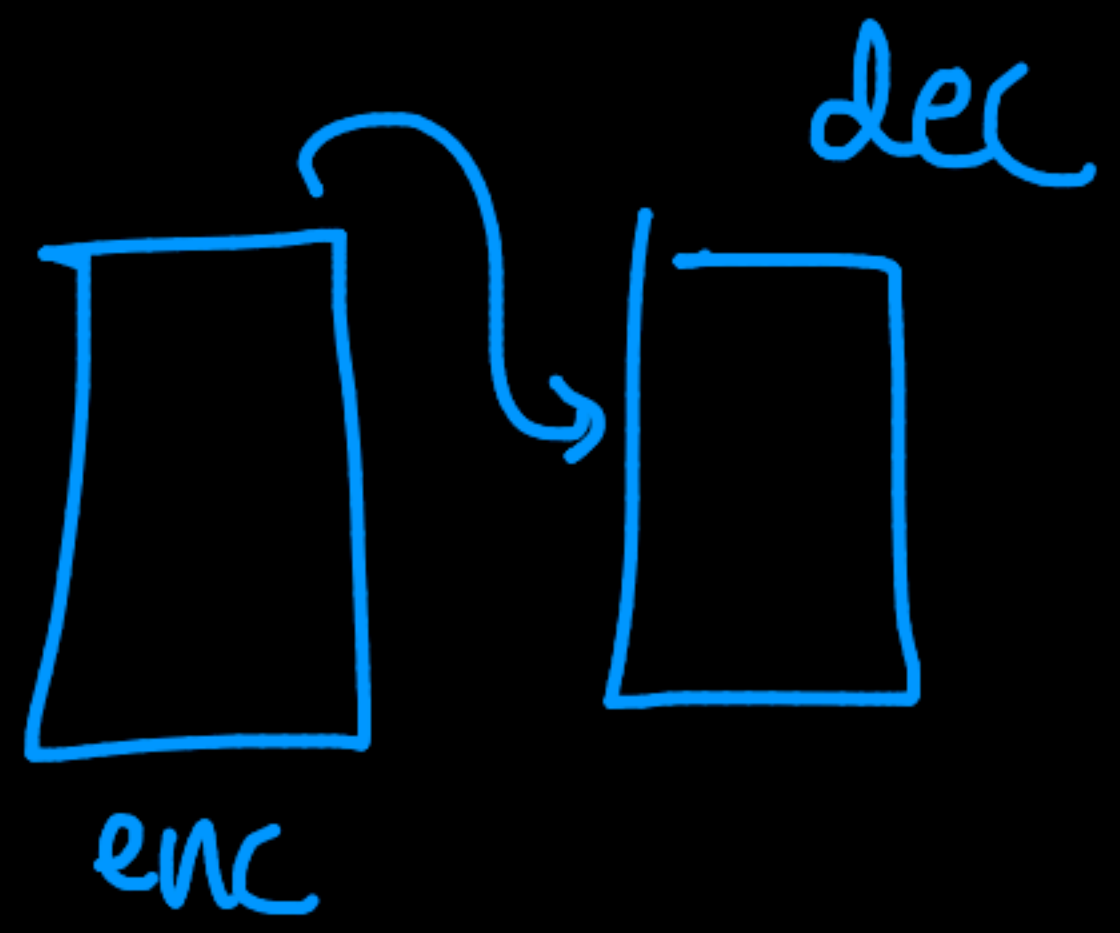
Agenda:

→ BERT - code

→ Two more problems + code (NLP)

(Q)

encoder-decoder
- → LSTM
- → Transformers ✓

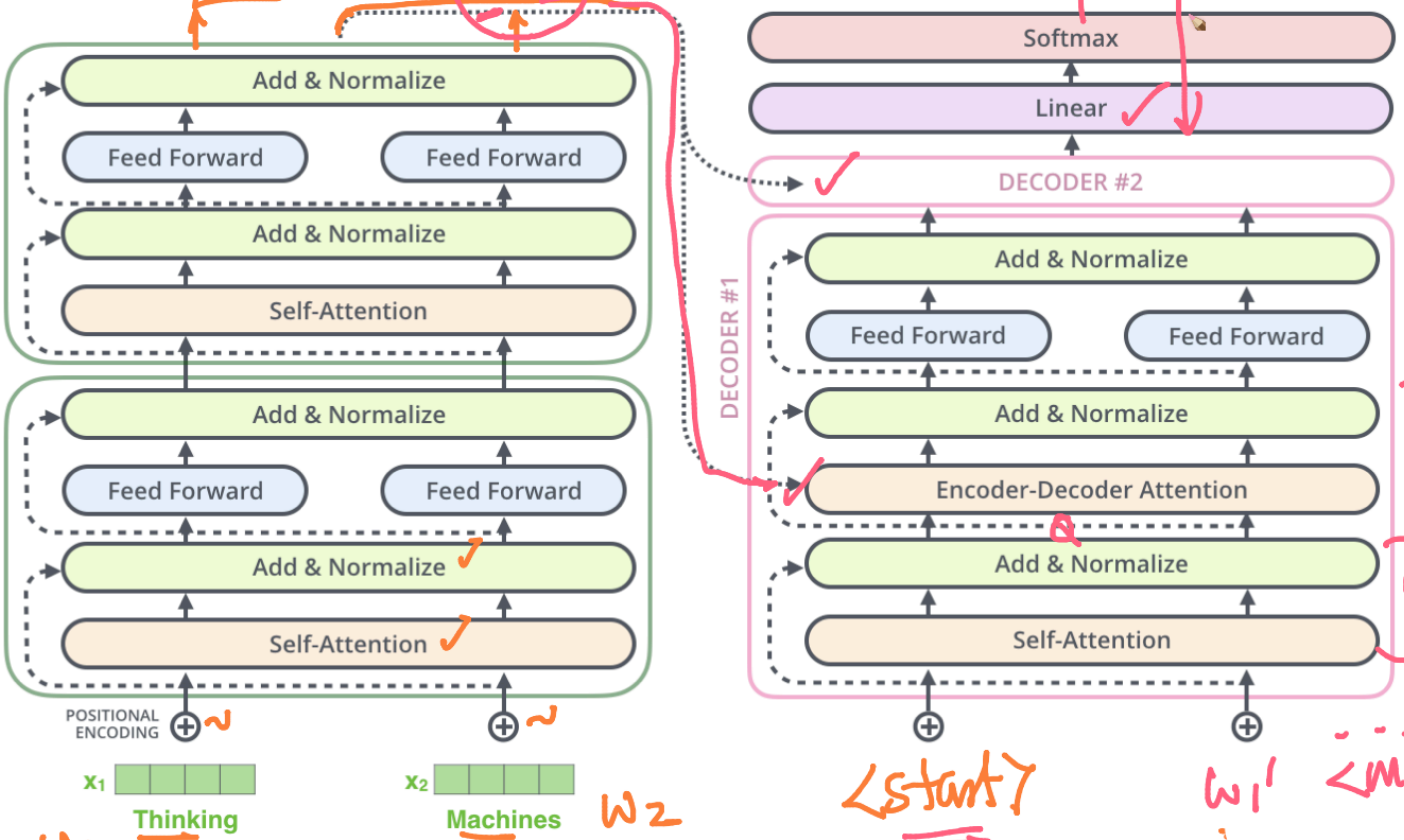


(Recap)

This goes for the sub-layers of the decoder as well. If we're to think of a Transformer of 2 stacked encoders and decoders, it would look something like this.

encdec K, V

Stack



① w_1'
② w_2'
③ w_3'

K words

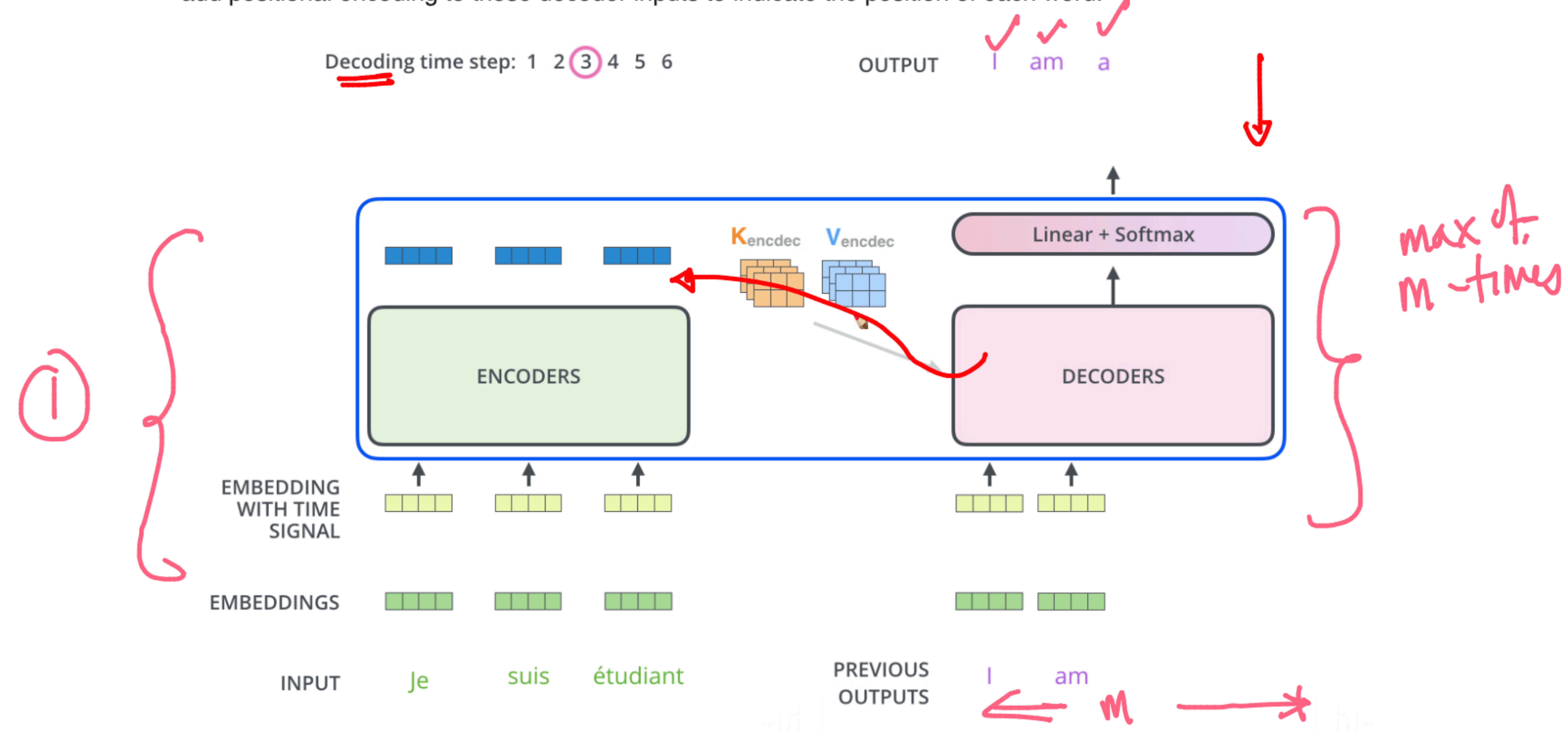
w_1 W_2

<start> w_1' <mask> m-words

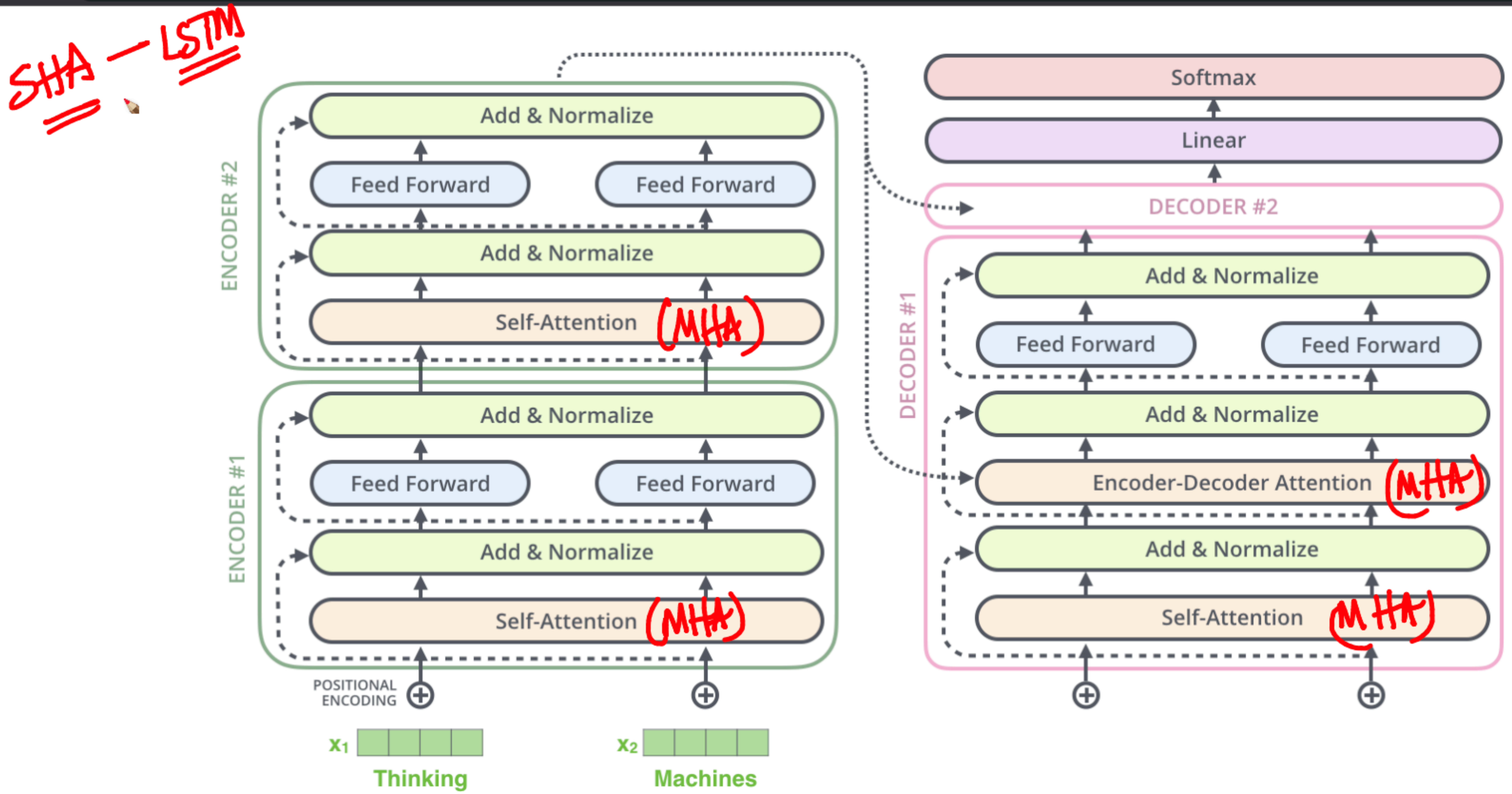
prev-o/ps

The Decoder Side

completed its output. The output of each step is fed to the bottom decoder in the next time step, and the decoders bubble up their decoding results just like the encoders did. And just like we did with the encoder inputs, we embed and add positional encoding to those decoder inputs to indicate the position of each word.



The self attention layers in the [illegible] in the encoder:

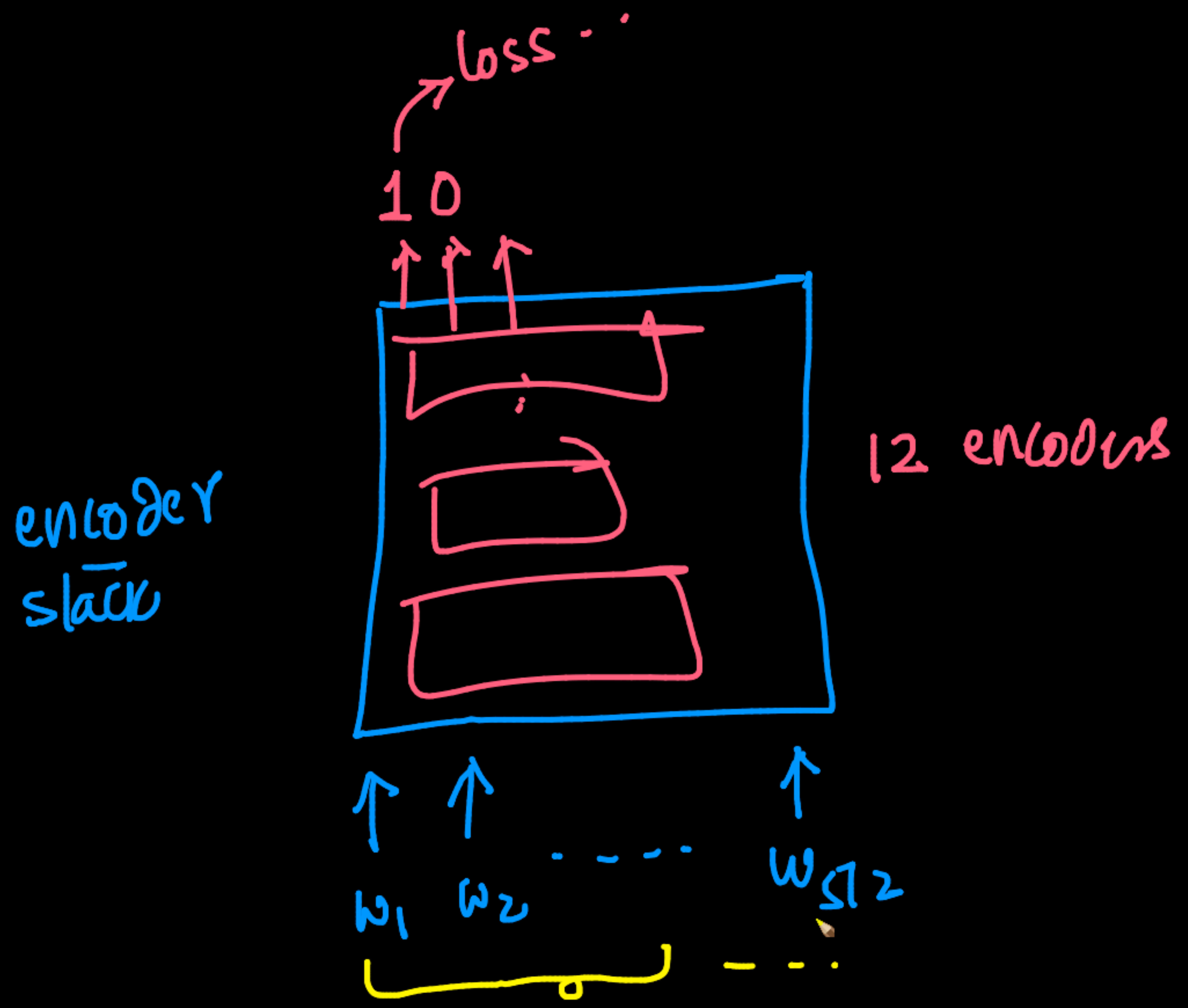


The Decoder Side

Now that we've covered most ... how the components of

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:22

7 / 7



{ 1: disease name
{ 0: non disease

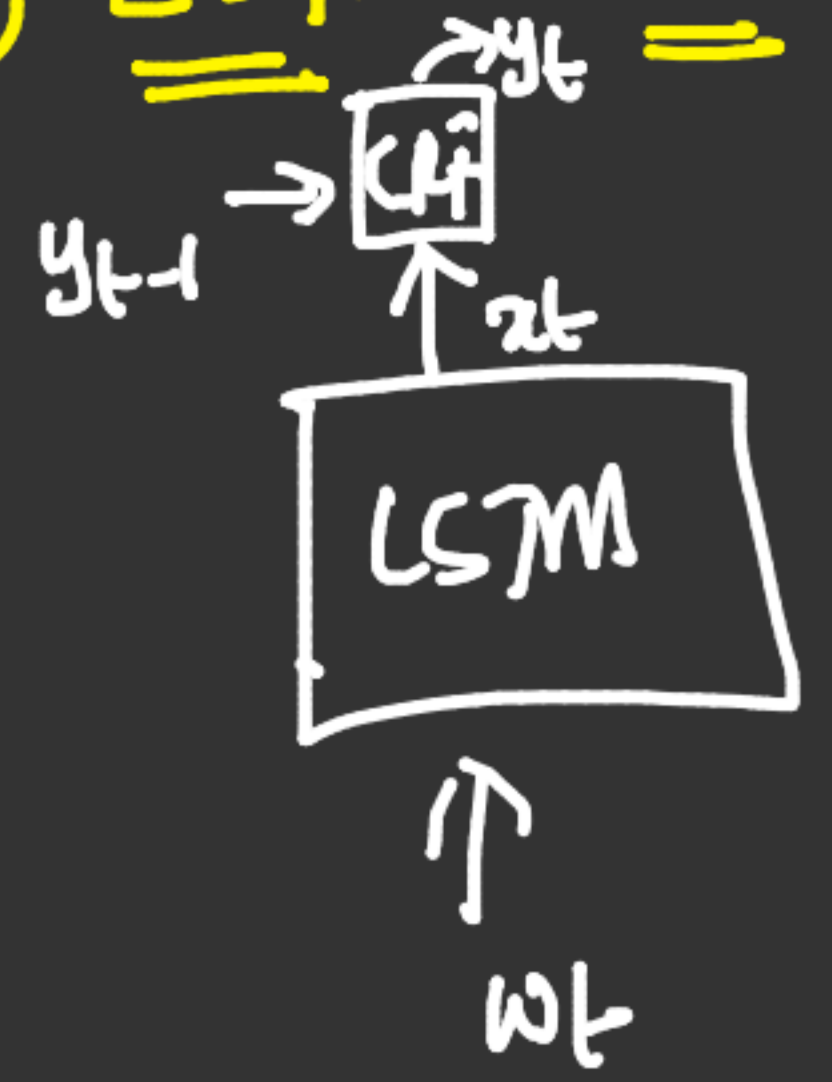
{ 0 : non-dis
{ 1 : dis.
{ 2 : padding

[ ]

	Category	Article
221	Technology	world tour for top video gamers two uk gamers ...
852	Entertainment	prince crowned top music earner prince earne...
80	Business	us company admits benin bribery a us defence a...
73	Sports	funding cut hits wales students the wales stud...
614	Entertainment	spears seeks aborted tour payment singer britn...
1507	Business	mexican in us send $16bn home mexican labourer...
993	Technology	us hacker breaks into t-mobile a man is facing...
256	Politics	lib dems new election pr chief the lib dems h...
2046	Politics	tory expert denies defeatism the conservatives...
1137	Sports	d arcy injury adds to ireland woe gordon d arc...

① w2Vec + RNN(?)
LSTM

② LSTM + CRF → POS



Data Exploration

First, let's check the shape of the dataset that we have.

Chrome File Edit View History Bookmarks Profiles Tab Window Help

BERT_V1.ipynb - Colaboratory | FlipItNews_News_Article_Cate | Twitter_NER.ipynb - Colaborat | The Illustrated Transformer – J

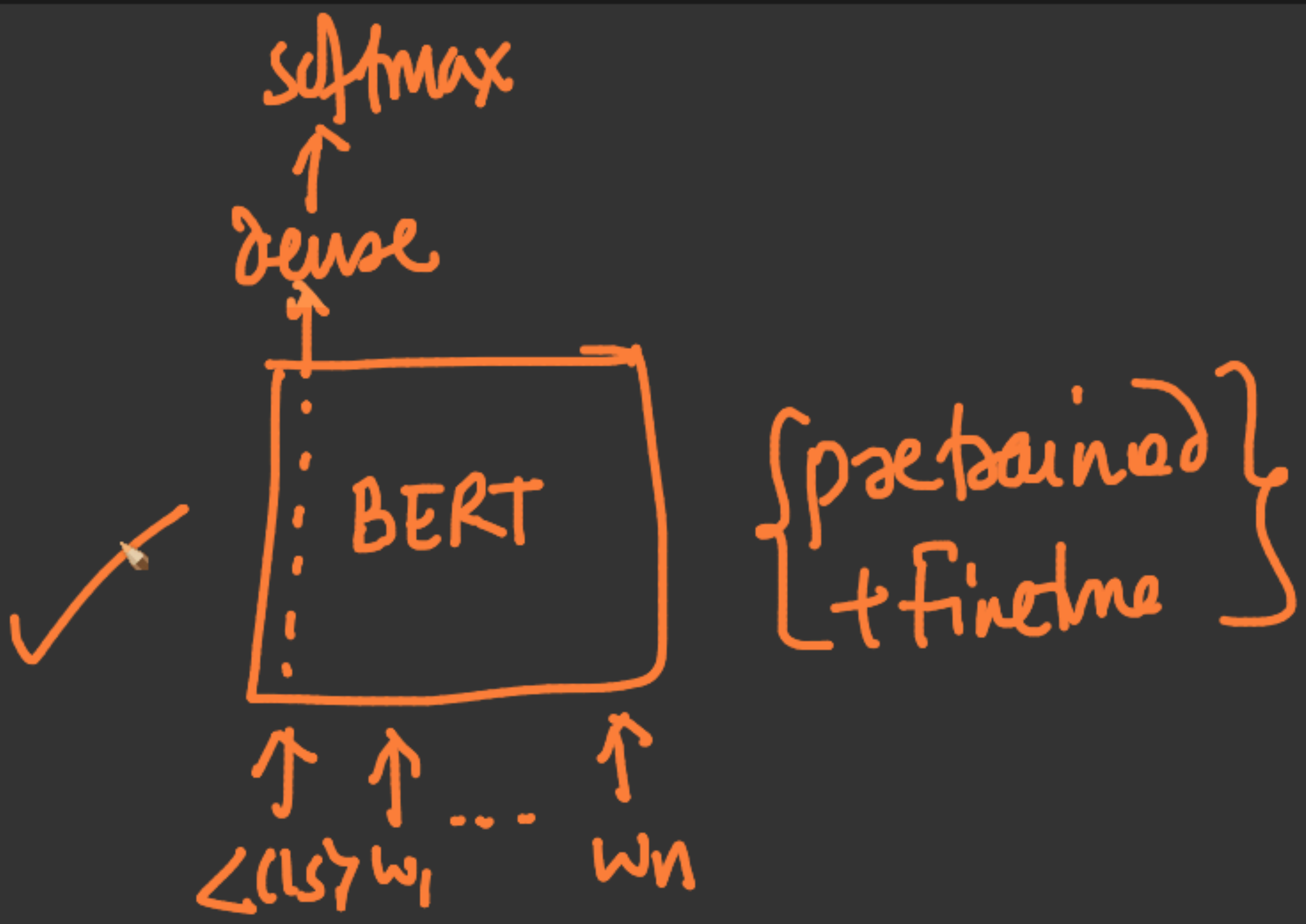
colab.research.google.com/drive/1zoZOQuJgapUXbs5V6jPSJ1KBu4A_mQJT

+ Code + Text Last edited on 10 November

Connect

[ ]

	Category	Article
221	Technology	world tour for top video gamers two uk gamers ...
852	Entertainment	prince crowned top music earner prince earne...
80	Business	us company admits benin bribery a us defence a...
73	Sports	funding cut hits wales students the wales stud...
614	Entertainment	spears seeks aborted tour payment singer britn...
1507	Business	mexican in us send $16bn home mexican labourer...
993	Technology	us hacker breaks into t-mobile a man is facing...
256	Politics	lib dems new election pr chief the lib dems h...
2046	Politics	tory expert denies defeatism the conservatives...
1137	Sports	d arcy injury adds to ireland woe gordon d arc...



Data Exploration

First, let's check the shape of the dataset that we have.

9 / 9

+ Code + Text Last edited on 10 November

	Category	Article
221	Technology	world tour for top video gamers two uk gamers ...
852	Entertainment	prince crowned top music earner prince earne...
80	Business	us company admits benin bribery a us defence a...
73	Sports	funding cut hits wales students the wales stud...
614	Entertainment	spears seeks aborted tour payment singer britn...
1507	Business	mexican in us send $16bn home mexican labourer...
993	Technology	us hacker breaks into t-mobile a man is facing...
256	Politics	lib dems new election pr chief the lib dems h...
2046	Politics	tory expert denies defeatism the conservatives...
1137	Sports	d arcy injury adds to ireland woe gordon d arc...

Simplest

(Let)

NB → 98% acc

BERT → 98.5% acc

110M

Data Exploration

First, let's check the shape of the dataset that we have.

NER on short-text

Problem statement

Context: Twitter is a microblogging and social networking service on which users post and interact with messages known as "tweets". Every second, on average, around 6,000 tweets are tweeted on Twitter, corresponding to over 350,000 tweets sent per minute, 500 million tweets per day. Twitter wants to automatically tag and analyze tweets for better understanding of the trends and topics without being dependent on the hashtags that the users use. Many users do not use hashtags or sometimes use wrong or mis-spelled tags, so they want to completely remove this problem and create a system of recognizing important content of the tweets.

Objective: You need to train a model that will be able to identify the various named entities.

① BiDir LSTM+CRF

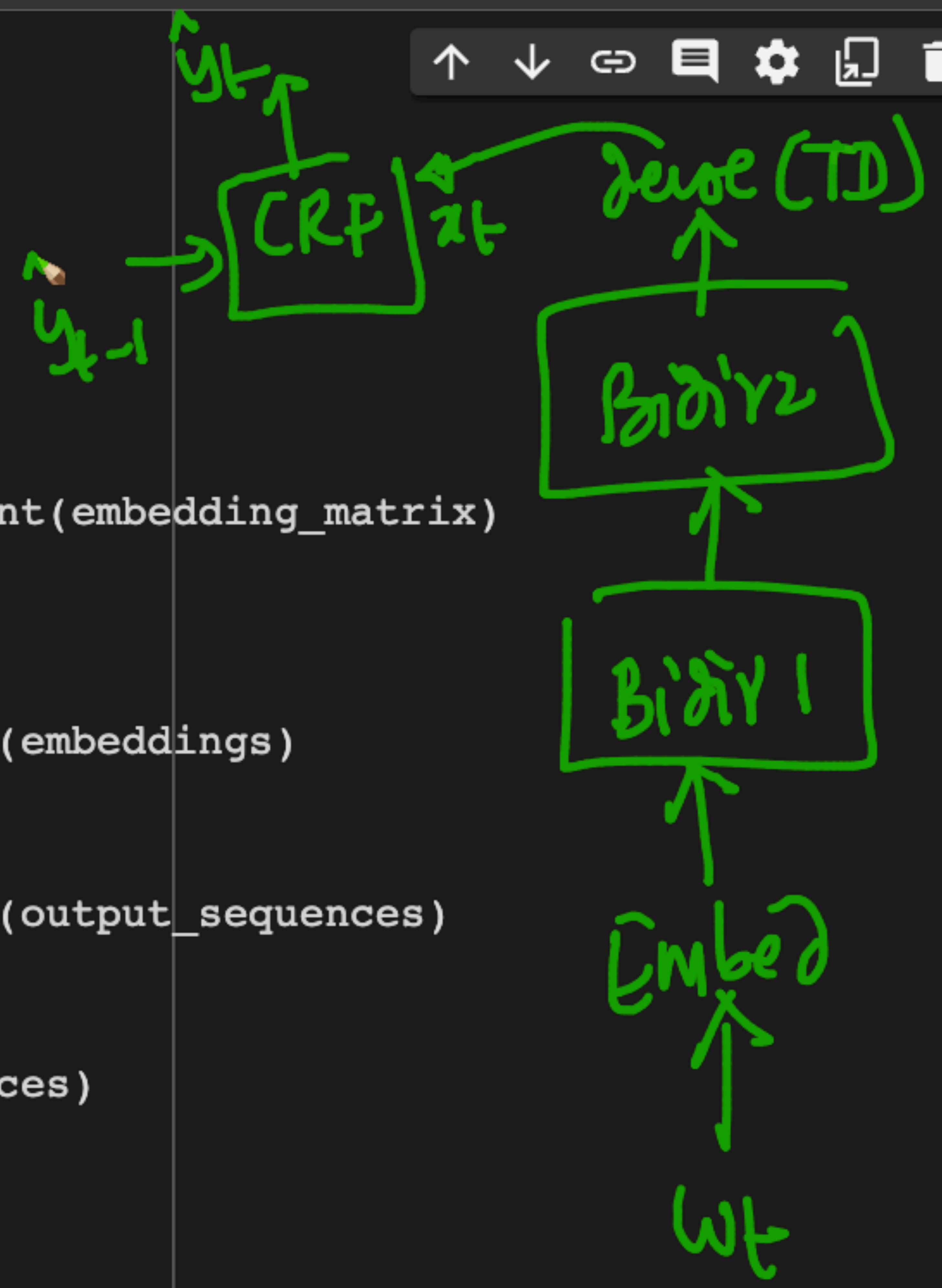
② BERT

Downloading data

```
!gdown 14_VHffl1qBUEnZ1IWFHnh6B9M5_A-Wf8
!gdown 1cnrGjppPOU_NtHNpGu0RJGg1CUNNsse_
```

```
Downloading...
From: https://drive.google.com/uc?id=14_VHffl1qBUEnZ1IWFHnh6B9M5_A-Wf8
To: /content/wnut 16.txt.conll
100% 403k/403k [00:00<00:00, 118MB/s]
Downloading...
From: https://drive.google.com/uc?id=1cnrGjppPOU_NtHNpGu0RJGg1CUNNsse_
To: /content/wnut 16test.txt.conll
```

```python
def build_model():
  # Model definition
  input = Input(shape=(max_len,))

  # Get embeddings
  embeddings = Embedding(input_dim=embedding_matrix.shape[0],
                         output_dim=embedding_dim,
                         input_length=max_len, mask_zero=True,
                         embeddings_initializer=tf.keras.initializers.Constant(embedding_matrix)
                        )(input)

  # variational biLSTM
  output_sequences = Bidirectional(LSTM(units=50, return_sequences=True))(embeddings)

  # Stacking
  output_sequences = Bidirectional(LSTM(units=50, return_sequences=True))(output_sequences)

  # Adding more non-linearity
  dense_out = TimeDistributed(Dense(25, activation="relu"))(output_sequences)

  # CRF layer
  crf = CRF(len(schema), name='crf')
  predicted_sequence, potentials, sequence_length, crf_kernel = crf(dense_out)

  model = Model(input, potentials)
  model.compile(
```



BPTT

Bi'dir LSTM + CRF (cheaper)

Max: 39 words

BERT

Run in Microsoft Learn | Run in Google Colab | Download Notebook | View on GitHub

Learn the Basics || **Quickstart** || Tensors || Datasets & DataLoaders || Transforms || Build Model || Autograd || Optimization || Save & Load Model

QUICKSTART

This section runs through the API for common tasks in machine learning. Refer to the links in each section to dive deeper.

Working with data

PyTorch has two primitives to work with data: `torch.utils.data.DataLoader` and `torch.utils.data.Dataset`. `Dataset` stores the samples and their corresponding labels, and `DataLoader` wraps an iterable around the `Dataset`.

```
import torch
from torch import nn
from torch.utils.data import DataLoader
from torchvision import datasets
from torchvision.transforms import ToTensor
```

PyTorch offers domain-specific libraries such as TorchText, TorchVision, and TorchAudio, all of which include datasets. For this tutorial, we will be using a TorchVision dataset.

The `torchvision.datasets` module contains `Dataset` objects for many real-world vision data like CIFAR, COCO (full list here). In this tutorial, we use the FashionMNIST dataset. Every TorchVision `Dataset` includes two arguments: `transform` and `target_transform` to modify the samples and labels respectively.

```
# Download training data from open datasets.
training_data = datasets.FashionMNIST(
    root="data",
    train=True,
```