PyTorch Notes

**00. PyTorch Fundamentals**

**Tensors**

Tensors are a fundamental building block of machine learning and represents data in a numerical way. For example, an image tensor is represented with the shape [3, 244, 244] corresponding to color channels, width, and height.

Creating Scalars

# Scalar

scalar = torch.tensor(7)

scalar.ndim

scalar.item()

Creating Vectors

# Vector

vector = torch.tensor([7, 7])

vector.ndim

vector.shape

Creating Matrices

# Matrix

MATRIX = torch.tensor([[7, 8], [9, 10]])

MATRIX.ndim

MATRIX.shape

Creating Tensors

# Tensor

TENSOR = torch.tensor([[[1, 2, 3], [3, 6, 9], [2, 4, 5]]])

TENSOR.ndim

TENSOR.shape

**Random Tensors**

When building machine learning models with PyTorch, it's rare you'll create tensors by hand (like what we've being doing).

Instead, a machine learning model often starts out with large random tensors of numbers and adjusts these random numbers as it works through data to better represent it.

In essence:

*Start with random numbers -> look at data -> update random numbers -> look at data -> update random numbers...*

Random Tensors

# Create random tensor of size (244, 244, 3)

random\_image\_size\_tensor = torch.rand(size=(244, 244, 3))

random\_image\_size\_tensor.shap, random\_image\_size\_tensor.ndim

**Zeros**

Sometimes we want to set this tensor to zeros and ones effectively masking which weighted parameters the model should not learn.

Zeros and Ones Tensors

zeros = torch.zeros(size=(3, 4))

zeros, zeros.dtype

ones = torch.ones(size=(3, 4))

ones, ones.dtype

Range Tensors

zeros\_to\_ten = torch.arange(start=0, end=10, step=1)

zeros\_to\_ten

Note: sometimes you’ll want one tensor of a certain type with the same shape as another tensor. To do so you can use torch.zeros\_like(input tensor) or torch.ones\_like(input tensor).

**Tensor Datatypes**

There are many different tensor datatypes and some are specific to CPU and some are better for GPU.

Types include: torch.cuda, torch.float32, torch.float, torch.float16, torch.half, torch.float64, torch.double

Of course, the higher the precision value, the more detail and data used to express a number. This matters in deep learning and numerical computing because of the shear volumes of operations. The more detail in each number, the more you have to compute. So, one would say to use lower precision datatypes since they are faster to compute. However, this sacrifices some performance on evaluation metrics like accuracy.

Range Tensors

# Default datatype for tensors is float32

float\_32\_tensor = torch.tensor([3.0, 6.0, 9.0],

dtype=None,

device=None,

requires\_grad=False)

float\_32\_tensor.shape, float\_32\_tensor.dtype, float\_32\_tensor.device

Tensor Properties

# Create a tensor

some\_tensor = torch.rand(3, 4)

# Find out details about it

print(some\_tensor)

print(f"Shape of tensor: {some\_tensor.shape}")

print(f"Datatype of tensor: {some\_tensor.dtype}")

print(f"Device tensor is stored on: {some\_tensor.device}") # will default to CPU

**Tensor Operations**

tensor = torch.tensor([1, 2, 3])

tensor + 10

tensor \* 10

torch.multiply(tensor, 10)

torch.matmul(tensor, tensor)

torch.mm(tensor, tensor)

# element wise multiplication

Tensor \* tensor

The torch.nn.Linear() module (we'll see this in action later on), also known as a feed-forward layer or fully connected layer, implements a matrix multiplication between an input x and a weights matrix A.

𝑦 = 𝑥 ⋅ 𝐴𝑇 + 𝑏

**Feed Forward Linear Layer**

# Since the linear layer starts with a random weight matrix, let's make it reproducible

torch.manual\_seed(42)

# This uses matrix multiplication

linear = torch.nn.Linear(in\_features=2, # in\_features = matches inner dimension of input

out\_features=6) # out\_features = describes outer value

x = tensor\_A

output = linear(x)

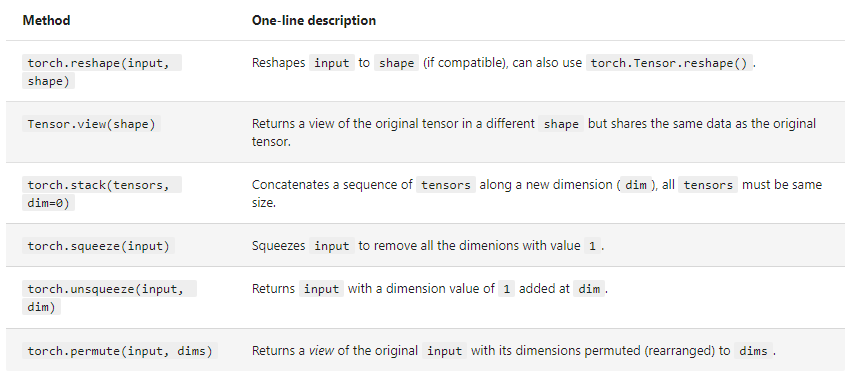
print(f"Input shape: {x.shape}\n")

print(f"Output:\n{output}\n\nOutput shape: {output.shape}")

**Converting Tensor Type**

tensor\_float16 = tensor.type(torch.float16)

**Reshape, Stacking, Squeezing, and Unsqueezing**

****

**PyTorch Tensors & NumPy**

# NumPy array to tensor

import torch

import numpy as np

array = np.arange(1.0, 8.0)

tensor = torch.from\_numpy(array)

array, tensor

# Tensor to NumPy array

tensor = torch.ones(7) # create a tensor of ones with dtype=float32

numpy\_tensor = tensor.numpy() # will be dtype=float32 unless changed

tensor, numpy\_tensor

**Reproducibility**

As you learn more about neural networks and machine learning, you'll start to discover how much randomness plays a part.

import torch

import random

# # Set the random seed

RANDOM\_SEED=42

torch.manual\_seed(seed=RANDOM\_SEED)

random\_tensor\_C = torch.rand(3, 4)

# Have to reset the seed every time a new rand() is called

# Without this, tensor\_D would be different to tensor\_C

torch.random.manual\_seed(seed=RANDOM\_SEED) # try commenting this line out and seeing what happens

random\_tensor\_D = torch.rand(3, 4)

print(f"Tensor C:\n{random\_tensor\_C}\n")

print(f"Tensor D:\n{random\_tensor\_D}\n")

print(f"Does Tensor C equal Tensor D? (anywhere)")

random\_tensor\_C == random\_tensor\_D

**Using Your GPU**

# Check for GPU

import torch

torch.cuda.is\_available()

device = “cuda” if torch.cuda.is\_available() else “cpu”

# Count number of devices

torch.cuda.device\_count()

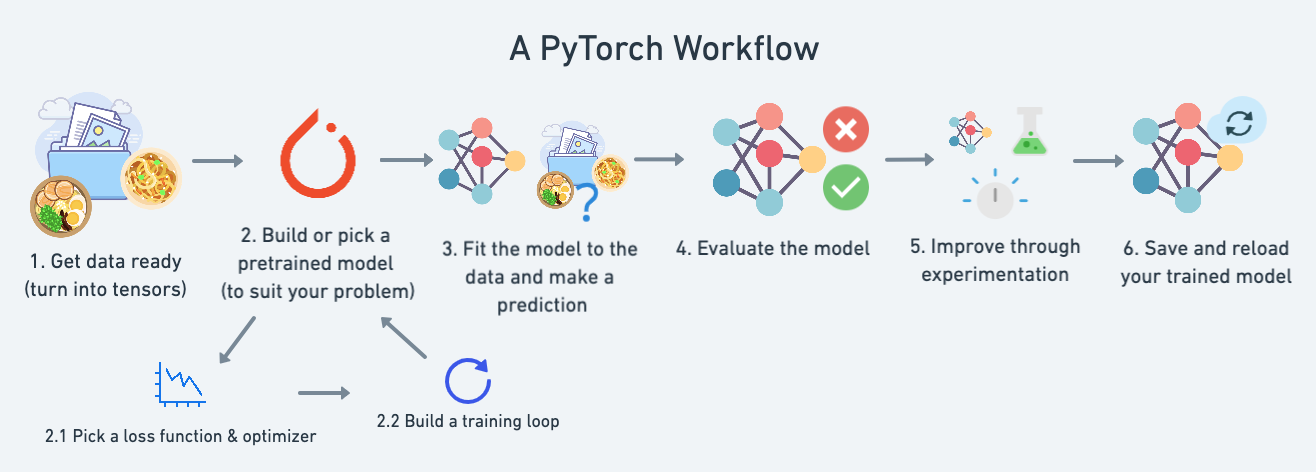
tensor = torch.tensor([1, 2, 3])

tensor\_on\_gpu = tensor.to(device)

# Note if tensor is on GPU, it can’t be transformed it to NumPy. To move back.

tensor\_back\_to\_cpu = torch\_on\_gpu.cpu()

**01. PyTorch Workflow Fundamentals**

C

**1. Creating Data**

# Create \*known\* parameters

weight = 0.7

bias = 0.3

# Create data

start = 0

end = 1

step = 0.02

X = torch.arange(start, end, step).unsqueeze(dim=1)

y = weight \* X + bias

X[:10], y[:10]

Split Data

We need to split the data up into training set, validation set, testing set.

# Create train/test split

train\_split = int(0.8 \* len(X)) # 80% of data used for training set, 20% for testing

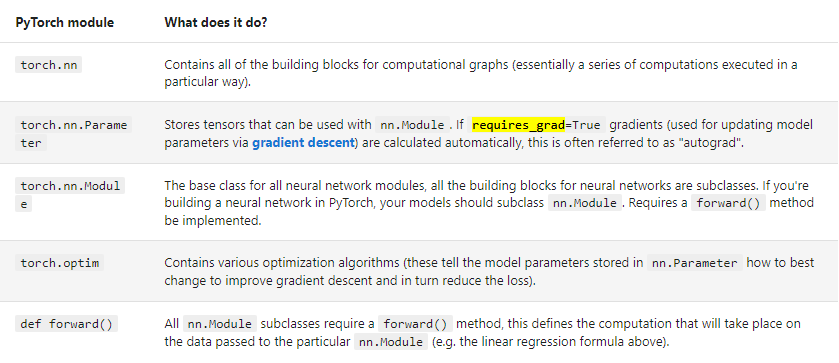
X\_train, y\_train = X[:train\_split], y[:train\_split]

X\_test, y\_test = X[train\_split:], y[train\_split:]

len(X\_train), len(y\_train), len(X\_test), len(y\_test)

**2. Building the Model**

*torch.nn*



# Create a Linear Regression model class

class LinearRegressionModel(nn.Module):

def \_\_init\_\_(self):

super().\_\_init\_\_()

self.weights = nn.Parameter(torch.randn(1, dtype=torch.float), requires\_grad=True)

self.bias = nn.Parameter(torch.randn(1, dtype=torch.float), requires\_grad=True)

def forward(self, x: torch.Tensor) -> torch.Tensor:

return self.weights \* x + self.bias # <- this is the linear regression formula (y = m\*x + b)

torch.manual\_seed(42)

model\_0 = LinearRegressionMode()

list(model\_0.parameters())

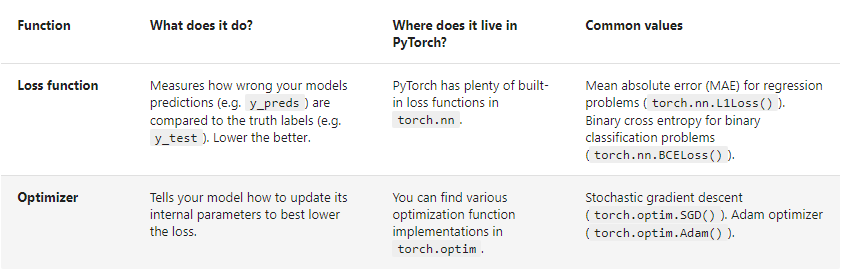
model\_0.state\_dict()

# Make predictions with model

with torch.inference\_mode():

y\_preds = model\_0(X\_test)

**3. Train Model**



Let’s use stochastic gradient descent as our optimizer and the mean absolute error as our loss function.

# Create the loss function

loss\_fn = nn.L1Loss() # MAE loss is same as L1Loss

# Create the optimizer

optimizer = torch.optim.SGD(params=model\_0.parameters(), lr=0.01)

# learning rate (how much the optimizer should change parameters at each step, higher=more (less stable), lower=less (might take a long time))

*Training Loop*

Now that we defined our model, loss function, and optimizer, we now want to create a training loop to help the model determine the relationship between the features and labels.

Forward pass -> calculate loss -> zero gradients -> backpropagation on loss -> update optimizer

\*Note: Back propagation is the process or computing the gradient of the loss with respect to every model parameter to be updated.

*Testing Loop*

Forward pass -> calculate loss -> calculate evaluation metrics for personal analysis

Putting it all together:

torch.manual\_seed(42)

epochs = 100

train\_loss\_values = []

test\_loss\_values = []

epoch\_count = []

for epoch in range(epochs):

### Training

# Put model in training mode (this is the default state of a model)

model\_0.train()

# 1. Forward pass on train data using the forward() method inside

y\_pred = model\_0(X\_train)

# print(y\_pred)

# 2. Calculate the loss (how different are our models predictions to the ground truth)

loss = loss\_fn(y\_pred, y\_train)

optimizer.zero\_grad()

loss.backward()

optimizer.step()

### Testing

# Put the model in evaluation mode

model\_0.eval()

with torch.inference\_mode():

test\_pred = model\_0(X\_test)

test\_loss = loss\_fn(test\_pred, y\_test.type(torch.float))

if epoch % 10 == 0:

epoch\_count.append(epoch)

train\_loss\_values.append(loss.detach().numpy())

test\_loss\_values.append(test\_loss.detach().numpy())

print(f"Epoch: {epoch} | MAE Train Loss: {loss} | MAE Test Loss: {test\_loss} ")

**4. Make Predictions**

model\_0.eval()

# 2. Setup the inference mode context manager

with torch.inference\_mode():

# 3. Make sure the calculations are done with the model and data on the same device

# in our case, we haven't setup device-agnostic code yet so our data and model are

# on the CPU by default.

# model\_0.to(device)

# X\_test = X\_test.to(device)

y\_preds = model\_0(X\_test)

**5. Saving and Loading a PyTorch Model**

If you’ve trained a PyTorch model, chances are you’ll want to save it and export it somewhere:

from pathlib import Path

MODEL\_PATH = Path("models")

MODEL\_PATH.mkdir(parents=True, exist\_ok=True)

MODEL\_NAME = "01\_pytorch\_workflow\_model\_0.pth"

MODEL\_SAVE\_PATH = MODEL\_PATH / MODEL\_NAME

torch.save(obj=model\_0.state\_dict(), f=MODEL\_SAVE\_PATH)

# Note we only saved the model’s parameters not the model itself

# Instantiate a new instance of our model (this will be instantiated with random weights)

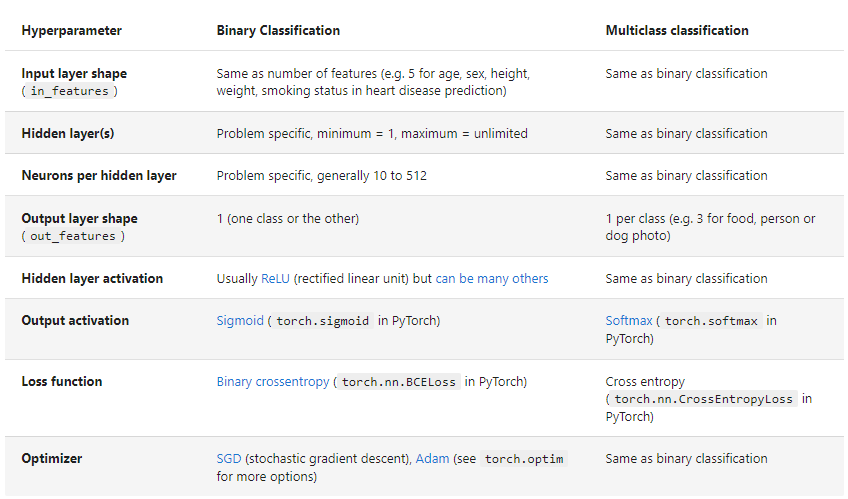
loaded\_model\_0 = LinearRegressionModel()

# Load the state\_dict of our saved model (this will update the new instance of our model with trained weights)

loaded\_model\_0.load\_state\_dict(torch.load(f=MODEL\_SAVE\_PATH))

**02. PyTorch Neural Network Classification**

The classification problem includes: binary, multi-class, and multi-label classification. Below is the architecture of a classification neural network:



Let’s generate some classification data:

from sklearn.datasets import make\_circles

import matplotlib.pyplot as plt

n\_samples = 1000

X, y = make\_circles(n\_samples, noise=0.03, random\_state=42)

# Check different labels

Circles.label.value\_counts()

Plt.scatter(x=X[:,0], y=X[:, 1], c=y, cmap=plt.cm.RdYlBy)

Let’s try and find a neural network to classify dots into class 1 or 2. First, we prepare the data in tensor format for PyTorch and we create our test data:

import torch

from sklearn.model\_selection import train\_test\_split

X = torch.from\_numpy(X).type(torch.float)

y = torch.from\_numpy(y).type(torch.float)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

Now, let’s create our model. This setup where you have features and labels is referred to as **supervised learning** because your data is telling your model what the outputs should be given a certain input.

class CircleModelV0(nn.Module):

def \_\_init\_\_(self):

super().\_\_init\_\_()

self.layer\_1 = nn.Linear(in\_features=2, out\_features=5)

self.layer\_2 = nn.Linear(in\_features=5, out\_features=1)

# the out\_features are the hidden units or neurons which become the input for layer 2

def forward(self, x):

return self.layer\_2(self.layer\_1(x))

model\_0 = CircleModelV0().to(device)

#==========================================================#

# You can also use **nn.Squential** to replicate what’s happening above and ensures it happens

# sequentially in a straight-forward computation. You’ll have to define your own nn.Model if

# you don’t want it to be straight-forward and sequential.

model\_0 = nn.Sequential(

nn.Linear(in\_features=2, out\_features=5),

nn.Linear(in\_features=5, out\_features=1)

).to(device)

# Make prediction with the model

untrained\_preds = model\_0(X\_test.to(device))

Now that we’ve defined our model, we need to create a loss function and an optimizer. See **torch.optim table** documentation for more loss function and optimizer options.

We can also add an **evaluation metric** to offer another perspective of how the model is doing (i.e. an evaluation metric can be a measure of how right you are).

# Calculate accuracy (a classification metric)

def accuracy\_fn(y\_true, y\_pred):

correct = torch.eq(y\_true, y\_pred).sum().item()

acc = (correct / len(y\_pred)) \* 100

return acc

With all that done, let’s develop our training loop:

You’ll notice the output of the code above is as good as guessing. Why? Well, with only two sequential operations our decision boundary looks like this:

import requests

from pathlib import Path

if Path("helper\_functions.py").is\_file():

print("helper\_functions.py already exists, skipping download")

else:

print("Downloading helper\_functions.py")

request = requests.get("https://raw.githubusercontent.com/mrdbourke/pytorch-deep-learning/main/helper\_functions.py")

with open("helper\_functions.py", "wb") as f:

f.write(request.content)

from helper\_functions import plot\_predictions, plot\_decision\_boundary

plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)

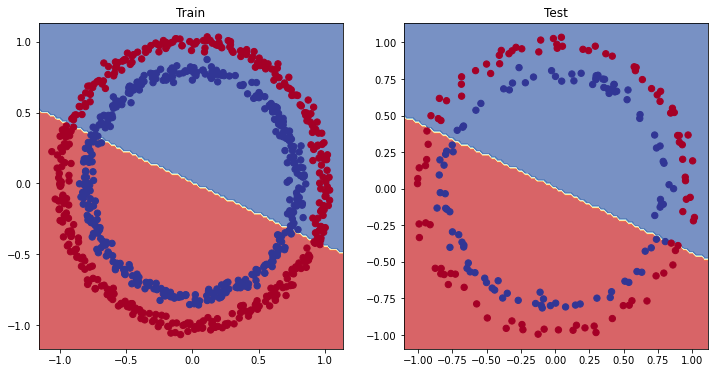
plt.title("Train")

plot\_decision\_boundary(model\_0, X\_train, y\_train)

plt.subplot(1, 2, 2)

plt.title("Test")

plot\_decision\_boundary(model\_0, X\_test, y\_test)



torch.manual\_seed(42)

epochs = 100

X\_train, y\_train = X\_train.to(device), y\_train.to(device)

X\_test, y\_test = X\_test.to(device), y\_test.to(device)

for epoch in range(epochs):

model\_0.train()

y\_logits = model\_0(X\_train).squeeze()

y\_pred = torch.round(torch.sigmoid(y\_logits))

# loss = loss\_fn(torch.sigmoid(y\_logits), y\_train)

loss = loss\_fn(y\_logits, y\_train)

acc = accuracy\_fn(y\_true=y\_train, y\_pred=y\_pred)

optimizer.zero\_grad()

loss.backward()

optimizer.step()

model\_0.eval()

with torch.inference\_mode():

test\_logits = model\_0(X\_test).squeeze()

test\_pred = torch.round(torch.sigmoid(test\_logits))

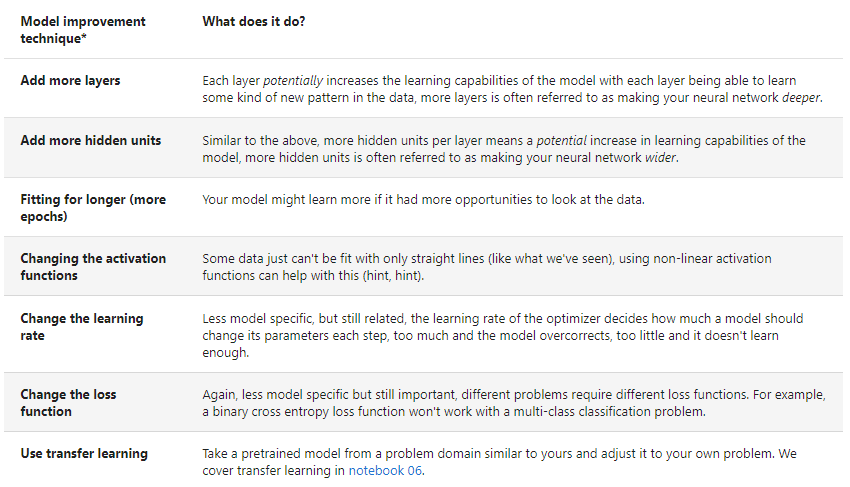
test\_loss = loss\_fn(test\_logits, y\_test)

test\_acc = accuracy\_fn(y\_true=y\_test, y\_pred=test\_pred)

if epoch % 10 == 0:

print(f"Epoch: {epoch} | Loss: {loss:.5f}, Accuracy: {acc:.2f}% | Test loss: {test\_loss:.5f}, Test acc: {test\_acc:.2f}%")

We need to improve our model! Here are some methods for improving our model:



One thing we add to our model to introduce non-linearity in a case such as this, we use non-linear functions such as **RELU** and **Sigmoid**. Combining non-linear and linear layers allows us to combine a set of straight and non-straight lines to achieve a model with more tailored boundaries.

**What about working with multiple classifications?**

We can use sklearn make blob to create multiple class datasets. Note, in multi-class models, we explore different functions such as the loss function nn.CrossEntropyLoss() and the softmax activation function.

The softmax function calculates the probability of each prediction class being the actual predicted class compared to all other possible classes.

**Generating multiple blobs**

# Import dependencies

import torch

import matplotlib.pyplot as plt

from sklearn.datasets import make\_blobs

from sklearn.model\_selection import train\_test\_split

# Set the hyperparameters for data creation

NUM\_CLASSES = 4

NUM\_FEATURES = 2

RANDOM\_SEED = 42

# 1. Create multi-class data

X\_blob, y\_blob = make\_blobs(n\_samples=1000,

n\_features=NUM\_FEATURES, # X features

centers=NUM\_CLASSES, # y labels

cluster\_std=1.5, # give the clusters a little shake up (try changing this to 1.0, the default)

random\_state=RANDOM\_SEED

)

# 2. Turn data into tensors

X\_blob = torch.from\_numpy(X\_blob).type(torch.float)

y\_blob = torch.from\_numpy(y\_blob).type(torch.LongTensor)

print(X\_blob[:5], y\_blob[:5])

# 3. Split into train and test sets

X\_blob\_train, X\_blob\_test, y\_blob\_train, y\_blob\_test = train\_test\_split(X\_blob,

y\_blob,

test\_size=0.2,

random\_state=RANDOM\_SEED

)

# 4. Plot data

plt.figure(figsize=(10, 7))

plt.scatter(X\_blob[:, 0], X\_blob[:, 1], c=y\_blob, cmap=plt.cm.RdYlBu);

**Building Multi-class Model**

# Fit the model

torch.manual\_seed(42)

# Set number of epochs

epochs = 100

# Put data to target device

X\_blob\_train, y\_blob\_train = X\_blob\_train.to(device), y\_blob\_train.to(device)

X\_blob\_test, y\_blob\_test = X\_blob\_test.to(device), y\_blob\_test.to(device)

for epoch in range(epochs):

### Training

model\_4.train()

# 1. Forward pass

y\_logits = model\_4(X\_blob\_train) # model outputs raw logits

y\_pred = torch.softmax(y\_logits, dim=1).argmax(dim=1) # go from logits -> prediction probabilities -> prediction labels

# print(y\_logits)

# 2. Calculate loss and accuracy

loss = loss\_fn(y\_logits, y\_blob\_train)

acc = accuracy\_fn(y\_true=y\_blob\_train,

y\_pred=y\_pred)

# 3. Optimizer zero grad

optimizer.zero\_grad()

# 4. Loss backwards

loss.backward()

# 5. Optimizer step

optimizer.step()

### Testing

model\_4.eval()

with torch.inference\_mode():

# 1. Forward pass

test\_logits = model\_4(X\_blob\_test)

test\_pred = torch.softmax(test\_logits, dim=1).argmax(dim=1)

# 2. Calculate test loss and accuracy

test\_loss = loss\_fn(test\_logits, y\_blob\_test)

test\_acc = accuracy\_fn(y\_true=y\_blob\_test, y\_pred=test\_pred)

# Print out what's happening

if epoch % 10 == 0:

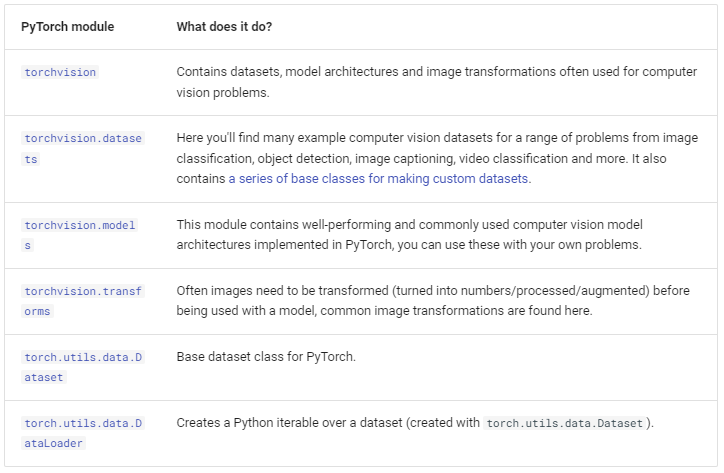
print(f"Epoch: {epoch} | Loss: {loss:.5f}, Acc: {acc:.2f}% | Test Loss: {test\_loss:.5f}, Test Acc: {test\_acc:.2f}%")

**More Classification Evaluation Metrics**



**03. PyTorch Computer Vision**

Computer vision is the art of teaching a computer to see (e.g. binary classification, multi-class classification, object detection, panoptic segmentation). First, we need to go over the computer vision library in PyTorch.



**Setup Training & Testing Data**

# Setup fashion training data

Training\_data = datasets.FashionMNIST(

Root=“data”

Train=True

Download=True

Transform=ToTensor()

Target\_transform=None

)

# Set up testing data

# Setup testing data

test\_data = datasets.FashionMNIST(

root="data",

train=False, # get test data

download=True,

transform=ToTensor()

)

The image data is often provided as color channels, height, width (CHW). You’ll often see NCHW where the N denotes a number of images.

To see the data attributes, check len(train\_data.data), len(train\_data.targets), or train\_data.classes.

**Visualizing the Data**

import matplotlib.pyplot as plt

image, label = train\_data[0]

print(f"Image shape: {image.shape}")

plt.imshow(image.squeeze()) # image shape is [1, 28, 28] (colour channels, height, width)

plt.title(label);

# For grayscale:

plt.imshow(image.squeeze(), cmap="gray")

plt.title(class\_names[label]);

Now that we can visualize our data, we want to use data loader to turn large datasets in python to batches or mini-batches to make computation more efficient and more informed gradient decent steps per epoch. 32 is a good place to start for batch sizes and typically we increase by multiples of two from there.

**Importing DataLoader**

from torch.utils.data import DataLoader

# Setup the batch size hyperparameter

BATCH\_SIZE = 32

# Turn datasets into iterables (batches)

train\_dataloader = DataLoader(train\_data, # dataset to turn into iterable

batch\_size=BATCH\_SIZE, # how many samples per batch?

shuffle=True # shuffle data every epoch?

)

test\_dataloader = DataLoader(test\_data,

batch\_size=BATCH\_SIZE,

shuffle=False # don't necessarily have to shuffle the testing data

)

# Let's check out what we've created

print(f"Dataloaders: {train\_dataloader, test\_dataloader}")

print(f"Length of train dataloader: {len(train\_dataloader)} batches of {BATCH\_SIZE}")

print(f"Length of test dataloader: {len(test\_dataloader)} batches of {BATCH\_SIZE}")

# Check out what's inside the training dataloader

train\_features\_batch, train\_labels\_batch = next(iter(train\_dataloader))

train\_features\_batch.shape, train\_labels\_batch.shape

**Flatten Image Data for Model**

# Create a flatten layer

flatten\_model = nn.Flatten() # all nn modules function as a model (can do a forward pass)

# Get a single sample

x = train\_features\_batch[0]

# Flatten the sample

output = flatten\_model(x) # perform forward pass

# Print out what happened

print(f"Shape before flattening: {x.shape} -> [color\_channels, height, width]")

print(f"Shape after flattening: {output.shape} -> [color\_channels, height\*width]")

# Try uncommenting below and see what happens

#print(x)

#print(output)

**Finally Our Baseline Model**

The input shape for the model is the number of pixels with the output shape being the number of classes expected in our problem.

**from** torch **import** nn

**class** FashionMNISTModelV0(nn**.**Module):

**def** \_\_init\_\_(self, input\_shape: int, hidden\_units: int, output\_shape: int):

super()**.**\_\_init\_\_()

self**.**layer\_stack **=** nn**.**Sequential(

nn**.**Flatten(), *# neural networks like their inputs in vector form*

nn**.**Linear(in\_features**=**input\_shape, out\_features**=**hidden\_units),

nn**.**Linear(in\_features**=**hidden\_units, out\_features**=**output\_shape)

)

**def** forward(self, x):

**return** self**.**layer\_stack(x)

torch.manual\_seed(42)

# Need to setup model with input parameters

model\_0 = FashionMNISTModelV0(input\_shape=784, # one for every pixel (28x28)

hidden\_units=10, # how many units in the hidden layer

output\_shape=len(class\_names) # one for every class

)

model\_0.to("cpu") # keep model on CPU to begin with

After we create the model, we need to set up the loss, optimizer, and evaluation. Since, our evaluation metrics will be calculated PER BATCH rather than across the whole data set, well have to divide our loss and accuracy values by the number of batches in each dataset’s respective dataloader. The whole thing put together looks like this:

# Import tqdm for progress bar

from tqdm.auto import tqdm

# Set the seed and start the timer

torch.manual\_seed(42)

train\_time\_start\_on\_cpu = timer()

# Set the number of epochs (we'll keep this small for faster training times)

epochs = 3

# Create training and testing loop

for epoch in tqdm(range(epochs)):

print(f"Epoch: {epoch}\n-------")

### Training

train\_loss = 0

# Add a loop to loop through training batches

for batch, (X, y) in enumerate(train\_dataloader):

model\_0.train()

# 1. Forward pass

y\_pred = model\_0(X)

# 2. Calculate loss (per batch)

loss = loss\_fn(y\_pred, y)

train\_loss += loss # accumulatively add up the loss per epoch

# 3. Optimizer zero grad

optimizer.zero\_grad()

# 4. Loss backward

loss.backward()

# 5. Optimizer step

optimizer.step()

# Print out how many samples have been seen

if batch % 400 == 0:

print(f"Looked at {batch \* len(X)}/{len(train\_dataloader.dataset)} samples")

# Divide total train loss by length of train dataloader (average loss per batch per epoch)

train\_loss /= len(train\_dataloader)

### Testing

# Setup variables for accumulatively adding up loss and accuracy

test\_loss, test\_acc = 0, 0

model\_0.eval()

with torch.inference\_mode():

for X, y in test\_dataloader:

# 1. Forward pass

test\_pred = model\_0(X)

# 2. Calculate loss (accumatively)

test\_loss += loss\_fn(test\_pred, y) # accumulatively add up the loss per epoch

# 3. Calculate accuracy (preds need to be same as y\_true)

test\_acc += accuracy\_fn(y\_true=y, y\_pred=test\_pred.argmax(dim=1))

# Calculations on test metrics need to happen inside torch.inference\_mode()

# Divide total test loss by length of test dataloader (per batch)

test\_loss /= len(test\_dataloader)

# Divide total accuracy by length of test dataloader (per batch)

test\_acc /= len(test\_dataloader)

## Print out what's happening

print(f"\nTrain loss: {train\_loss:.5f} | Test loss: {test\_loss:.5f}, Test acc: {test\_acc:.2f}%\n")

# Calculate training time

train\_time\_end\_on\_cpu = timer()

total\_train\_time\_model\_0 = print\_train\_time(start=train\_time\_start\_on\_cpu,

end=train\_time\_end\_on\_cpu,

device=str(next(model\_0.parameters()).device))

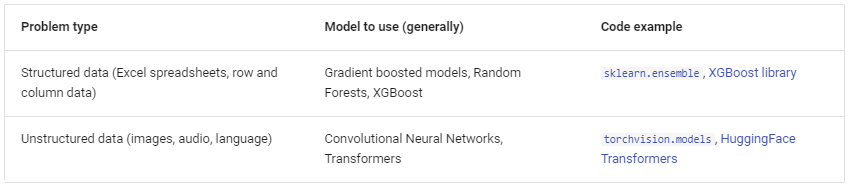
**CNN**

The typical CNN follows this structure

Input layer -> [Convolutional layer -> activation layer -> pooling layer] -> Output layer

Where the convolutional layer, activation layer, and pooling layer can be upscaled and repeated multiple times, depending on requirements.

What model should we use?



Note: Performance-speed trade-off! Something to be aware of in machine learning is the performance-speed trade-off. Generally, you get better performance out of a larger, more complex model (like we did with model\_2). However, this performance increase often comes at a sacrifice of training speed and inference speed.

Essentially, **every layer in a neural network is trying to compress data from higher dimensional space to lower dimensional space**.

See the link below for CNN model design, confusion matrix creation, and saving/loading best performing model:

<https://www.learnpytorch.io/03_pytorch_computer_vision/>