

Отчет по БДЗ 2

Большая часть кода была взята из

https://pytorch.org/tutorials/beginner/translation_transformer, для которой я дописал кастомный датасет для данных из задачи.

Эксперименты я запускал на 1/4 от всех данных на стандартных настройках из tutorials и на кастомных оптимайзерах, но получал переобучение к последним эпохам

766

Epoch: 1, Train loss: 5.837, Val loss: 5.044, Epoch time = 187.923s

766

Epoch: 2, Train loss: 4.723, Val loss: 4.454, Epoch time = 184.548s

766

Epoch: 3, Train loss: 4.217, Val loss: 4.088, Epoch time = 184.644s

766

Epoch: 4, Train loss: 3.856, Val loss: 3.829, Epoch time = 184.620s

766

Epoch: 5, Train loss: 3.557, Val loss: 3.616, Epoch time = 185.680s

766

Epoch: 6, Train loss: 3.302, Val loss: 3.476, Epoch time = 184.831s

766

Epoch: 7, Train loss: 3.080, Val loss: 3.401, Epoch time = 184.698s

766

Epoch: 8, Train loss: 2.891, Val loss: 3.342, Epoch time = 184.706s

766

Epoch: 9, Train loss: 2.719, Val loss: 3.321, Epoch time = 185.455s

766

Epoch: 10, Train loss: 2.561, Val loss: 3.345, Epoch time = 184.706s

766

Epoch: 11, Train loss: 2.419, Val loss: 3.360, Epoch time = 184.590s

766

Epoch: 12, Train loss: 2.292, Val loss: 3.356, Epoch time = 185.818s

766

Epoch: 13, Train loss: 2.169, Val loss: 3.326, Epoch time = 186.716s

766

Epoch: 14, Train loss: 2.060, Val loss: 3.310, Epoch time = 185.503s

766

Epoch: 15, Train loss: 1.957, Val loss: 3.315, Epoch time = 185.345s

Также я сделал запуск с кастомным оптимайзером, который был указан в статье по трансформеру(<https://arxiv.org/pdf/1706.03762.pdf>), но он очень медленно обучался.

```
Epoch: 1, Train loss: 6.266, Val loss: 5.718, Epoch time = 110.369s
766
Epoch: 2, Train loss: 5.607, Val loss: 5.475, Epoch time = 112.615s
766
Epoch: 3, Train loss: 5.382, Val loss: 5.326, Epoch time = 113.456s
766
Epoch: 4, Train loss: 5.238, Val loss: 5.213, Epoch time = 107.640s
766
Epoch: 5, Train loss: 5.133, Val loss: 5.131, Epoch time = 103.083s
766
Epoch: 6, Train loss: 5.050, Val loss: 5.063, Epoch time = 108.068s
766
Epoch: 7, Train loss: 4.980, Val loss: 5.006, Epoch time = 109.158s
766
Epoch: 8, Train loss: 4.919, Val loss: 4.954, Epoch time = 110.008s
766
Epoch: 9, Train loss: 4.866, Val loss: 4.906, Epoch time = 110.192s
766
Epoch: 10, Train loss: 4.816, Val loss: 4.859, Epoch time = 110.470s
766
Epoch: 11, Train loss: 4.773, Val loss: 4.822, Epoch time = 112.364s
766
Epoch: 12, Train loss: 4.733, Val loss: 4.782, Epoch time = 110.725s
766
```

Для получения лучшего результата, обучался на всех данных и словил переобучение на 11 эпохе

```
3062
Epoch: 1, Train loss: 4.898, Val loss: 3.916, Epoch time = 1176.200s
3062
Epoch: 2, Train loss: 3.665, Val loss: 3.206, Epoch time = 1175.403s
3062
Epoch: 3, Train loss: 3.141, Val loss: 2.897, Epoch time = 1175.666s
3062
Epoch: 4, Train loss: 2.821, Val loss: 2.743, Epoch time = 1176.370s
3062
Epoch: 5, Train loss: 2.598, Val loss: 2.650, Epoch time = 1177.502s
3062
Epoch: 6, Train loss: 2.430, Val loss: 2.599, Epoch time = 1177.879s
3062
Epoch: 7, Train loss: 2.301, Val loss: 2.572, Epoch time = 1177.552s
3062
Epoch: 8, Train loss: 2.196, Val loss: 2.555, Epoch time = 1178.872s
3062
Epoch: 9, Train loss: 2.106, Val loss: 2.545, Epoch time = 1178.822s
3062
Epoch: 10, Train loss: 2.026, Val loss: 2.566, Epoch time = 1178.770s
3062
Epoch: 11, Train loss: 1.954, Val loss: 2.552, Epoch time = 1178.132s
```