Transformer Model

https://github.com/jenegger/transformer_model

Default Settings:

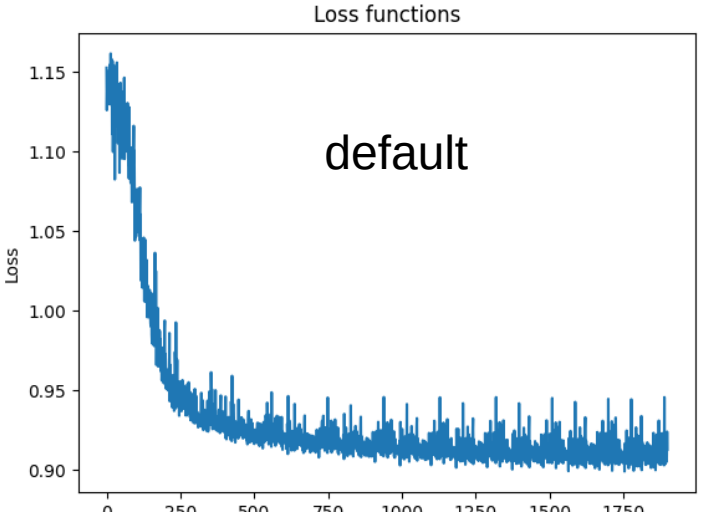
Batch size = 64
Epochs = 10
feature_nr = 32
loss_rate = 2e-4
loss_function = nn.BCEWithLogitsLoss()
Optimizer = optim.SGD

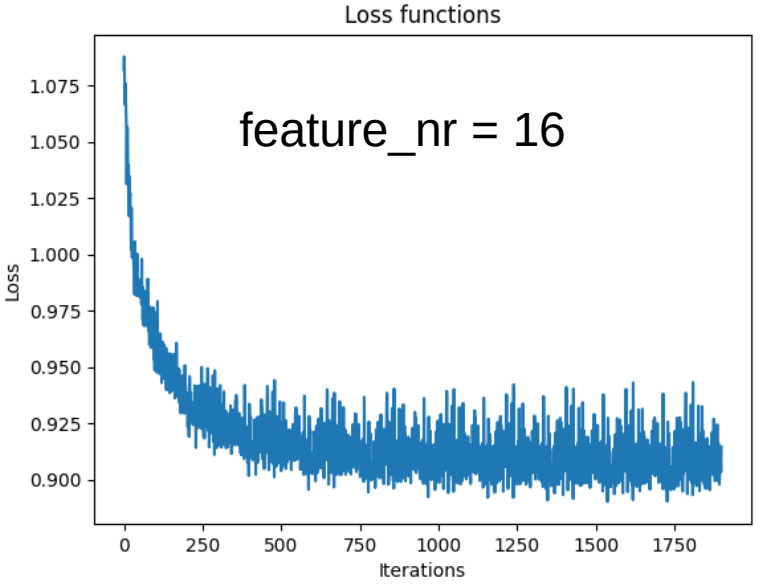
Forward step:

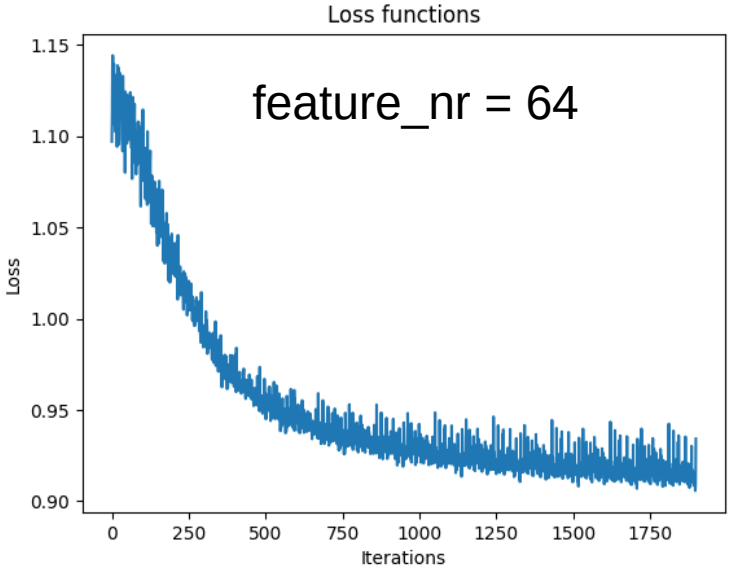
```
def forward(self, x,in_hitnr):
        x = self.linear_embedding(x)
        out = self.transformer_encoder(x)
        ...
```

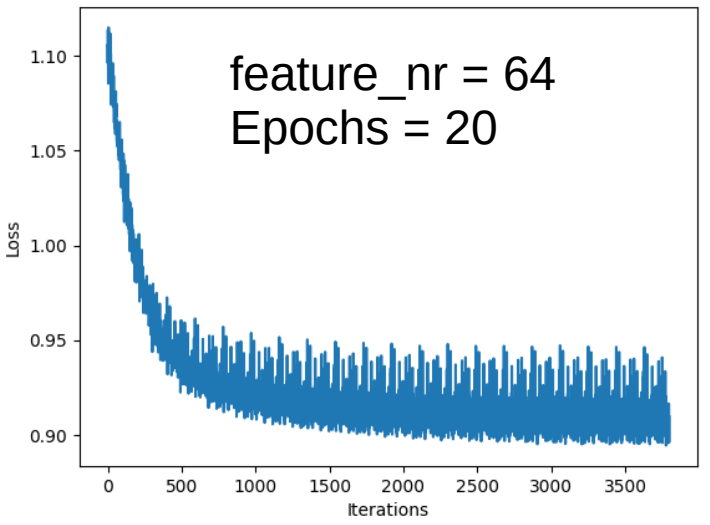
= torch.nn.Linear(4,feature_nr)

With:
heads = 6
num_layers = 6 (I.e sublayers)

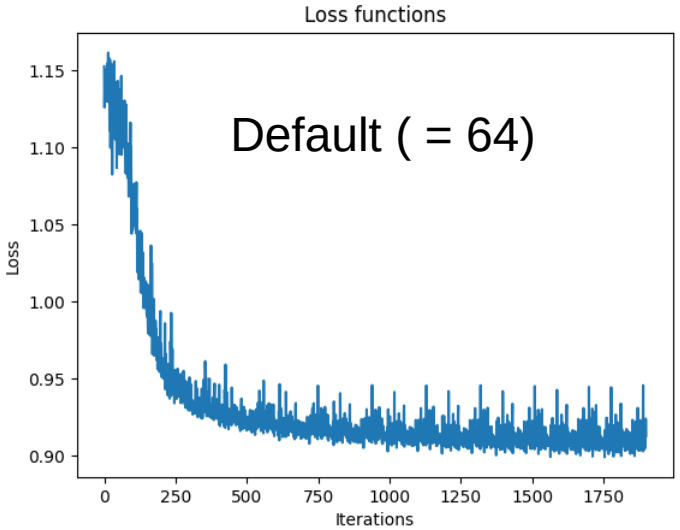

Loss functions
default
Loss
1.15
1.10
1.05
1.00
0.95
0.90

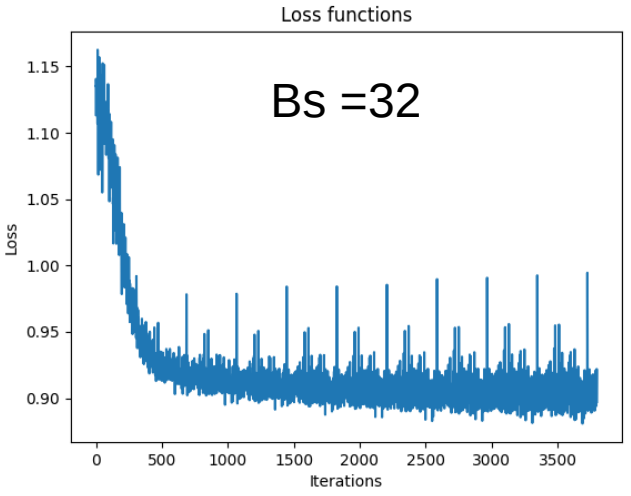

Loss functions
feature_nr = 16
Loss
1.075
1.050
1.025
1.000
0.975
0.950
0.925
0.900
0
250
500
750
1000
1250
1500
1750
Iterations

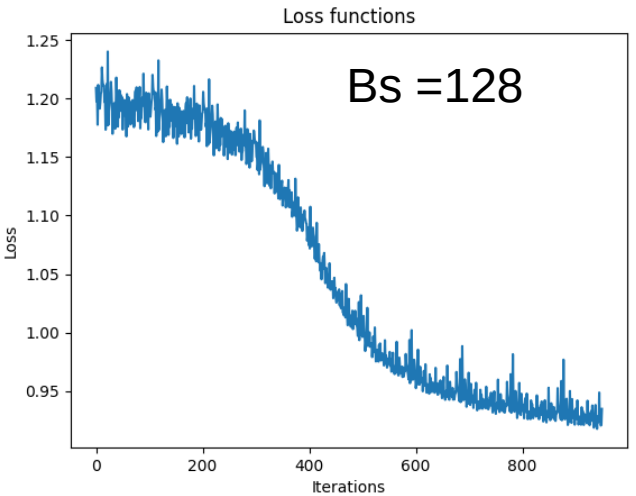

Loss functions
feature_nr = 64
Loss
1.15
1.10
1.05
1.00
0.95
0.90
0
250
500
750
1000
1250
1500
1750
Iterations


feature_nr = 64
Epochs = 20
Loss
1.10
1.05
1.00
0.95
0.90
0
500
1000
1500
2000
2500
3000
3500
Iterations

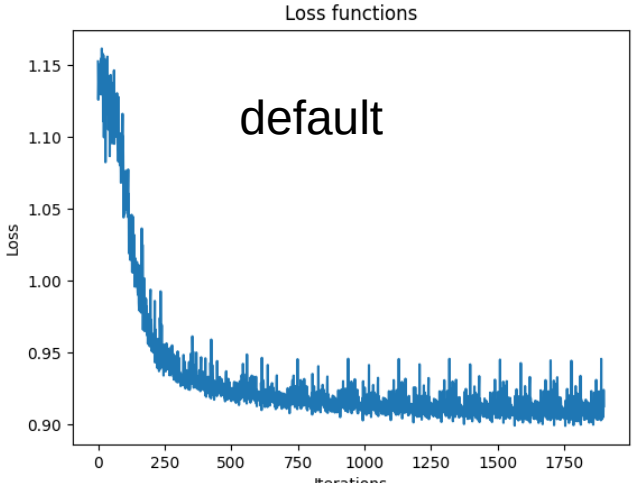
Different batchsizes

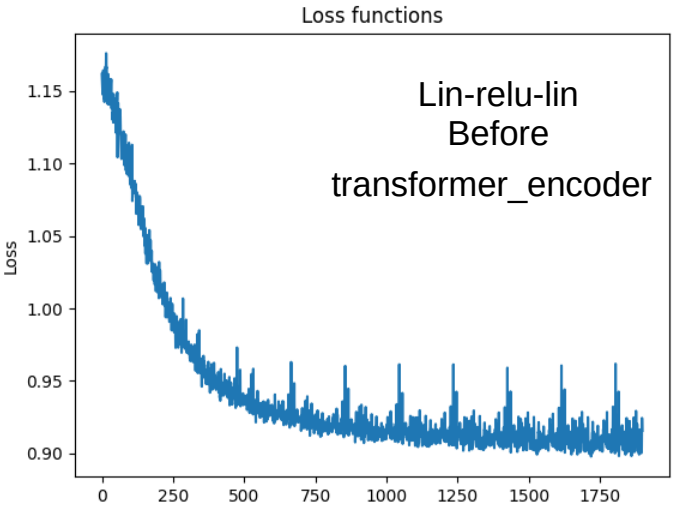


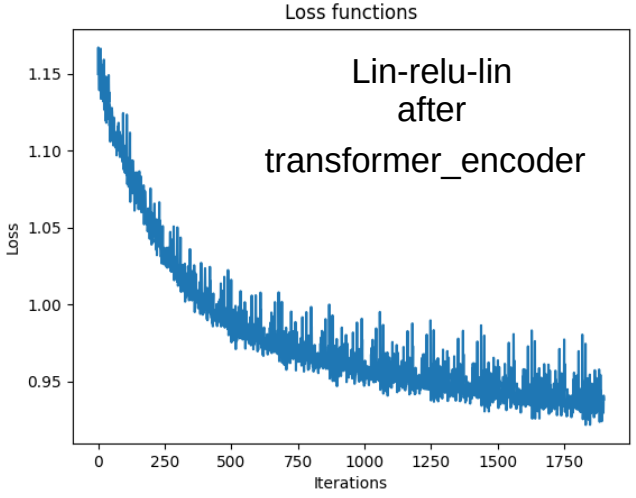


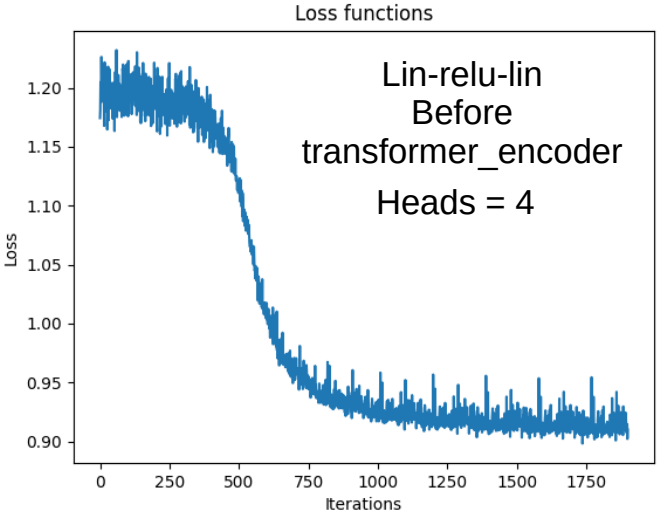


Additional Layer lin → RELU → lin









Different Loss Functions