

Question Duplicates

How old are you?

Question Duplicates

How old are you?

What is your age?

Question Duplicates

How old are you? = What is your age?

Question Duplicates

How old are you? = What is your age?

Where are you from?

Question Duplicates

How old are you? = What is your age?

Where are you from? Where are you going?

Question Duplicates

How old are you?

=

What is your age?

Where are you from?

Where are you going?

Question Duplicates

How old are you? = What is your age?

Where are you from? ≠

Where are you going?

What do Siamese Networks learn?

What do Siamese Networks learn?

I am happy because I am learning



What do Siamese Networks learn?

I am happy because I am learning



Classification: categorize things

What do Siamese Networks learn?

I am happy because I am learning



Classification: categorize things

Siamese Networks: Identify similarity between things

What do Siamese Networks learn?

I am happy because I am learning



Classification: categorize things

Siamese Networks: Identify similarity between things



What is your age?
How old are you?

What do Siamese Networks learn?

I am happy because I am learning



Classification: categorize things

Siamese Networks: Identify similarity between things



What is your age?
How old are you?



Difference or
Similarity

What do Siamese Networks learn?

I am happy because I am learning



Classification: categorize things

Siamese Networks: Identify similarity between things



What is your age?
How old are you?



Difference or
Similarity

Siamese Networks in NLP

Siamese Networks in NLP

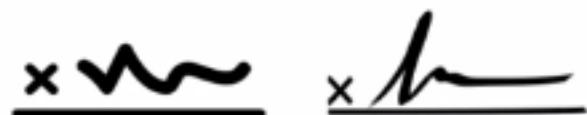


Handwritten checks

Siamese Networks in NLP



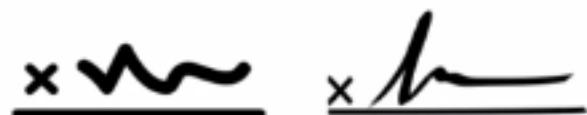
Handwritten checks



Siamese Networks in NLP



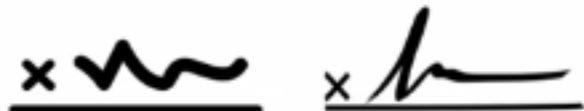
Handwritten checks



Siamese Networks in NLP



Handwritten checks



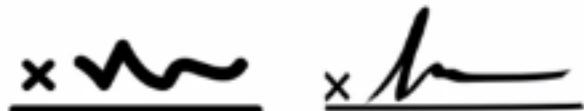
What is your age?
How old are you?

Question duplicates

Siamese Networks in NLP



Handwritten checks



What is your age?
How old are you?

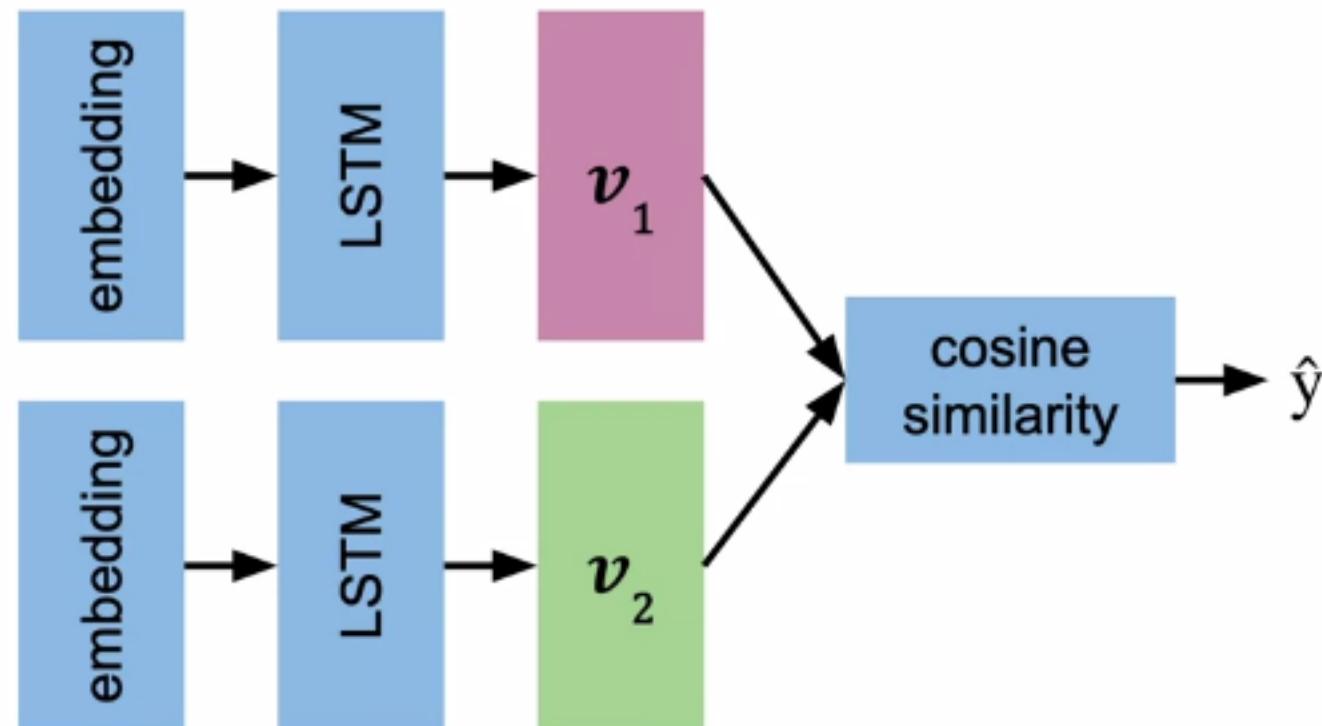
Question duplicates



Queries

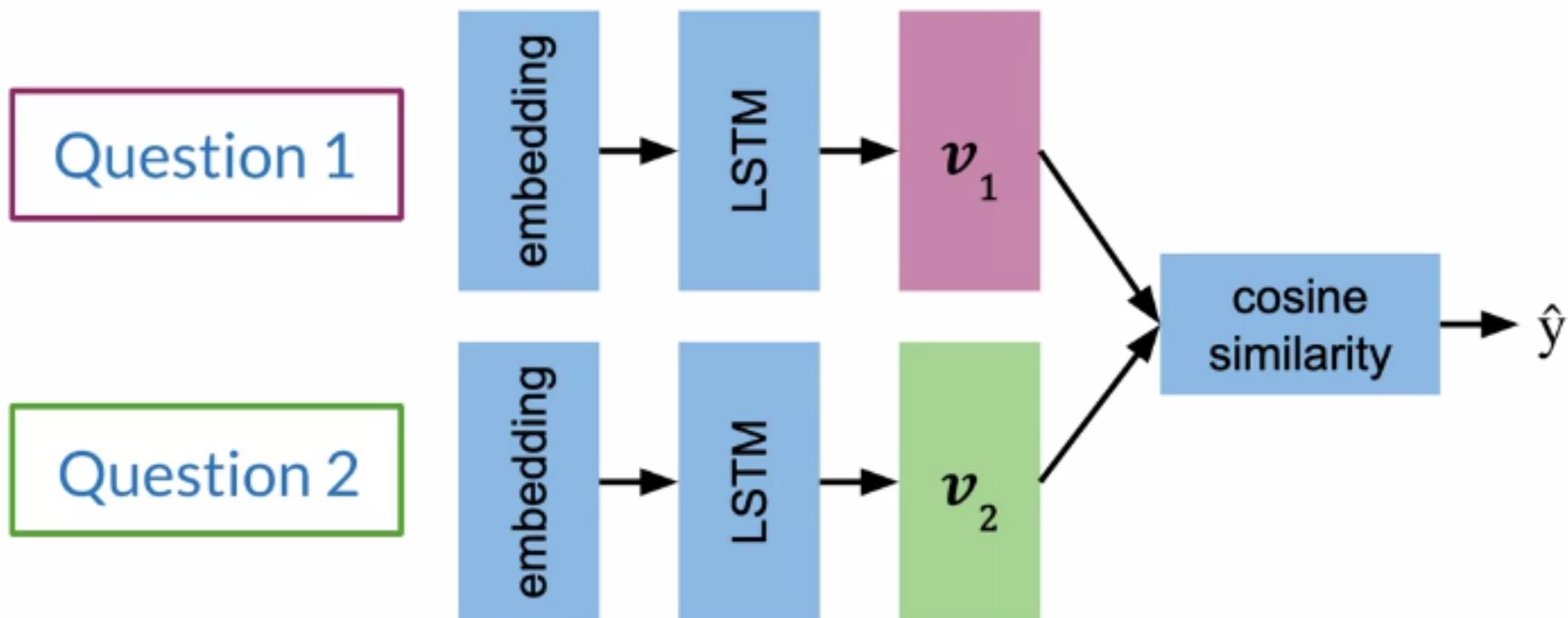
Model Architecture

Question 1



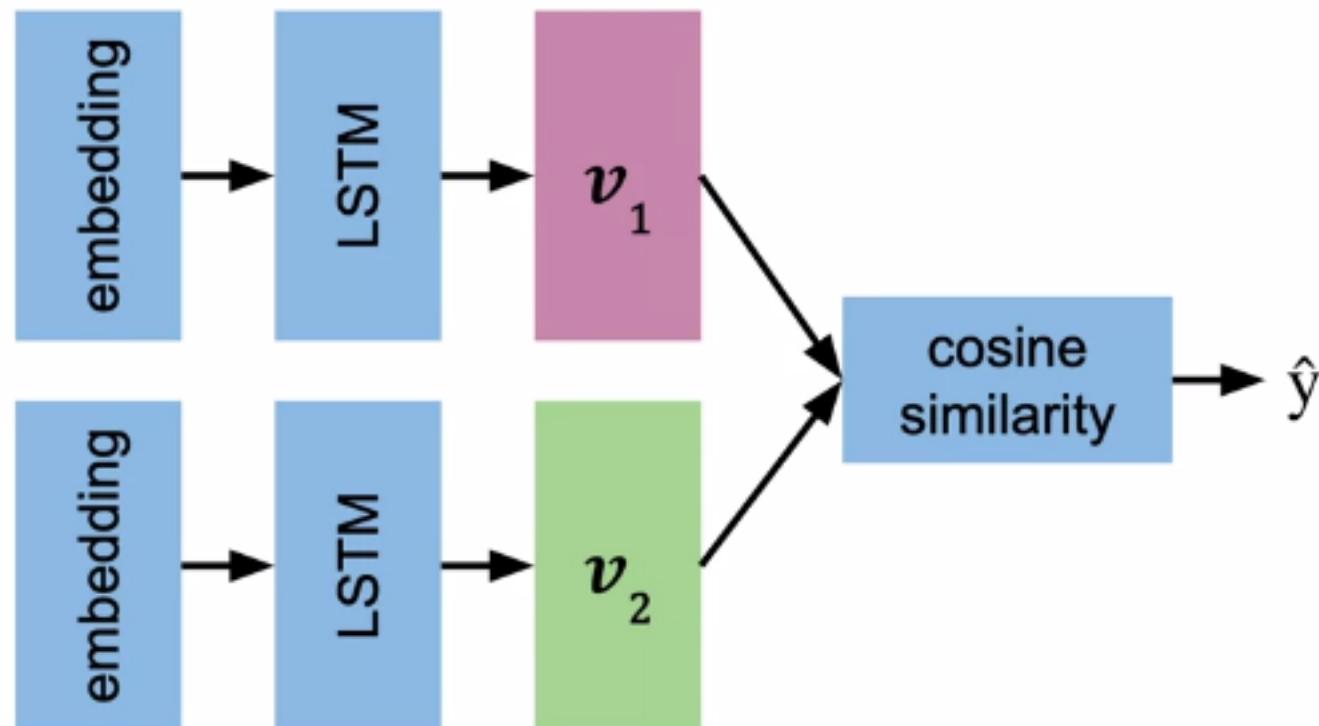
Question 2

Model Architecture



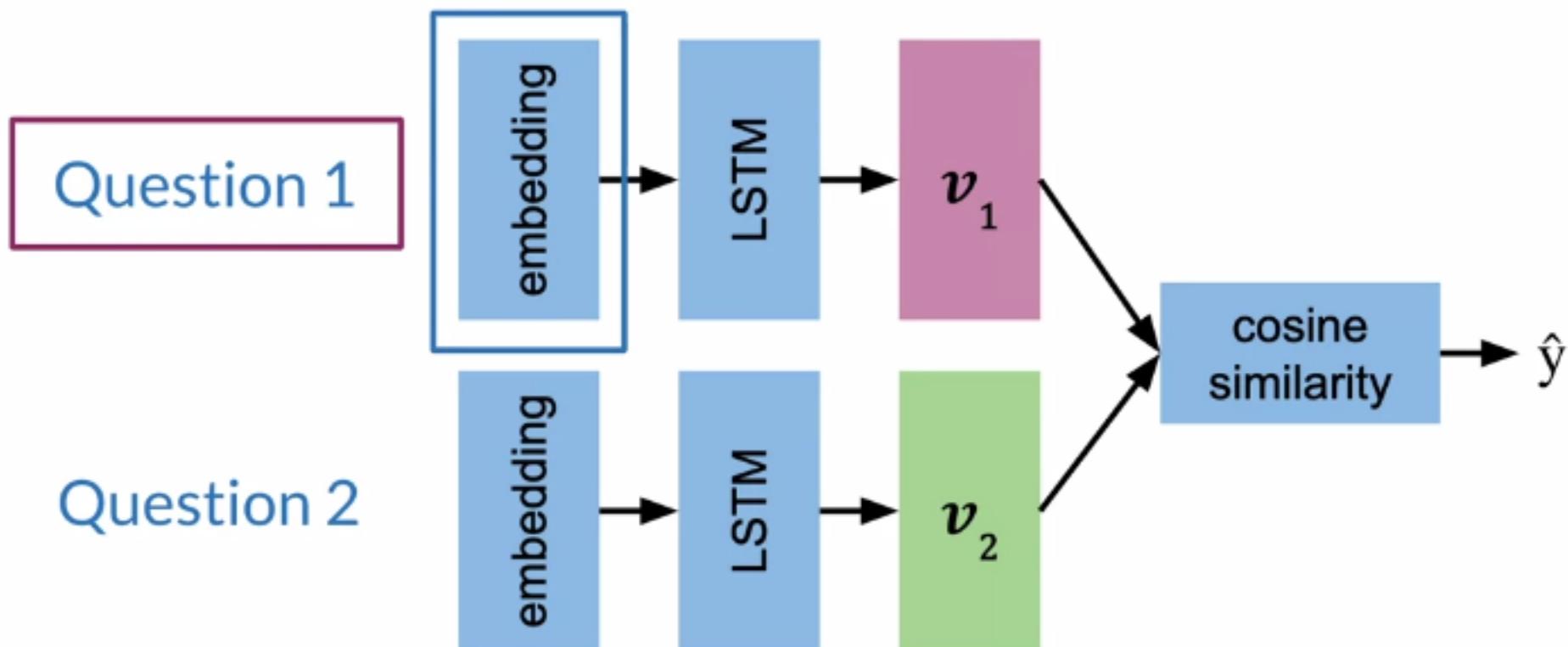
Model Architecture

Question 1



Question 2

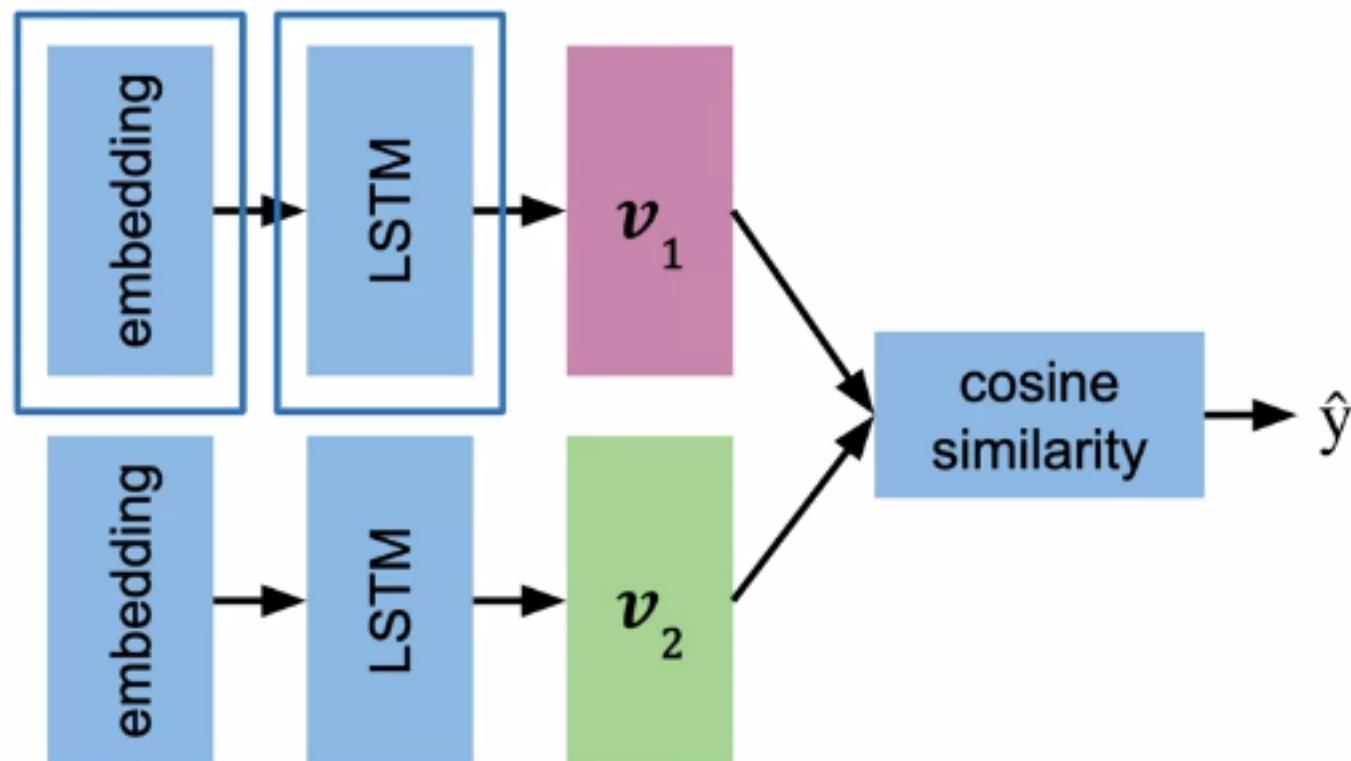
Model Architecture



Model Architecture

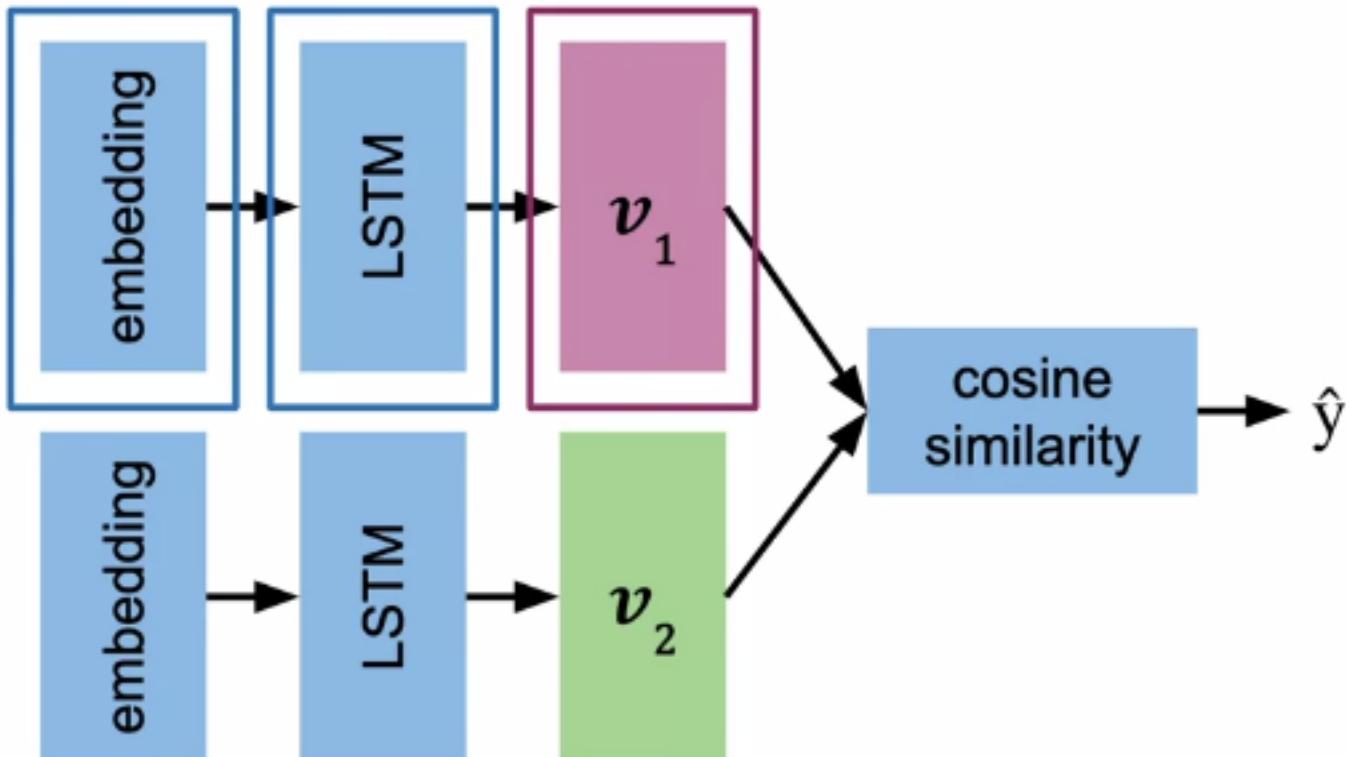
Question 1

Question 2



Model Architecture

Question 1

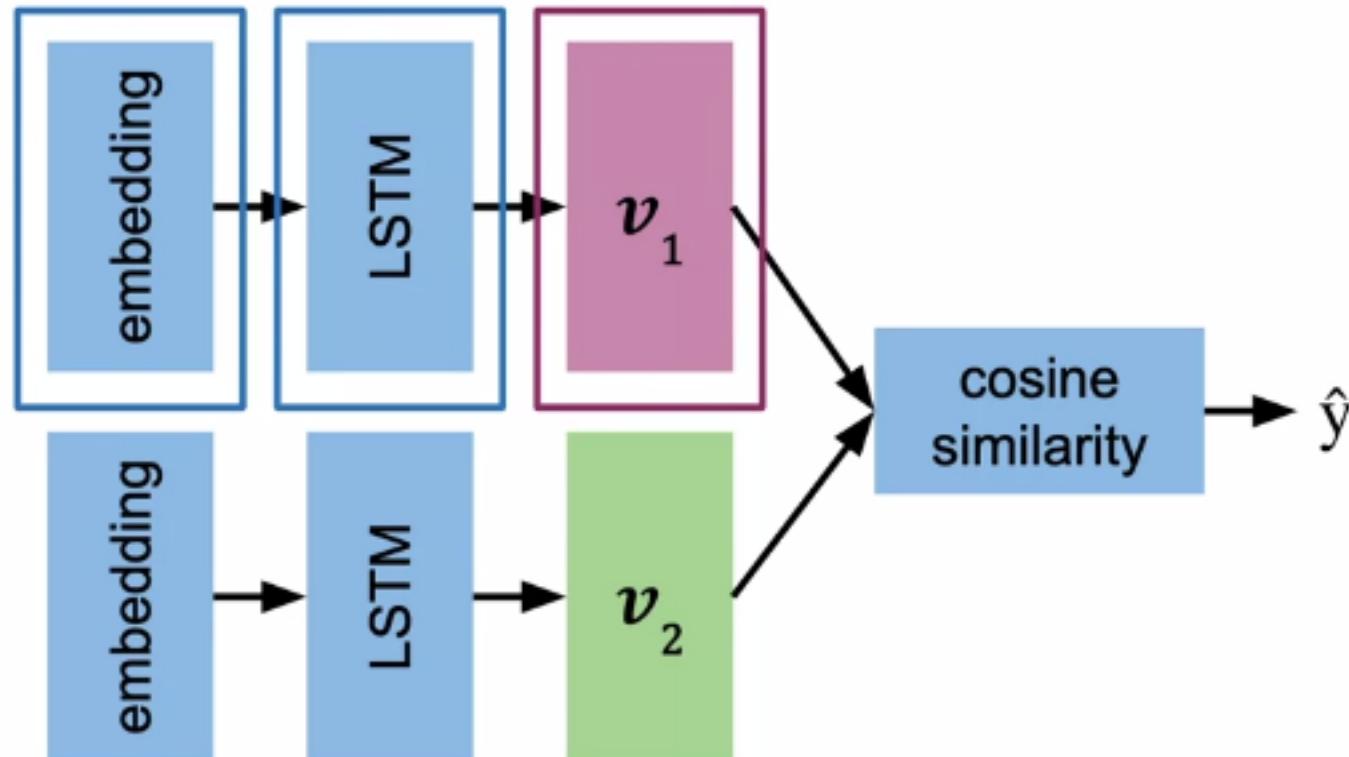


Question 2

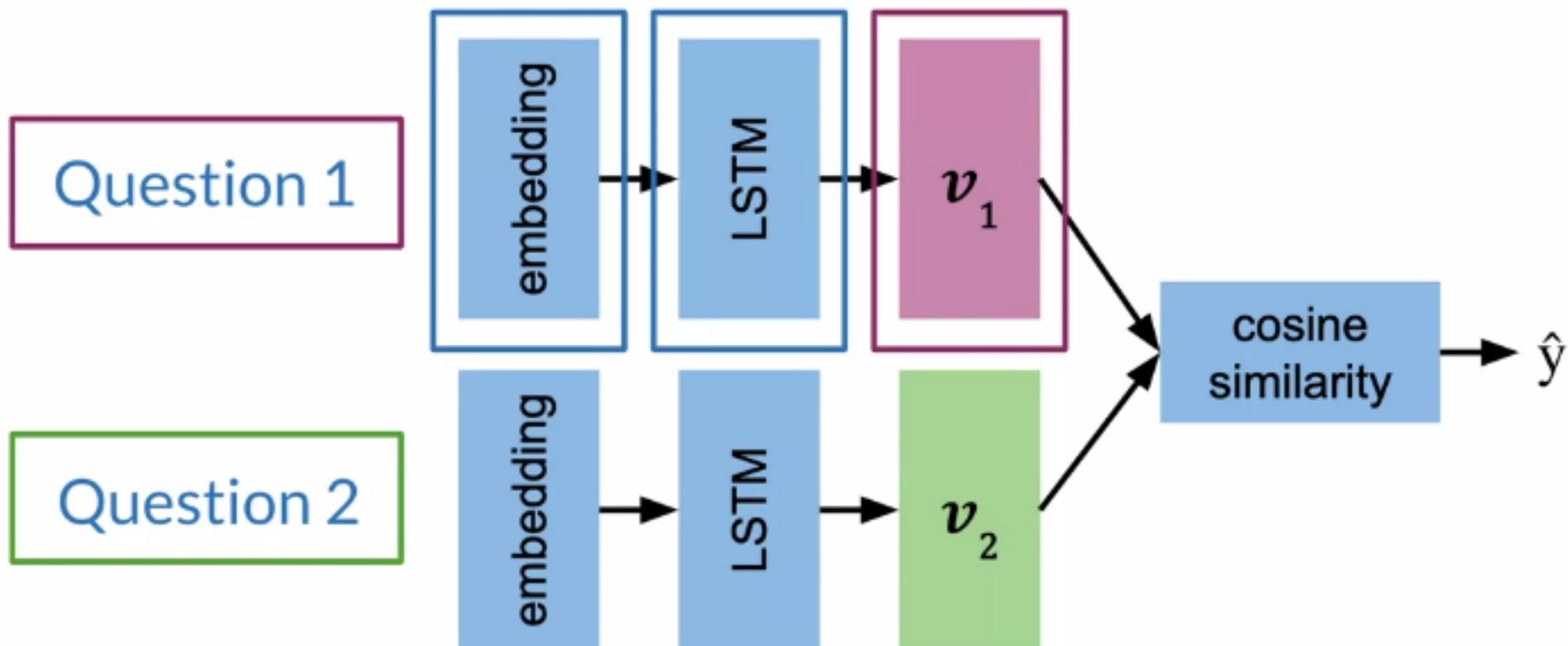
Model Architecture

Question 1

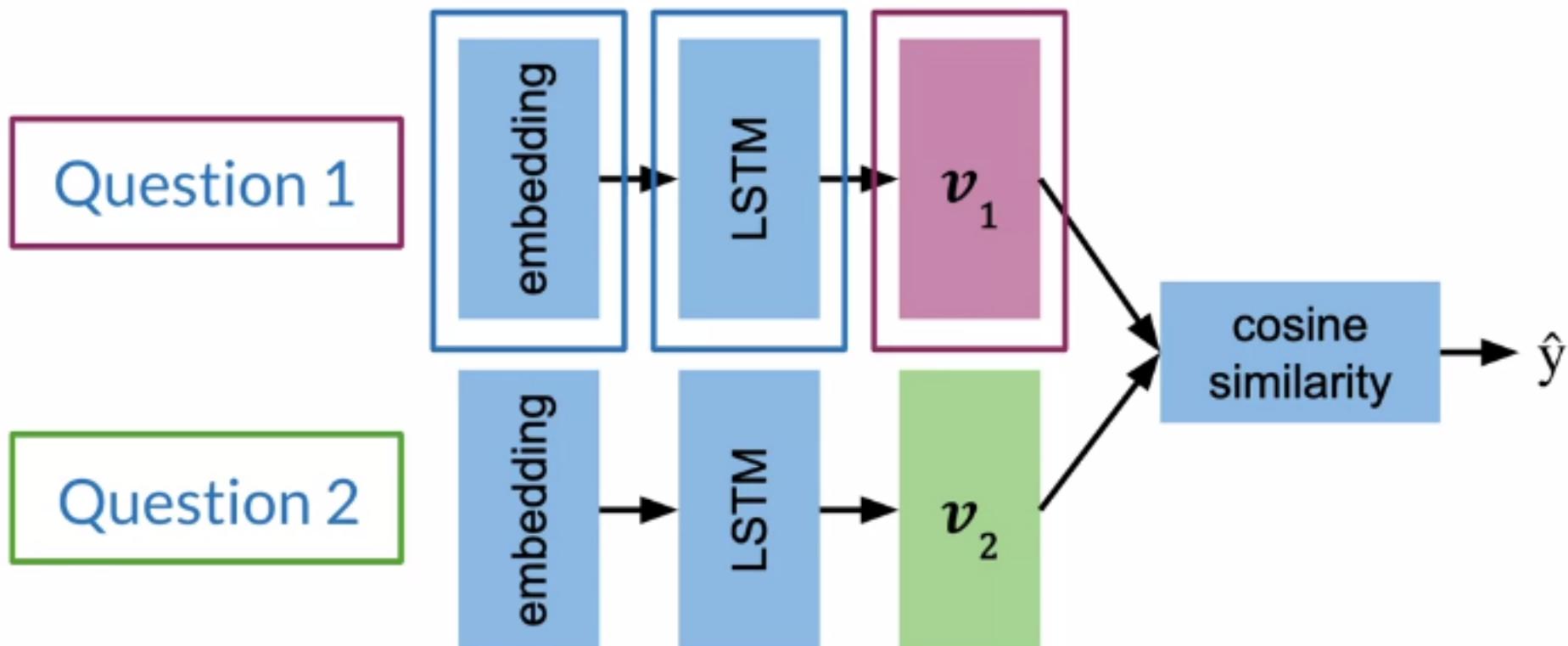
Question 2



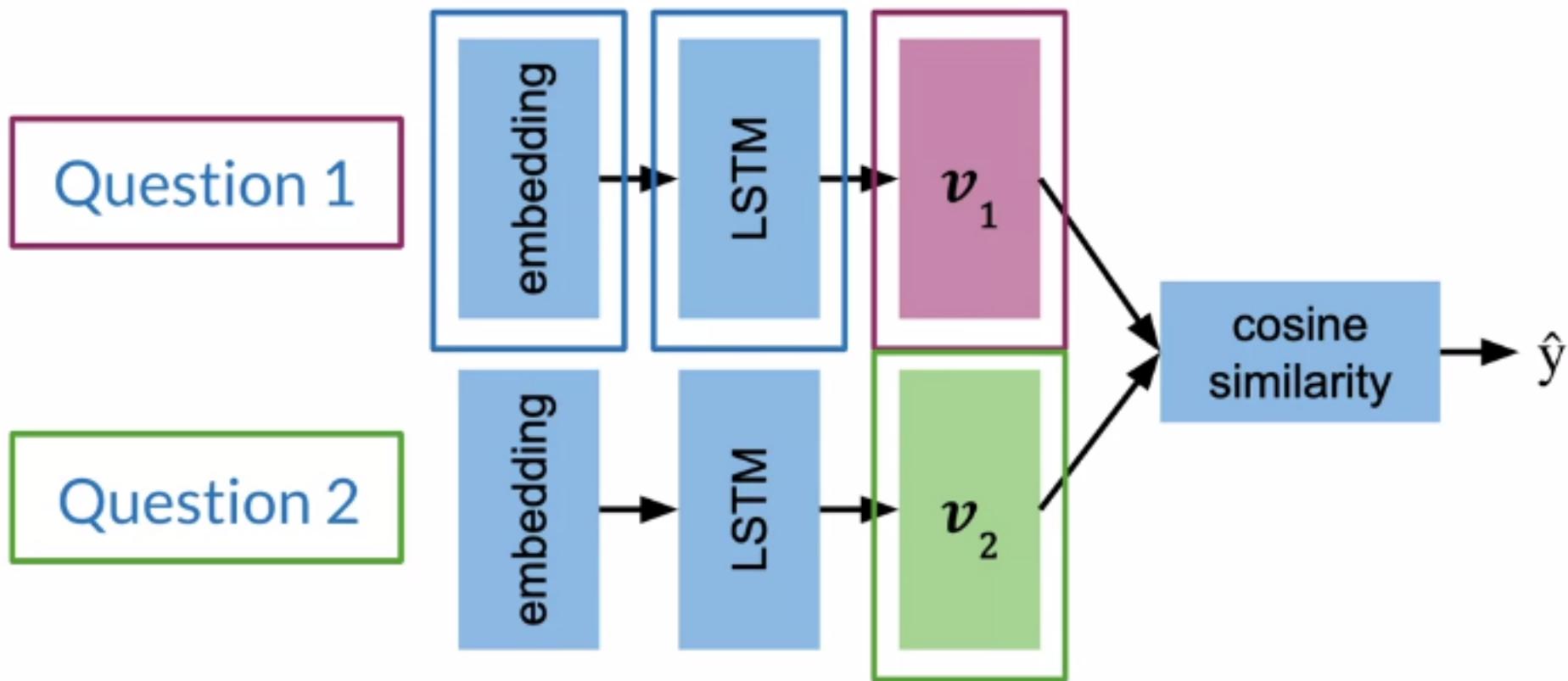
Model Architecture



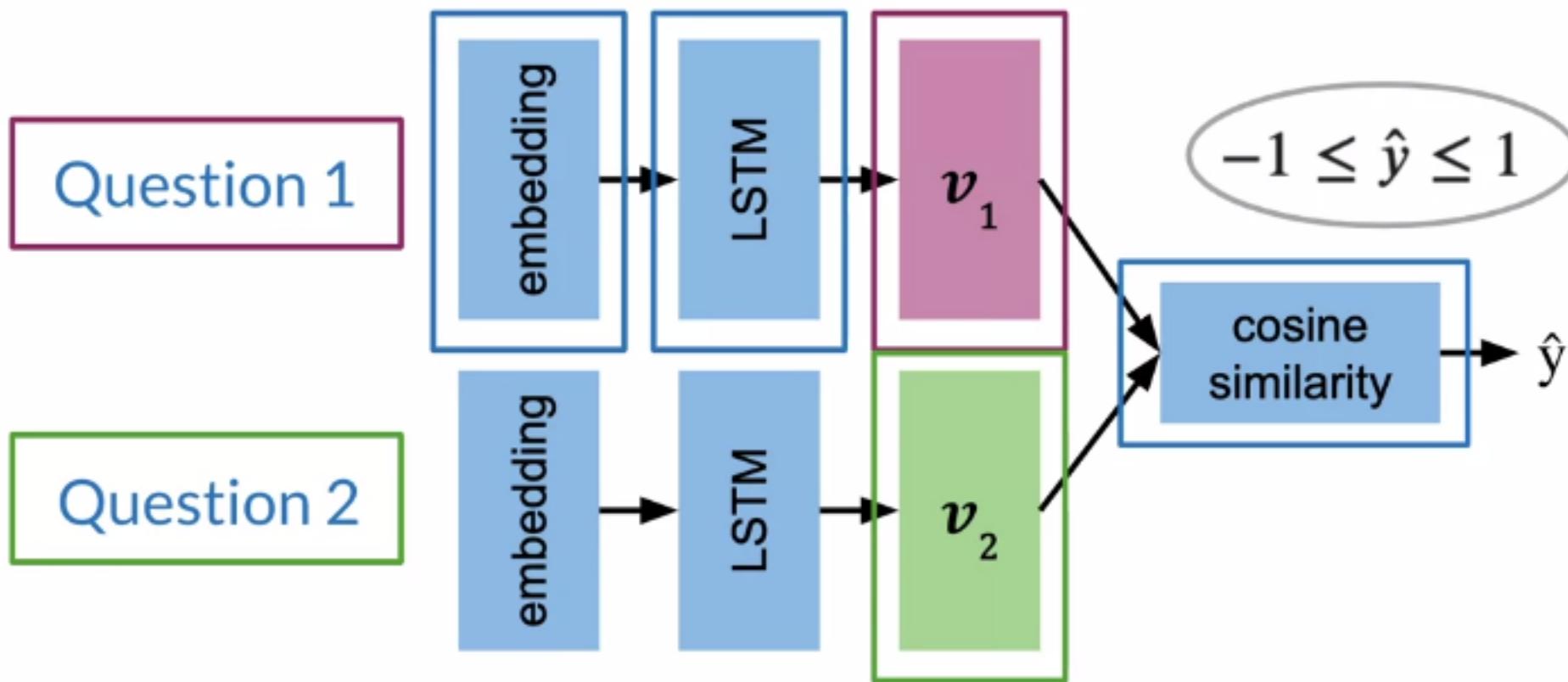
Model Architecture



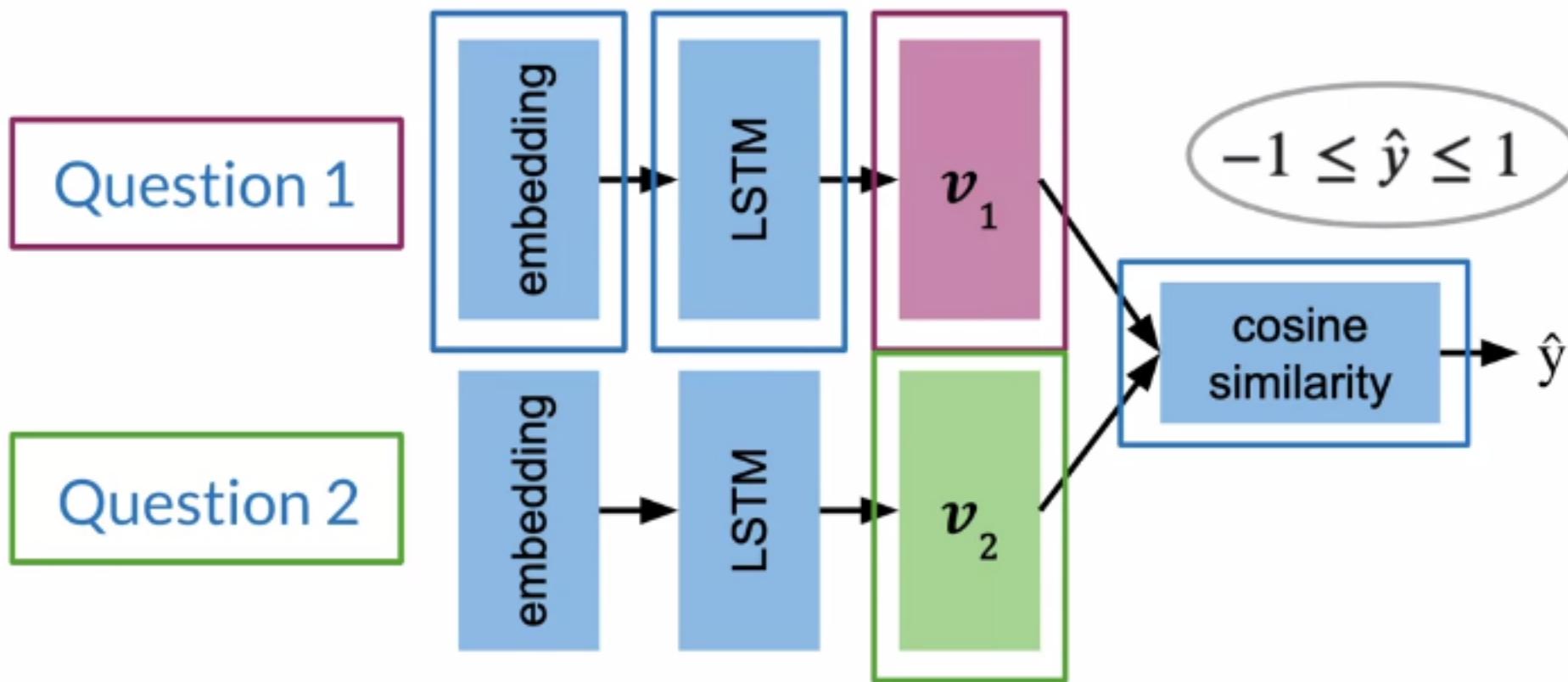
Model Architecture



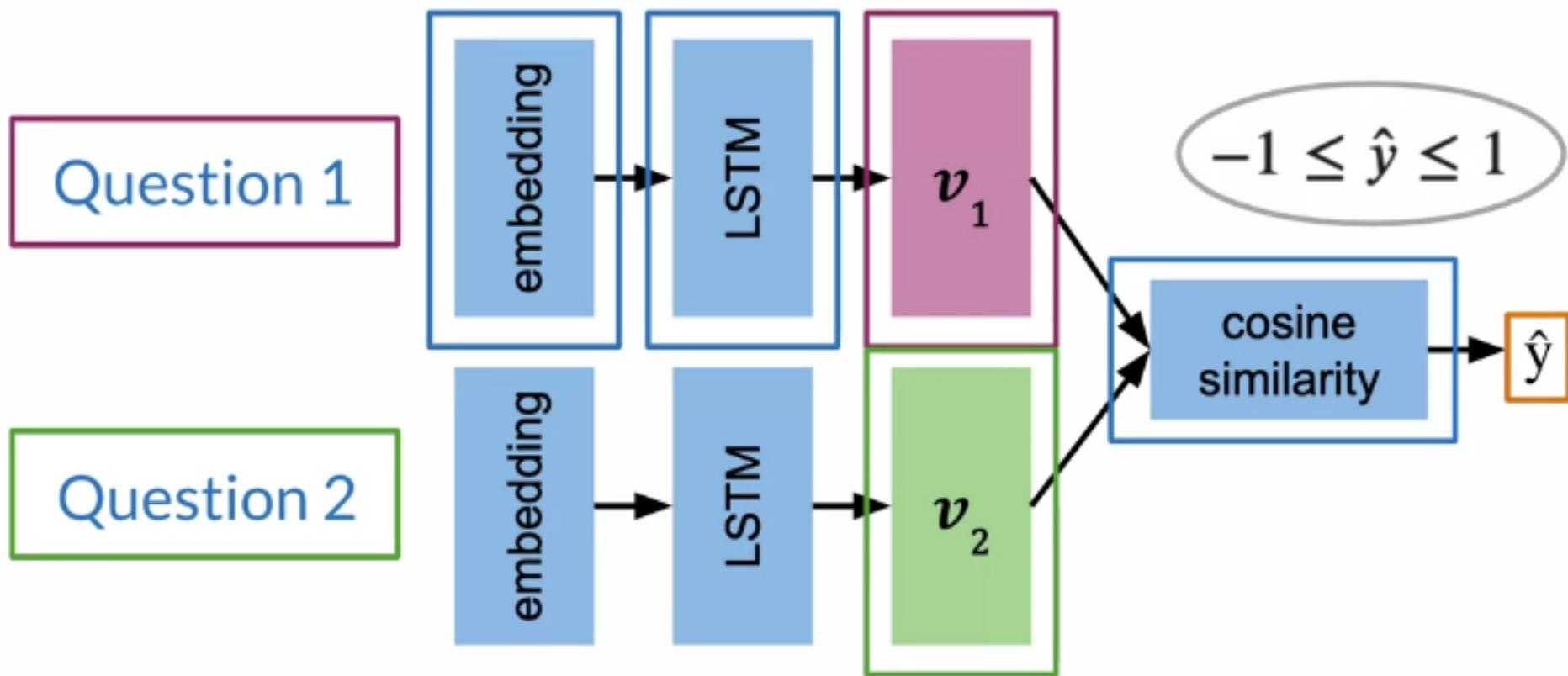
Model Architecture



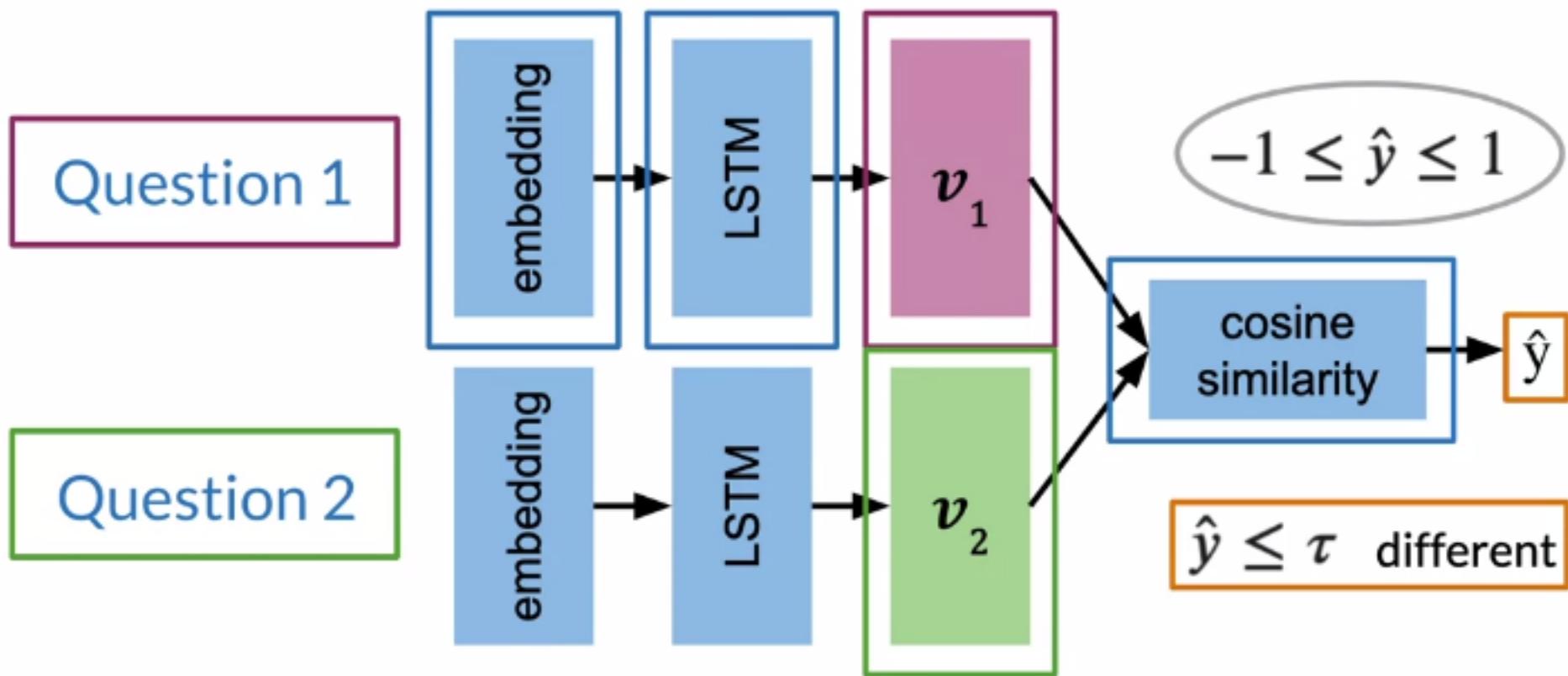
Model Architecture



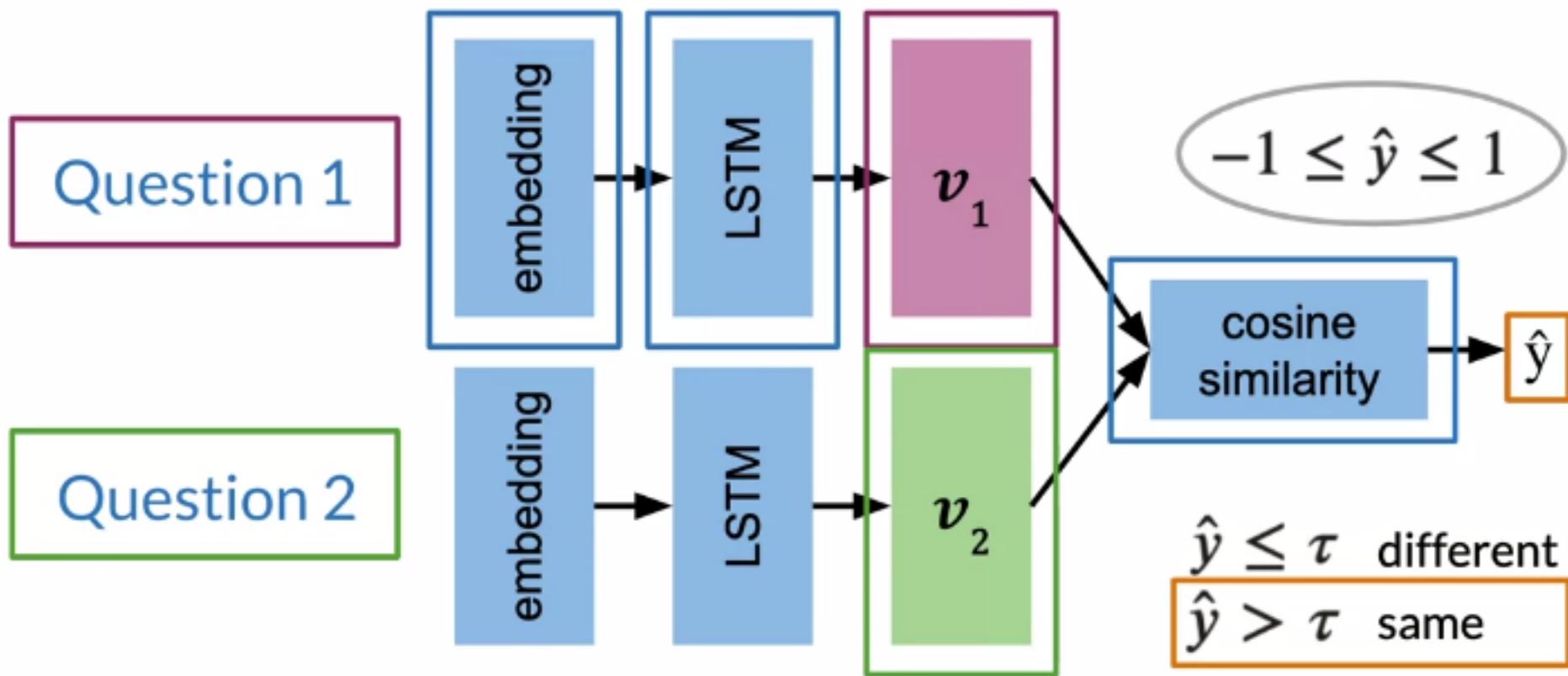
Model Architecture



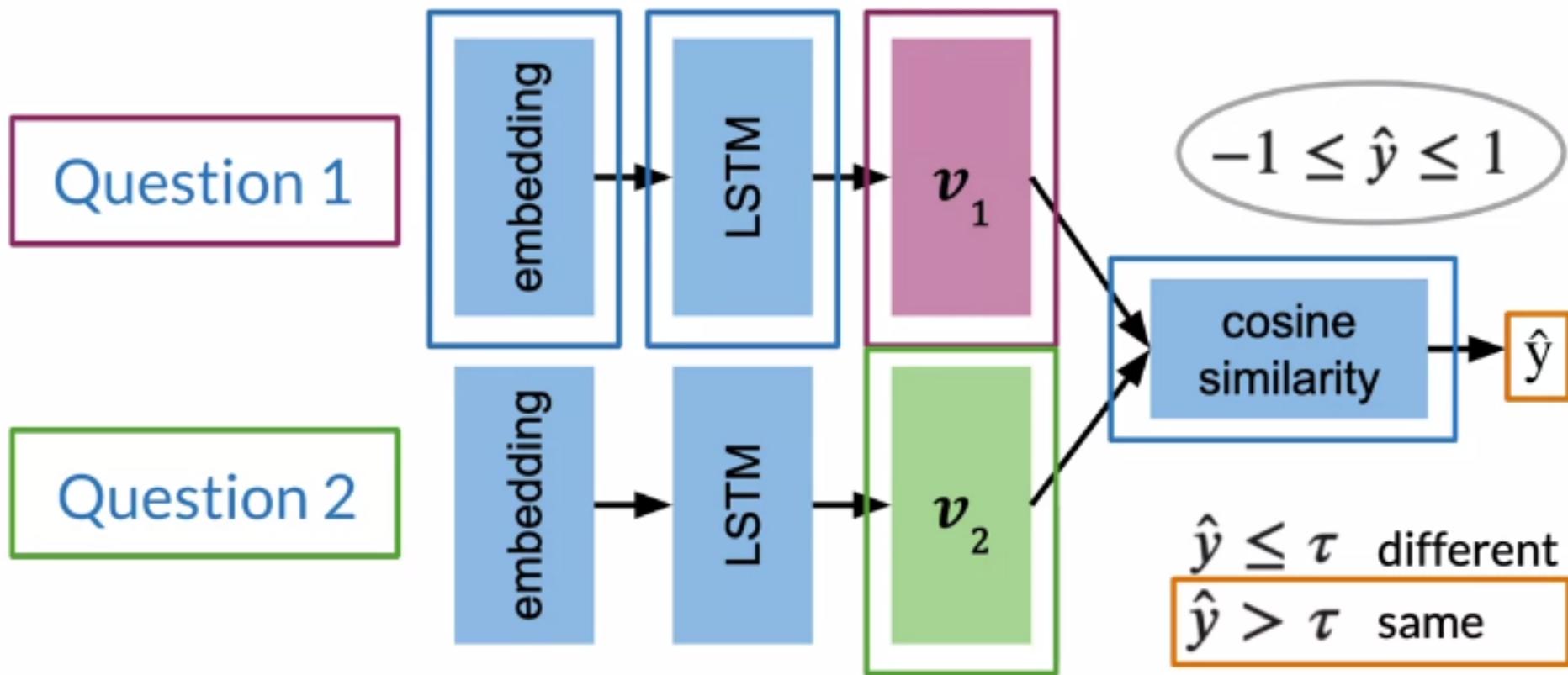
Model Architecture



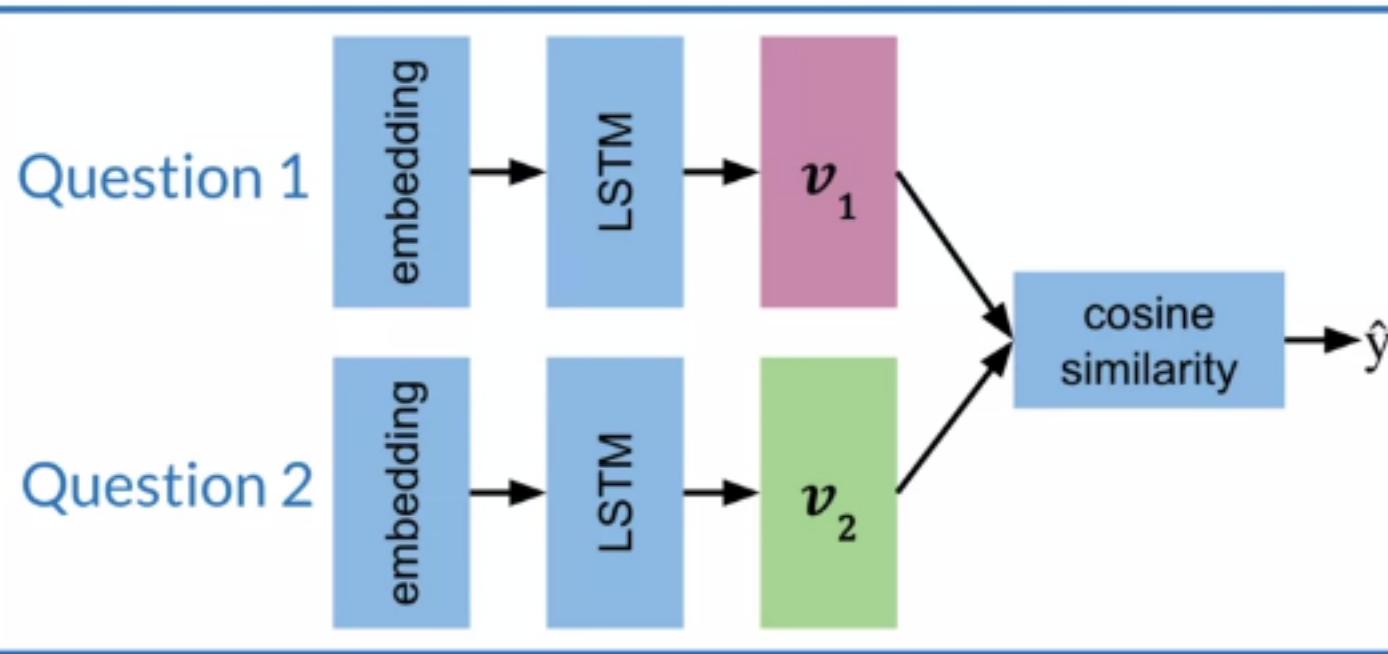
Model Architecture



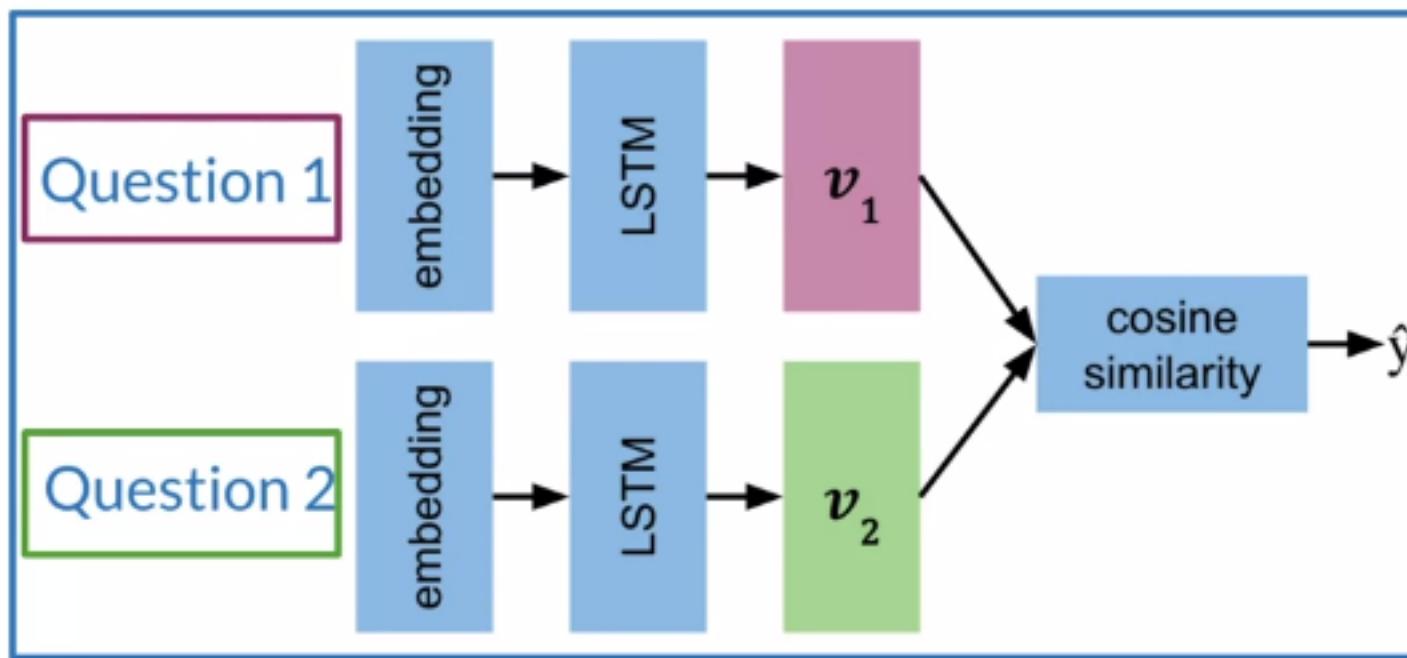
Model Architecture



Model Architecture

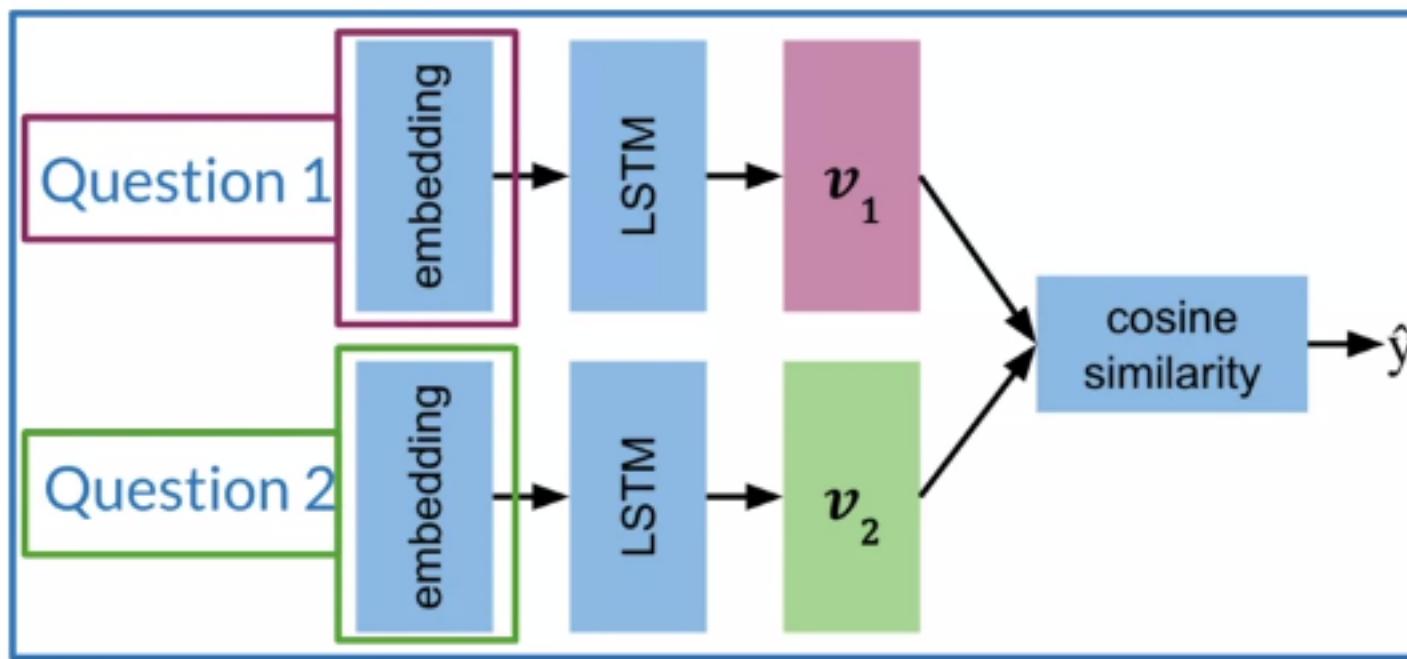


Model Architecture



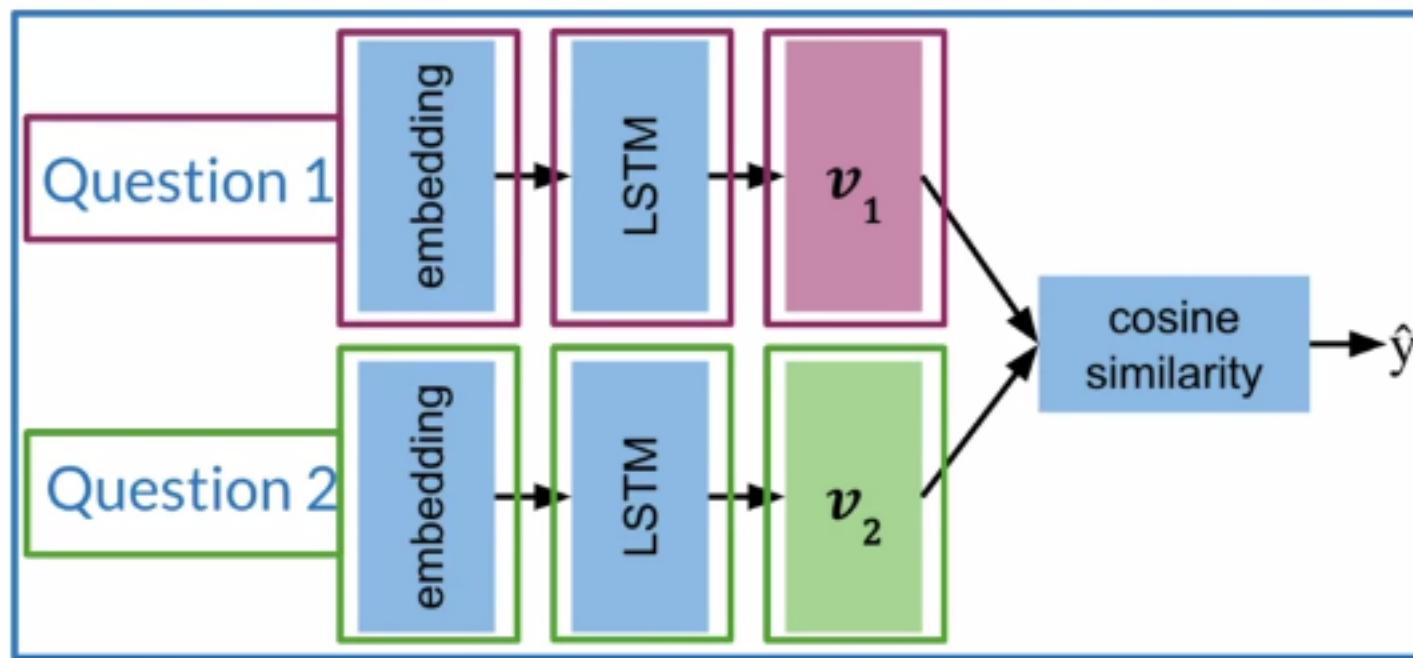
1) Inputs

Model Architecture



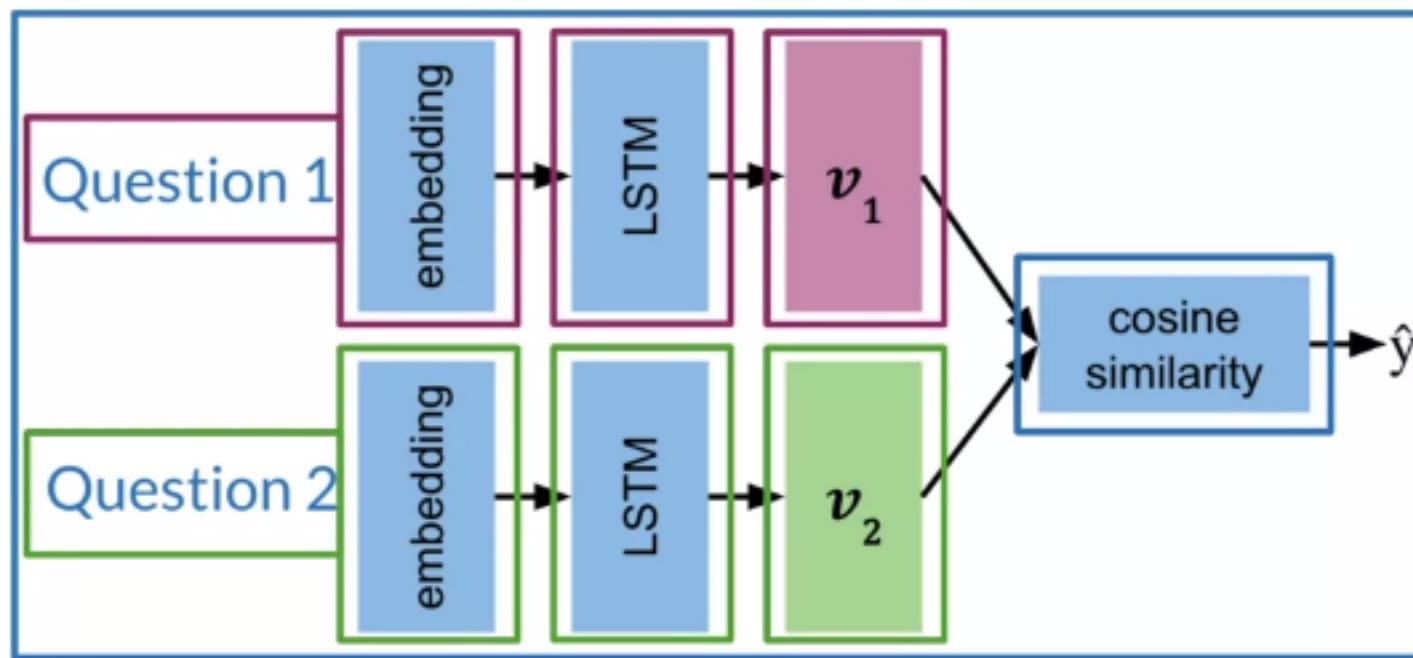
- 1) Inputs
- 2) Embedding

Model Architecture



- 1) Inputs
- 2) Embedding
- 3) LSTM
- 4) Vectors

Model Architecture



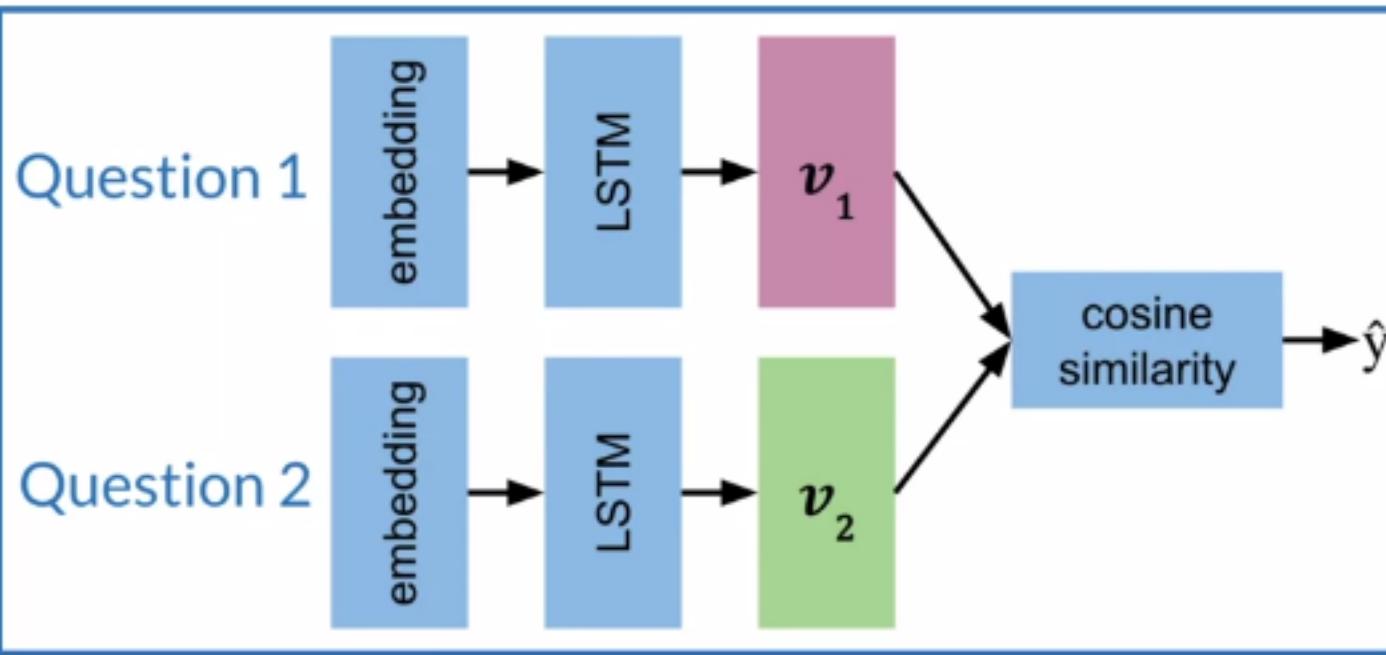
- 1) Inputs
- 2) Embedding
- 3) LSTM
- 4) Vectors
- 5) Cosine Similarity



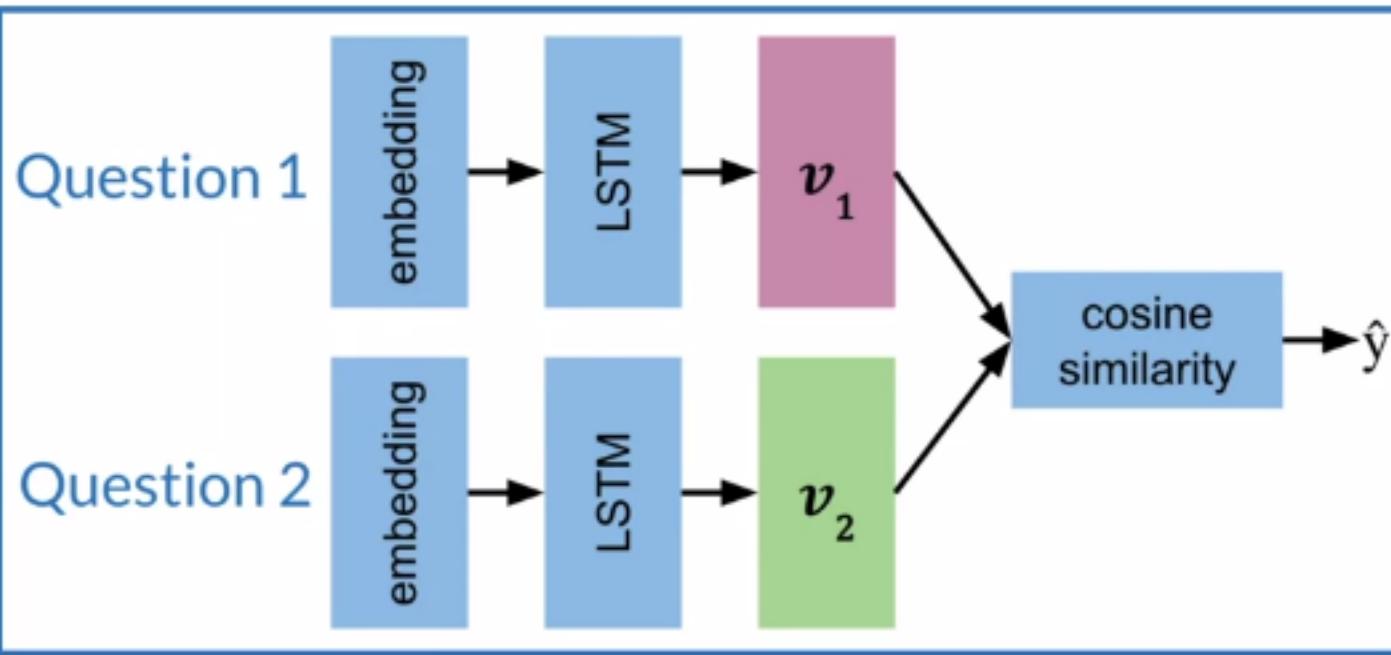
deeplearning.ai

Loss Function

Loss Function

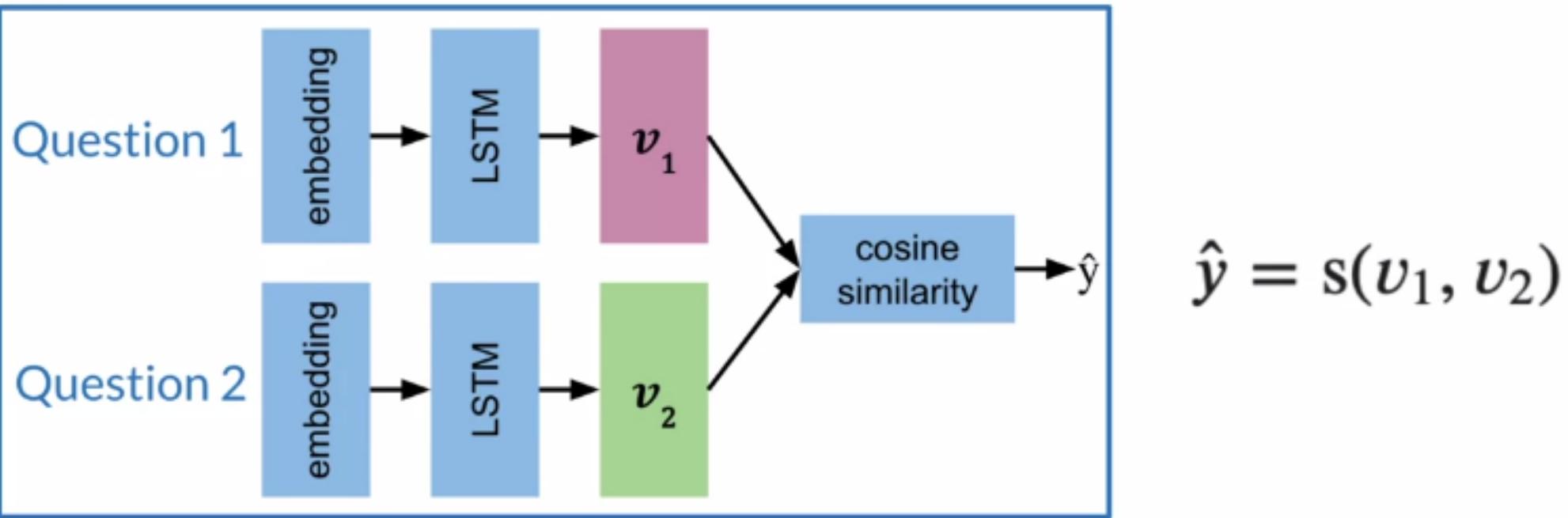


Loss Function



$$\hat{y} = s(v_1, v_2)$$

Loss Function



Loss Function

How old are you?

Loss Function

How old are you?

Loss Function

How old are you?

Anchor

Loss Function

How old are you? Anchor

What is your age? Positive

Loss Function

How old are you? Anchor

What is your age? Positive

Where are you from? Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$

How old are you?

Anchor

What is your age?

Positive

Where are you from?

Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$

How old are you?

Anchor

$s(v_1, v_2)$

What is your age?

Positive

Where are you from?

Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$

How old are you?

Anchor

$s(v_1, v_2)$

What is your age?

Positive

$s(A, P)$

Where are you from? Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$

How old are you?

Anchor

$s(v_1, v_2)$

What is your age?

Positive

$s(A, P)$

Where are you from? Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$

How old are you?

Anchor

$s(v_1, v_2)$

What is your age?

Positive

$s(A, P)$

Where are you from? Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$

s(v_1, v_2)

How old are you?

Anchor

What is your age?

Positive

s(A , P)

Where are you from? Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$
$$s(v_1, v_2)$$

How old are you?

Anchor

What is your age?

Positive

s(A, P)

Where are you from? Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$
$$s(v_1, v_2)$$

How old are you?

Anchor

What is your age?

Positive

s(A, P) ≈ 1

Where are you from? Negative

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$
$$s(v_1, v_2)$$

How old are you?

Anchor

What is your age?

Positive

$$s(A, P) \approx 1$$

Where are you from? Negative

$$s(A, N) \approx -1$$

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$
$$s(v_1, v_2)$$

How old are you?

Anchor

What is your age?

Positive

$$s(A, P) \approx 1$$

Where are you from?

Negative

$$s(A, N) \approx -1$$

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$
$$s(v_1, v_2)$$

How old are you?

Anchor

What is your age?

Positive

$$s(A, P) \approx 1$$

Where are you from? Negative

$$s(A, N) \approx -1$$

$$s(A, N) - s(A, P)$$

Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}$$
$$s(v_1, v_2)$$

How old are you?

Anchor

What is your age?

Positive

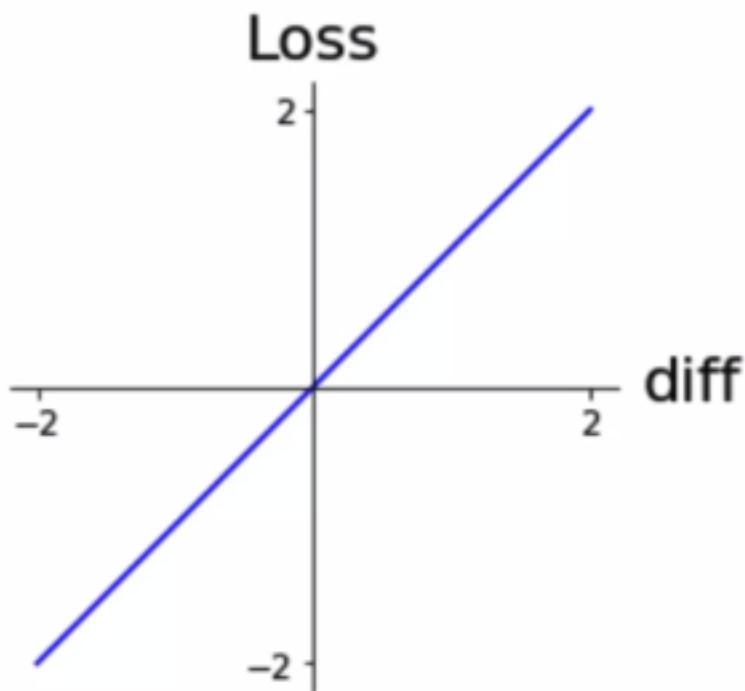
$s(A, P) \approx 1$

Where are you from? Negative

$s(A, N) \approx -1$

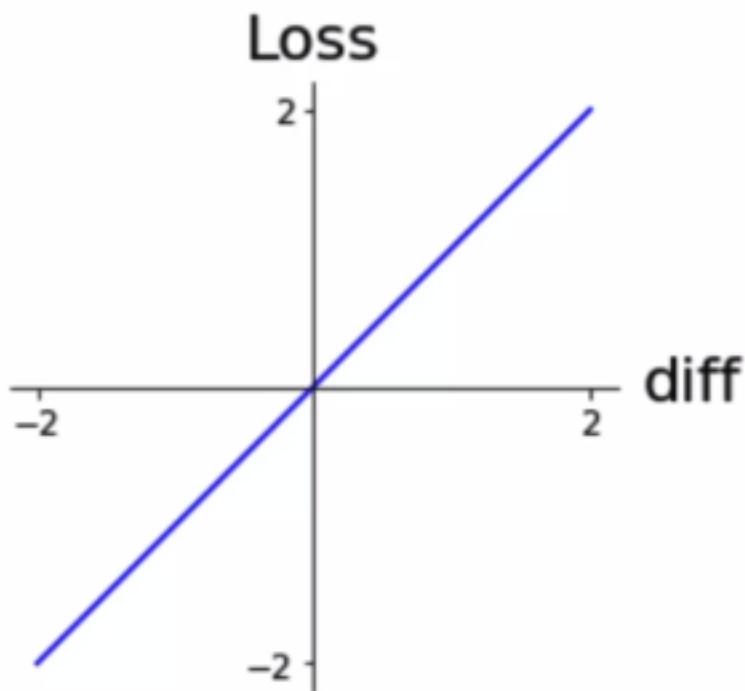
$$\text{diff} = s(A, N) - s(A, P)$$

Loss Function



$$\text{diff} = s(A, N) - s(A, P)$$

Loss Function



$$\text{diff} = s(A, N) - s(A, P)$$

Triplets

How old are you?

What is your age?

Where are you from?

Triplets

How old are you?

What is your age?

Where are you from?

A
P
N

Triplets !!!



Triplet Loss

Simple loss:

How old are you? A

What is your age? P

Where are you from? N

Triplet Loss

Simple loss:

$$\text{diff} = s(A, N) - s(A, P)$$

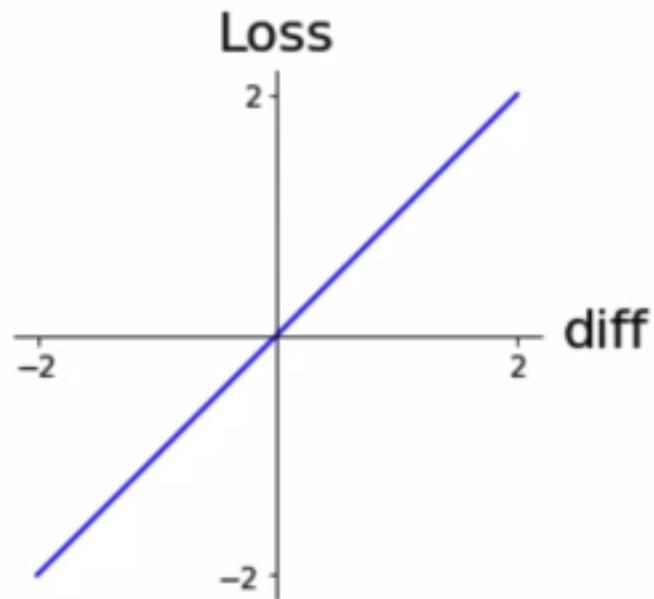
How old are you? A

What is your age? P

Where are you from? N

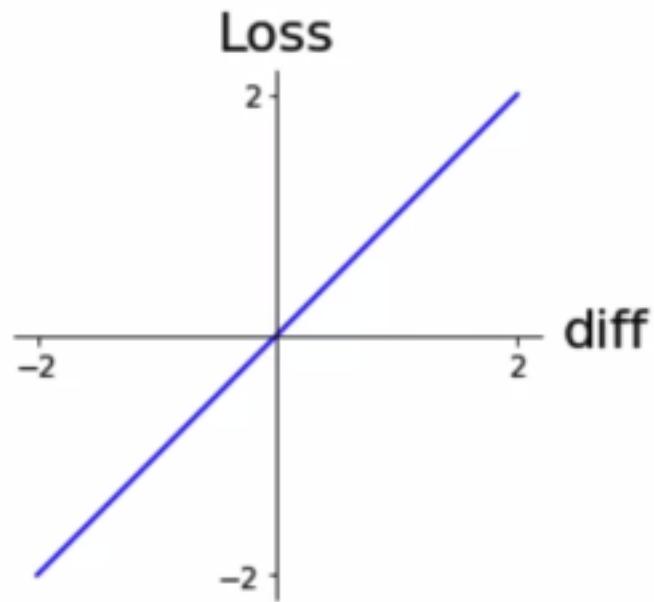
Triplet Loss

Simple loss:
 $\text{diff} = s(A, N) - s(A, P)$

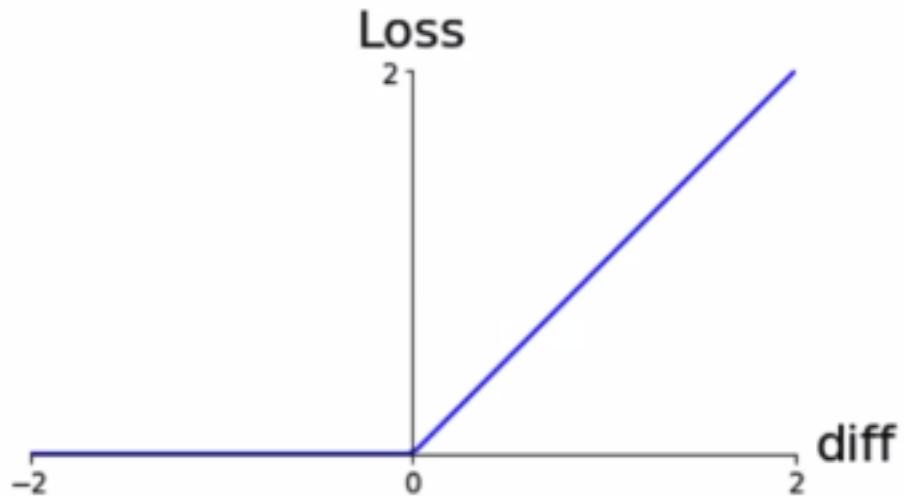


Triplet Loss

Simple loss:
 $\text{diff} = s(A, N) - s(A, P)$



Triplet Loss



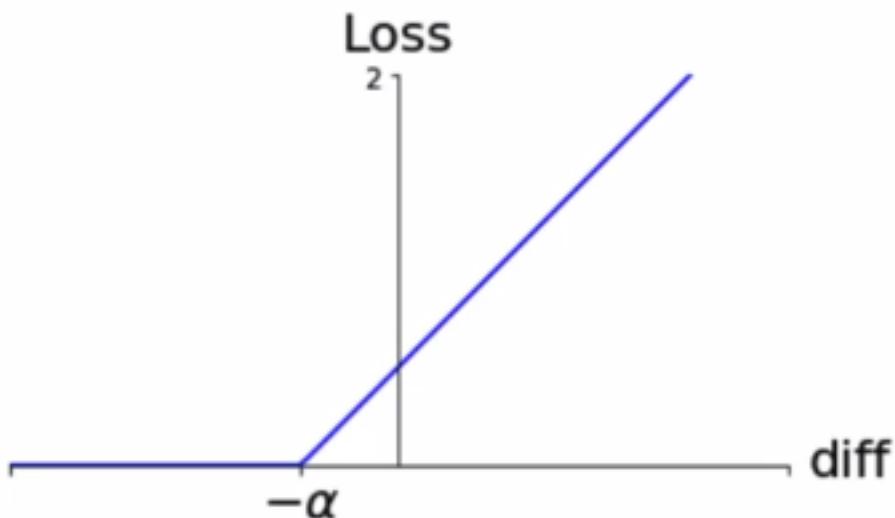
Simple loss:

$$\text{diff} = s(A, N) - s(A, P)$$

Non linearity:

$$\mathcal{L} = \begin{cases} 0; & \text{if } \text{diff} \leq 0 \\ \text{diff}; & \text{if } \text{diff} > 0 \end{cases}$$

Triplet Loss



Simple loss:

$$\text{diff} = s(A, N) - s(A, P)$$

Non linearity:

$$\mathcal{L} = \begin{cases} 0; & \text{if } \text{diff} \leq 0 \\ \text{diff}; & \text{if } \text{diff} > 0 \end{cases}$$

Alpha margin:

$$\mathcal{L} = \begin{cases} 0; & \text{if } \text{diff} + \alpha \leq 0 \\ \text{diff} + \alpha; & \text{if } \text{diff} + \alpha > 0 \end{cases}$$

Triplet Loss

$$\mathcal{L} = \begin{cases} 0; & \text{if } diff + \alpha \leq 0 \\ diff; & \text{if } diff + \alpha > 0 \end{cases}$$

Triplet Loss

$$\mathcal{L} = \begin{cases} 0; & \text{if } diff + \alpha \leq 0 \\ diff; & \text{if } diff + \alpha > 0 \end{cases}$$

$$\mathcal{L} = \max (diff + \alpha, 0)$$

Triplet Loss

$s(v_1, v_2)$ $d(v_1, v_2)$ 

$$\mathcal{L} = \begin{cases} 0; & \text{if } diff + \alpha \leq 0 \\ diff; & \text{if } diff + \alpha > 0 \end{cases}$$

$$\mathcal{L}(A, P, N) = \max (diff + \alpha, 0)$$

Triplet Selection

Hard triplets are better for training !

Triplet A, P, N:



duplicate set: A, P
non-duplicate set: A, N



Triplet Selection

Hard triplets are better for training !

Triplet A, P, N:



duplicate set: A, P
non-duplicate set: A, N

Random:

$$\mathcal{L} = \max (diff + \alpha, 0)$$

$$diff = s(A, N) - s(A, P)$$

Easy to satisfy. Little to learn



Triplet Selection

Hard triplets are better for training !

Triplet A, P, N:



duplicate set: A, P
non-duplicate set: A, N

Random:

$$\mathcal{L} = \max (diff + \alpha, 0)$$

$$diff = s(A, N) - s(A, P)$$

Easy to satisfy. Little to learn



Hard:

$$s(A, N) \approx s(A, P)$$

Harder to train. More to learn



Triplet Selection

Hard triplets are better for training !

Triplet A, P, N:



duplicate set: A, P
non-duplicate set: A, N



Random:

$$\mathcal{L} = \max (diff + \alpha, 0)$$

$$diff = s(A, N) - s(A, P)$$

Easy to satisfy. Little to learn



Hard:

$$s(A, N) \approx s(A, P)$$

Harder to train. More to learn



Computing The Cost

Prepare the batches as follows:

Computing The Cost

Prepare the batches as follows:

What is your age?

How old are you?

Can you see me?

Are you seeing me?

Computing The Cost

Prepare the batches as follows:

What is your age?

How old are you?

Can you see me?

Are you seeing me?

Where are thou?

Where are you?

Computing The Cost

Prepare the batches as follows:

What is your age?

How old are you?

Can you see me?

Are you seeing me?

Where are thou?

Where are you?

When is the game?

What time is the game?

Computing The Cost

Prepare the batches as follows:

What is your age?

How old are you?

Can you see me?

Are you seeing me?

Where are thou?

Where are you?

When is the game?

What time is the game?

$$b = 4$$

Computing The Cost

Prepare the batches as follows:



What is your age?

How old are you?



Can you see me?

Are you seeing me?

Where are thou?

Where are you?

When is the game?

What time is the game?

$$b = 4$$

Computing The Cost

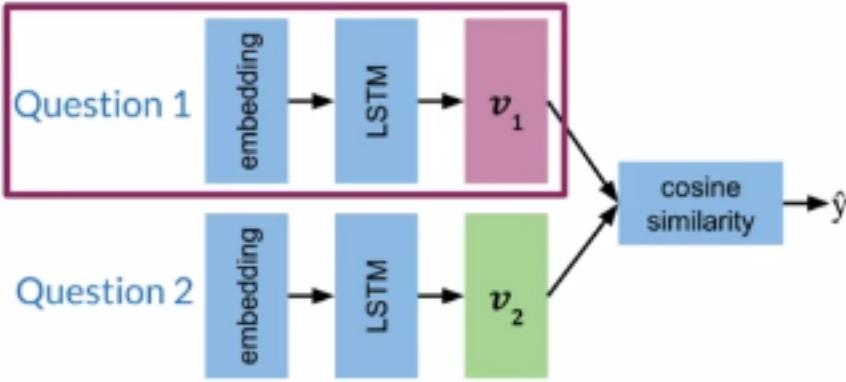
Batch 1

What is your age?

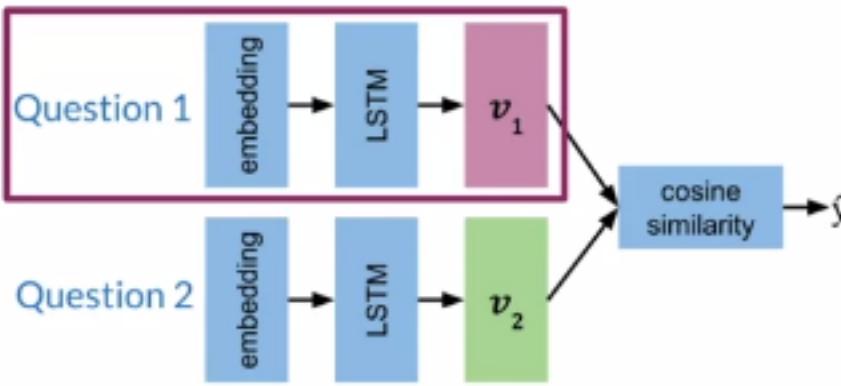
Can you see me?

Where are thou?

When is the game?



Computing The Cost



Batch 1

What is your age?

Can you see me?

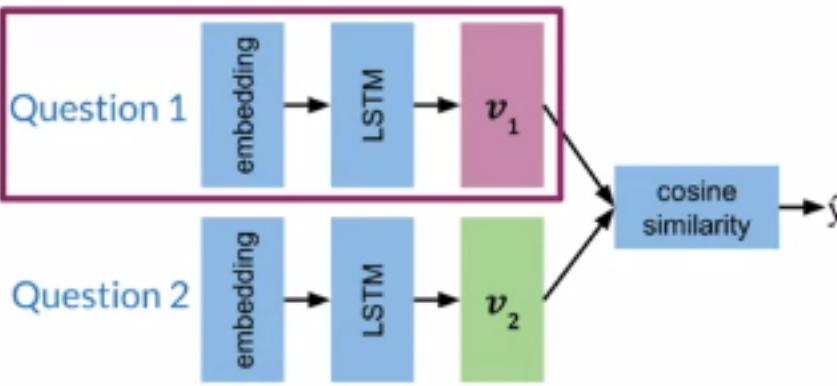
Where are thou?

When is the game?

$$\mathbf{v}_1 = (1, d_{\text{model}})$$

--	--	--	--	--

Computing The Cost



Batch 1

What is your age?

Can you see me?

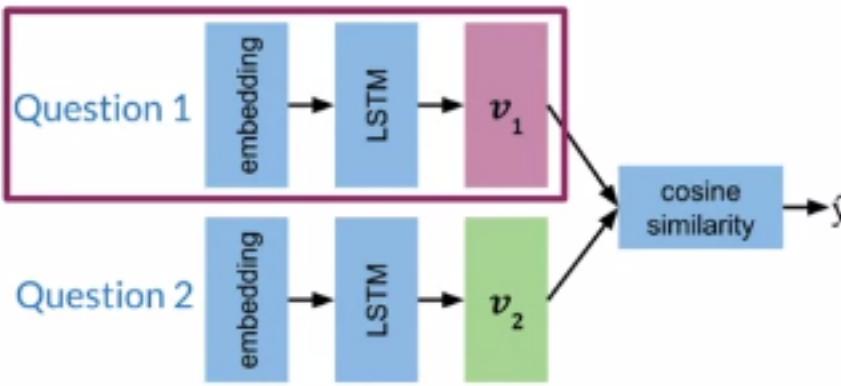
Where are thou?

When is the game?

$$v_1 = (1, d_{\text{model}})$$

--	--	--	--	--

Computing The Cost

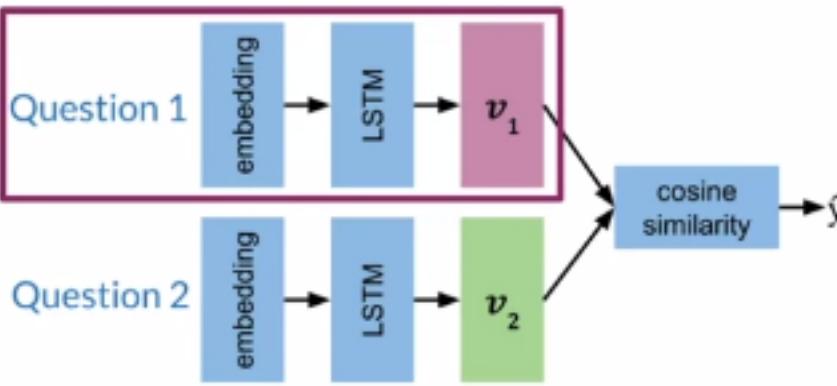


Batch 1

What is your age?
Can you see me?
Where are thou?
When is the game?

$$v_1 = (1, d_{\text{model}})$$

Computing The Cost



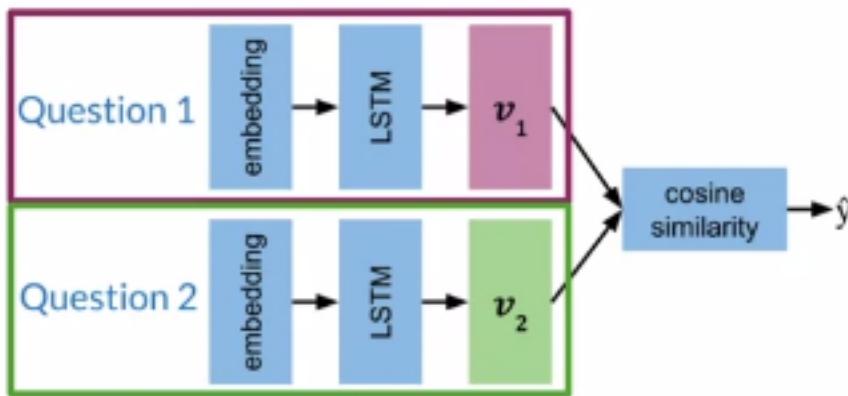
Batch 1

What is your age?
Can you see me?
Where are thou?
When is the game?

$$\mathbf{v}_1 = (1, \text{d_model})$$

$\mathbf{v}_{1,1}$				
$\mathbf{v}_{1,2}$				
$\mathbf{v}_{1,3}$				
$\mathbf{v}_{1,4}$				

Computing The Cost



Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

How old are you?

Are you seeing me?

Where are you?

What time is the game?

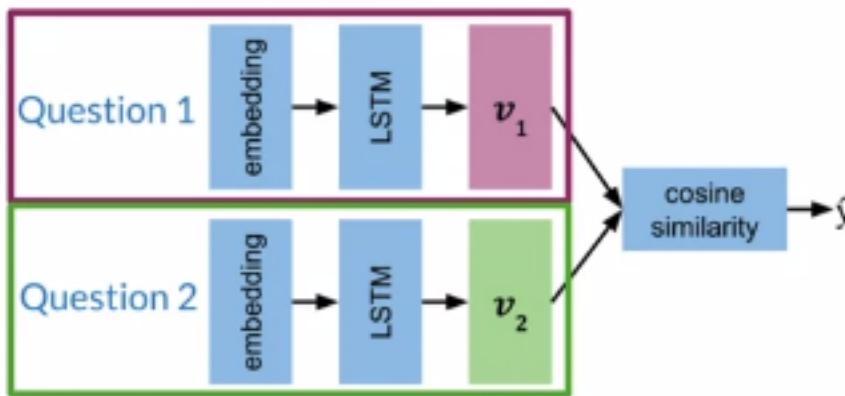
$$\mathbf{v}_1 = (1, \mathbf{d}_{\text{model}})$$

\mathbf{v}_{1_1}				
\mathbf{v}_{1_2}				
\mathbf{v}_{1_3}				
\mathbf{v}_{1_4}				

$$\mathbf{v}_2$$

\mathbf{v}_{2_1}				
\mathbf{v}_{2_2}				
\mathbf{v}_{2_3}				
\mathbf{v}_{2_4}				

Computing The Cost



Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

How old are you?

Are you seeing me?

Where are you?

What time is the game?

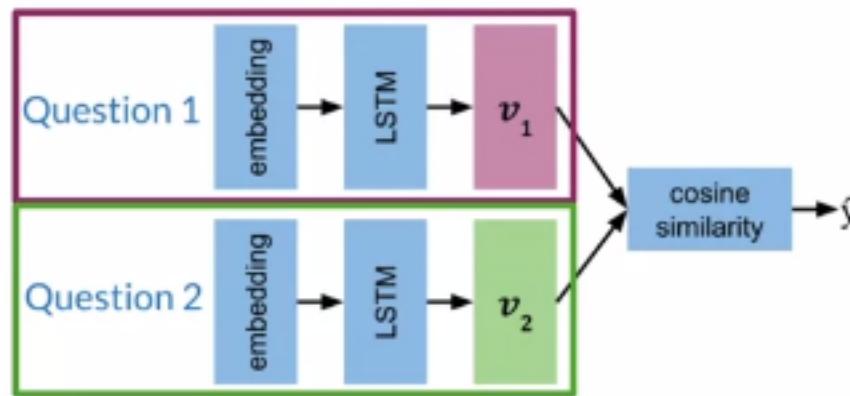
$$\mathbf{v}_1 = (1, \mathbf{d}_{\text{model}})$$

\mathbf{v}_{1_1}			
\mathbf{v}_{1_2}			
\mathbf{v}_{1_3}			
\mathbf{v}_{1_4}			

$$\mathbf{v}_2$$

\mathbf{v}_{2_1}			
\mathbf{v}_{2_2}			
\mathbf{v}_{2_3}			
\mathbf{v}_{2_4}			

Computing The Cost



Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

How old are you?

Are you seeing me?

Where are you?

What time is the game?

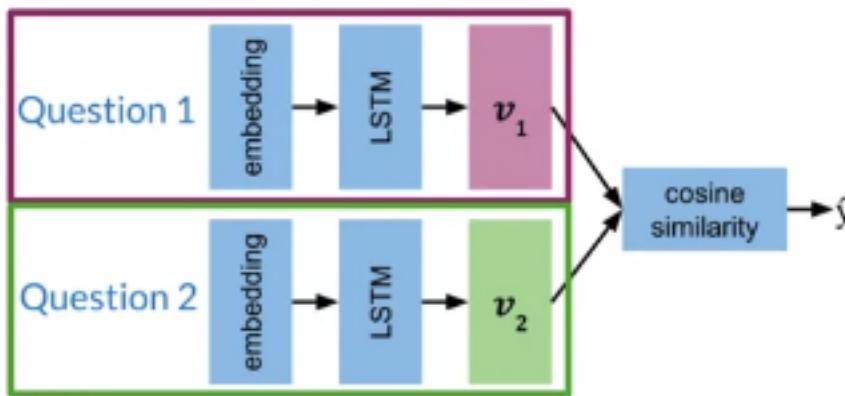
$$\mathbf{v}_1 = (1, \text{d_model})$$

\mathbf{v}_{1_1}				
\mathbf{v}_{1_2}				
\mathbf{v}_{1_3}				
\mathbf{v}_{1_4}				

$$\mathbf{v}_2$$

\mathbf{v}_{2_1}				
\mathbf{v}_{2_2}				
\mathbf{v}_{2_3}				
\mathbf{v}_{2_4}				

Computing The Cost



Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

How old are you?

Are you seeing me?

Where are you?

What time is the game?

$$\mathbf{v}_1 = (1, \text{d_model})$$

\mathbf{v}_{1_1}					
\mathbf{v}_{1_2}					
\mathbf{v}_{1_3}					
\mathbf{v}_{1_4}					

$$\mathbf{v}_2$$

\mathbf{v}_{2_1}					
\mathbf{v}_{2_2}					
\mathbf{v}_{2_3}					
\mathbf{v}_{2_4}					

Computing The Cost

$s(v_1, v_2)$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Computing The Cost

$$s(v_1, v_2)$$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Computing The Cost

$$s(v_1, v_2)$$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Computing The Cost

$$s(v_1, v_2)$$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Computing The Cost

$$s(v_1, v_2)$$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Computing The Cost

		s(v_1, v_2)				
		v_1	-1	-2	-3	-4
v_2		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

$$\mathcal{L}(A, P, N) = \max (diff + \alpha, 0)$$
$$diff = s(A, N) - s(A, P)$$

Computing The Cost

		s(v_1, v_2)				
		v_1	-1	-2	-3	-4
v_2		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

$$\mathcal{L}(A, P, N) = \max (diff + \alpha, 0)$$
$$diff = s(A, N) - s(A, P)$$

$$\mathcal{J} = \sum_{i=1}^m \mathcal{L}(A^{(i)}, P^{(i)}, N^{(i)})$$

Computing The Cost

		s(v_1, v_2)				
		v_1	-1	-2	-3	-4
v_2		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

$$\mathcal{L}(A, P, N) = \max (diff + \alpha, 0)$$
$$diff = s(A, N) - s(A, P)$$

$$\mathcal{J} = \sum_{i=1}^m \mathcal{L}(A^{(i)}, P^{(i)}, N^{(i)})$$

Computing The Cost

		s(v_1, v_2)				
		v_1	-1	-2	-3	-4
v_2		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

$$\mathcal{L}(A, P, N) = \max (diff + \alpha, 0)$$
$$diff = s(A, N) - s(A, P)$$

$$\mathcal{J} = \sum_{i=1}^m \mathcal{L}(A^{(i)}, P^{(i)}, N^{(i)})$$

Computing The Cost

Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

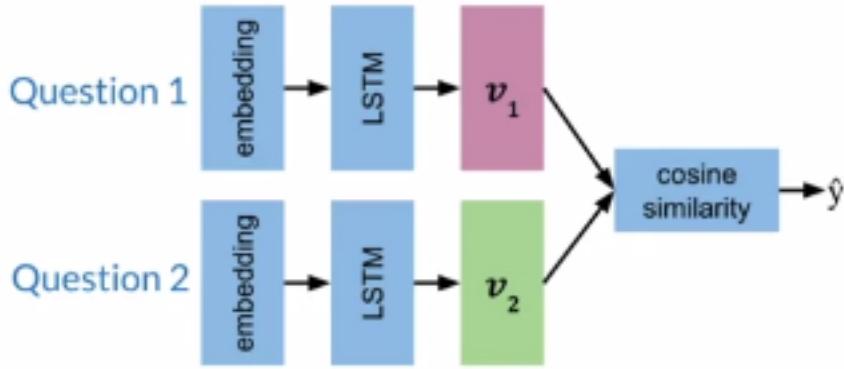
How old are you?

Are you seeing me?

Where are you?

What time is the game?

Computing The Cost



Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

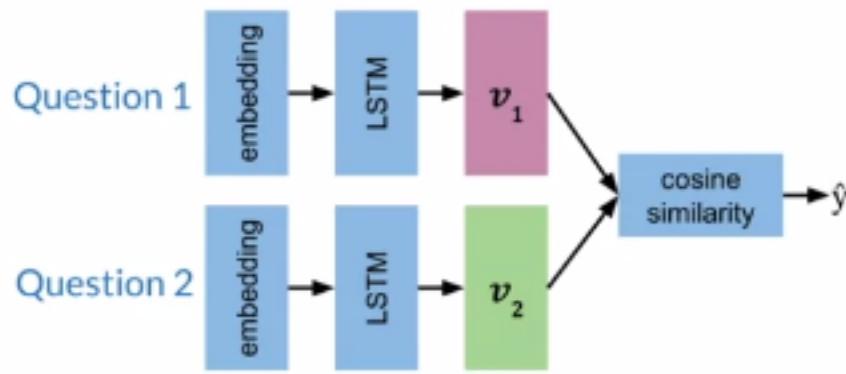
How old are you?

Are you seeing me?

Where are you?

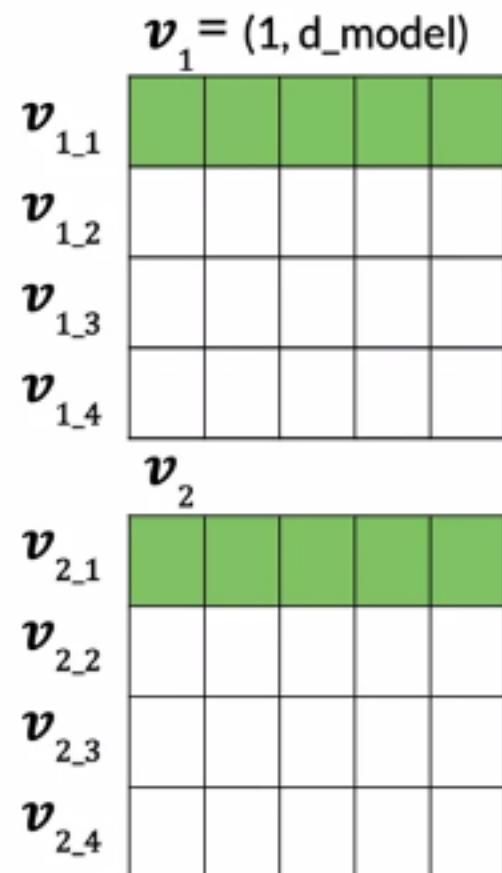
What time is the game?

Computing The Cost

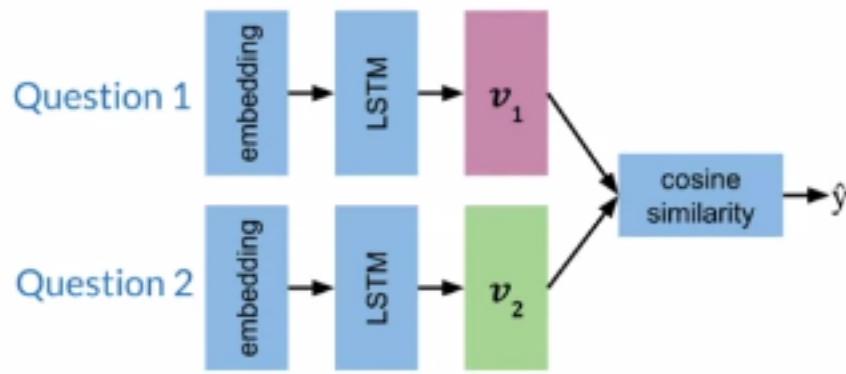


Batch 1
What is your age?
Can you see me?
Where are thou?
When is the game?

Batch 2
How old are you?
Are you seeing me?
Where are you?
What time is the game?



Computing The Cost

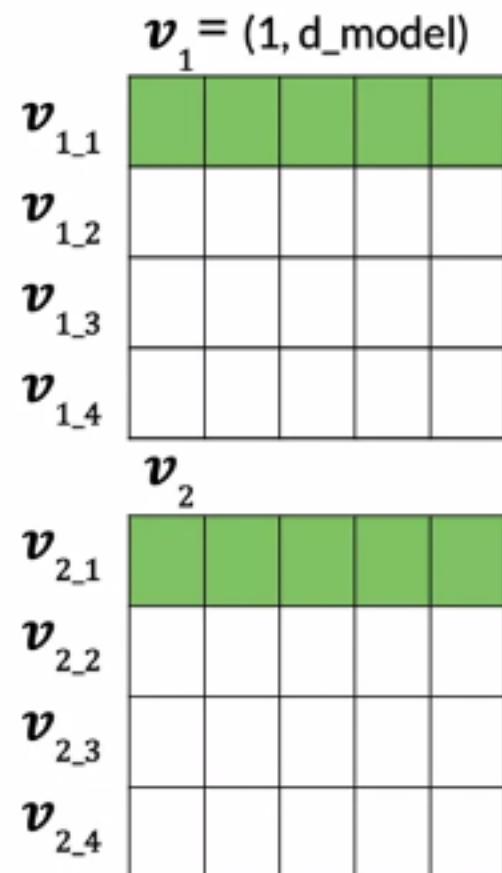


Batch 1

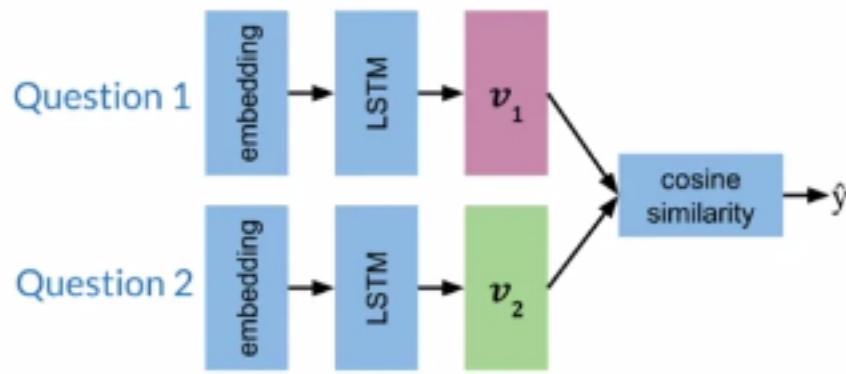
What is your age?
Can you see me?
Where are thou?
When is the game?

Batch 2

How old are you?
Are you seeing me?
Where are you?
What time is the game?

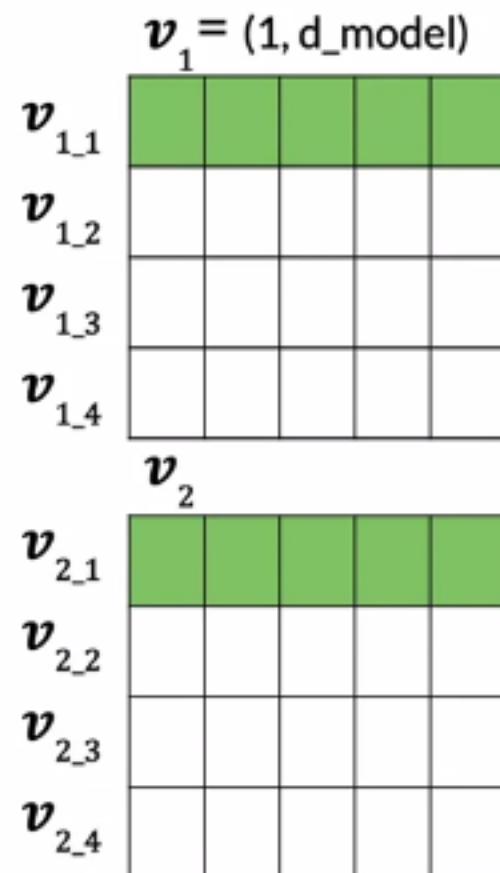


Computing The Cost

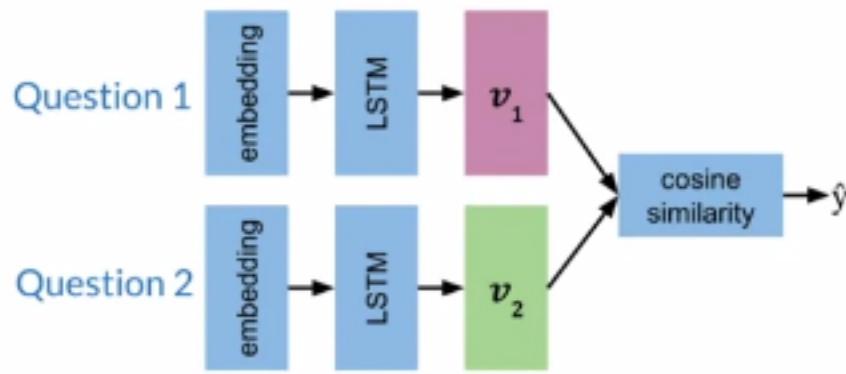


Batch 1
What is your age?
Can you see me?
Where are thou?
When is the game?

Batch 2
How old are you?
Are you seeing me?
Where are you?
What time is the game?



Computing The Cost

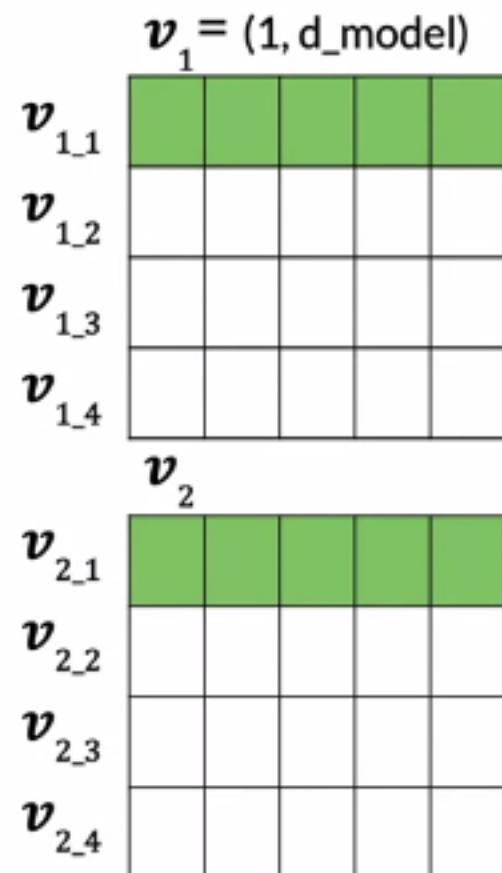


Batch 1

What is your age?
Can you see me?
Where are thou?
When is the game?

Batch 2

How old are you?
Are you seeing me?
Where are you?
What time is the game?



Computing The Cost

$s(v_1, v_2)$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Computing The Cost

$$s(v_1, v_2)$$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Computing The Cost

$$s(v_1, v_2)$$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Computing The Cost

		s(v_1, v_2)				
		v_1	-1	-2	-3	-4
v_2		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

$$\mathcal{L}(A, P, N) = \max (diff + \alpha, 0)$$
$$diff = s(A, N) - s(A, P)$$

Computing The Cost

		s(v_1, v_2)				
		v_1	-1	-2	-3	-4
v_2		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

$$\mathcal{L}(A, P, N) = \max (diff + \alpha, 0)$$
$$diff = s(A, N) - s(A, P)$$

$$\mathcal{J} = \sum_{i=1}^m \mathcal{L}(A^{(i)}, P^{(i)}, N^{(i)})$$

Computing The Cost

Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

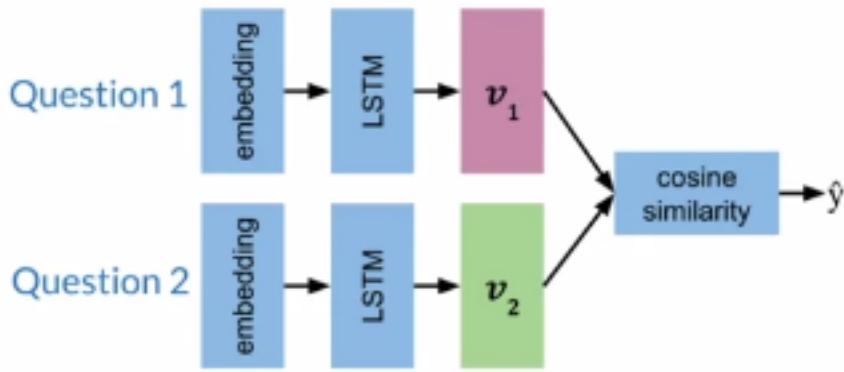
How old are you?

Are you seeing me?

Where are you?

What time is the game?

Computing The Cost



Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

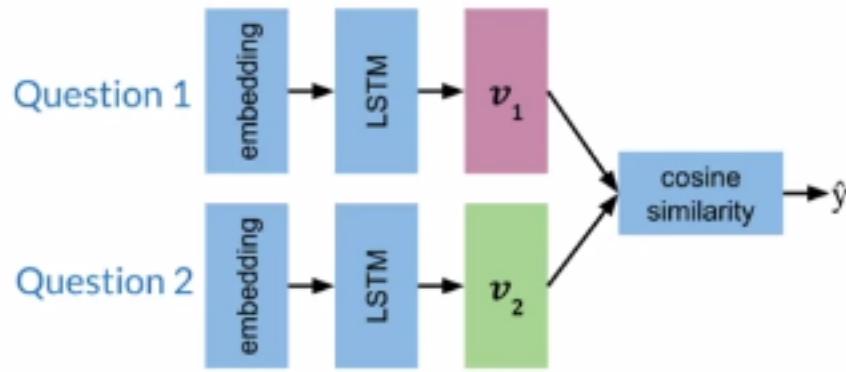
How old are you?

Are you seeing me?

Where are you?

What time is the game?

Computing The Cost



Batch 1
What is your age?
Can you see me?
Where are thou?
When is the game?

Batch 2
How old are you?
Are you seeing me?
Where are you?
What time is the game?

$v_1 = (1, d_{\text{model}})$
v_{1_1}
v_{1_2}
v_{1_3}
v_{1_4}
v_2
v_{2_1}
v_{2_2}
v_{2_3}
v_{2_4}

Hard Negative Mining

$$s(v_1, v_2)$$

		v_1				
		-1	-2	-3	-4	
		-1	0.9	-0.8	0.3	-0.5
		-2	-0.8	0.5	0.1	-0.2
		-3	0.3	0.1	0.7	-0.8
		-4	-0.5	-0.2	-0.8	1.0

Hard Negative Mining

$s(v_1, v_2)$

v_1

_1 _2 _3 _4

-1	0.9	-0.8	0.3	-0.5
-2	-0.8	0.5	0.1	-0.2
-3	0.3	0.1	0.7	-0.8
-4	-0.5	-0.2	-0.8	1.0

mean negative:

mean of off-diagonal values in each row

Hard Negative Mining

$$s(v_1, v_2)$$

		v_1	v_2			
		_1	_2	_3		
		_1	0.9	-0.8	0.3	-0.5
		_2	-0.8	0.5	0.1	-0.2
		_3	0.3	0.1	0.7	-0.8
		_4	-0.5	-0.2	-0.8	1.0

mean negative:

mean of off-diagonal values in each row

Hard Negative Mining

$s(v_1, v_2)$

v_1

_1 _2 _3 _4

-1	0.9	-0.8	0.3	-0.5
-2	-0.8	0.5	0.1	-0.2
-3	0.3	0.1	0.7	-0.8
-4	-0.5	-0.2	-0.8	1.0

v_2

mean negative:

mean of off-diagonal values in each row

closest negative:

off-diagonal value closest to (but less than) the value on diagonal in each row

Hard Negative Mining

$$\mathcal{L}_{\text{Original}} = \max(s(A, N) - s(A, P) + \alpha, 0)$$

Hard Negative Mining

$$\mathcal{L}_{\text{Original}} = \max(\underbrace{s(A, N) - s(A, P) + \alpha}_{\text{diff}}, 0)$$

Hard Negative Mining

$$\mathcal{L}_{\text{Original}} = \max (\underbrace{s(A, N) - s(A, P) + \alpha}_{\text{diff}}, 0)$$

$$\mathcal{L}_1 = \max (mean_neg - s(A, P) + \alpha, 0)$$

Hard Negative Mining

$$\mathcal{L}_{\text{Original}} = \max (\underbrace{s(A, N) - s(A, P) + \alpha}_{\text{diff}}, 0)$$

$$\mathcal{L}_1 = \max (mean_neg - s(A, P) + \alpha, 0)$$

Hard Negative Mining

mean negative: mean of off-diagonal values

$$\mathcal{L}_{\text{Original}} = \max(\underbrace{s(A, N) - s(A, P) + \alpha}_{\text{diff}}, 0)$$

$$\mathcal{L}_1 = \max(\text{mean_neg} - s(A, P) + \alpha, 0)$$

Hard Negative Mining

mean negative: mean of off-diagonal values

$$\mathcal{L}_{\text{Original}} = \max (\underbrace{s(A, N) - s(A, P)}_{\text{diff}} + \alpha, 0)$$

$$\mathcal{L}_1 = \max (mean_neg - s(A, P) + \alpha, 0)$$

$$\mathcal{L}_2 = \max (closest_neg - s(A, P) + \alpha, 0)$$

Hard Negative Mining

mean negative: mean of off-diagonal values

closest negative: closest off-diagonal value

$$\mathcal{L}_{\text{Original}} = \max(\underbrace{s(A, N) - s(A, P) + \alpha}_{\text{diff}}, 0)$$

$$\mathcal{L}_1 = \max(mean_neg - s(A, P) + \alpha, 0)$$

$$\mathcal{L}_2 = \max(closest_neg - s(A, P) + \alpha, 0)$$

Hard Negative Mining

mean negative: mean of off-diagonal values

closest negative: closest off-diagonal value

$$\mathcal{L}_{\text{Original}} = \max(\underbrace{s(A, N) - s(A, P) + \alpha}_{\text{diff}}, 0)$$

$$\mathcal{L}_1 = \max(mean_neg - s(A, P) + \alpha, 0)$$

$$\mathcal{L}_2 = \max(closest_neg - s(A, P) + \alpha, 0)$$

$$\mathcal{L}_{\text{Full}} = \mathcal{L}_1 + \mathcal{L}_2$$

Hard Negative Mining

$$\mathcal{L}_{\text{Full}}(A, P, N) = \mathcal{L}_1 + \mathcal{L}_2$$

Hard Negative Mining

$$\mathcal{L}_{\text{Full}}(A, P, N) = \mathcal{L}_1 + \mathcal{L}_2$$

$$\mathcal{J} = \sum_{i=1}^m \mathcal{L}_{\text{Full}}(A^{(i)}, P^{(i)}, N^{(i)})$$

Classification vs One Shot Learning

Classification vs One Shot Learning

Classification



Classify as 1 of K classes

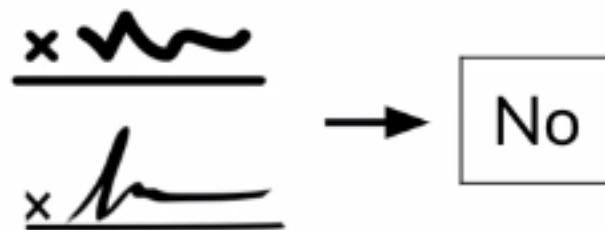
Classification vs One Shot Learning

Classification



Classify as 1 of K classes

One Shot Learning



Measure similarity between
2 classes

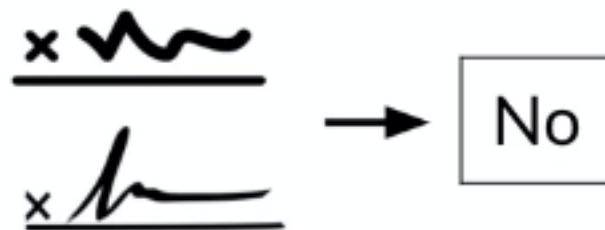
Classification vs One Shot Learning

Classification



Classify as 1 of K classes

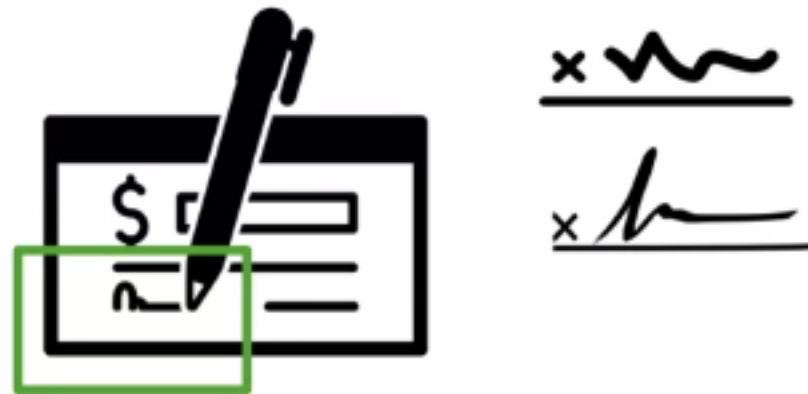
One Shot Learning



Measure similarity between
2 classes

One Shot Learning

No need for retraining !



One Shot Learning

No need for retraining !



One Shot Learning

No need for retraining !



Learn a similarity score!

One Shot Learning

No need for retraining !



Learn a similarity score!

$$s(sig1, sig2) > \tau$$

One Shot Learning

No need for retraining !



Learn a similarity score!

$$s(sig1, sig2) > \tau \quad \checkmark$$

One Shot Learning

No need for retraining !



Learn a similarity score!

$$s(sig1, sig2) > \tau \quad \checkmark$$

$$s(sig1, sig2) \leq \tau$$

One Shot Learning

No need for retraining !



Learn a similarity score!

$s(sig1, sig2) > \tau$

$s(sig1, sig2) \leq \tau$

Dataset

Question 1	Question 2	is_duplicate
What is your age?	How old are you?	true
Where are you from?	Where are you going?	false
:	:	:

Dataset

Question 1	Question 2	is_duplicate
What is your age?	How old are you?	true
Where are you from?	Where are you going?	false
:	:	:

Prepare Batches

Batch 1

What is your age?

Can you see me?

Where are thou?

When is the game?

Batch 2

How old are you?

Are you seeing me?

Where are you?

What time is the game?

Prepare Batches

Question 1:
batch size b

Question 2:
batch size b

Batch 1

What is your age?
Can you see me?
Where are thou?
When is the game?

Batch 2

How old are you?
Are you seeing me?
Where are you?
What time is the game?

Prepare Batches

Question 1:
batch size b

Question 2:
batch size b

- | | |
|---|-------------|
| <p>Batch 1</p> <ul style="list-style-type: none">What is your age?Can you see me?Where are thou?When is the game? <p>Batch 2</p> <ul style="list-style-type: none">How old are you?Are you seeing me?Where are you?What time is the game? | <p>q1_a</p> |
|---|-------------|

Prepare Batches

Question 1:
batch size b

Question 2:
batch size b

- | | | |
|---------|------------------------|------|
| Batch 1 | What is your age? | q1_a |
| | Can you see me? | |
| | Where are thou? | |
| | When is the game? | |
| Batch 2 | How old are you? | q2_a |
| | Are you seeing me? | |
| | Where are you? | |
| | What time is the game? | |

Prepare Batches

Question 1:
batch size b

Batch 1

{ What is your age? q1_a
Can you see me? q1_b
Where are thou?
When is the game?

Question 2:
batch size b

Batch 2

{ How old are you? q2_a
Are you seeing me?
Where are you?
What time is the game?

Prepare Batches

Question 1:
batch size b

Batch 1

What is your age?	q1_a
Can you see me?	q1_b
Where are thou?	
When is the game?	

Question 2:
batch size b

Batch 2

How old are you?	q2_a
Are you seeing me?	q2_b
Where are you?	
What time is the game?	

Prepare Batches

Question 1:
batch size b

Batch 1

{ What is your age?
Can you see me?
Where are thou?
When is the game?

q1_a
q1_b

Question 2:
batch size b

Batch 2

{ How old are you? q2_a
Are you seeing me? q2_b
Where are you?
What time is the game?

Prepare Batches

Question 1:
batch size b

Batch 1

What is your age?	q1_a
Can you see me?	q1_b
Where are thou?	
When is the game?	

Question 2:
batch size b

Batch 2

How old are you?	q2_a
Are you seeing me?	q2_b
Where are you?	
What time is the game?	

Prepare Batches

Question 1:
batch size b

Question 2:
batch size b

- Batch 1
- What is your age?
 - Can you see me?
 - Where are thou?
 - When is the game?
- Batch 2
- How old are you?
 - Are you seeing me?
 - Where are you?
 - What time is the game?

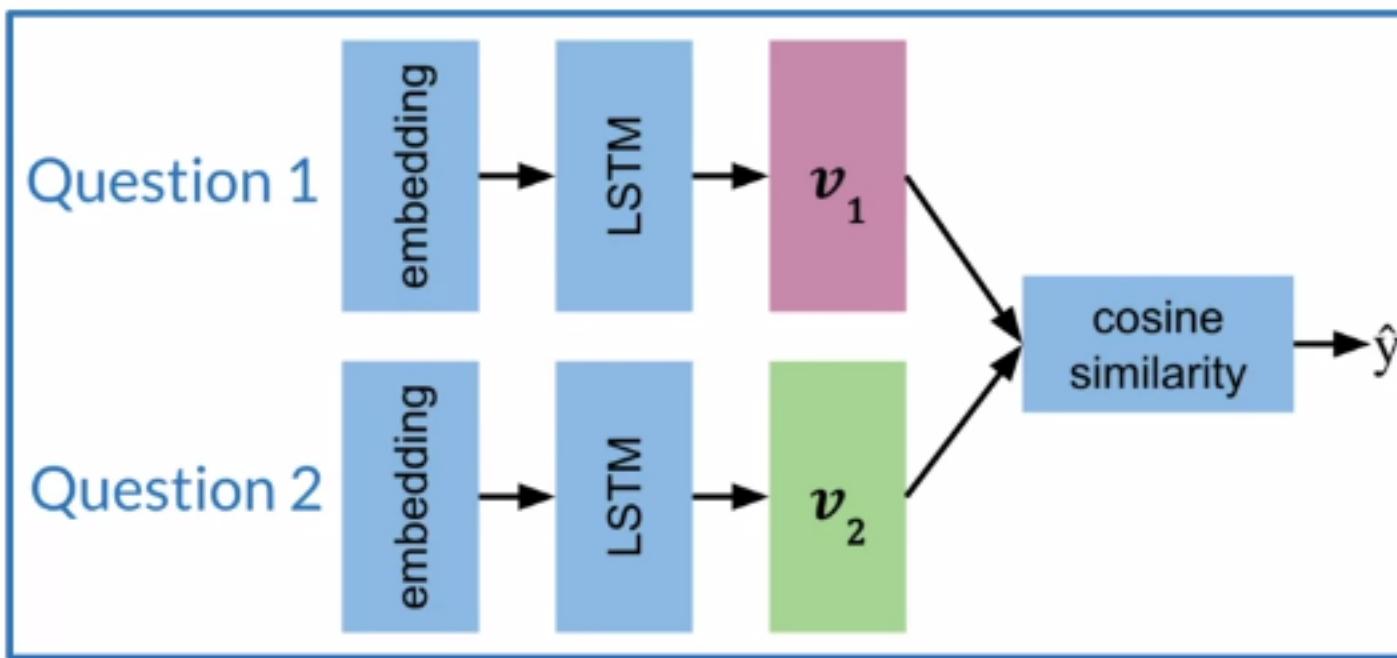
$$\boldsymbol{v}_1 = (1, d_{\text{model}})$$

v_{1_1}				
v_{1_2}				
v_{1_3}				
v_{1_4}				

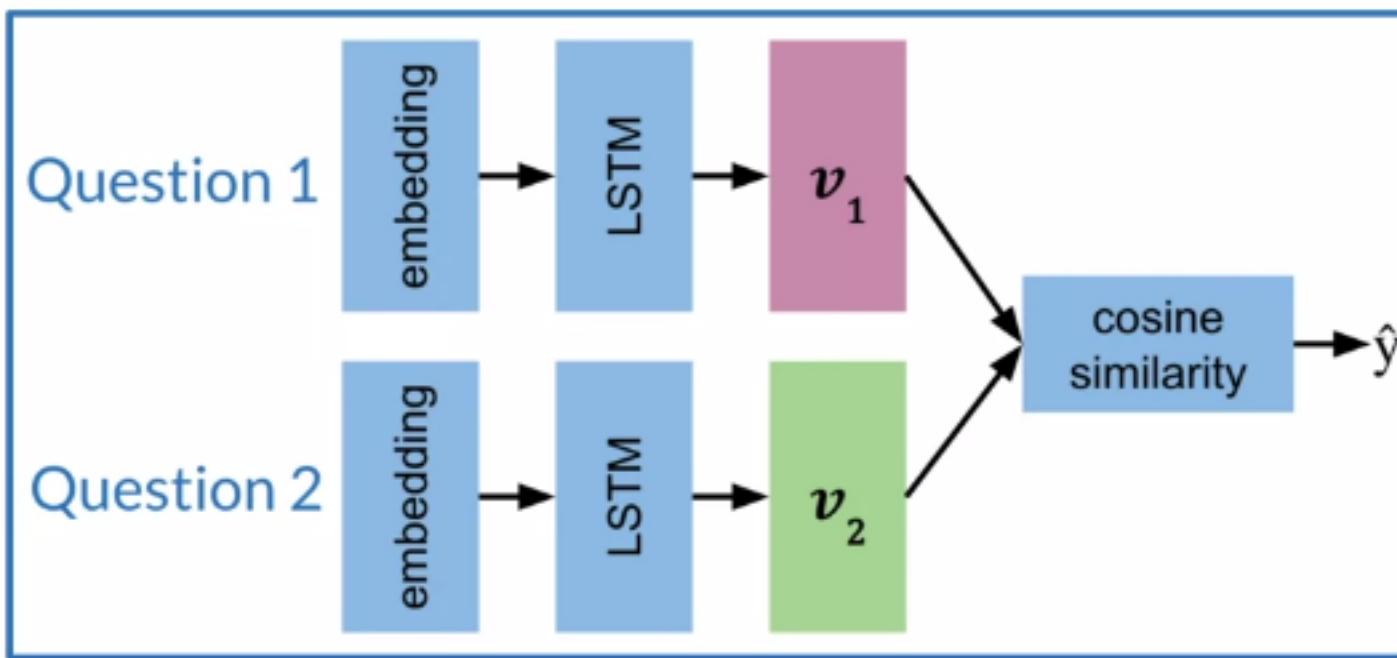
$$\boldsymbol{v}_2$$

v_{2_1}				
v_{2_2}				
v_{2_3}				
v_{2_4}				

Siamese Model



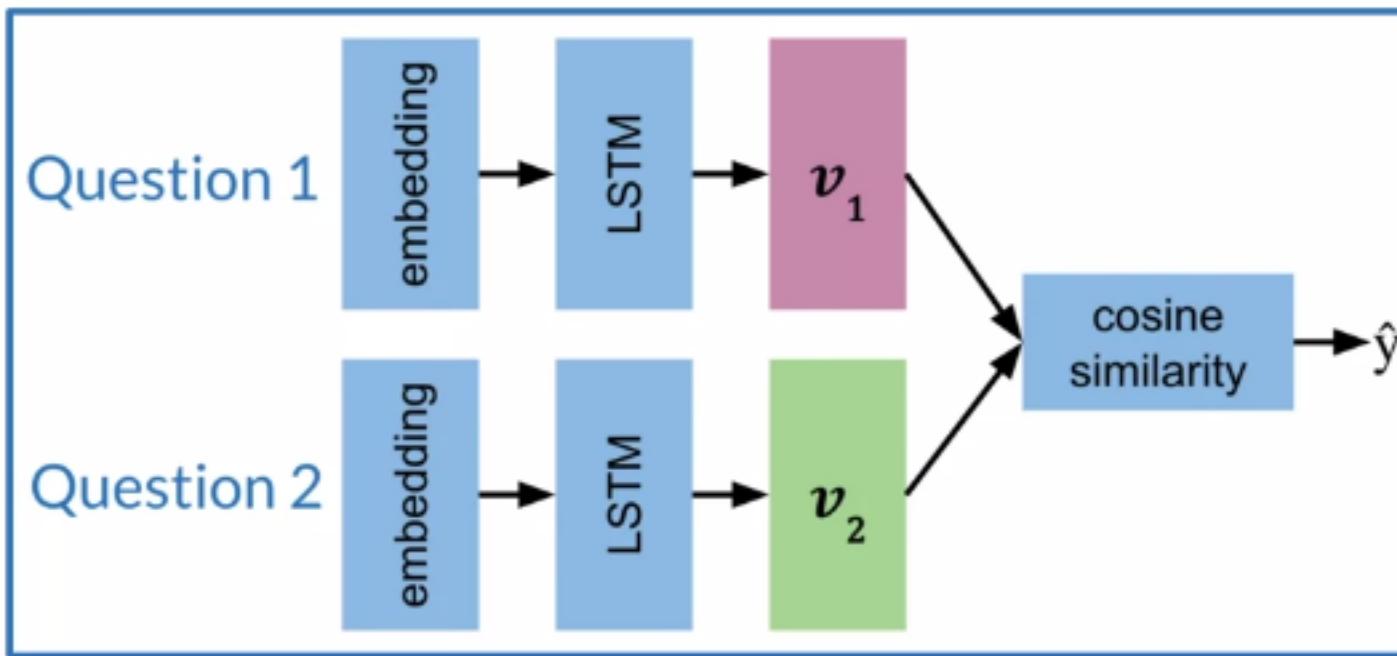
Siamese Model



Create a subnetwork:

- 1) Embedding
- 2) LSTM

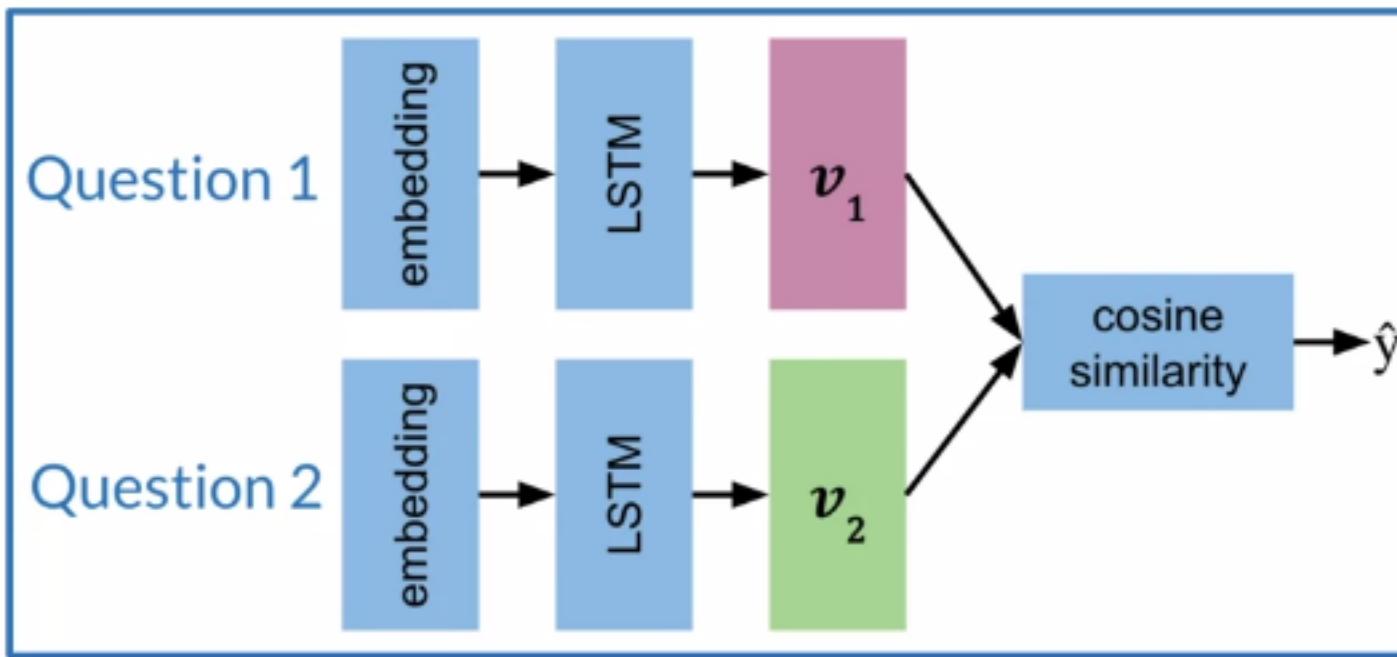
Siamese Model



Create a subnetwork:

- 1) Embedding
- 2) LSTM
- 3) Vectors

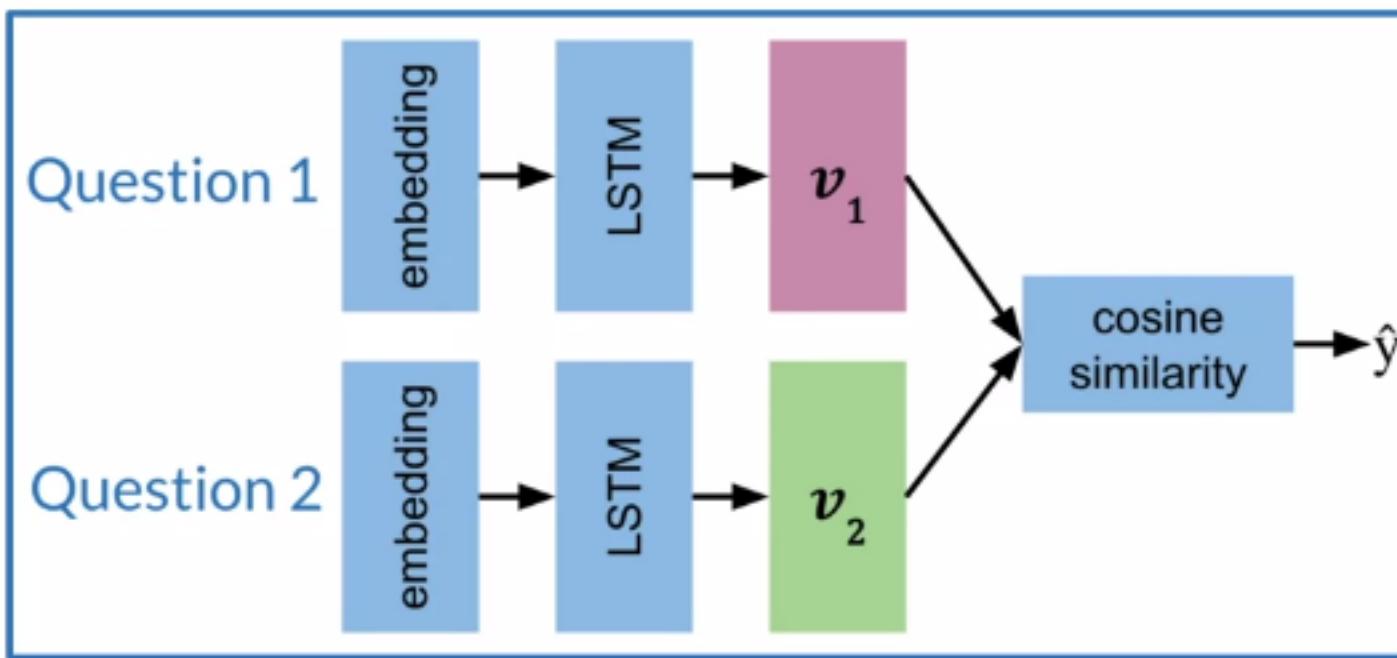
Siamese Model



Create a subnetwork:

- 1) Embedding
- 2) LSTM
- 3) Vectors
- 4) Cosine Similarity

Siamese Model



Create a subnetwork:

- 1) Embedding
- 2) LSTM
- 3) Vectors
- 4) Cosine Similarity

Testing

Testing

1. Convert each input into an array of numbers

Testing

1. Convert each input into an array of numbers
2. Feed arrays into your model

Testing

1. Convert each input into an array of numbers
2. Feed arrays into your model
3. Compare v_1, v_2 using cosine similarity

Testing

1. Convert each input into an array of numbers
2. Feed arrays into your model
3. Compare v_1, v_2 using cosine similarity
4. Test against a threshold τ