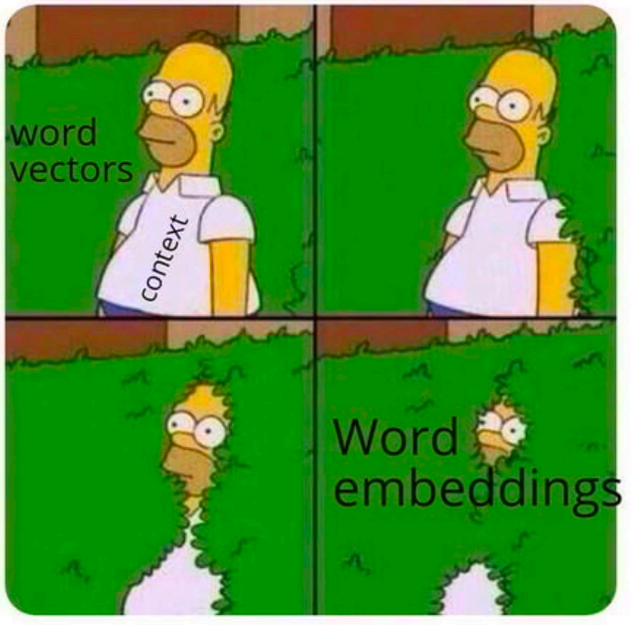
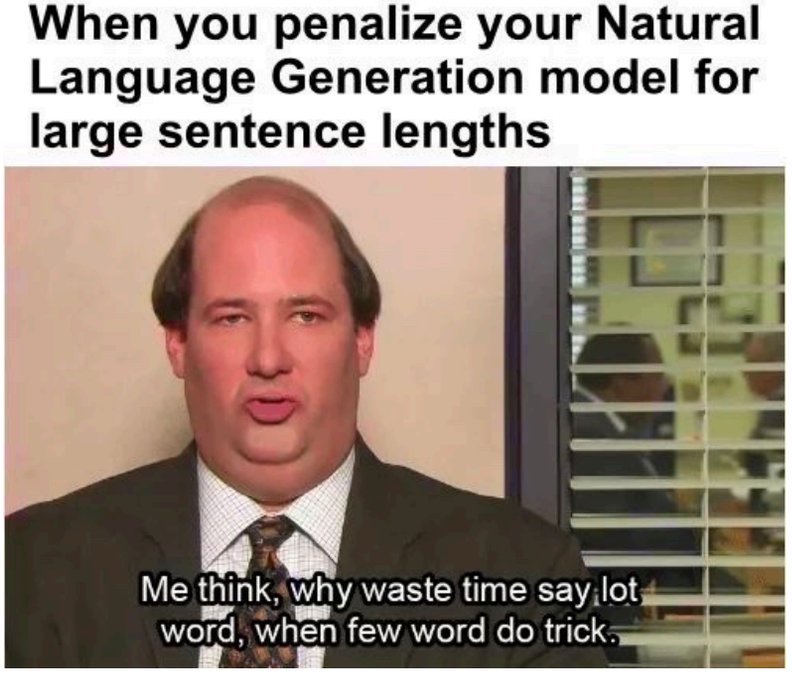
May 12, 2023.

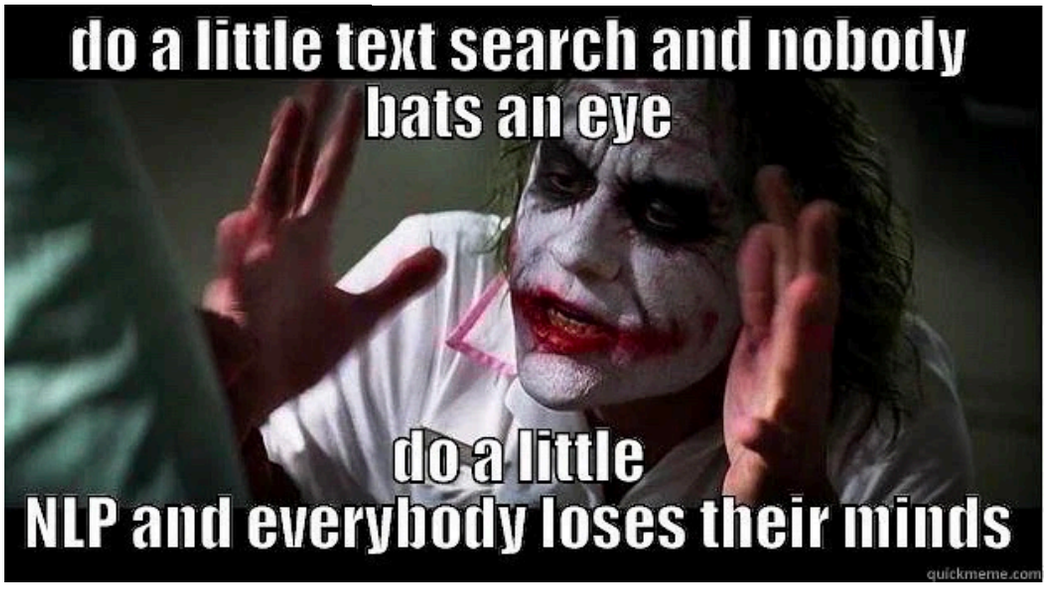
DSML : NLP module.

Attention → Transformers

Class starts @ 9:05

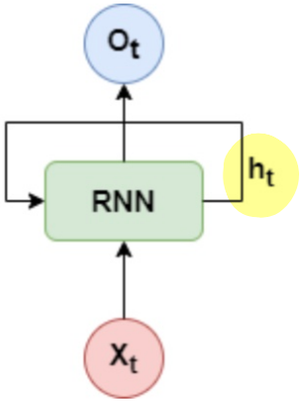
## Recap:

* The problem: capture correct contextual information from a previously seen text, and apply it correctly when needed.

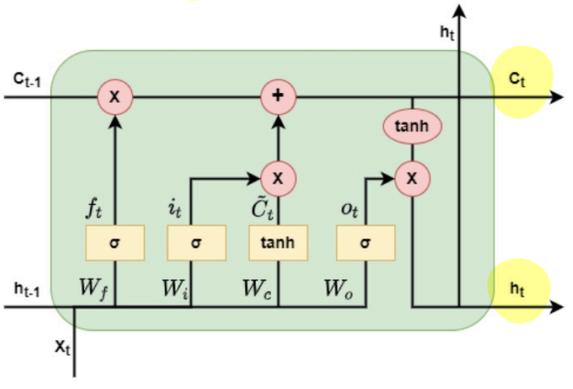
weakness of

* Approaches we have seen so far:

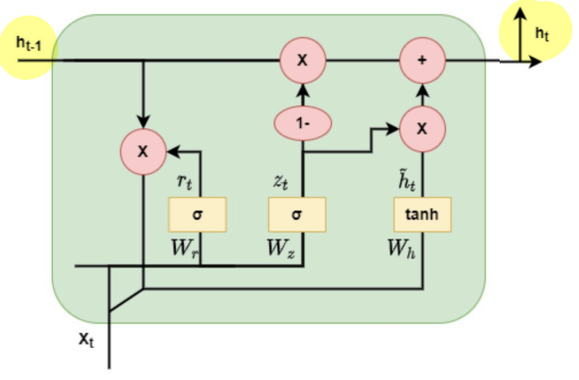
(a) RNNs

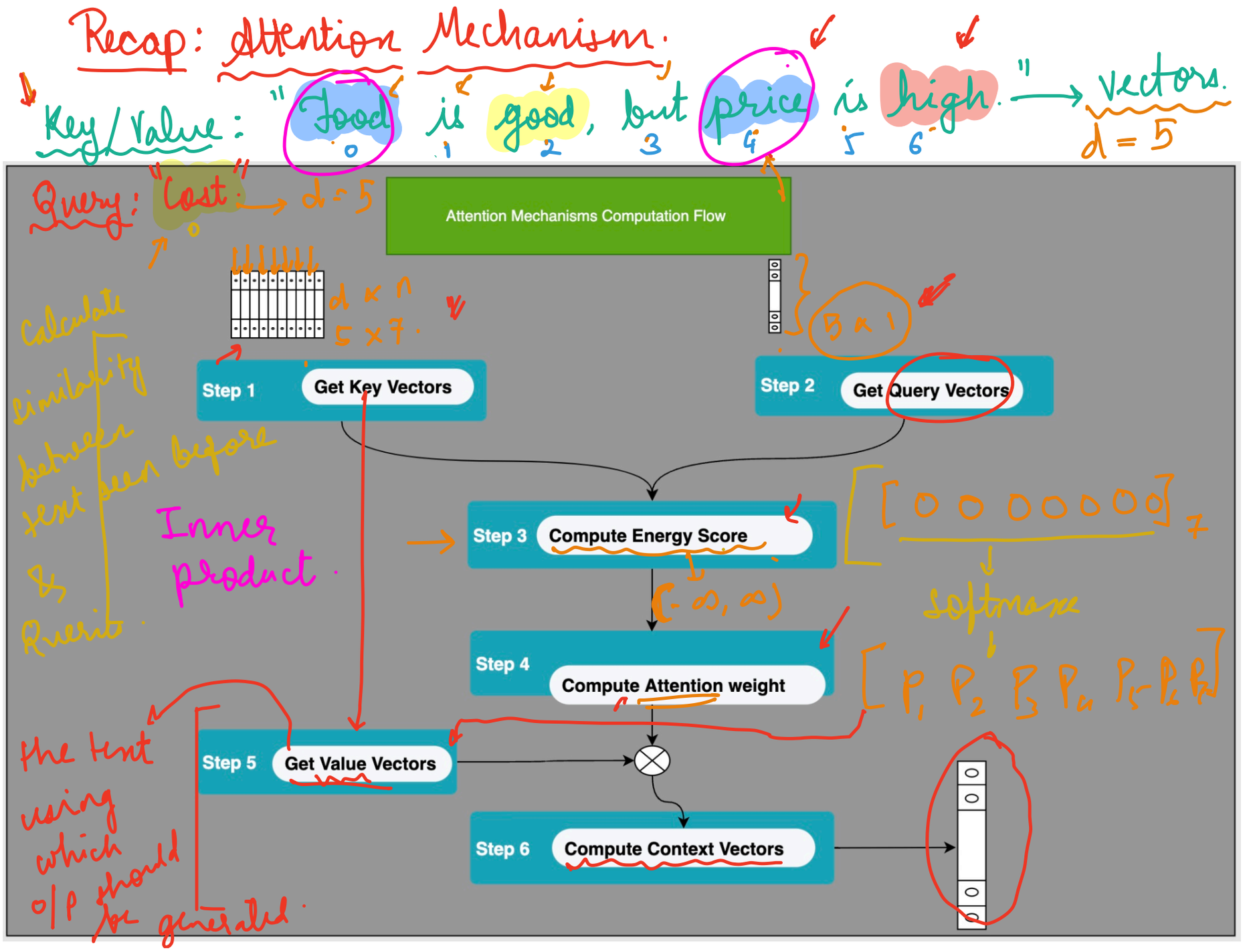



(b) LSTMs




(c) GRUs




Recap: Attention Mechanism.
Key/Value: "Food is good, but price is high." → vectors. d = 5
Query: "Cost" → d = 5
Attention Mechanisms Computation Flow
d x n
5 x 7
B x 1
Step 1 Get Key Vectors
Step 2 Get Query Vectors
Calculate similarity between text seen before & Query
Inner product
Step 3 Compute Energy Score
(-∞, ∞)
softmax
Step 4 Compute Attention weight
Step 5 Get Value Vectors
Step 6 Compute Context Vectors
the text using which o/p should be generated.

# Recap: Deep Dive into LSTM Results:

1

```
Review: facebook has suspended canadian firm claiming it may be affiliated with cambridge
analytica and improperly received its users data in response to facebook s action said
that it has never been and is not a part of cambridge analytica earlier it was revealed
that cambridge analytica exploited the data of facebook users to meddle in the us elections

Original summary: start facebook suspends canadian firm amid data scandal probe end
Predicted summary:  start facebook suspends facebook over fake news data scandal end
```

2

```
Review: siddaramaiah on thursday accepted a nine member committee s design for the
proposed official state flag the committee was set up last year to examine the legal
feasibility of an official state flag after demands by pro kannada activists if the centre
approves the proposal karnataka will be the second state to have a separate flag after jammu and kashmir

Original summary: start karnataka government unveils official state flag end
Predicted summary:  start karnataka govt to launch new political entity end
```

## Problems:

1] Forgetting necessary information.

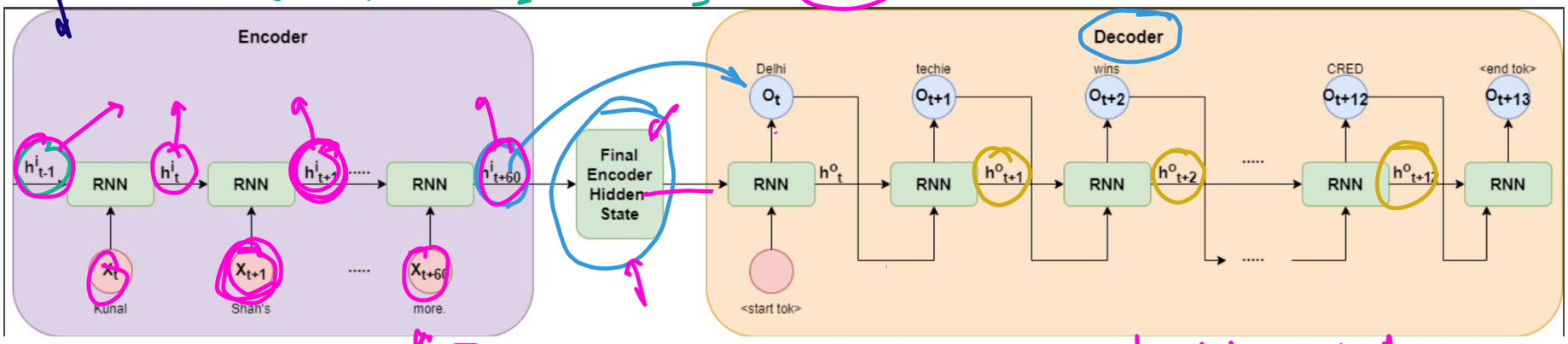
2] Focuses on the wrong parts of the text while generating summary.

Agenda:

* How to use attention to fix problems with RNN / LSTM / GRUs?
* Different types of attention mechanisms.
* Code: Using attention for Aspect based sentiment analysis.
* Transformers: Do we even need an RNN?

# LSTM Encoder Decoder Architecture.

→ food is good but price is high.
0 1 2 3 4 5 6




① Maybe key passing all hidden states forward.
② Maybe also key passing info of the input word.

## Attention:

- Key Vectors. ← hidden states from encoder.
- Query Vectors. ← hidden states from decoder.

Energy Score → Attention weights (softmax)

$\langle \cdot \rangle$ → inner product.

Value Vectors ← hidden states from encoder.

Attention weights ⊗ Value Vectors → Context Vector $c_t$

Before Attention (RNN math)

$$o_t = f(h_{t-1}^{dec}, o_{t-1})$$

After attention.

$$o_t = f(h_{t-1}^{dec}, o_{t-1}, c_t)$$

$c_t$: context vector from attention block.

## Designing attention mechanisms:

Q1] How to set Key, Query, Value?

Ans. Highly dependent on use case / task that we are performing.

Q2] How to generate energy score & Attention weights?

Ans. Choose one of many options.

Q3) How to use content vector?

Ans. Highly dependent on use case / task that we are performing.

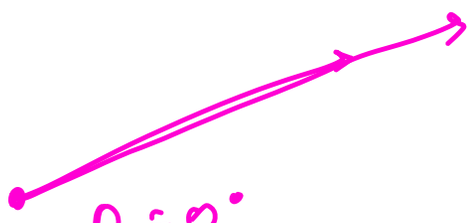
Key → $n \times 1$ Query is $k \times 1$

where $k \neq n$

$$\underbrace{Q^T_{1 \times k} \; W_{k \times n} \; K_{n \times 1}}$$

cosine similarity = 1 - cosine_distance.

✓ cosine distance = 1 - cosine similarity.



$\theta = 0°$

$\cos(\theta) = 1$

$-1 \qquad 1.$

distance → 0 — ∞.

0 – 2.

Which of the following statements is/are True:

A. Magnitude of vector affects Cosine similarity. F
B. Range of Cosine Similarity is from [-1, 1] and that of Cosine Distance is [0, ∞) ✗ F
C. Cosine Similarity is used to capture the relationship of the words within a corpus, with other words. T
D. Cosine Similarity = 1 + Cosine Distance T

will get this fixed.