Model Persistence - Saving

Auxiliary materials

In [1]:

```
import numpy as np
import torch
import lightning.pytorch as pl
import torchmetrics
import torchvision
from torchinfo import summary
from torchview import draw_graph
import pandas as pd
import matplotlib.pyplot as plt
```

Iris Net

In [2]:

```
class MultiClassModule(pl.LightningModule):
    def __init__(self,
                 output_size,
                 **kwargs):
        super().__init__(**kwargs)
        self.mc_acc = torchmetrics.classification.Accuracy(task='multiclass',
                                                           num_classes=output_si
        self.cce_loss = torch.nn.CrossEntropyLoss()

    def predict(self, x):
        return torch.softmax(self(x),-1)

    def configure_optimizers(self):
        optimizer = torch.optim.Adam(self.parameters(), lr=0.001)
        return optimizer

    def training_step(self, train_batch, batch_idx):
        x, y_true = train_batch
        y_pred = self(x)
        perm = (0,-1) + tuple(range(y_pred.ndim))[1:-1]
        acc = self.mc_acc(y_pred.permute(*perm),y_true)
        loss = self.cce_loss(y_pred.permute(*perm),y_true)
        self.log('train_acc', acc, on_step=False, on_epoch=True)
        self.log('train_loss', loss, on_step=False, on_epoch=True)
        return loss

    def validation_step(self, val_batch, batch_idx):
        x, y_true = val_batch
        y_pred = self(x)
        perm = (0,-1) + tuple(range(y_pred.ndim))[1:-1]
        acc = self.mc_acc(y_pred.permute(*perm),y_true)
        loss = self.cce_loss(y_pred.permute(*perm),y_true)
        self.log('val_acc', acc, on_step=False, on_epoch=True)
        self.log('val_loss', loss, on_step=False, on_epoch=True)
        return loss

    def test_step(self, test_batch, batch_idx):
        x, y_true = test_batch
        y_pred = self(x)
        perm = (0,-1) + tuple(range(y_pred.ndim))[1:-1]
        acc = self.mc_acc(y_pred.permute(*perm),y_true)
        loss = self.cce_loss(y_pred.permute(*perm),y_true)
        self.log('test_acc', acc, on_step=False, on_epoch=True)
        self.log('test_loss', loss, on_step=False, on_epoch=True)
        return loss
```

In [3]:
```
data = np.loadtxt("https://www.cs.mtsu.edu/~jphillips/courses/CSCI4850-5850/publ
np.random.seed(0)
np.random.shuffle(data)
X = data[:,:-1]
Y = data[:,-1]
split_point = int(X.shape[0] * 0.8)
# The dataloaders handle shuffling, batching, etc...
xy_train = torch.utils.data.DataLoader(list(zip(torch.tensor(X[:split_point]).fl
                                                torch.tensor(Y[:split_point]).lo
                                       shuffle=True, batch_size=32,num_workers=4
xy_val = torch.utils.data.DataLoader(list(zip(torch.tensor(X[split_point:]).floa
                                              torch.tensor(Y[split_point:]).long
                                     shuffle=False, batch_size=32,num_workers=
```

```
/opt/conda/lib/python3.11/site-packages/torch/utils/data/dataloader.py:557: UserW
arning: This DataLoader will create 4 worker processes in total. Our suggested ma
x number of worker in current system is 2, which is smaller than what this DataLo
ader is going to create. Please be aware that excessive worker creation might get
DataLoader running slow or even freeze, lower the worker number to avoid potentia
l slowness/freeze if necessary.
  warnings.warn(_create_warning_msg(
```

In [4]:
```
class MultiLayerNetwork(MultiClassModule):
    def __init__(self,
                 input_size,
                 output_size,
                 hidden_size = 64,
                 **kwargs):
        super().__init__(output_size=output_size,
                         **kwargs)
        self.save_hyperparameters()
        self.hidden = torch.nn.Linear(input_size,
                                      hidden_size)
        self.output = torch.nn.Linear(hidden_size,
                                      output_size)

    def forward(self, x):
        y = x
        y = torch.tanh(self.hidden(y))
        y = self.output(y)
        return y
```

In [5]:
```
iris_net = MultiLayerNetwork(X.shape[1],
                             len(np.unique(Y)))
```

In [6]:
```
x = torch.Tensor(X[0:1]).to(iris_net.device)
x
```

Out[6]:
```
tensor([[6.4000, 2.9000, 4.3000, 1.3000]])
```

In [7]:
```
iris_net.predict(x).cpu().detach().numpy()
```

```
Out[7]:  array([[0.22635038, 0.43114087, 0.34250873]], dtype=float32)
```

```
In [8]:  Y[1]
```

```
Out[8]:  0.0
```

```
In [9]:  logger = pl.loggers.CSVLogger("logs",
                                      name="persistence",
                                      version="iris_net-0")
```

```
In [10]: trainer = pl.Trainer(logger=logger,
                              max_epochs=100,
                              enable_progress_bar=True,
                              log_every_n_steps=0,
                              enable_checkpointing=True, # Notice this!
                              callbacks=[pl.callbacks.TQDMProgressBar(refresh_rate=50)])
```

```
GPU available: True (cuda), used: True
TPU available: False, using: 0 TPU cores
IPU available: False, using: 0 IPUs
HPU available: False, using: 0 HPUs
```

```
In [12]: trainer.validate(iris_net, xy_val)
```

```
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]
SLURM auto-requeueing enabled. Setting signal handlers.
Validation: |          | 0/? [00:00<?, ?it/s]
```

```
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━┓
┃      Validate metric      ┃       DataLoader 0        ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━┩
│          val_acc          │    0.2666666805744171     │
│         val_loss          │    1.0410279035568237     │
└───────────────────────────┴───────────────────────────┘
```

```
Out[12]: [{'val_acc': 0.2666666805744171, 'val_loss': 1.0410279035568237}]
```

```
In [13]: trainer.fit(iris_net, xy_train, xy_val)
```

```
/opt/conda/lib/python3.11/site-packages/lightning/pytorch/callbacks/model_checkpo
int.py:639: Checkpoint directory logs/persistence/iris_net-0/checkpoints exists a
nd is not empty.
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]

  | Name     | Type               | Params
------------------------------------------------
0 | mc_acc   | MulticlassAccuracy | 0
1 | cce_loss | CrossEntropyLoss   | 0
2 | hidden   | Linear             | 320
3 | output   | Linear             | 195
------------------------------------------------
515       Trainable params
0         Non-trainable params
515       Total params
0.002     Total estimated model params size (MB)
SLURM auto-requeueing enabled. Setting signal handlers.
Sanity Checking: |          | 0/? [00:00<?, ?it/s]
Training: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
```

```
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
```

```
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
`Trainer.fit` stopped: `max_epochs=100` reached.
```

Keep track of the following result for comparison later...

In [14]:

```
trainer.validate(iris_net, xy_val)
```

```
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]
SLURM auto-requeueing enabled. Setting signal handlers.
Validation: |          | 0/? [00:00<?, ?it/s]
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━┓
┃      Validate metric      ┃       DataLoader 0        ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━┩
│          val_acc          │            1.0            │
│         val_loss          │    0.12450060248374939    │
└───────────────────────────┴───────────────────────────┘
```

Out[14]:

```
[{'val_acc': 1.0, 'val_loss': 0.12450060248374939}]
```

In [18]:

```
x = torch.Tensor(X[0:1]).to(iris_net.device)
x
```

Out[18]:
```
tensor([[6.4000, 2.9000, 4.3000, 1.3000]])
```

In [19]:
```
iris_net.predict(x).cpu().detach().numpy()
```

Out[19]:
```
array([[0.02313578, 0.9282432 , 0.048621  ]], dtype=float32)
```

In [20]:
```
Y[0:1]
```

Out[20]:
```
array([1.])
```

Here's the magic word:

In [21]:
```
trainer.save_checkpoint("iris_net.ckpt")
```

Switch to Loading example notebook...

Encoder-Decoder

In [22]:
```
# Tokenization Tools
import os
os.environ["TOKENIZERS_PARALLELISM"] = "false"
import tokenizers
import io
```

In [23]:
```
import urllib
data = []
my_url = "https://raw.githubusercontent.com/luisroque/deep-learning-articles/mai
with urllib.request.urlopen(my_url) as raw_data:
    for line in raw_data:
        data.append(line.decode("utf-8").split('\t')[0:2])
data = np.array(data)
```

In [24]:
```
data.shape
```

Out[24]:
```
(170304, 2)
```

In [25]:
```
# Subset? - All of the data will take some time...
n_seq = data.shape[0]
n_seq = 10000
data = data[0:n_seq]
split_point = int(data.shape[0] * 0.8) # Keep 80/20 split
np.random.seed(0)
np.random.shuffle(data) # In-place modification
```

In [26]:
```
data[0]
```

Out[26]:
```
array(['These are real.', 'Estas são autênticas.'], dtype='<U184')
```

In [27]:
```
eng = data[:,0]
por = data[:,1]
```

In [28]:
```
eng.shape
```

Out[28]:

```
(10000,)
```

In [29]:

```
eng[0:5]
```

Out[29]:

```
array(['These are real.', "I'm sorry.", 'I have wine.', "I won't do it.",
       'I eat bread.'], dtype='<U184')
```

In [30]:

```
por.shape
```

Out[30]:

```
(10000,)
```

In [31]:

```
por[0:5]
```

Out[31]:

```
array(['Estas são autênticas.', 'Desculpe!', 'Tenho vinho.',
       'Eu não irei fazer isso.', 'Eu como pão.'], dtype='<U184')
```

In [32]:

```
unknown_token = "<UNK>"  # token for unknown words
special_tokens = [unknown_token, "<START>","<STOP>"]  # special tokens

eng_tokenizer = tokenizers.Tokenizer(tokenizers.models.BPE(unk_token=unknown_tok
eng_token_trainer = tokenizers.trainers.BpeTrainer(vocab_size=100000,special_tok
eng_tokenizer.pre_tokenizer = tokenizers.pre_tokenizers.Whitespace()

por_tokenizer = tokenizers.Tokenizer(tokenizers.models.BPE(unk_token=unknown_tok
por_token_trainer = tokenizers.trainers.BpeTrainer(vocab_size=100000,special_tok
por_tokenizer.pre_tokenizer = tokenizers.pre_tokenizers.Whitespace()
```

In [33]:

```
with open("eng_strings.txt","w") as f:
    for s in eng:
        f.write(s)
        f.write("\n")
with open("por_strings.txt","w") as f:
    for s in por:
        f.write(s)
        f.write("\n")
```

In [34]:

```
eng_tokenizer.train(["eng_strings.txt"],eng_token_trainer)
por_tokenizer.train(["por_strings.txt"],por_token_trainer)
```

Persist the tokenizer...

In [35]:

```
eng_tokenizer.save("eng_trained.json")
por_tokenizer.save("por_trained.json")
```

In [36]:

```
eng_tokenizer.encode("Here is a test.").tokens
```

Out[36]:

```
['Here', 'is', 'a', 'test', '.']
```

In [37]:

```
temp = eng_tokenizer.encode("Here is a test.").ids
temp
```

```
Out[37]: [443, 78, 46, 3165, 9]
```

```
In [38]: temp = eng_tokenizer.decode(temp + [0,0])
         temp
```

```
Out[38]: 'Here is a test .'
```

```
In [39]: eng_tokenizer.get_vocab_size()
```

```
Out[39]: 3694
```

```
In [40]: por_tokenizer.get_vocab_size()
```

```
Out[40]: 6459
```

```
In [41]: eng_recoded = np.array([eng_tokenizer.decode(eng_tokenizer.encode(s).ids) for s
         por_recoded = np.array([por_tokenizer.decode(por_tokenizer.encode(s).ids) for s
```

```
In [42]: eng_recoded[0]
```

```
Out[42]: 'These are real .'
```

```
In [43]: eng[0]
```

```
Out[43]: 'These are real.'
```

```
In [44]: por_recoded[0]
```

```
Out[44]: 'Estas são autênticas .'
```

```
In [45]: por[0]
```

```
Out[45]: 'Estas são autênticas.'
```

```
In [46]: def encode_seq(x,tokenizer,max_length=0):
             # String to integer
             x = tokenizer.encode("<START>"+x+"<STOP>").ids
             x += [0]*(max_length-len(x))
             return x

         def decode_seq(x,tokenizer):
             return tokenizer.decode(x)
```

```
In [47]: temp = encode_seq(eng_recoded[0],eng_tokenizer,20)
         temp
```

```
Out[47]: [1, 425, 140, 442, 9, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

```
In [48]: len(temp)
```

```
Out[48]: 20
```

```
In [49]: decode_seq(temp,eng_tokenizer)
```

```
Out[49]: 'These are real .'
```

In [50]:
```
temp = encode_seq(por_recoded[0],por_tokenizer,20)
temp
```

Out[50]:
```
[1, 862, 229, 6063, 8, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

In [51]:
```
decode_seq(temp,por_tokenizer)
```

Out[51]:
```
'Estas são autênticas .'
```

In [52]:
```
max_eng = np.max([len(encode_seq(i,eng_tokenizer)) for i in eng_recoded])
max_eng
```

Out[52]:
```
12
```

In [53]:
```
max_por = np.max([len(encode_seq(i,por_tokenizer)) for i in por_recoded])
max_por
```

Out[53]:
```
11
```

In [54]:
```
X = np.vstack([encode_seq(x,eng_tokenizer,max_eng) for x in eng_recoded])
Y = np.vstack([encode_seq(x,por_tokenizer,max_por) for x in por_recoded])
```

In [55]:
```
enc_x_train = X[:split_point]
enc_x_val = X[split_point:]
enc_x_train
```

Out[55]:
```
array([[   1,  425,  140, ...,    0,    0,    0],
       [   1,   30,    6, ...,    0,    0,    0],
       [   1,   30,  152, ...,    0,    0,    0],
       ...,
       [   1,   90,    6, ...,    0,    0,    0],
       [   1,   30,    6, ...,    0,    0,    0],
       [   1, 1592,   21, ...,    0,    0,    0]])
```

In [56]:
```
dec_x_train = Y[:,0:-1][:split_point]
dec_x_val = Y[:,0:-1][split_point:]
dec_x_train
```

Out[56]:
```
array([[   1,  862,  229, ...,    0,    0,    0],
       [   1, 3279,    3, ...,    0,    0,    0],
       [   1,  352,  719, ...,    0,    0,    0],
       ...,
       [   1,  141,  416, ...,    0,    0,    0],
       [   1,  188,  850, ...,    0,    0,    0],
       [   1, 3776,   19, ...,    0,    0,    0]])
```

In [57]:
```
dec_y_train = Y[:,1:][:split_point]
dec_y_val = Y[:,1:][split_point:]
dec_y_train
```

Out[57]:
```
array([[ 862,  229, 6063, ...,    0,    0,    0],
       [3279,    3,    2, ...,    0,    0,    0],
       [ 352,  719,    8, ...,    0,    0,    0],
       ...,
       [ 141,  416, 3597, ...,    0,    0,    0],
       [ 188,  850,    8, ...,    0,    0,    0],
       [3776,   19,    2, ...,    0,    0,    0]])
```

In [58]:

```
print(enc_x_train.shape)
print(dec_x_train.shape)
print(dec_y_train.shape)
```

```
(8000, 12)
(8000, 10)
(8000, 10)
```

In [59]:

```
print(enc_x_val.shape)
print(dec_x_val.shape)
print(dec_y_val.shape)
```

```
(2000, 12)
(2000, 10)
(2000, 10)
```

In [60]:

```
batch_size = 256
xy_train = torch.utils.data.DataLoader(list(zip(torch.Tensor(enc_x_train).long()
                                                torch.Tensor(dec_x_train).long()
                                                torch.Tensor(dec_y_train).long()
                                       shuffle=True, batch_size=batch_size,
                                       num_workers=4)
xy_val = torch.utils.data.DataLoader(list(zip(torch.Tensor(enc_x_val).long(),
                                              torch.Tensor(dec_x_val).long(),
                                              torch.Tensor(dec_y_val).long())),
                                     shuffle=False, batch_size=batch_size,
                                     num_workers=4)
```

Encoder-Decoder Network

In [61]:

```
class RecurrentResidual(pl.LightningModule):
    def __init__(self,
                 latent_size = 256,
                 bidirectional = False,
                 **kwargs):
        super().__init__(**kwargs)
        self.layer_norm = torch.nn.LayerNorm(latent_size)
        self.rnn_layer = torch.nn.LSTM(latent_size,
                                       latent_size // 2 if bidirectional else la
                                       bidirectional=bidirectional,
                                       batch_first=True)
    def forward(self, x):
        return x + self.rnn_layer(self.layer_norm(x))[0]
```

In [62]:

```
class EncoderNetwork(pl.LightningModule):
    def __init__(self,
                 num_tokens,
                 latent_size = 256, # Use something divisible by 2
                 n_layers = 8,
                 **kwargs):
        super().__init__(**kwargs)
        self.embedding = torch.nn.Embedding(num_tokens,
                                            latent_size,
                                            padding_idx=0)
        self.dropout = torch.nn.Dropout1d(0.05) # Whole token dropped
        self.rnn_layers = torch.nn.Sequential(*[
            RecurrentResidual(latent_size,True) for _ in range(n_layers)
        ])

    def forward(self, x):
        y = x
        y = self.embedding(y)
        y = self.dropout(y)
        y = self.rnn_layers(y)[:,-1]
        return y
```

In [63]:

```
class DecoderNetwork(pl.LightningModule):
    def __init__(self,
                 num_tokens,
                 latent_size = 256, # Use something divisible by 2
                 n_layers = 8,
                 **kwargs):
        super().__init__(**kwargs)
        self.embedding = torch.nn.Embedding(num_tokens,
                                            latent_size,
                                            padding_idx=0)
        # self.dropout = torch.nn.Dropout1d(0.1) # Whole token dropped
        self.linear = torch.nn.Linear(latent_size*2,
                                      latent_size)
        self.rnn_layers = torch.nn.Sequential(*[
            RecurrentResidual(latent_size,False) for _ in range(n_layers)
        ])
        self.output_layer = torch.nn.Linear(latent_size,
                                            num_tokens)

    def forward(self, x_enc, x_dec):
        y_enc = x_enc.unsqueeze(1).repeat(1,x_dec.shape[1],1)
        y_dec = self.embedding(x_dec)
        # y_dec = self.dropout(y_dec)
        y = y_enc
        y = torch.concatenate([y_enc,y_dec],-1)
        y = self.linear(y)
        y = self.rnn_layers(y)
        y = self.output_layer(y)
        return y
```

In [64]:

```
class EncDecLightningModule(pl.LightningModule):
    def __init__(self,
                 output_size,
                 **kwargs):
        super().__init__(**kwargs)
        self.mc_acc = torchmetrics.classification.Accuracy(task='multiclass',
                                                           num_classes=output_si
                                                           ignore_index=0)
        self.cce_loss = torch.nn.CrossEntropyLoss(ignore_index=0)

    def predict(self, x):
        return torch.softmax(self(x),-1)

    def configure_optimizers(self):
        optimizer = torch.optim.Adam(self.parameters(), lr=0.001)
        return optimizer

    def training_step(self, train_batch, batch_idx):
        x_enc, x_dec, y_dec = train_batch
        y_pred = self(x_enc, x_dec)
        perm = (0,-1) + tuple(range(y_pred.ndim))[1:-1]
        acc = self.mc_acc(y_pred.permute(*perm),y_dec)
        loss = self.cce_loss(y_pred.permute(*perm),y_dec)
        self.log('train_acc', acc, on_step=False, on_epoch=True)
        self.log('train_loss', loss, on_step=False, on_epoch=True)
        return loss

    # Validate used for Teacher Forcing
    def validation_step(self, val_batch, batch_idx):
        x_enc, x_dec, y_dec = val_batch
        y_pred = self(x_enc, x_dec)
        perm = (0,-1) + tuple(range(y_pred.ndim))[1:-1]
        acc = self.mc_acc(y_pred.permute(*perm),y_dec)
        loss = self.cce_loss(y_pred.permute(*perm),y_dec)
        self.log('val_acc', acc, on_step=False, on_epoch=True)
        self.log('val_loss', loss, on_step=False, on_epoch=True)
        return loss

    # Test used for Non-Teacher Forcing
    def test_step(self, test_batch, batch_idx):
        x_enc, x_dec, y_dec = test_batch
        context = self.enc_net(x_enc)
        tokens = torch.zeros_like(x_dec).long()
        tokens[:,0] = 1
        for i in range(y_dec.shape[1]-1):
            tokens[:,i+1] = self.dec_net(context, tokens).argmax(-1)[:,i]
        y_pred = self(x_enc, tokens)
        perm = (0,-1) + tuple(range(y_pred.ndim))[1:-1]
        acc = self.mc_acc(y_pred.permute(*perm),y_dec)
        loss = self.cce_loss(y_pred.permute(*perm),y_dec)
        self.log('test_acc', acc, on_step=False, on_epoch=True)
        self.log('test_loss', loss, on_step=False, on_epoch=True)
        return loss

    def predict_step(self, predict_batch, batch_idx):
        x_enc, x_dec, y_dec = test_batch
        context = self.enc_net(x_enc)
        tokens = torch.zeros_like(x_dec).long()
        tokens[:,0] = 1
        for i in range(y_dec.shape[1]-1):
            tokens[:,i+1] = self.dec_net(context, tokens).argmax(-1)[:,i]
        y_pred = self(x_enc, tokens)
        return y_pred
```

In [65]:

```
class EncDecNetwork(EncDecLightningModule):
    def __init__(self,
                 num_enc_tokens,
                 num_dec_tokens,
                 latent_size = 256, # Use something divisible by 2
                 n_layers = 8,
                 **kwargs):
        super().__init__(output_size=num_dec_tokens,
                         **kwargs)
        self.save_hyperparameters()
        self.enc_net = EncoderNetwork(num_enc_tokens,latent_size,n_layers)
        self.dec_net = DecoderNetwork(num_dec_tokens,latent_size,n_layers)

    def forward(self, x_enc, x_dec):
        return self.dec_net(self.enc_net(x_enc), x_dec)
```

```
In [66]: enc_dec_net = EncDecNetwork(num_enc_tokens=eng_tokenizer.get_vocab_size(),
                                     num_dec_tokens=por_tokenizer.get_vocab_size())
```

Training Time

```
In [67]: logger = pl.loggers.CSVLogger("logs",
                                       name="persistence",
                                       version="encdec-0")
```

```
In [68]: trainer = pl.Trainer(logger=logger,
                              max_epochs=30,
                              enable_progress_bar=True,
                              log_every_n_steps=0,
                              enable_checkpointing=True, # Notice this here!
                              callbacks=[pl.callbacks.TQDMProgressBar(refresh_rate=50)])
```

```
GPU available: True (cuda), used: True
TPU available: False, using: 0 TPU cores
IPU available: False, using: 0 IPUs
HPU available: False, using: 0 HPUs
```

```
In [70]: trainer.validate(enc_dec_net, xy_val)
```

```
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]
SLURM auto-requeueing enabled. Setting signal handlers.
Validation: |          | 0/? [00:00<?, ?it/s]
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┓
┃       Validate metric       ┃        DataLoader 0         ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┩
│           val_acc           │    9.861325816018507e-05    │
│          val_loss           │      8.858915328979492      │
└─────────────────────────────┴─────────────────────────────┘
```

```
Out[70]: [{'val_acc': 9.861325816018507e-05, 'val_loss': 8.858915328979492}]
```

```
In [71]: trainer.test(enc_dec_net, xy_val)
```

```
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]
SLURM auto-requeueing enabled. Setting signal handlers.
Testing: |          | 0/? [00:00<?, ?it/s]
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┓
┃         Test metric         ┃        DataLoader 0         ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┩
│          test_acc           │             0.0             │
│          test_loss          │      8.799009323120117      │
└─────────────────────────────┴─────────────────────────────┘
```

```
Out[71]: [{'test_acc': 0.0, 'test_loss': 8.799009323120117}]
```

```
In [72]: trainer.fit(enc_dec_net, xy_train, xy_val)
```

```
/opt/conda/lib/python3.11/site-packages/lightning/pytorch/callbacks/model_checkpo
int.py:639: Checkpoint directory logs/persistence/encdec-0/checkpoints exists and
is not empty.
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]

  | Name     | Type               | Params
------------------------------------------------
0 | mc_acc   | MulticlassAccuracy | 0
1 | cce_loss | CrossEntropyLoss   | 0
2 | enc_net  | EncoderNetwork     | 4.1 M
3 | dec_net  | DecoderNetwork     | 7.7 M
------------------------------------------------
11.8 M    Trainable params
0         Non-trainable params
11.8 M    Total params
47.086    Total estimated model params size (MB)
SLURM auto-requeueing enabled. Setting signal handlers.
Sanity Checking: |          | 0/? [00:00<?, ?it/s]
Training: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
Validation: |          | 0/? [00:00<?, ?it/s]
`Trainer.fit` stopped: `max_epochs=30` reached.
```

In [73]:

```
results = pd.read_csv(logger.log_dir+"/metrics.csv")
results
```

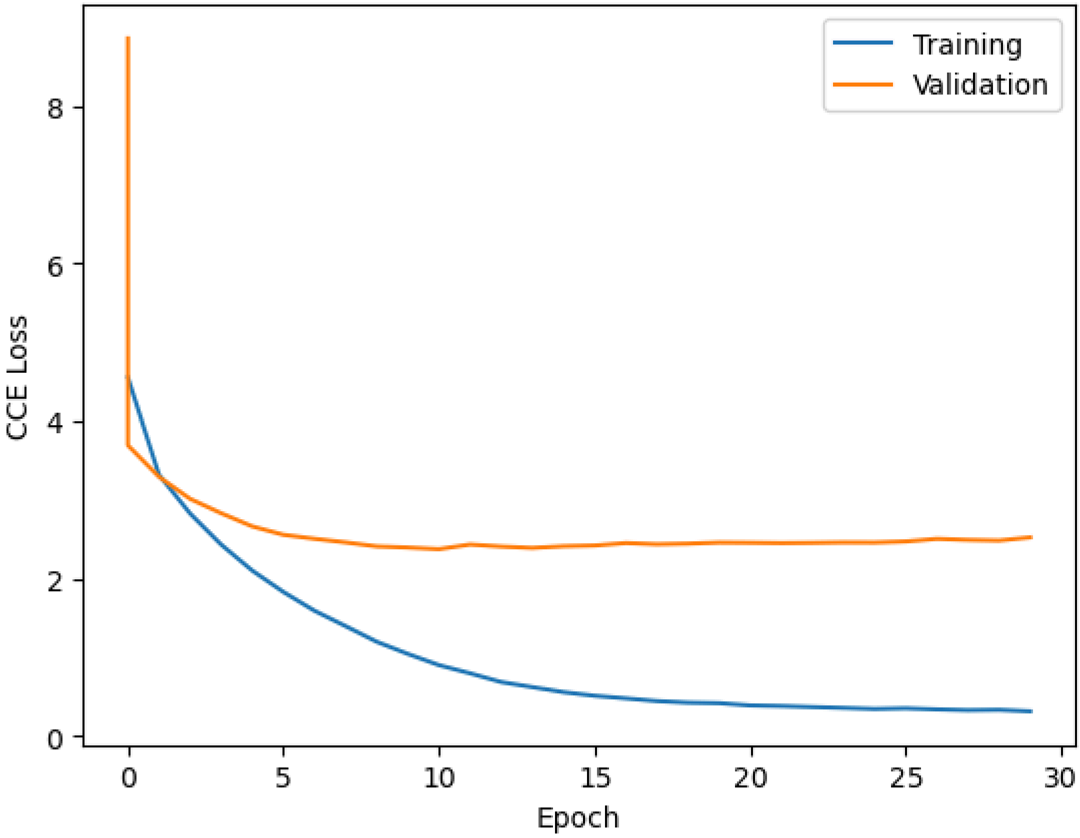
Out[73]:

	val_acc	val_loss	epoch	step	test_acc	test_loss	train_acc	train_loss
0	0.000099	8.858916	0	0	NaN	NaN	NaN	NaN
1	0.000099	8.858915	0	0	NaN	NaN	NaN	NaN
2	NaN	NaN	0	0	0.0	8.799009	NaN	NaN
3	0.456022	3.693235	0	31	NaN	NaN	NaN	NaN
4	NaN	NaN	0	31	NaN	NaN	0.378197	4.549609
...	...	...	...	...	...	...	...	...
58	NaN	NaN	27	895	NaN	NaN	0.893199	0.327084
59	0.670228	2.483268	28	927	NaN	NaN	NaN	NaN
60	NaN	NaN	28	927	NaN	NaN	0.891603	0.331121
61	0.671111	2.522851	29	959	NaN	NaN	NaN	NaN
62	NaN	NaN	29	959	NaN	NaN	0.894371	0.313119

63 rows × 8 columns

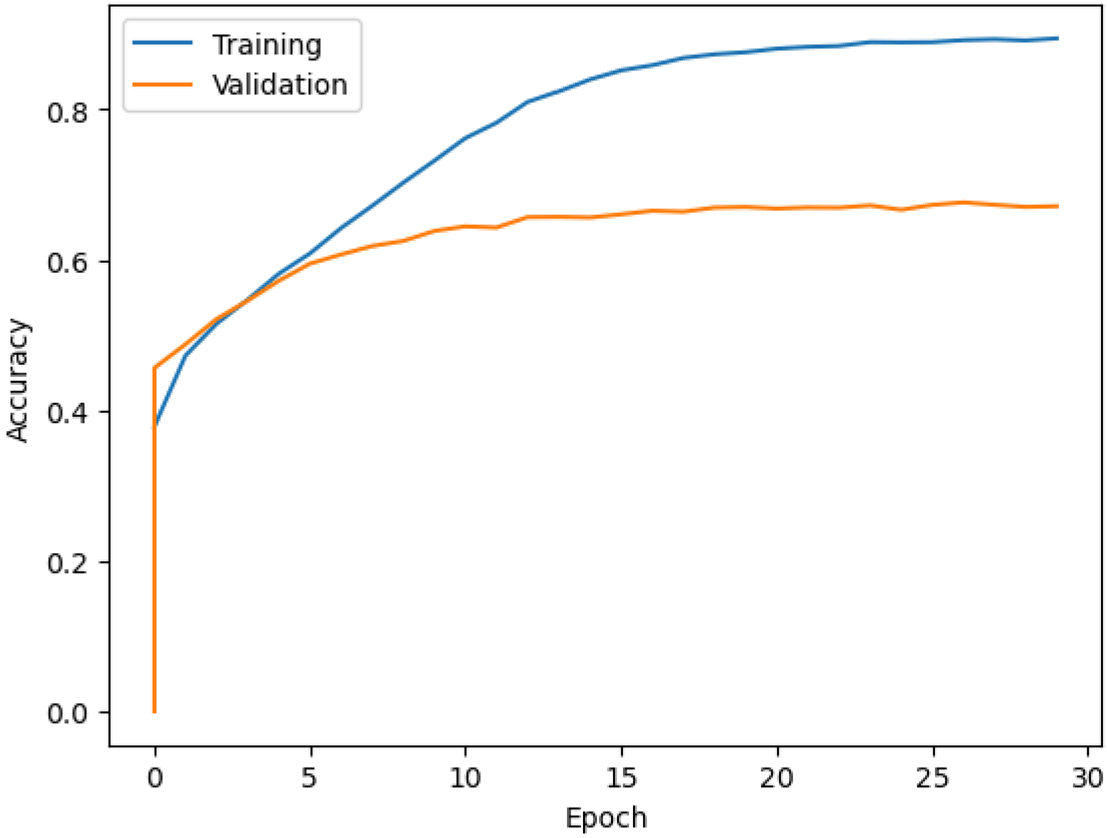
In [74]:

```python
plt.plot(results["epoch"][np.logical_not(np.isnan(results["train_loss"]))],
         results["train_loss"][np.logical_not(np.isnan(results["train_loss"]))],
         label="Training")
plt.plot(results["epoch"][np.logical_not(np.isnan(results["val_loss"]))],
         results["val_loss"][np.logical_not(np.isnan(results["val_loss"]))],
         label="Validation")
plt.legend()
plt.ylabel("CCE Loss")
plt.xlabel("Epoch")
plt.show()
```



In [75]:

```
plt.plot(results["epoch"][np.logical_not(np.isnan(results["train_acc"]))],
         results["train_acc"][np.logical_not(np.isnan(results["train_acc"]))],
         label="Training")
plt.plot(results["epoch"][np.logical_not(np.isnan(results["val_acc"]))],
         results["val_acc"][np.logical_not(np.isnan(results["val_acc"]))],
         label="Validation")
plt.legend()
plt.ylabel("Accuracy")
plt.xlabel("Epoch")
plt.show()
```



Test without Teacher Forcing

In [76]:
```
# Complete max_length cycles with the decoder
i = 0
enc_dec_net.to("cpu")
context = enc_dec_net.enc_net(torch.Tensor(enc_x_val[i:i+1]).long())
token = torch.zeros((1,dec_y_val.shape[1])).long()
token[0,0] = 1

for x in range(dec_y_val.shape[1]-1):
    result = enc_dec_net.dec_net(context,token).argmax(-1)
    token[0,x+1] = result[0,x]
    if result[0,x] == 2:
        break
result = token.cpu().detach().numpy()[0]
result
```

Out[76]:
```
array([   1, 5549,  514,   19,    2,    0,    0,    0,    0,    0])
```

English input...

In [77]:
```
decode_seq(enc_x_val[i],eng_tokenizer)
```

Out[77]:
```
'Can you swim ?'
```

Portuguese translation from network...

In [78]:
```
decode_seq(result,por_tokenizer)
```

Out[78]:
```
'Sabes nadar ?'
```

Target translation from the data set...

```
In [79]: decode_seq(dec_y_val[i],por_tokenizer)
```

```
Out[79]: 'Sabe nadar ?'
```

```
In [80]: result.shape
```

```
Out[80]: (10,)
```

```
In [81]: dec_y_val.shape
```

```
Out[81]: (2000, 10)
```

Accuracy **without** teacher forcing...

Keep track of the following result for comparison later...

```
In [82]: trainer.test(enc_dec_net, xy_val)
```

```
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]
SLURM auto-requeueing enabled. Setting signal handlers.
/opt/conda/lib/python3.11/site-packages/torch/utils/data/dataloader.py:557: UserW
arning: This DataLoader will create 4 worker processes in total. Our suggested ma
x number of worker in current system is 2, which is smaller than what this DataLo
ader is going to create. Please be aware that excessive worker creation might get
DataLoader running slow or even freeze, lower the worker number to avoid potentia
l slowness/freeze if necessary.
  warnings.warn(_create_warning_msg(
Testing: |          | 0/? [00:00<?, ?it/s]
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┓
┃        Test metric          ┃        DataLoader 0         ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┩
│          test_acc           │     0.4213317930698395      │
│          test_loss          │      6.924009323120117      │
└─────────────────────────────┴─────────────────────────────┘
```

```
Out[82]: [{'test_acc': 0.4213317930698395, 'test_loss': 6.924009323120117}]
```

Here's the magic word:

```
In [83]: trainer.save_checkpoint("enc_dec_net.ckpt")
```

Switch to Loading example notebook...