

PMLDL HW-4

tags: `pmldl` , `hw-4`

[model weights](#)

Team

- Evgeny Sorokin
- Vitaly Volobuev
- Daniil Botnarenku

Model

Transformer model on tensorflow.

Hyperparameters:

- loss: sparse categorical cross entropy
- learning rate (scheduler from paper)
- number of layers: 4
- d_model: 256 (embedding)
- dff: 1024 (hidden size)
- number of heads: 8
- dropout: 0.1
- MAX_SEQ_LENGTH: 50
- BATCH_SIZE: 128
- BUFFER: 20000 (buffer on gpu)
- tokenizer-ru size: 32768
- tokenizer-en size: 16384

Resources:

- gtx 1080 (8gb VRAM)
- 3.6 MHz CPU

One epoch takes ~ 11 minutes.

Dataset

[Ted HRLR translate dataset](#)

`ru_to_en dataset` : ~140 mb

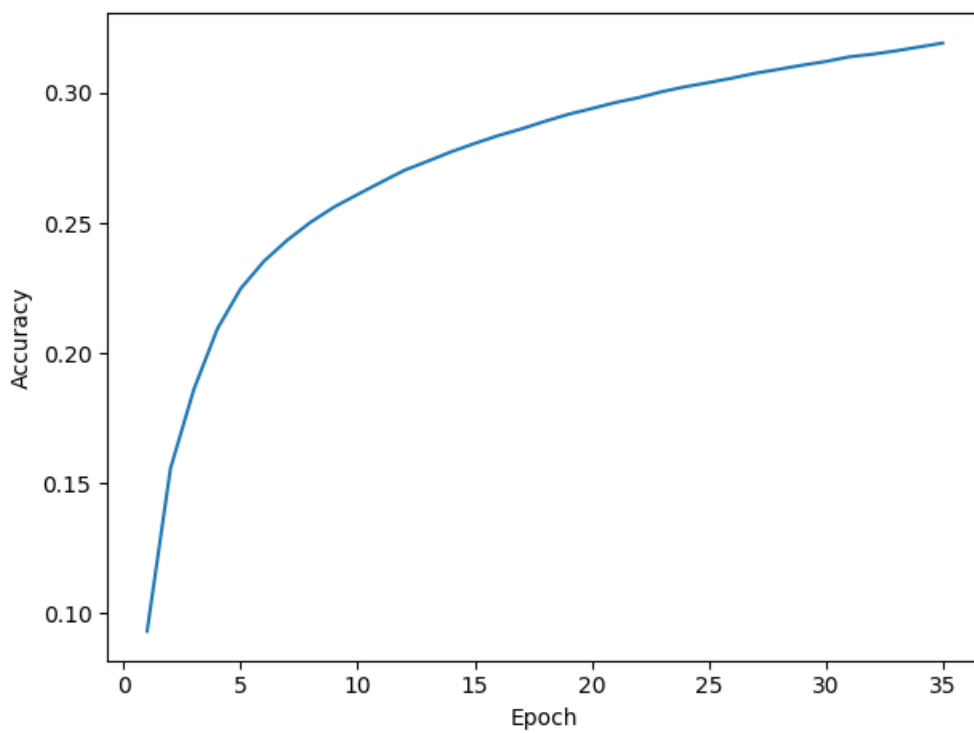
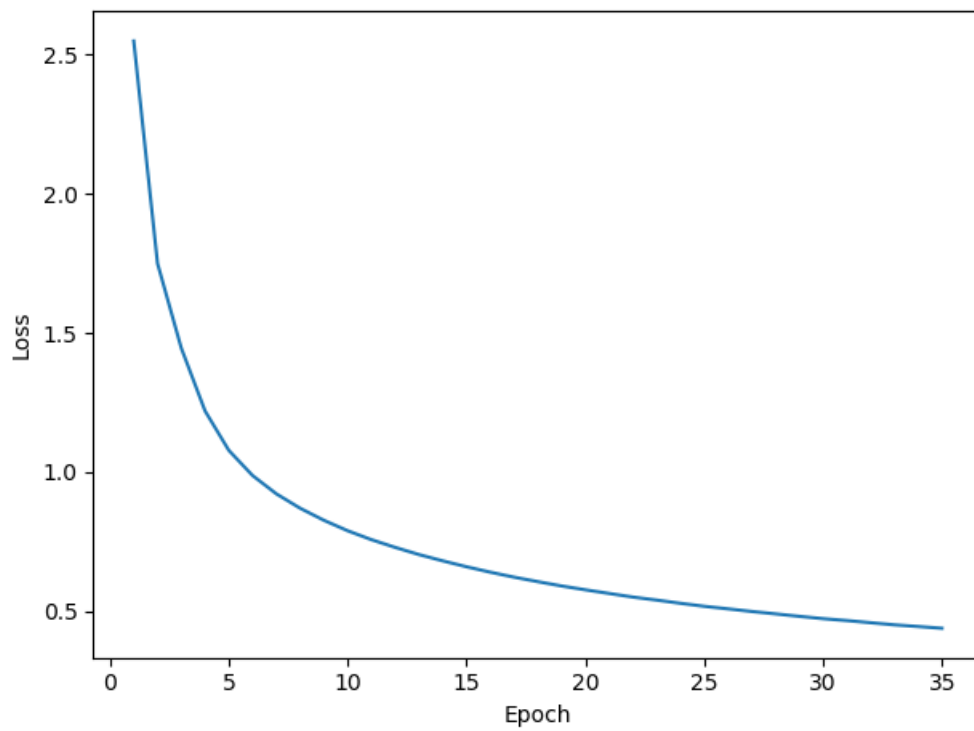
- **train**: 208106 examples
- **test**: 5476 examples
- **validation**: 4805 examples

Example:

Results

Train accuracy after 35 epochs: **31.9%**

Graphics of training:

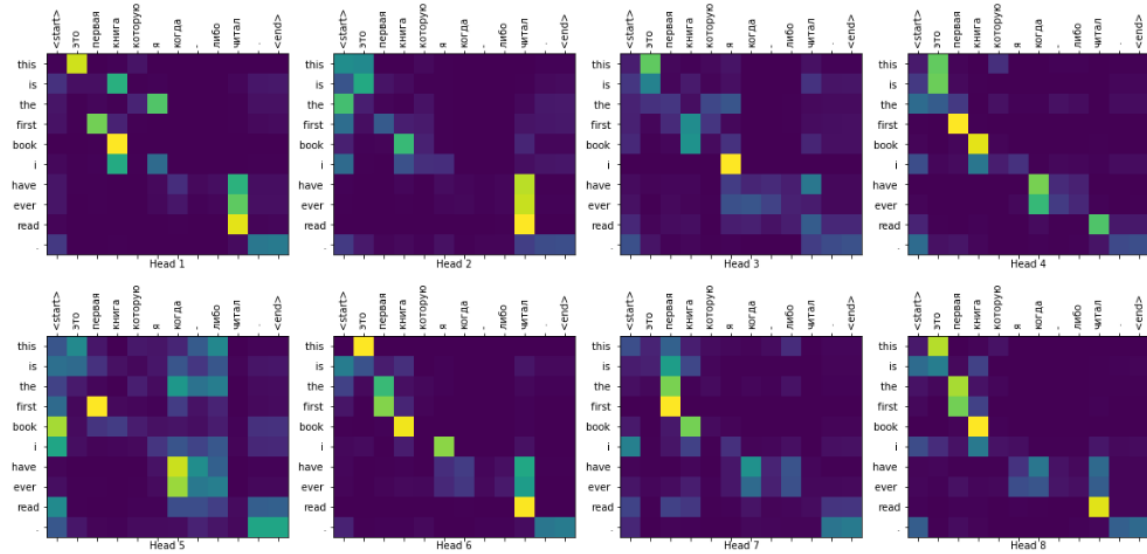


Predictions:

```
translate("это первая книга которую я когда-либо читал .", plot='decoder_layer4_block2')
print ("Real translation: this is the first book i've ever done.")
```

Input: это первая книга которую я когда-либо читал .

Predicted translation: this is the first book i have ever read .



Real translation: this is the first book i've ever done.

References

- [Attention is all you need](#)
- [Training Tips for the Transformer Model](#)
- [Transformer model for language understanding](#)