Assignment 2: Fashion Dataset and Architecture Comparisons

In this assignment, you are to perform the following:

1. Study the `FashionMNIST` dataset, and understand the images and their respective labels.
2. Implement a `Trainer` class in the file `trainer_lib.py` for reusable and reproducible training loops.
3. Implement the following architectures:

- Linear model
- MLP with one hidden layer
- Simple convolution with pooling
- Deep convolution with pooling, followed by MLP with one hidden layer.

In [128…
```
"🔒"
import torch
from torch import nn
from torch import optim
from torch.utils.data import DataLoader, random_split
import torchvision
from torchsummaryX import summary
import numpy as np
import pandas as pd
import os
import matplotlib.pyplot as plt
from importlib import reload

import warnings
warnings.filterwarnings('ignore')
```

Loading the data

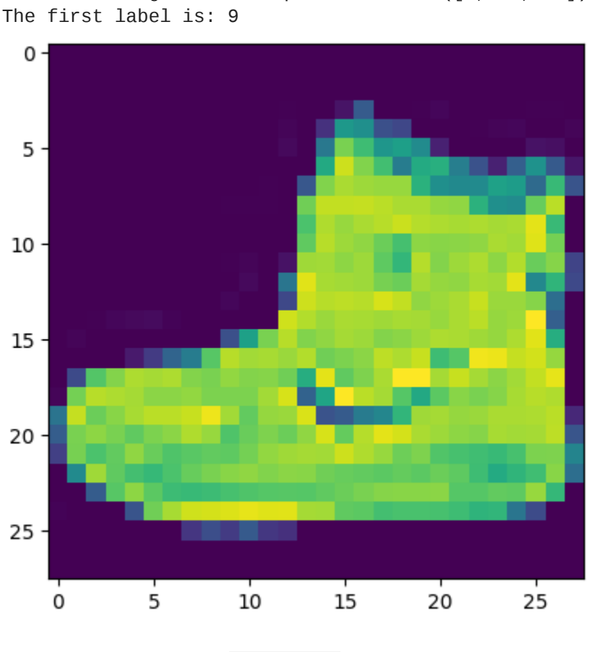
In this section, you will simply run the following data loading cells to familiarize yourself with the layout and semantics of the training dataset.

In [129…
```
"🔒"
#
# Loading data
#
home = os.environ.get('HOME')
root = os.path.join(home, 'public/data')
dataset = torchvision.datasets.FashionMNIST(
    root,
    train=True,
    transform=torchvision.transforms.ToTensor())

dataset
```

Out[129]:
```
Dataset FashionMNIST
    Number of datapoints: 60000
    Root location: /home/jovyan/public/data
    Split: Train
    StandardTransform
Transform: ToTensor()
```

In [130…
```
"🔒"
#
# Print out important stats about the dataset
#
image_tensor, label = dataset[0]
print("The first image is of shape:", image_tensor.shape)
print("The first label is:", label)

plt.imshow(np.transpose(image_tensor, (1, 2, 0)));
```

```
The first image is of shape: torch.Size([1, 28, 28])
The first label is: 9
```



The dataset object has a `.classes` field that contains the names of the different labels. It has 10 classes ranging from **T-shirt** to **Ankle boot**.

In [131…
```
"🔒"
#
# Print the lookup
#
lookup = pd.Series({x: i for (i,x) in enumerate(dataset.classes)})
lookup
```

Out[131]:
```
T-shirt/top    0
Trouser        1
Pullover       2
Dress          3
Coat           4
Sandal         5
Shirt          6
Sneaker        7
Bag            8
Ankle boot     9
dtype: int64
```

In fact, we can print the first 36 entries in the dataset, and plot them as a grid. Below is the result of that.

In [132…
```
"🔒"
xs = dataset.data[:36]
xs = xs.reshape(36, 1, 28, 28)
mosiac = torchvision.utils.make_grid(xs, nrow=6)
mosiac = np.transpose(mosiac, (1, 2, 0))
plt.imshow(mosiac)
plt.xticks([])
plt.yticks([]);
```

Getting ready for training and validation

We will be using the same training data for all three neural network architectures.

The `train_dataloader` is a dataloader for the training dataset, and `val_dataloader` is the dataloader for the validation dataset.

In [133…
```
"🔒"
train_dataset, val_dataset = random_split(dataset, (0.8, 0.2))
train_dataloader = DataLoader(train_dataset, batch_size=32, shuffle=True)
val_dataloader = DataLoader(val_dataset, batch_size=len(val_dataset), shuffle=False)
```

Linear classifier

In this section, you will implement a simple linear model. The linear model is to be implemented as a `nn.Module` and should have a field `model` which is a `nn.Sequential` module.

You are to complete the implementation by creating two layers in the `nn.Sequence(...)`:

- A layer to flatten the input from `(1,28,28)` to a vector.
- The second layer perform linear classification to 10 dimensional logit. You are to use the `nn.LazyLinear` layer to implement the linear layer. The advantage of the `LazyLinear` layer is that you do not need to compute the input dimension explicitly.

In [134…
```
"✍"
# @workUnit

class MyLinear(torch.nn.Module):
    def __init__(self):
        super().__init__()
        self.model = nn.Sequential(
            nn.Flatten(),
            nn.LazyLinear(10)
        )
    def forward(self, image):
        return self.model(image)
```

In [135…
```
"🔒"
# @check
# @title: check linear model architecture

m = MyLinear()
xs, ys = next(iter(train_dataloader))
summary(m, xs);
```

```
=================================================================
                     Kernel Shape Output Shape Params Mult-Adds
Layer
0_model.Flatten_0               -    [32, 784]      -         -
1_model.LazyLinear_1    [784, 10]     [32, 10]  7.85k     7.84k
-----------------------------------------------------------------
                      Totals
Total params           7.85k
Trainable params       7.85k
Non-trainable params     0.0
Mult-Adds              7.84k
=================================================================
```

A trainer class

In this section, you will be implementing a trainer class that will be used throughout the assignment.

The trainer class has the following methods:

Constructor:

- Input:
  - `model`: a `nn.Module` that is the neural network model.
  - `train_dataloader`: a `DataLoader` that is the training data
  - `val_dataloader`: is the `Dataloader` for validation
  - The constructor uses the `optim.Adam(...)` as the optimizer.
  - The constructor uses the `nn.CrossEntropy` as the loss function.

`Trainer.train_one_epoch`:

It performs training of `model` for **one epoch** by iterating over the batches from `train_dataloader`.

- Input:
  - `max_batches` is the maximum batches taken from the dataloader. **For performance reasons, we will only sample 10 batches.**
- Output: it returns two float numbers:
  - mean loss over the batches
  - mean accuracy over the batches

`Trainer.val_one_epoch`:

It performs validation by iterating over all batches from the validation dataloader.

- Output: it returns two float numbers:
  - mean loss
  - mean accuracy

`Trainer.train`

This method performs training.

- Input:
  - `epochs`: the number of epochs.
  - `max_batches`: the maximum batches to take per epoch.
- Output: returns a `DataFrame` containing training loss, training accuracy, validation loss, validation accuracy, and the time per epoch.

`Trainer.reset`

This method resets the trainable parameters of the model.

A skeleton implementation is given in `trainer_lib.py`. You are to complete the implementation.

In [155…
```
"🔒"
import trainer_lib
```

Below are some basic sanity checks of the trainer implementation.

In [156…
```
"🔒"
# @check
# @title: check trainer construction
reload(trainer_lib)
trainer = trainer_lib.Trainer(m, train_dataloader, val_dataloader)
print(trainer is not None)
```

```
True
```

In [157…
```
"🔒"
# @check
# @title: check trainer methods
for method in ['train_one_epoch', 'val_one_epoch', 'train', 'reset']:
    print(f"trainer.{method}?", hasattr(trainer, method))
```

```
trainer.train_one_epoch? True
trainer.val_one_epoch? True
trainer.train? True
trainer.reset? True
```

In [158…
```
"🔒"
# @check
# @title: check trainer.train_one_epoch

(loss, acc) = trainer.train_one_epoch(max_batches=1)
print('loss is numeric', isinstance(loss, float))
print('acc is numeric', isinstance(acc, float))
```

```
ouputs torch.Size([32, 10])
xs torch.Size([32, 1, 28, 28])
targets torch.Size([32])
loss is numeric True
acc is numeric True
```

In [159…
```
"🔒"
# @check
# @title: check trainer.val_one_epoch

(loss, acc) = trainer.val_one_epoch()
print('loss is numeric', isinstance(loss, float))
print('acc is numeric', isinstance(acc, float))
```

```
loss is numeric True
acc is numeric True
```

In [160…
```
"🔒"
# @check
# @title: check trainer.reset

trainer.reset()
print("Ok")
```

```
Ok
```

Training the linear model

In this section, we will train the linear model. We will take 10 batches per epoch, and train for 20 epochs.

The training history is stored in `model_linear_log`.

This will take approximately **70 seconds**.

In [76]:
```
"🔒"
reload(trainer_lib)

model_linear = MyLinear()
trainer = trainer_lib.Trainer(model_linear, train_dataloader, val_dataloader)
trainer.reset()
model_linear_log = trainer.train(epochs=20, max_batches=10)
model_linear_log.round(2)
```

```
[0 (2.65s)]: train_loss=2.07 train_acc=0.24, val_loss=1.85 val_acc=0.36
[1 (3.45s)]: train_loss=1.69 train_acc=0.51, val_loss=1.47 val_acc=0.63
[2 (3.36s)]: train_loss=1.42 train_acc=0.59, val_loss=1.27 val_acc=0.60
[3 (3.36s)]: train_loss=1.23 train_acc=0.60, val_loss=1.14 val_acc=0.65
[4 (3.57s)]: train_loss=1.11 train_acc=0.66, val_loss=1.08 val_acc=0.66
[5 (3.27s)]: train_loss=1.03 train_acc=0.70, val_loss=1.00 val_acc=0.67
[6 (4.29s)]: train_loss=1.02 train_acc=0.66, val_loss=0.96 val_acc=0.70
[7 (4.19s)]: train_loss=0.94 train_acc=0.69, val_loss=0.92 val_acc=0.72
[8 (3.30s)]: train_loss=0.95 train_acc=0.72, val_loss=0.89 val_acc=0.71
[9 (3.76s)]: train_loss=0.87 train_acc=0.73, val_loss=0.86 val_acc=0.71
[10 (3.45s)]: train_loss=0.73 train_acc=0.78, val_loss=0.84 val_acc=0.72
[11 (3.48s)]: train_loss=0.80 train_acc=0.73, val_loss=0.82 val_acc=0.71
[12 (3.21s)]: train_loss=0.80 train_acc=0.76, val_loss=0.81 val_acc=0.74
[13 (3.79s)]: train_loss=0.81 train_acc=0.74, val_loss=0.79 val_acc=0.74
[14 (3.01s)]: train_loss=0.71 train_acc=0.79, val_loss=0.78 val_acc=0.73
[15 (3.50s)]: train_loss=0.74 train_acc=0.72, val_loss=0.78 val_acc=0.73
[16 (3.51s)]: train_loss=0.77 train_acc=0.74, val_loss=0.79 val_acc=0.73
[17 (3.09s)]: train_loss=0.79 train_acc=0.73, val_loss=0.75 val_acc=0.75
[18 (3.53s)]: train_loss=0.74 train_acc=0.75, val_loss=0.74 val_acc=0.75
[19 (3.75s)]: train_loss=0.76 train_acc=0.76, val_loss=0.74 val_acc=0.75
== Total training time 69.54 seconds ==
```

Out[76]:

	train_loss	train_accuracy	val_loss	val_accuracy	epoch_duration
0	2.07	0.24	1.85	0.36	2.65
1	1.69	0.51	1.47	0.63	3.45
2	1.42	0.59	1.27	0.60	3.36
3	1.23	0.60	1.14	0.65	3.36
4	1.11	0.66	1.08	0.66	3.57
5	1.03	0.70	1.00	0.67	3.27
6	1.02	0.66	0.96	0.70	4.29
7	0.94	0.69	0.92	0.72	4.19
8	0.95	0.72	0.89	0.71	3.30
9	0.87	0.73	0.86	0.71	3.76
10	0.73	0.78	0.84	0.72	3.45
11	0.80	0.73	0.82	0.71	3.48
12	0.80	0.76	0.81	0.74	3.21
13	0.81	0.74	0.79	0.74	3.79
14	0.71	0.79	0.78	0.73	3.01
15	0.74	0.72	0.78	0.73	3.50
16	0.77	0.74	0.79	0.73	3.51
17	0.79	0.73	0.75	0.75	3.09
18	0.74	0.75	0.74	0.75	3.53
19	0.76	0.76	0.74	0.75	3.75

MLP Models

In this section, you are to construct a MLP model. It uses `LazyLinear` to create a hidden dimension of 50, with activation `nn.ReLU` for the hidden layer.

In [186…
```
"✍"
# @workUnit

class MLPModel(torch.nn.Module):
    def __init__(self):
        super().__init__()
        self.model = nn.Sequential(
            nn.Flatten(), #Just needed to add this
            nn.LazyLinear(50),
            nn.ReLU(),
            nn.Linear(50,10)
        )
    def forward(self, x):
        return self.model(x)
```

In [187…
```
"🔒"
# @check
# @title: MLP model architecture

model_mlp = MLPModel()
summary(model_mlp, xs);
```

```
==================================================================
                     Kernel Shape Output Shape  Params Mult-Adds
Layer
0_model.Flatten_0               -    [32, 784]       -         -
1_model.LazyLinear_1    [784, 50]     [32, 50]  39.25k     39.2k
2_model.ReLU_2                  -     [32, 50]       -         -
3_model.Linear_3         [50, 10]     [32, 10]   510.0     500.0
------------------------------------------------------------------
                      Totals
Total params          39.76k
Trainable params      39.76k
Non-trainable params     0.0
Mult-Adds              39.7k
==================================================================
```

Training MLP

We will train the MLP model in 20 epochs with 10 batches per epoch.

You should expect:

- Total training time: 80 seconds
- Final validation accuracy approximately 75% or higher.

In [166…
```
"🔒"
model_mlp = MLPModel()
reload(trainer_lib)
trainer = trainer_lib.Trainer(model_mlp, train_dataloader, val_dataloader)
trainer.reset()
model_mlp_log = trainer.train(epochs=20, max_batches=10)
model_mlp_log.round(2)
```

```
[0 (4.24s)]: train_loss=2.11 train_acc=0.38, val_loss=1.89 val_acc=0.48
[1 (3.61s)]: train_loss=1.70 train_acc=0.54, val_loss=1.52 val_acc=0.60
[2 (4.80s)]: train_loss=1.42 train_acc=0.60, val_loss=1.30 val_acc=0.56
[3 (4.39s)]: train_loss=1.23 train_acc=0.62, val_loss=1.12 val_acc=0.65
[4 (4.20s)]: train_loss=1.06 train_acc=0.65, val_loss=1.04 val_acc=0.63
[5 (4.12s)]: train_loss=1.05 train_acc=0.65, val_loss=0.96 val_acc=0.68
[6 (3.58s)]: train_loss=0.94 train_acc=0.69, val_loss=0.92 val_acc=0.67
[7 (3.41s)]: train_loss=0.88 train_acc=0.70, val_loss=0.86 val_acc=0.70
[8 (3.29s)]: train_loss=0.83 train_acc=0.69, val_loss=0.84 val_acc=0.72
[9 (3.41s)]: train_loss=0.82 train_acc=0.72, val_loss=0.81 val_acc=0.71
[10 (3.70s)]: train_loss=0.76 train_acc=0.73, val_loss=0.79 val_acc=0.72
[11 (3.73s)]: train_loss=0.74 train_acc=0.76, val_loss=0.77 val_acc=0.72
[12 (3.35s)]: train_loss=0.76 train_acc=0.71, val_loss=0.75 val_acc=0.73
[13 (3.60s)]: train_loss=0.73 train_acc=0.73, val_loss=0.75 val_acc=0.73
[14 (4.81s)]: train_loss=0.69 train_acc=0.74, val_loss=0.76 val_acc=0.70
[15 (3.69s)]: train_loss=0.77 train_acc=0.72, val_loss=0.72 val_acc=0.77
[16 (3.90s)]: train_loss=0.82 train_acc=0.73, val_loss=0.71 val_acc=0.77
[17 (4.00s)]: train_loss=0.72 train_acc=0.75, val_loss=0.69 val_acc=0.76
[18 (3.50s)]: train_loss=0.69 train_acc=0.76, val_loss=0.67 val_acc=0.77
[19 (3.80s)]: train_loss=0.66 train_acc=0.78, val_loss=0.66 val_acc=0.78
== Total training time 77.14 seconds ==
```

Out[166]:

	train_loss	train_accuracy	val_loss	val_accuracy	epoch_duration
0	2.11	0.38	1.89	0.48	4.24
1	1.70	0.54	1.52	0.60	3.61
2	1.42	0.60	1.30	0.56	4.80
3	1.23	0.62	1.12	0.65	4.39
4	1.06	0.65	1.04	0.63	4.20
5	1.05	0.65	0.96	0.68	4.12
6	0.94	0.69	0.92	0.67	3.58
7	0.88	0.70	0.86	0.70	3.41
8	0.83	0.69	0.84	0.72	3.29
9	0.82	0.72	0.81	0.71	3.41
10	0.76	0.73	0.79	0.72	3.70
11	0.74	0.76	0.77	0.72	3.73
12	0.76	0.71	0.75	0.73	3.35
13	0.73	0.73	0.75	0.73	3.60
14	0.69	0.74	0.76	0.70	4.81
15	0.77	0.72	0.72	0.77	3.69
16	0.82	0.73	0.71	0.77	3.90
17	0.72	0.75	0.69	0.76	4.00
18	0.69	0.76	0.67	0.77	3.50
19	0.66	0.78	0.66	0.78	3.80

Convolution based networks

In this section, we will construct a simple convolutional network. It consists of the following layers:

- 2D convolution with the `num_kernels` kernels of size `kernel_size`.
- Max pooling with pooling size `pool_size`.
- ReLU activation
- Flatten the max pooling output to a vector
- A (lazy) linear layer that maps to 10 dimensional logits vector

In [188…
```
"✍"
# @workUnit

class LinearCNNModel(nn.Module):
    def __init__(self, num_kernels, kernel_size, pool_size):
        super().__init__()
        self.model = nn.Sequential(
            nn.Conv2d(in_channels=1, out_channels=num_kernels, kernel_size=kernel_size),
            nn.MaxPool2d(kernel_size=pool_size),
            nn.ReLU(),
            nn.Flatten(),
            nn.Linear(num_kernels * ((28-kernel_size+1)//pool_size)**2, 10)
        )
    def forward(self, x):
        return self.model(x)
```

We will construct a convolutional network with 5 kernels, with kernel size 3, and pooling size of 2.

In [189…
```
"🔒"
# @check
# @title: CNN Linear architecture

model_cnn_linear = LinearCNNModel(5, 3, 2)
summary(model_cnn_linear, xs);
```

```
=====================================================================
                     Kernel Shape     Output Shape Params Mult-Adds
Layer
0_model.Conv2d_0     [1, 5, 3, 3]  [32, 5, 26, 26]   50.0    30.42k
1_model.MaxPool2d_1             -  [32, 5, 13, 13]      -         -
2_model.ReLU_2                  -  [32, 5, 13, 13]      -         -
3_model.Flatten_3               -        [32, 845]      -         -
4_model.Linear_4        [845, 10]         [32, 10]  8.46k     8.45k
---------------------------------------------------------------------
                      Totals
Total params           8.51k
Trainable params       8.51k
Non-trainable params     0.0
Mult-Adds             38.87k
=====================================================================
```

Training of Convolutional Network

We will train a convolutional network using 20 epochs, and 10 batches per epoch.

You should expect:

- Total training time: 120 seconds
- Validation accuracy approximately 76% or higher.

In [190…
```
"🔒"
reload(trainer_lib)
trainer = trainer_lib.Trainer(model_cnn_linear, train_dataloader, val_dataloader)
trainer.reset()
model_cnn_linear_log = trainer.train(epochs=20, max_batches=10)
model_cnn_linear_log.round(2)
```

```
[0 (7.00s)]: train_loss=2.25 train_acc=0.17, val_loss=2.07 val_acc=0.26
[1 (6.09s)]: train_loss=1.95 train_acc=0.36, val_loss=1.81 val_acc=0.44
[2 (6.00s)]: train_loss=1.67 train_acc=0.52, val_loss=1.54 val_acc=0.58
[3 (6.59s)]: train_loss=1.38 train_acc=0.62, val_loss=1.33 val_acc=0.59
[4 (5.80s)]: train_loss=1.26 train_acc=0.62, val_loss=1.18 val_acc=0.65
[5 (6.40s)]: train_loss=1.13 train_acc=0.66, val_loss=1.07 val_acc=0.68
[6 (5.81s)]: train_loss=1.01 train_acc=0.70, val_loss=0.98 val_acc=0.67
[7 (5.80s)]: train_loss=0.95 train_acc=0.68, val_loss=0.91 val_acc=0.72
[8 (5.89s)]: train_loss=0.82 train_acc=0.74, val_loss=0.87 val_acc=0.72
[9 (5.81s)]: train_loss=0.88 train_acc=0.70, val_loss=0.85 val_acc=0.69
[10 (5.49s)]: train_loss=0.82 train_acc=0.72, val_loss=0.82 val_acc=0.72
[11 (5.80s)]: train_loss=0.86 train_acc=0.73, val_loss=0.78 val_acc=0.74
[12 (6.31s)]: train_loss=0.79 train_acc=0.73, val_loss=0.77 val_acc=0.74
[13 (6.59s)]: train_loss=0.76 train_acc=0.75, val_loss=0.76 val_acc=0.74
[14 (6.50s)]: train_loss=0.79 train_acc=0.71, val_loss=0.74 val_acc=0.74
[15 (6.50s)]: train_loss=0.69 train_acc=0.78, val_loss=0.74 val_acc=0.74
[16 (6.11s)]: train_loss=0.78 train_acc=0.70, val_loss=0.73 val_acc=0.73
[17 (6.89s)]: train_loss=0.68 train_acc=0.74, val_loss=0.73 val_acc=0.73
[18 (6.30s)]: train_loss=0.74 train_acc=0.72, val_loss=0.71 val_acc=0.75
[19 (6.31s)]: train_loss=0.68 train_acc=0.75, val_loss=0.69 val_acc=0.75
== Total training time 124.00 seconds ==
```

Out[190]:

	train_loss	train_accuracy	val_loss	val_accuracy	epoch_duration
0	2.25	0.17	2.07	0.26	7.00
1	1.95	0.36	1.81	0.44	6.09
2	1.67	0.52	1.54	0.58	6.00
3	1.38	0.62	1.33	0.59	6.59
4	1.26	0.62	1.18	0.65	5.80
5	1.13	0.66	1.07	0.68	6.40
6	1.01	0.70	0.98	0.67	5.81
7	0.95	0.68	0.91	0.72	5.80
8	0.82	0.74	0.87	0.72	5.89
9	0.88	0.70	0.85	0.69	5.81
10	0.82	0.72	0.82	0.72	5.49
11	0.86	0.73	0.78	0.74	5.80
12	0.79	0.73	0.77	0.74	6.31
13	0.76	0.75	0.76	0.74	6.59
14	0.79	0.71	0.74	0.74	6.50
15	0.69	0.78	0.74	0.74	6.50
16	0.78	0.70	0.73	0.73	6.11
17	0.68	0.74	0.73	0.73	6.89
18	0.74	0.72	0.71	0.75	6.30
19	0.68	0.75	0.69	0.75	6.31

Deep Conv model

We will construct a **deep convolutional** network with the following layers:

- Conv2d with 16 kernels of size 5, with `padding='same'`
- MaxPool2d with pooling size of 2
- Conv2d with 32 kernels of size 3, with padding
- MaxPool2d with pooling size of 2
- Flatten the maxpooling output to a vector
- a MLP with hidden layer of 50, and ReLU is used as the activation function, and its output layer produces the 10 dimensional logits vector

In [171…
```
"✍"
# @workUnit

class DeepCNNModel(nn.Module):
    def __init__(self):
        super().__init__()
        self.model = nn.Sequential(
            nn.Conv2d(1, 16, kernel_size=5, padding='same'),
            nn.MaxPool2d(kernel_size=2),
            nn.Conv2d(16, 32, kernel_size=3, padding='same'),
            nn.MaxPool2d(kernel_size=2),
            nn.Flatten(),
            nn.Linear(32 * 7 * 7, 50),
            nn.ReLU(),
            nn.Linear(50, 10)
        )

    def forward(self, x):
        return self.model(x)
```

In [172…
```
"🔒"
# @check
# @title: deep cnn architecture

model_cnn_deep = DeepCNNModel()
summary(model_cnn_deep, xs);
```

```
==========================================================================
                       Kernel Shape      Output Shape  Params Mult-Adds
Layer
0_model.Conv2d_0      [1, 16, 5, 5]  [32, 16, 28, 28]   416.0    313.6k
1_model.MaxPool2d_1               -  [32, 16, 14, 14]       -         -
2_model.Conv2d_2     [16, 32, 3, 3]  [32, 32, 14, 14]   4.64k  903.168k
3_model.MaxPool2d_3               -    [32, 32, 7, 7]       -         -
4_model.Flatten_4                 -        [32, 1568]       -         -
5_model.Linear_5         [1568, 50]          [32, 50]  78.45k     78.4k
6_model.ReLU_6                    -          [32, 50]       -         -
7_model.Linear_7           [50, 10]          [32, 10]   510.0     500.0
--------------------------------------------------------------------------
                          Totals
Total params             84.016k
Trainable params         84.016k
Non-trainable params         0.0
Mult-Adds              1.295668M
==========================================================================
```

Training the deep CNN network

We train the deep convolutional network.

You should expect:

- Total training time: 250 seconds
- Validation accuracy should be close to or exceeds 80%.

In [173…
```
"🔒"
reload(trainer_lib)
trainer = trainer_lib.Trainer(model_cnn_deep, train_dataloader, val_dataloader)
trainer.reset()
model_cnn_deep_log = trainer.train(epochs=20, max_batches=10)
model_cnn_deep_log.round(2)
```

```
[0 (11.78s)]: train_loss=2.09 train_acc=0.41, val_loss=1.76 val_acc=0.45
[1 (11.31s)]: train_loss=1.43 train_acc=0.55, val_loss=1.24 val_acc=0.55
[2 (13.59s)]: train_loss=1.08 train_acc=0.58, val_loss=0.93 val_acc=0.66
[3 (12.40s)]: train_loss=0.96 train_acc=0.59, val_loss=0.91 val_acc=0.67
[4 (11.50s)]: train_loss=0.87 train_acc=0.69, val_loss=0.78 val_acc=0.71
[5 (12.89s)]: train_loss=0.76 train_acc=0.73, val_loss=0.78 val_acc=0.71
[6 (12.79s)]: train_loss=0.70 train_acc=0.74, val_loss=0.70 val_acc=0.75
[7 (12.00s)]: train_loss=0.63 train_acc=0.78, val_loss=0.67 val_acc=0.75
[8 (11.60s)]: train_loss=0.76 train_acc=0.75, val_loss=0.69 val_acc=0.76
[9 (12.50s)]: train_loss=0.69 train_acc=0.73, val_loss=0.69 val_acc=0.75
[10 (12.00s)]: train_loss=0.76 train_acc=0.73, val_loss=0.71 val_acc=0.72
[11 (12.70s)]: train_loss=0.70 train_acc=0.73, val_loss=0.65 val_acc=0.78
[12 (11.70s)]: train_loss=0.70 train_acc=0.74, val_loss=0.72 val_acc=0.76
[13 (11.90s)]: train_loss=0.69 train_acc=0.75, val_loss=0.63 val_acc=0.76
[14 (13.31s)]: train_loss=0.64 train_acc=0.76, val_loss=0.61 val_acc=0.77
[15 (13.89s)]: train_loss=0.59 train_acc=0.78, val_loss=0.60 val_acc=0.78
[16 (12.30s)]: train_loss=0.66 train_acc=0.74, val_loss=0.65 val_acc=0.75
[17 (13.01s)]: train_loss=0.61 train_acc=0.74, val_loss=0.60 val_acc=0.78
[18 (13.00s)]: train_loss=0.57 train_acc=0.79, val_loss=0.57 val_acc=0.79
[19 (13.20s)]: train_loss=0.55 train_acc=0.79, val_loss=0.62 val_acc=0.76
== Total training time 249.39 seconds ==
```

Out[173]:

	train_loss	train_accuracy	val_loss	val_accuracy	epoch_duration
0	2.09	0.41	1.76	0.45	11.78
1	1.43	0.55	1.24	0.55	11.31
2	1.08	0.58	0.93	0.66	13.59
3	0.96	0.59	0.91	0.67	12.40
4	0.87	0.69	0.78	0.71	11.50
5	0.76	0.73	0.78	0.71	12.89
6	0.70	0.74	0.70	0.75	12.79
7	0.63	0.78	0.67	0.75	12.00
8	0.76	0.75	0.69	0.76	11.60
9	0.69	0.73	0.69	0.75	12.50
10	0.76	0.73	0.71	0.72	12.00
11	0.70	0.73	0.65	0.78	12.70
12	0.70	0.74	0.72	0.76	11.70
13	0.69	0.75	0.63	0.76	11.90
14	0.64	0.76	0.61	0.77	13.31
15	0.59	0.78	0.60	0.78	13.89
16	0.66	0.74	0.65	0.75	12.30
17	0.61	0.74	0.60	0.78	13.01
18	0.57	0.79	0.57	0.79	13.00
19	0.55	0.79	0.62	0.76	13.20

Plot

The following plots the training accuracy.

In [191…
```
"🔒"
plt.figure()
plt.plot(model_linear_log.index, model_linear_log.train_accuracy)
plt.plot(model_mlp_log.index, model_mlp_log.train_accuracy)
plt.plot(model_cnn_linear_log.index, model_cnn_linear_log.train_accuracy)
plt.plot(model_cnn_deep_log.index, model_cnn_deep_log.train_accuracy)

plt.title('Training Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend(['Linear', 'MLP', 'CNN', 'Deep CNN']);
```

Plot the validation accuracy

Make sure you properly label the plot and the axes.

In [193…
```
"✍"
# @workUnit

#
# Generate the validation accuracy of the different models
plt.figure()
plt.plot(model_linear_log.index, model_linear_log.val_accuracy)
plt.plot(model_mlp_log.index, model_mlp_log.val_accuracy)
plt.plot(model_cnn_linear_log.index, model_cnn_linear_log.val_accuracy)
plt.plot(model_cnn_deep_log.index, model_cnn_deep_log.val_accuracy)

plt.title('Validation Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend(['Linear', 'MLP', 'CNN', 'Deep CNN']);
#
```


Validation Accuracy
Epoch
Value
Linear
ReLU
Sigmoid
Deep CNN