PH451, PH551

March 25, 2025

Announcements

- **Proposals due on 04/03**
- **Midterm - Tuesday 04/10**

Proposal

- **Should focus on your main idea**
- **Contain motivation and a literature search**
- **Brief outline and planned plots**
- **Discuss metrics that will be used to estimate performance and determine success**

Last time: Sequential Data

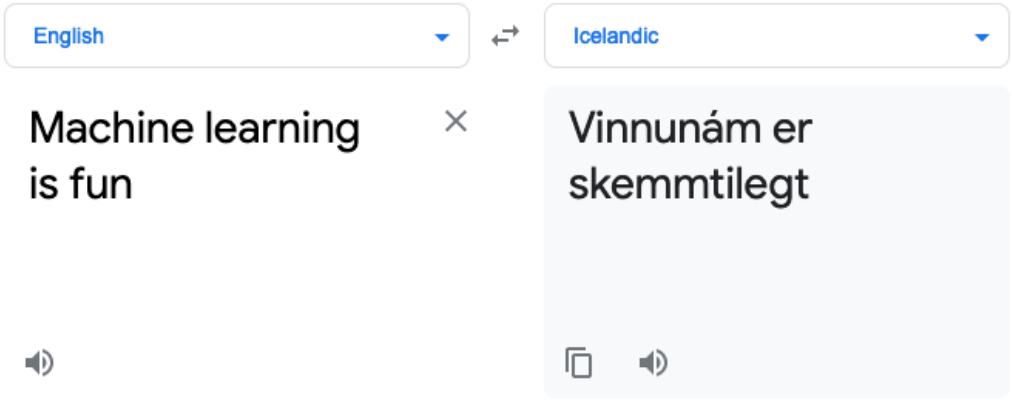
- **Input:**
  - Fixed size
- **Output**
  - Sequence

The man in grey swings a bat while the man in black looks on.

- **Example: image captioning**

Some Applications

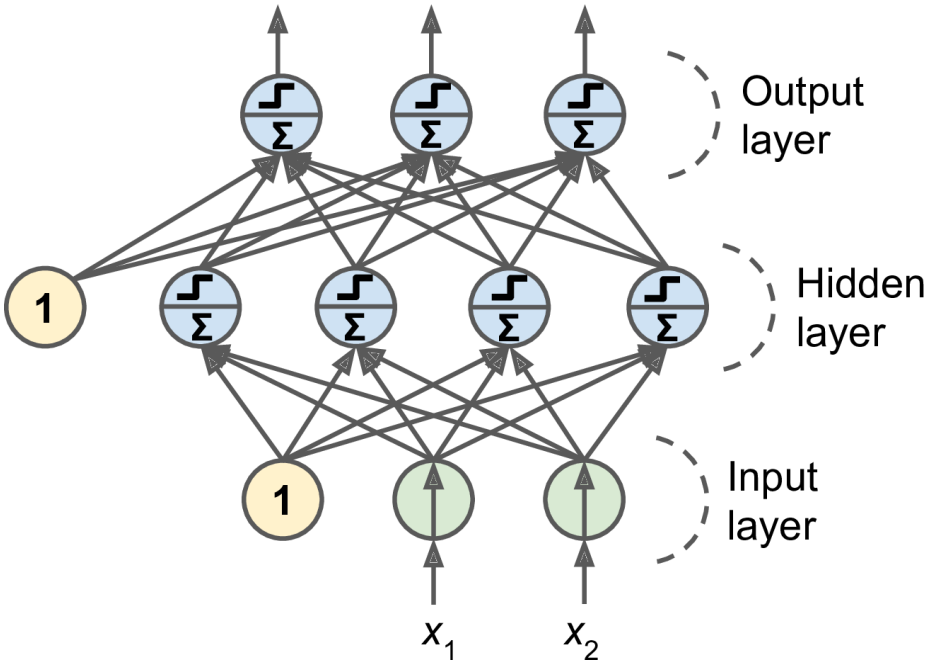
- **Input:**
  - Sequence
- **Output**
  - Sequence



- **Example: Google Translate**

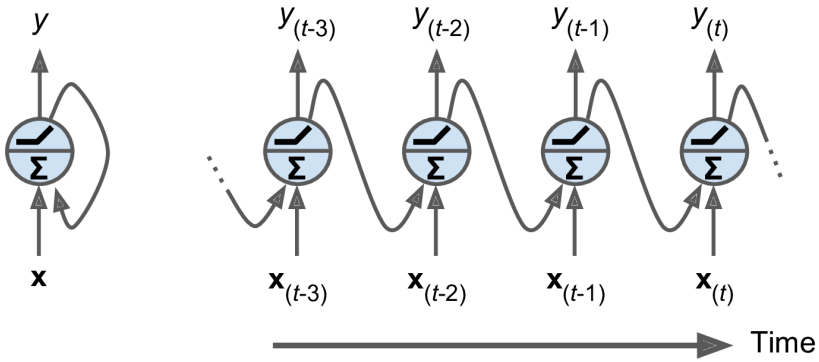
Last time: RNN vs MLP

MLP

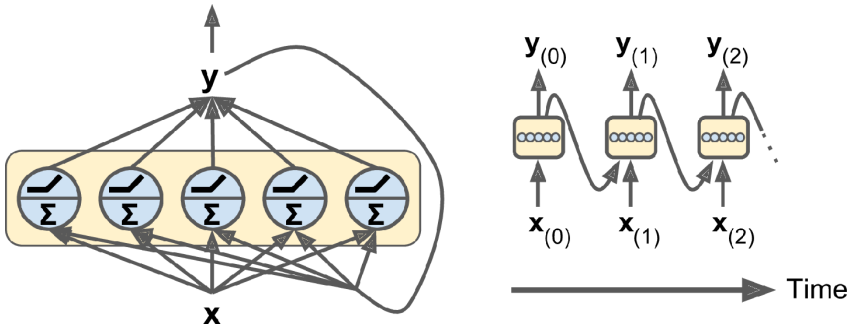


No loops

RNN neuron (unrolled)



RNN layer (unrolled)



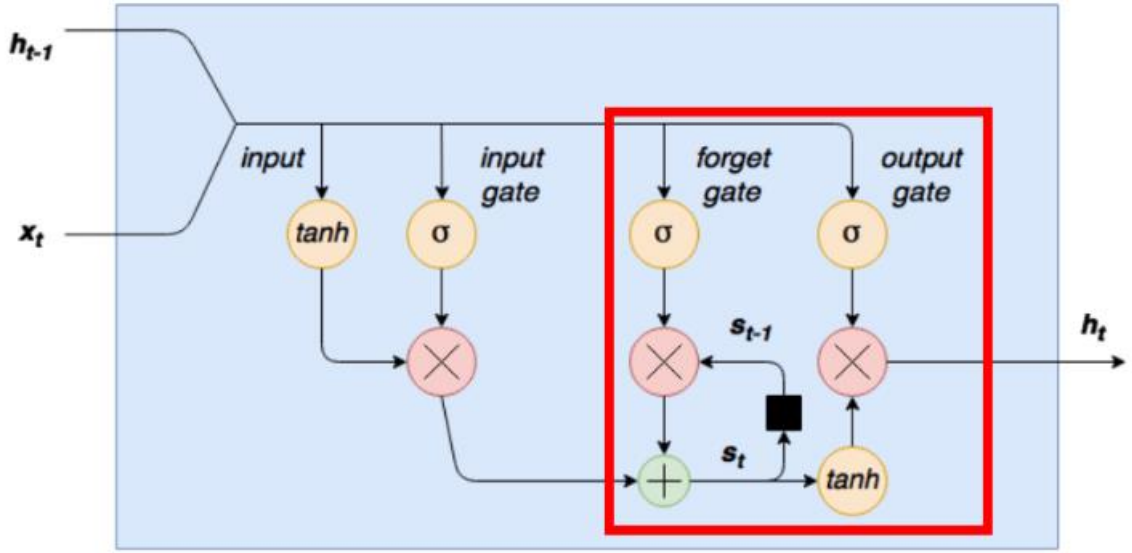
Outline

- **Recurrent Neural Networks**
- **LSTMs**
- **Transformers**

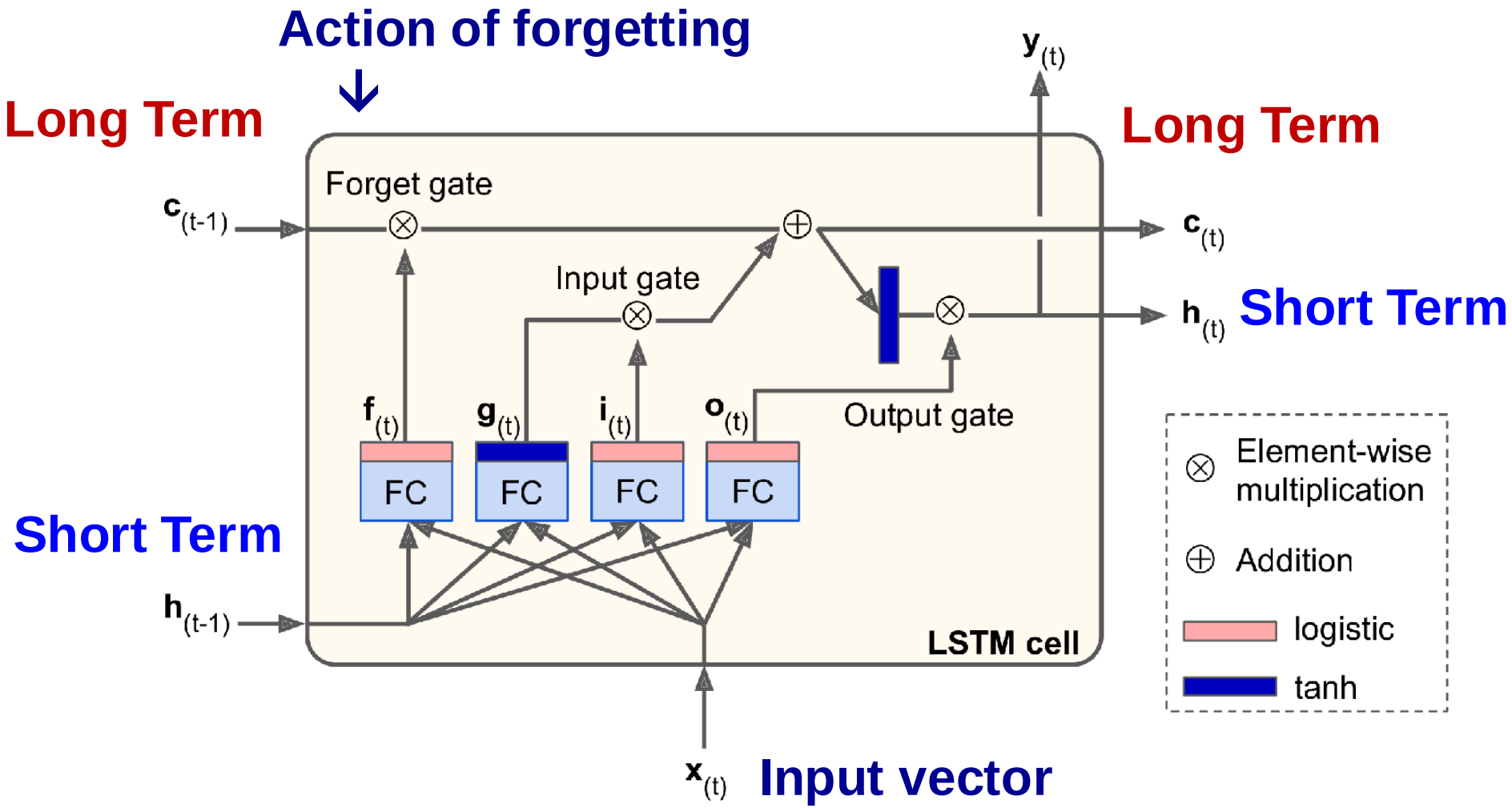
RNN Variants

Long Short Term Memory (LSTM)

- Hochreiter and Schmidhuber (1997)
- Modification of basic RNN preserving memory over time



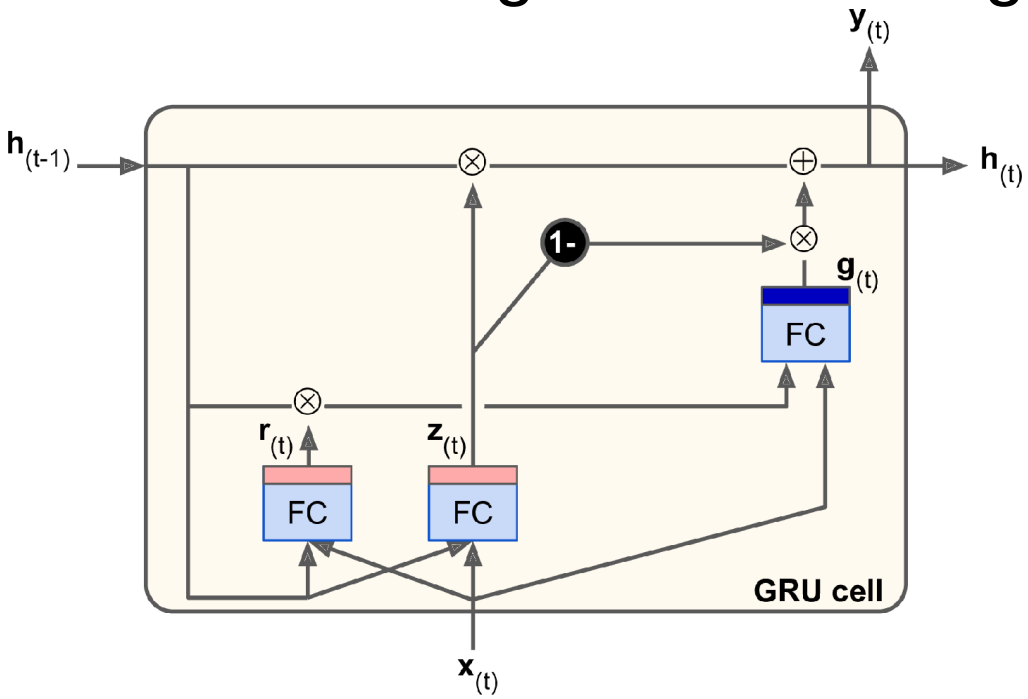
LSTM


Action of forgetting
Long Term
Long Term
Short Term
Short Term
Input vector
Forget gate
Input gate
Output gate
LSTM cell
FC
FC
FC
FC
Element-wise multiplication
Addition
logistic
tanh

GRU

Gated Recrurrent Unit (GRU)

- Cho et al. (2014)
- Simplified LSTM in a single vector and gate controller

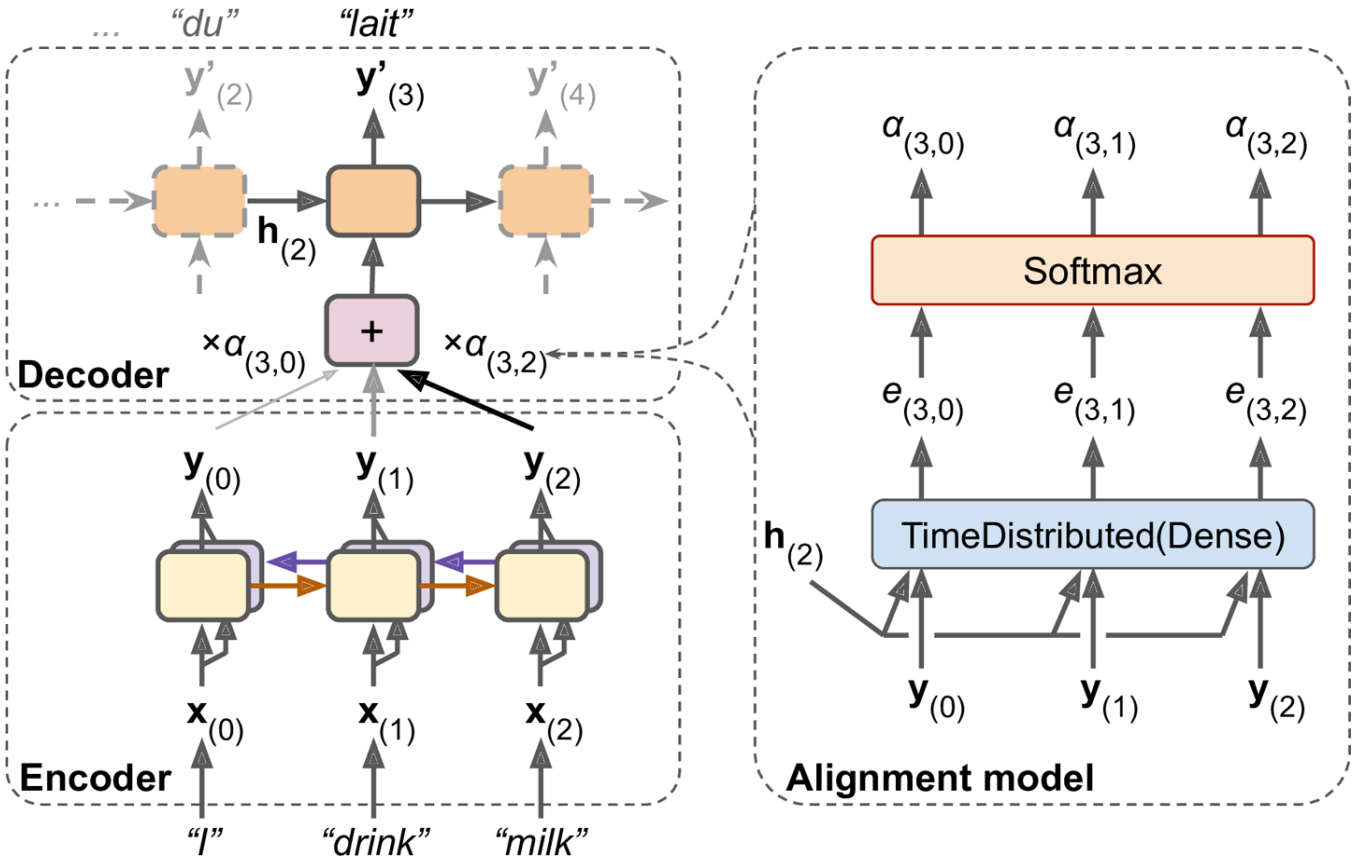


Attention Mechanisms

Attention

- Bahdanau et al. (2014)
- Encoder-decoder architecture that can focus its attention on specific input embeddings, shortening the path to output
- Idea that led to significant improvement in RNNs and turned out to be useful beyond RNNs

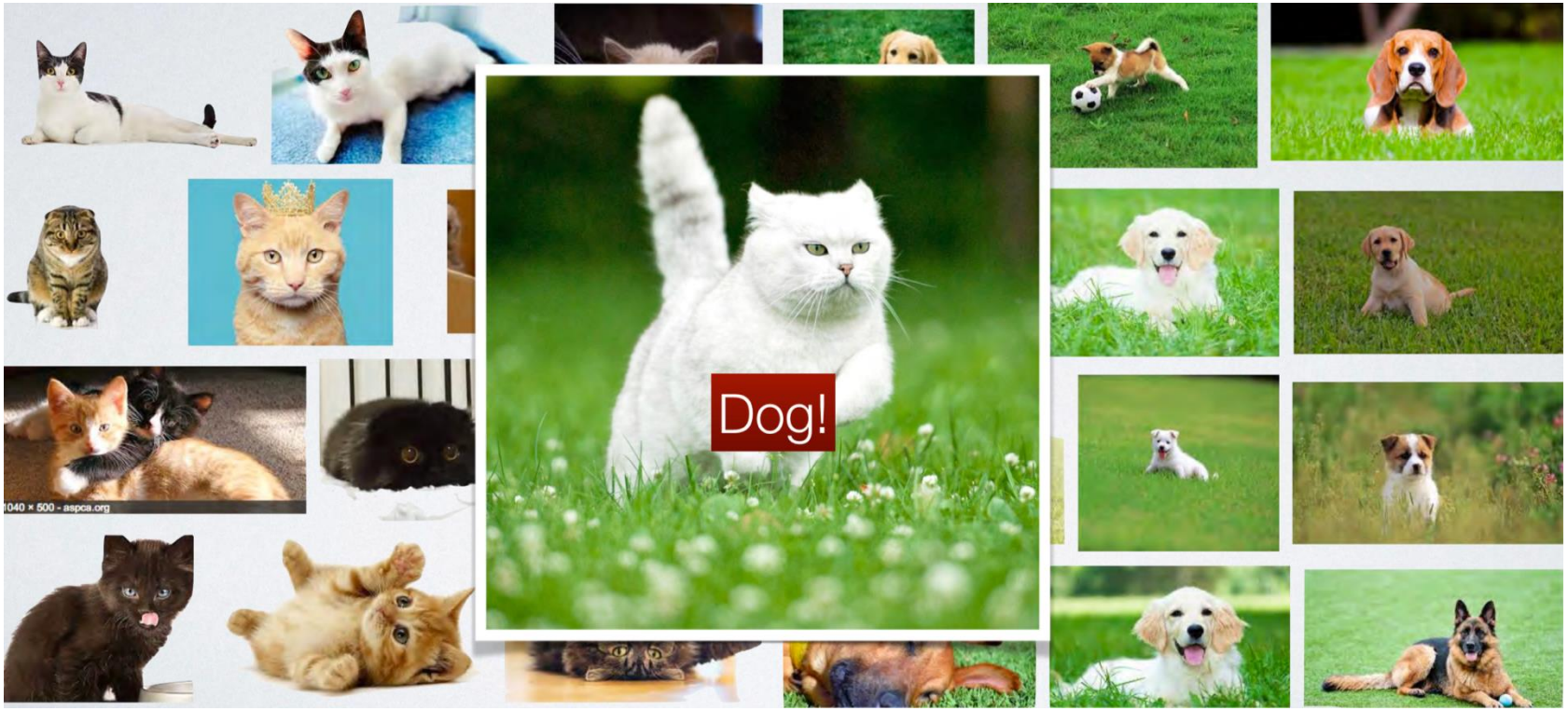
Attention Mechanisms



Attention Layer

Weights α tell you how much decoder should pay attention to each input "word"

Back to Explainability

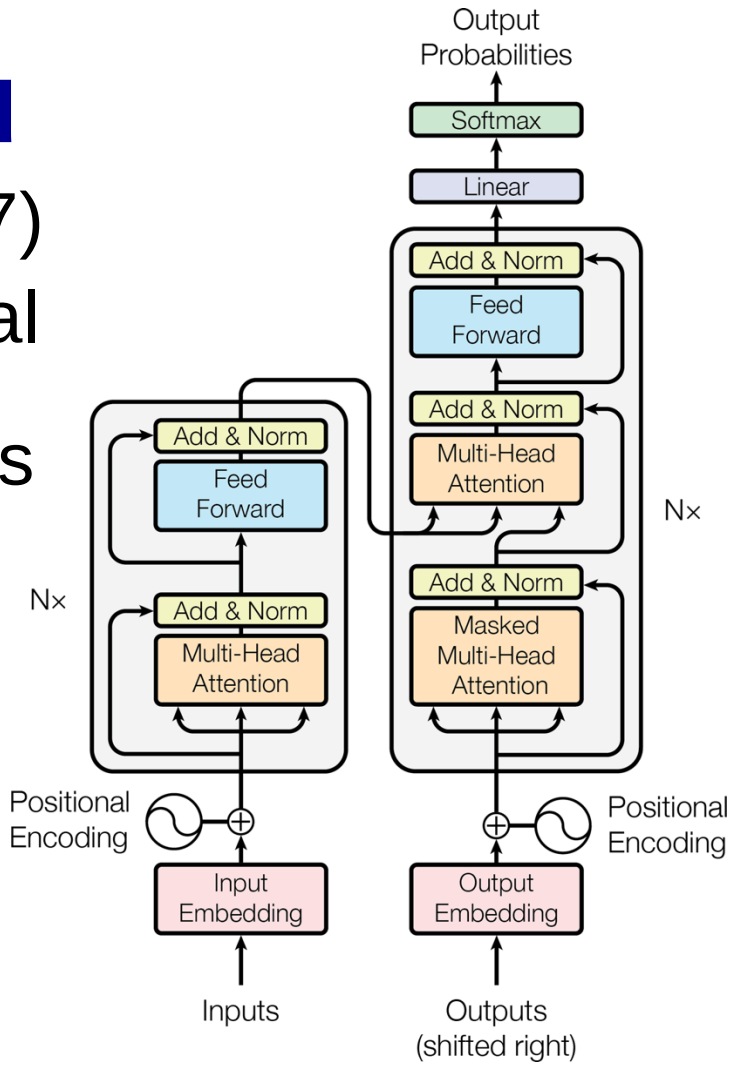


Attention can help you figure out what the model is trying to do

Transformers

Attention is all you need

- Vasvani et al. (Google, 2017)
- No recurrent or convolutional layers, just attention + normalization + dense layers
- Works on sets of vectors
- Significant improvement on RNNs for machine translation and other NLP tasks



BERT, GPT 1-3

Generative Pre-Training (GPT)

- Radford et al. (OpenAI, 2018-present)
- Unsupervised pre-training with transformer-like architectures
- GPT-3 Larger network trained for NLP (175 billion parameters), marked improvement compared to GPT-2
- GPT-3.5 improved on GPT-3 - Chat GPT
- GPT-4 (multimodal+vision), RLHF

BERT

- Devlin et al (Google 2018)
- Bi-directional encoder with transformers