

# Transformer Model

[https://github.com/jenegger/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

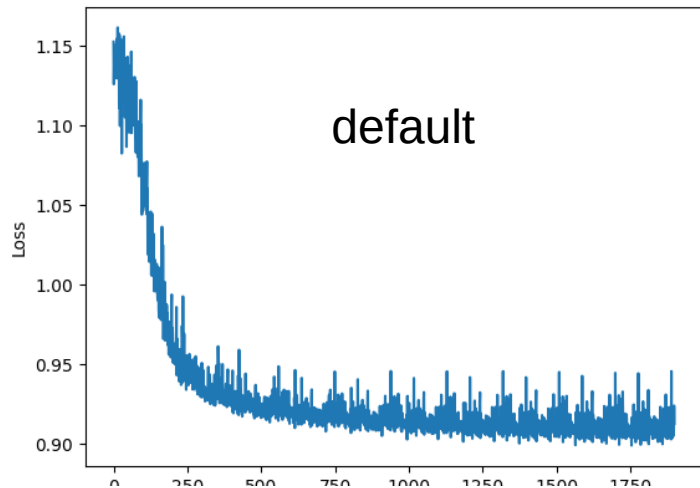
### Forward step:

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

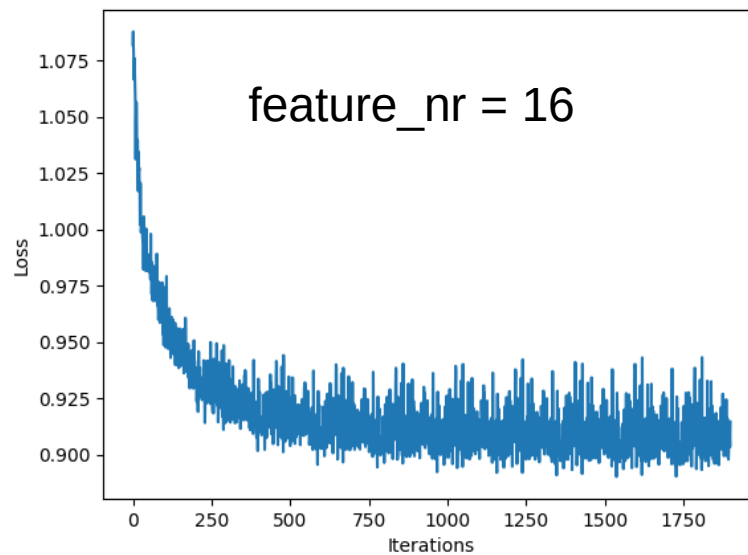
= torch.nn.Linear(4, feature\_nr)

With:  
heads = 6  
num\_layers = 6 (l.e sublayers)

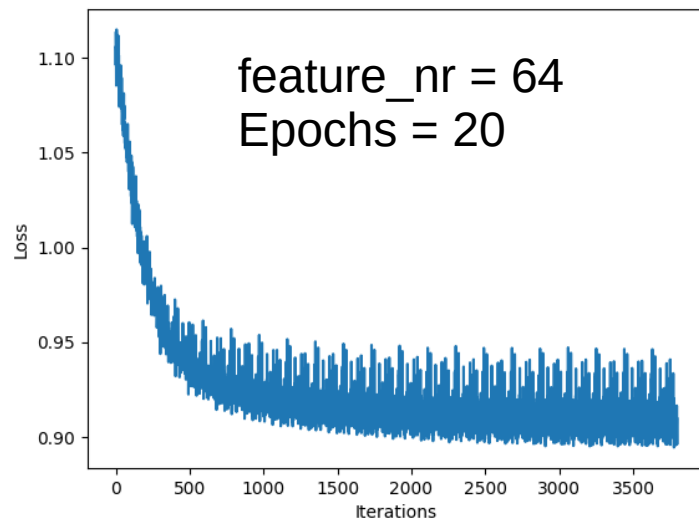
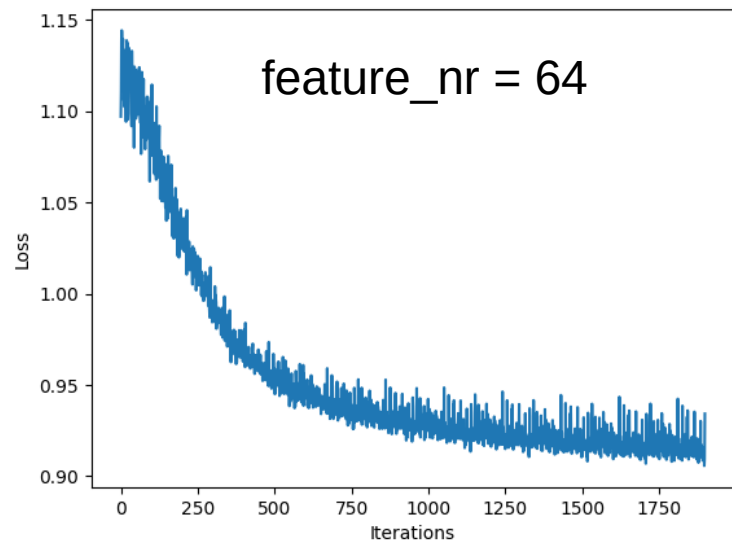
Loss functions



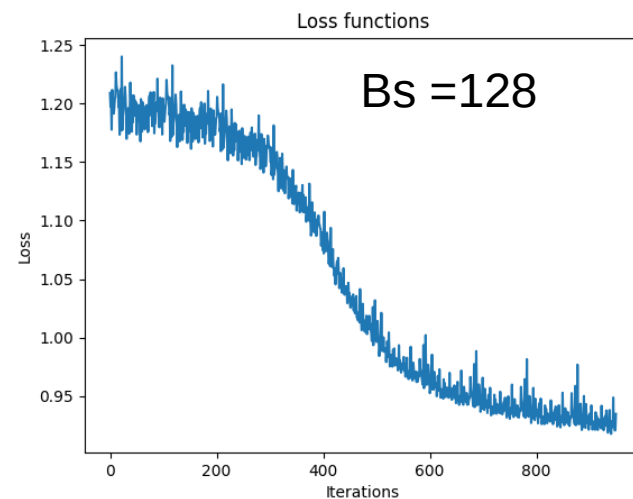
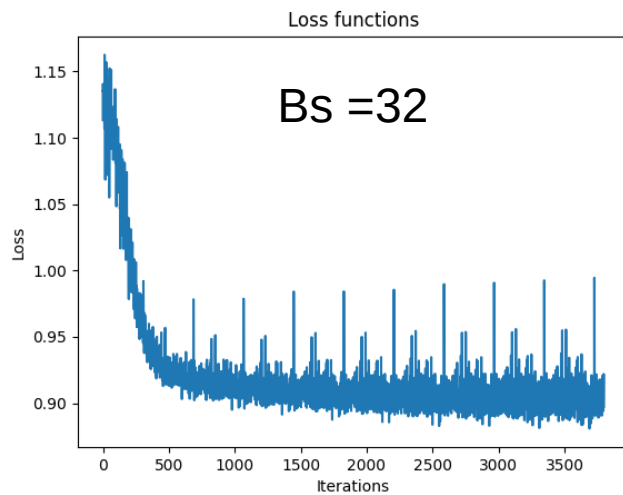
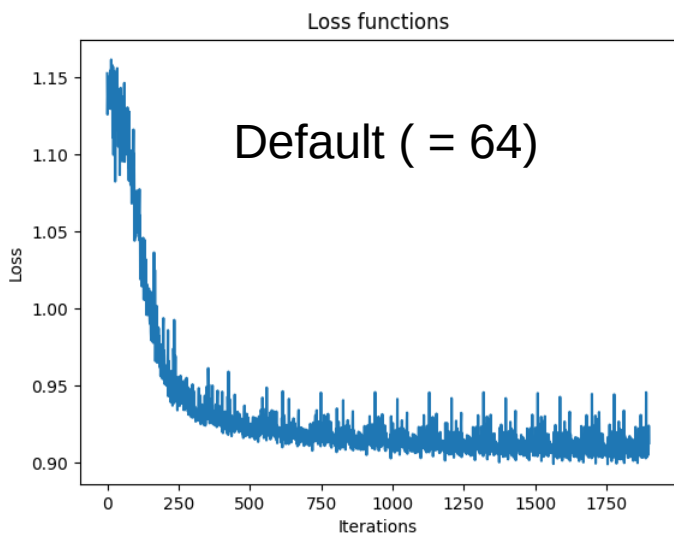
Loss functions



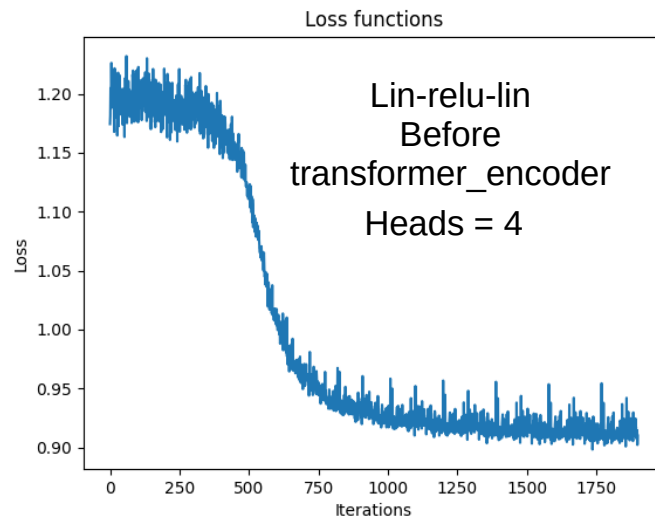
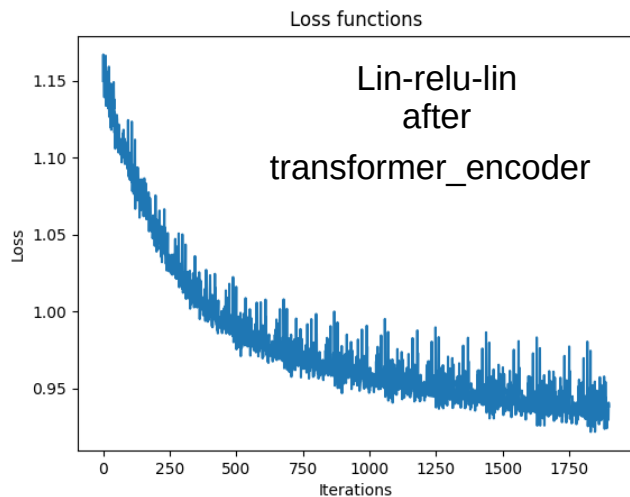
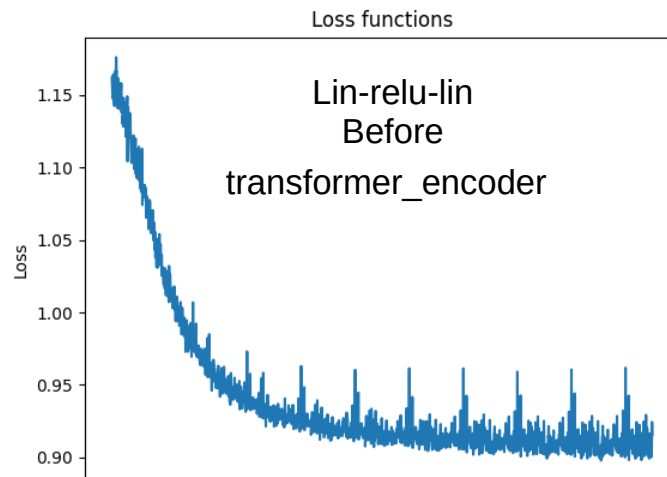
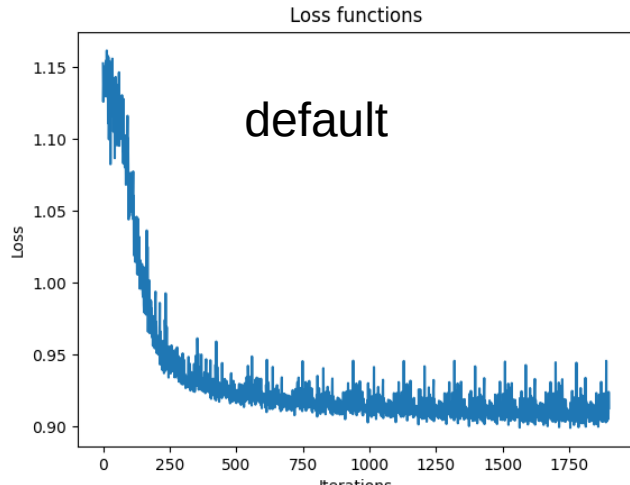
Loss functions



# Different batchsizes



## Additional Layer lin $\rightarrow$ RELU $\rightarrow$ lin

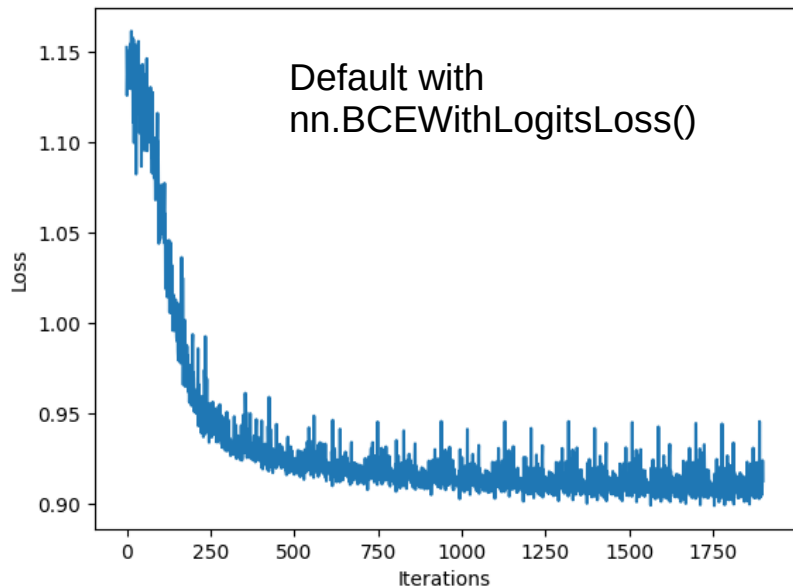


lin  $\rightarrow$  RELU  $\rightarrow$  Lin looks  
in forward function like  
this:

```
....  
x = self.linear_embedding(x)  
x = self.activation(x)  
x = self.additional_linear_layer(x)  
....
```

# Different Loss Functions

Loss functions



Loss functions

