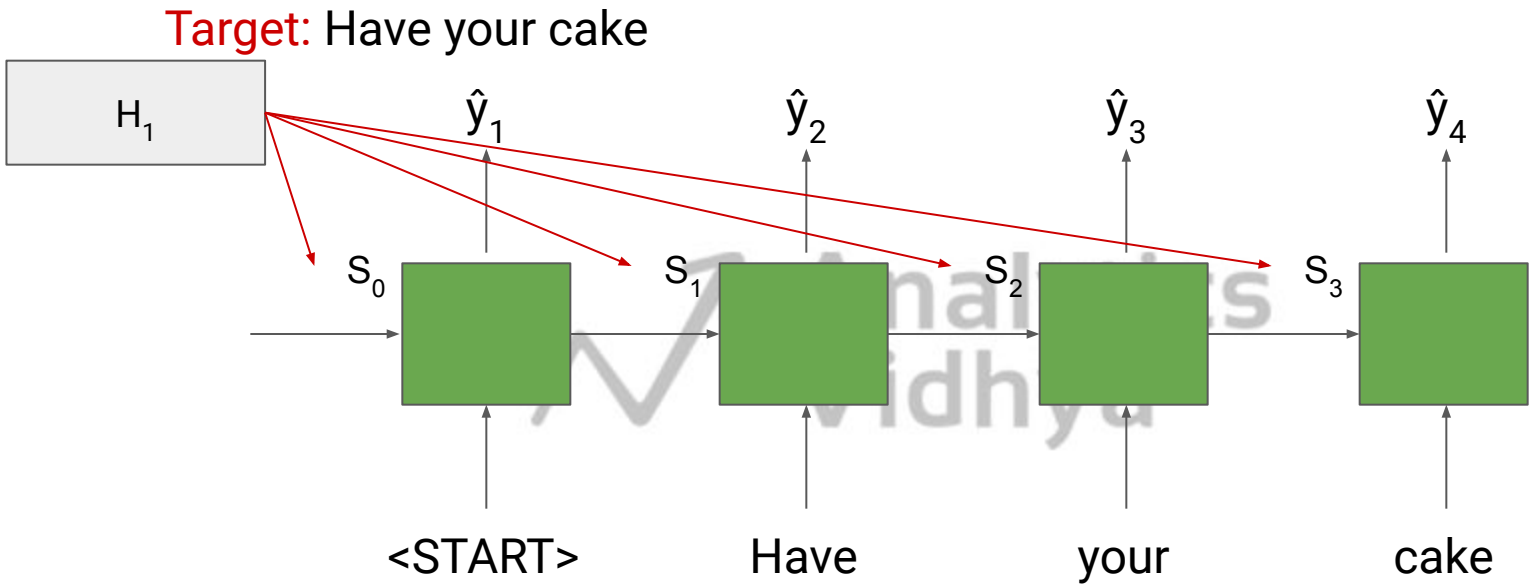


Decoder during Training Phase

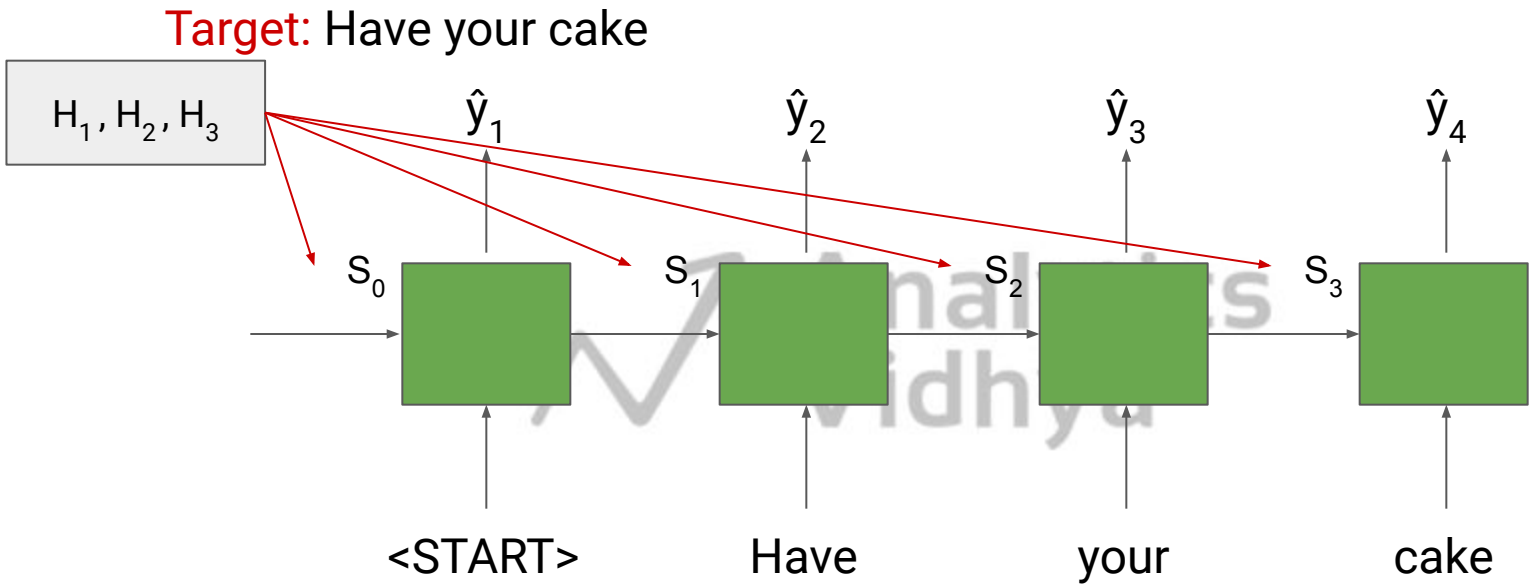
Target: Have your cake



Decoder during Training Phase

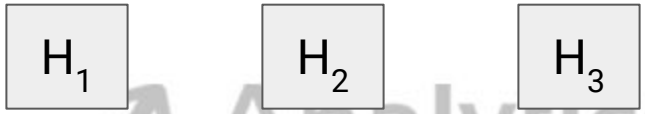


Decoder during Training Phase



Attention Mechanism

Encoder Hidden States



Attention Mechanism

Encoder Hidden States	H_1	H_2	H_3
Scores	7	3	5

Attention Mechanism

Encoder Hidden States	H_1	H_2	H_3
Scores	7	3	5
Softmax Output	0.87	0.01	0.12

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 x H_1	0.01 x H_2	0.12 x 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 x H_1	0.01 x H_2	0.12 x H_3

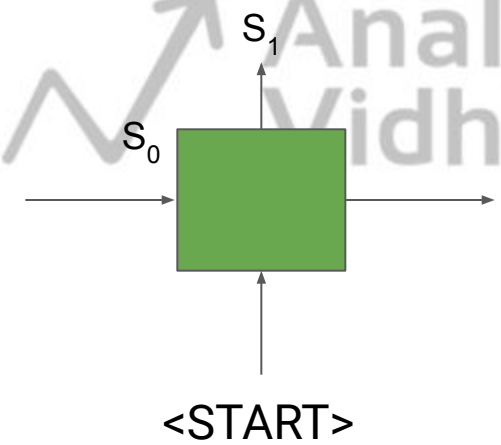
Attention Score Calculation

Encoder Hidden States	H_1	H_2	H_3
Scores	$S_1 . H_1$	$S_1 . H_2$	$S_1 . H_3$

Decoder

Attention Scores = $f(S_1, H_t)$

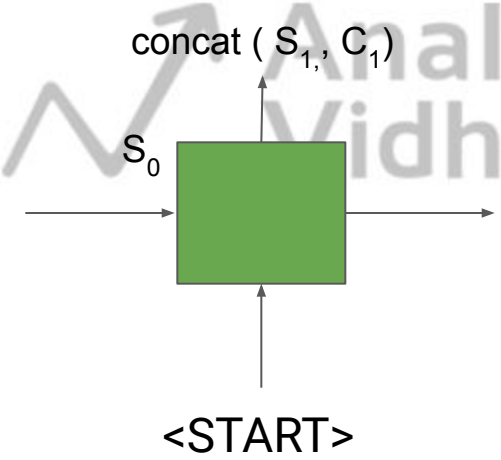
Context Vector = C_1



Decoder

Attention Scores = $f(S_1, H_t)$

Context Vector = C_1

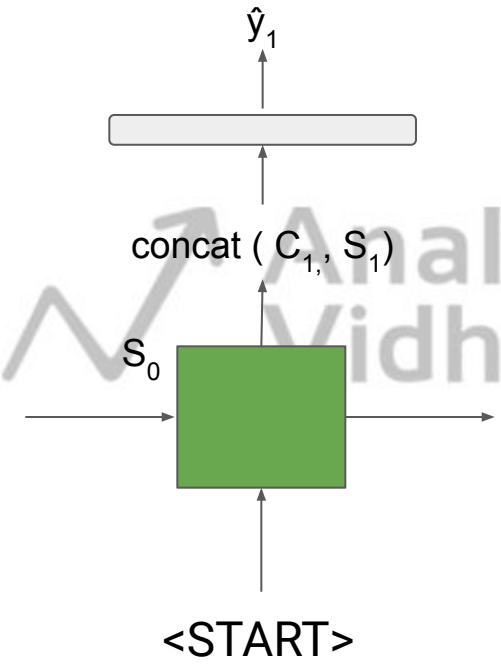


Decoder

Attention Scores = $f(S_1, H_t)$

Context Vector = C_1

$\hat{y}1 = \text{Softmax}(\tanh([\, C_1 : S_1 \,]))$

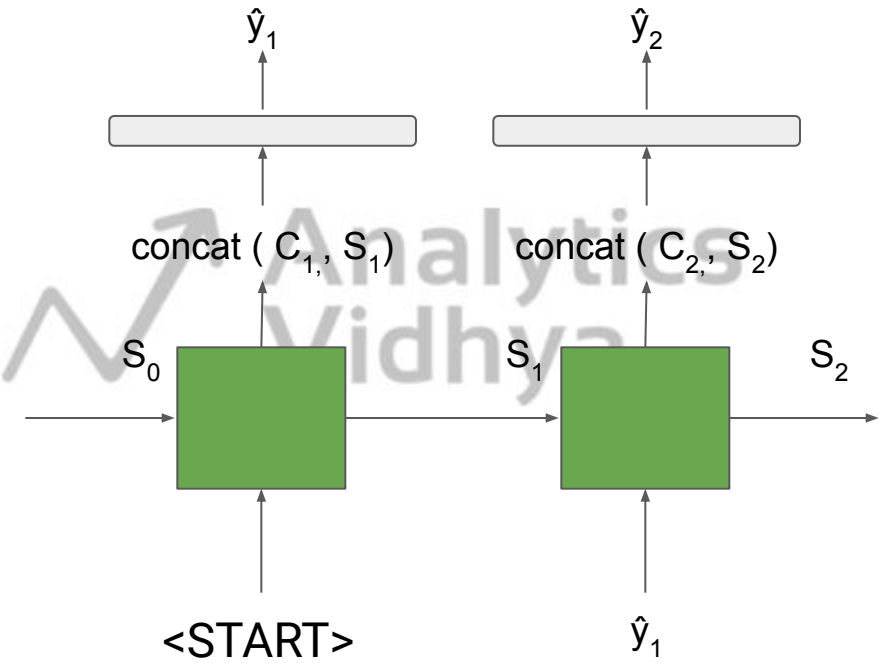


Decoder

Attention Scores = $f(S_2, H_t)$

Context Vector = C_2

$\hat{y}2 = \text{Softmax}(\tanh([\, C_2 : S_2 \,]))$



Thank You

Analytics
Vidhya