



3. Convolutional Neural Networks

In the previous section, we built and trained a simple model to classify ASL images. The model was able to learn how to correctly classify the training dataset with very high accuracy, but, it did not perform nearly as well on validation dataset. This behavior of not generalizing well to non-training data is called [overfitting](#), and in this section, we will introduce a popular kind of model called a [convolutional neural network](#) that is especially good for reading images and classifying them.

3.1 Objectives

- Prep data specifically for a CNN
- Create a more sophisticated CNN model, understanding a greater variety of model layers
- Train a CNN model and observe its performance

```
In [1]: import torch.nn as nn
import pandas as pd
import torch
from torch.optim import Adam
from torch.utils.data import Dataset, DataLoader

device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
torch.cuda.is_available()
```

Out[1]: True

3.2 Loading and Preparing the Data

3.2.1 Preparing Images

Let's begin by loading our DataFrames like we did in the previous lab:

```
In [2]: train_df = pd.read_csv("data/asl_data/sign_mnist_train.csv")
        valid_df = pd.read_csv("data/asl_data/sign_mnist_valid.csv")
```

This ASL data is already flattened.

```
In [3]: sample_df = train_df.head().copy() # Grab the top 5 rows
        sample_df.pop('label')
        sample_x = sample_df.values
        sample_x
```

```
Out[3]: array([[107, 118, 127, ..., 204, 203, 202],
               [155, 157, 156, ..., 103, 135, 149],
               [187, 188, 188, ..., 195, 194, 195],
               [211, 211, 212, ..., 222, 229, 163],
               [164, 167, 170, ..., 163, 164, 179]])
```

```
In [4]: sample_x.shape
```

```
Out[4]: (5, 784)
```

In this format, we don't have all the information about which pixels are near each other. Because of this, we can't apply convolutions that will detect features. Let's [reshape](#) our dataset so that they are in a 28x28 pixel format. This will allow our convolutions to associate groups of pixels and detect important features.

Note that for the first convolutional layer of our model, we need to have not only the height and width of the image, but also the number of [color channels](#). Our images are grayscale, so we'll just have 1 channel.

That means that we need to convert the current shape `(5, 784)` to `(5, 1, 28, 28)`. With [NumPy](#) arrays, we can pass a `-1` for any dimension we wish to remain the same.

```
In [5]: IMG_HEIGHT = 28
        IMG_WIDTH = 28
        IMG_CHS = 1
```

```
sample_x = sample_x.reshape(-1, IMG_CHS, IMG_HEIGHT, IMG_WIDTH)
sample_x.shape
```

Out[5]: (5, 1, 28, 28)

3.2.2 Create a Dataset

Let's add the steps above into our `MyDataset` class.

Exercise

There are 4 `FIXME`s in the class definition below. Can you replace them with the correct values?

```
In [9]: class MyDataset(Dataset):
        def __init__(self, base_df):
            x_df = base_df.copy() # Some operations below are in-place
            y_df = x_df.pop('label')
            x_df = x_df.values / 255 # Normalize values from 0 to 1
            x_df = x_df.reshape(-1, IMG_CHS, IMG_HEIGHT, IMG_WIDTH)
            self.xs = torch.tensor(x_df).float().to(device)
            self.ys = torch.tensor(y_df).to(device)

        def __getitem__(self, idx):
            x = self.xs[idx]
            y = self.ys[idx]
            return x, y

        def __len__(self):
            return len(self.xs)
```

Solution

Click the `...` below for the solution.

```
In [10]: # SOLUTION
class MyDataset(Dataset):
    def __init__(self, base_df):
```

```
x_df = base_df.copy() # Some operations below are in-place
y_df = x_df.pop('label')
x_df = x_df.values / 255 # Normalize values from 0 to 1
x_df = x_df.reshape(-1, IMG_CHS, IMG_WIDTH, IMG_HEIGHT)
self.xs = torch.tensor(x_df).float().to(device)
self.ys = torch.tensor(y_df).to(device)

def __getitem__(self, idx):
    x = self.xs[idx]
    y = self.ys[idx]
    return x, y

def __len__(self):
    return len(self.xs)
```

3.2.3 Create a DataLoader

Next, let's create the DataLoader from the Dataset

Exercise

One of these function calls is missing the `shuffle=True` argument. Can you remember which one it is and add it back in?

```
In [11]: BATCH_SIZE = 32

train_data = MyDataset(train_df)
train_loader = DataLoader(train_data, batch_size=BATCH_SIZE, shuffle=True)
train_N = len(train_loader.dataset)

valid_data = MyDataset(valid_df)
valid_loader = DataLoader(valid_data, batch_size=BATCH_SIZE)
valid_N = len(valid_loader.dataset)
```

Solution

Click the `...` below for the solution.

```
In [ ]: # SOLUTION
train_loader = DataLoader(train_data, batch_size=BATCH_SIZE, shuffle=True)
```

Let's grab a batch from the DataLoader to make sure it works.

```
In [12]: batch = next(iter(train_loader))
batch
```

```

Out[12]: [tensor([[[[0.3961, 0.4039, 0.4118, ..., 0.5020, 0.5059, 0.5059],
                    [0.4000, 0.4078, 0.4157, ..., 0.5137, 0.5098, 0.5098],
                    [0.4039, 0.4118, 0.4235, ..., 0.5137, 0.5137, 0.5098],
                    ...,
                    [0.4863, 0.5020, 0.5020, ..., 0.6275, 0.6235, 0.6235],
                    [0.4824, 0.5059, 0.4902, ..., 0.6314, 0.6275, 0.6275],
                    [0.4941, 0.5020, 0.6588, ..., 0.6353, 0.6353, 0.6314]]]],

          [[[0.7333, 0.1882, 0.3098, ..., 0.9333, 0.9373, 0.9412],
              [0.7020, 0.1647, 0.3176, ..., 0.9412, 0.9490, 0.9451],
              [0.6667, 0.1529, 0.3137, ..., 0.9412, 0.9490, 0.9529],
              ...,
              [0.1882, 0.3882, 0.4745, ..., 0.4471, 0.3922, 0.3490],
              [0.7176, 0.7608, 0.8078, ..., 0.2235, 0.1451, 0.1686],
              [0.6039, 0.5373, 0.5804, ..., 0.2078, 0.1333, 0.1137]]]],

          [[[0.7373, 0.7373, 0.7412, ..., 0.1961, 0.1804, 0.1804],
              [0.7412, 0.7451, 0.7490, ..., 0.1843, 0.1608, 0.1569],
              [0.7490, 0.7569, 0.7608, ..., 0.2039, 0.1412, 0.1569],
              ...,
              [0.2235, 0.2118, 0.2000, ..., 0.7686, 0.7647, 0.7569],
              [0.2157, 0.1922, 0.1804, ..., 0.7725, 0.7686, 0.7451],
              [0.1961, 0.1725, 0.1725, ..., 0.7804, 0.7765, 0.5765]]]],

          ...,

          [[[0.7569, 0.7569, 0.7608, ..., 0.7804, 0.7765, 0.7765],
              [0.7569, 0.7608, 0.7647, ..., 0.7843, 0.7843, 0.7804],
              [0.7529, 0.7608, 0.7686, ..., 0.7882, 0.7882, 0.7882],
              ...,
              [0.3922, 0.4000, 0.4000, ..., 0.8745, 0.8314, 0.7922],
              [0.3882, 0.3922, 0.3961, ..., 0.8510, 0.8510, 0.8235],
              [0.3922, 0.3922, 0.4000, ..., 0.8157, 0.8627, 0.8706]]]],

          [[[0.5843, 0.6000, 0.6118, ..., 0.6353, 0.6314, 0.6314],

```

```

[0.5961, 0.6039, 0.6196, ..., 0.6431, 0.6392, 0.6353],
[0.6000, 0.6157, 0.6353, ..., 0.6588, 0.6549, 0.6431],
...,
[0.7059, 0.7216, 0.7373, ..., 0.8000, 0.7922, 0.7882],
[0.7059, 0.7216, 0.7373, ..., 0.8039, 0.7961, 0.7922],
[0.7059, 0.7216, 0.7373, ..., 0.8039, 0.7961, 0.7922]]],

[[[0.4431, 0.4745, 0.4902, ..., 0.7294, 0.7333, 0.7412],
  [0.4510, 0.4745, 0.4863, ..., 0.7333, 0.7412, 0.7451],
  [0.4549, 0.4784, 0.4902, ..., 0.7412, 0.7529, 0.7490],
  ...,
  [0.5451, 0.5569, 0.5647, ..., 0.6549, 0.9255, 0.8549],
  [0.5451, 0.5569, 0.5647, ..., 0.6706, 0.9216, 0.8588],
  [0.5451, 0.5569, 0.5686, ..., 0.7608, 0.9098, 0.8745]]]],
device='cuda:0'),
tensor([13, 14,  2,  2,  7, 23, 21,  0, 14,  1, 20,  3, 15,  3, 20, 21, 10, 13,
        7, 16,  0, 18, 12, 13, 22,  1, 10, 10,  0, 12,  5, 15],
device='cuda:0')]
```

It looks different, but let's check the `shape`s to be sure.

```
In [13]: batch[0].shape
```

```
Out[13]: torch.Size([32, 1, 28, 28])
```

```
In [14]: batch[1].shape
```

```
Out[14]: torch.Size([32])
```

3.3 Creating a Convolutional Model

These days, many data scientists start their projects by borrowing model properties from a similar project. Assuming the problem is not totally unique, there's a great chance that people have created models that will perform well which are posted in online repositories like [TensorFlow Hub](#) and the [NGC Catalog](#). Today, we'll provide a model that will work well for this problem.




No description has been provided for this image

We covered many of the different kinds of layers in the lecture, and we will go over them all here with links to their documentation. When in doubt, read the official documentation (or ask [Stack Overflow](#)).

```
In [15]: n_classes = 24
kernel_size = 3
flattened_img_size = 75 * 3 * 3

model = nn.Sequential(
    # First convolution
    nn.Conv2d(IMG_CHS, 25, kernel_size, stride=1, padding=1), # 25 x 28 x 28
    nn.BatchNorm2d(25),
    nn.ReLU(),
    nn.MaxPool2d(2, stride=2), # 25 x 14 x 14
    # Second convolution
    nn.Conv2d(25, 50, kernel_size, stride=1, padding=1), # 50 x 14 x 14
    nn.BatchNorm2d(50),
    nn.ReLU(),
    nn.Dropout(.2),
    nn.MaxPool2d(2, stride=2), # 50 x 7 x 7
    # Third convolution
    nn.Conv2d(50, 75, kernel_size, stride=1, padding=1), # 75 x 7 x 7
    nn.BatchNorm2d(75),
    nn.ReLU(),
    nn.MaxPool2d(2, stride=2), # 75 x 3 x 3
    # Flatten to Dense
    nn.Flatten(),
    nn.Linear(flattened_img_size, 512),
    nn.Dropout(.3),
    nn.ReLU(),
    nn.Linear(512, n_classes)
)
```

3.3.1 Conv2D

No description has been provided for this image

These are our 2D convolutional layers. Small kernels will go over the input image and detect features that are important for classification. Earlier convolutions in the model will detect simple features such as lines. Later convolutions will detect more complex features. Let's look at

our first Conv2D layer:

```
nn.Conv2d(IMG_CHS, 25, kernel_size, stride=1, padding=1)
```

25 refers to the number of filters that will be learned. Even though `kernel_size = 3`, PyTorch will assume we want 3 x 3 filters. Stride refers to the step size that the filter will take as it passes over the image. Padding refers to whether the output image that's created from the filter will match the size of the input image.

3.3.2 BatchNormalization

Like normalizing our inputs, batch normalization scales the values in the hidden layers to improve training. [Read more about it in detail here.](#)

There is a debate on best where to put the batch normalization layer. [This Stack Overflow post](#) compiles many perspectives.

3.3.3 MaxPool2D



No description has been provided for this image Max pooling takes an image and essentially shrinks it to a lower resolution. It does this to help the model be robust to translation (objects moving side to side), and also makes our model faster.

3.3.4 Dropout



No description has been provided for this image Dropout is a technique for preventing overfitting. Dropout randomly selects a subset of neurons and turns them off, so that they do not participate in forward or backward propagation in that particular pass. This helps to make sure that the network is robust and redundant, and does not rely on any one area to come up with answers.

3.3.5 Flatten

Flatten takes the output of one layer which is multidimensional, and flattens it into a one-dimensional array. The output is called a feature vector and will be connected to the final classification layer.

3.3.6 Linear

We have seen dense linear layers before in our earlier models. Our first dense layer (512 units) takes the feature vector as input and learns which features will contribute to a particular classification. The second dense layer (24 units) is the final classification layer that outputs our prediction.

3.4 Summarizing the Model

This may feel like a lot of information, but don't worry. It's not critical that to understand everything right now in order to effectively train convolutional models. Most importantly we know that they can help with extracting useful information from images, and can be used in classification tasks.

```
In [16]: model = torch.compile(model.to(device))
         model
```

```
Out[16]: OptimizedModule(
  (_orig_mod): Sequential(
    (0): Conv2d(1, 25, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): BatchNorm2d(25, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): ReLU()
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (4): Conv2d(25, 50, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (5): BatchNorm2d(50, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (6): ReLU()
    (7): Dropout(p=0.2, inplace=False)
    (8): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (9): Conv2d(50, 75, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (10): BatchNorm2d(75, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (11): ReLU()
    (12): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (13): Flatten(start_dim=1, end_dim=-1)
    (14): Linear(in_features=675, out_features=512, bias=True)
    (15): Dropout(p=0.3, inplace=False)
    (16): ReLU()
    (17): Linear(in_features=512, out_features=24, bias=True)
  )
)
```

Since the problem we are trying to solve is still the same (classifying ASL images), we will continue to use the same `loss_function` and `accuracy` metric.

```
In [17]: loss_function = nn.CrossEntropyLoss()
optimizer = Adam(model.parameters())
```

```
In [18]: def get_batch_accuracy(output, y, N):
    pred = output.argmax(dim=1, keepdim=True)
    correct = pred.eq(y.view_as(pred)).sum().item()
    return correct / N
```

3.5 Training the Model

Despite the very different model architecture, the training looks exactly the same.

Exercise

These are the same `train` and `validate` functions as before, but they have been mixed up. Can you correctly name each function and replace the `FIXME` s?

One of them should have `model.train` and the other should have `model.eval` .

```
In [19]: def train():
    loss = 0
    accuracy = 0

    model.train()
    with torch.no_grad():
        for x, y in train_loader:
            output = model(x)

            loss += loss_function(output, y).item()
            accuracy += get_batch_accuracy(output, y, valid_N)
    print('FIXME - Loss: {:.4f} Accuracy: {:.4f}'.format(loss, accuracy))
```

```
In [22]: def validate():
    loss = 0
    accuracy = 0

    model.eval()
    for x, y in valid_loader:
        output = model(x)
        optimizer.zero_grad()
        batch_loss = loss_function(output, y)
        batch_loss.backward()
        optimizer.step()

        loss += batch_loss.item()
        accuracy += get_batch_accuracy(output, y, train_N)
    print('FIXME - Loss: {:.4f} Accuracy: {:.4f}'.format(loss, accuracy))
```

Solution

Click the two `...`s below for the solution.

```
In [ ]: # SOLUTION
def validate():
    loss = 0
    accuracy = 0

    model.eval()
    with torch.no_grad():
        for x, y in valid_loader:
            output = model(x)

            loss += loss_function(output, y).item()
            accuracy += get_batch_accuracy(output, y, valid_N)
    print('Valid - Loss: {:.4f} Accuracy: {:.4f}'.format(loss, accuracy))
```

```
In [ ]: # SOLUTION
def train():
    loss = 0
    accuracy = 0
```

```
model.train()
for x, y in train_loader:
    output = model(x)
    optimizer.zero_grad()
    batch_loss = loss_function(output, y)
    batch_loss.backward()
    optimizer.step()

    loss += batch_loss.item()
    accuracy += get_batch_accuracy(output, y, train_N)
print('Train - Loss: {:.4f} Accuracy: {:.4f}'.format(loss, accuracy))
```

In [23]: epochs = 20

```
for epoch in range(epochs):
    print('Epoch: {}'.format(epoch))
    train()
    validate()
```

Epoch: 0
FIXME - Loss: 2782.0439 Accuracy: 0.1482
FIXME - Loss: 154.1871 Accuracy: 0.2097
Epoch: 1
FIXME - Loss: 1208.9497 Accuracy: 2.3893
FIXME - Loss: 18.6420 Accuracy: 0.2586
Epoch: 2
FIXME - Loss: 1705.7598 Accuracy: 1.7455
FIXME - Loss: 9.6144 Accuracy: 0.2596
Epoch: 3
FIXME - Loss: 1455.4534 Accuracy: 2.1648
FIXME - Loss: 6.1028 Accuracy: 0.2605
Epoch: 4
FIXME - Loss: 1807.6431 Accuracy: 1.9254
FIXME - Loss: 7.9174 Accuracy: 0.2593
Epoch: 5
FIXME - Loss: 2045.8038 Accuracy: 1.9996
FIXME - Loss: 6.5056 Accuracy: 0.2600
Epoch: 6
FIXME - Loss: 2083.9521 Accuracy: 2.1694
FIXME - Loss: 6.3304 Accuracy: 0.2596
Epoch: 7
FIXME - Loss: 2770.1481 Accuracy: 1.8670
FIXME - Loss: 7.6571 Accuracy: 0.2598
Epoch: 8
FIXME - Loss: 2463.3709 Accuracy: 2.0089
FIXME - Loss: 9.2565 Accuracy: 0.2594
Epoch: 9
FIXME - Loss: 1544.9415 Accuracy: 2.3974
FIXME - Loss: 3.1741 Accuracy: 0.2605
Epoch: 10
FIXME - Loss: 1569.3760 Accuracy: 2.3756
FIXME - Loss: 2.4585 Accuracy: 0.2609
Epoch: 11
FIXME - Loss: 1402.3621 Accuracy: 2.4948
FIXME - Loss: 2.8435 Accuracy: 0.2606
Epoch: 12
FIXME - Loss: 1622.4240 Accuracy: 2.4438
FIXME - Loss: 5.8331 Accuracy: 0.2594
Epoch: 13
FIXME - Loss: 1551.6457 Accuracy: 2.5595

```
FIXME - Loss: 1.5718 Accuracy: 0.2610
Epoch: 14
FIXME - Loss: 1658.7028 Accuracy: 2.4940
FIXME - Loss: 1.7969 Accuracy: 0.2608
Epoch: 15
FIXME - Loss: 1755.1830 Accuracy: 2.4781
FIXME - Loss: 5.1343 Accuracy: 0.2593
Epoch: 16
FIXME - Loss: 1650.5110 Accuracy: 2.6398
FIXME - Loss: 3.3553 Accuracy: 0.2600
Epoch: 17
FIXME - Loss: 1796.6201 Accuracy: 2.5602
FIXME - Loss: 1.5597 Accuracy: 0.2608
Epoch: 18
FIXME - Loss: 1981.4730 Accuracy: 2.4993
FIXME - Loss: 1.7940 Accuracy: 0.2608
Epoch: 19
FIXME - Loss: 1816.9048 Accuracy: 2.5983
FIXME - Loss: 2.4913 Accuracy: 0.2606
```

3.5.1 Discussion of Results

It looks like this model is significantly improved! The training accuracy is very high, and the validation accuracy has improved as well. This is a great result, as all we had to do was swap in a new model.

You may have noticed the validation accuracy jumping around. This is an indication that our model is still not generalizing perfectly. Fortunately, there's more that we can do. Let's talk about it in the next lecture.

3.6 Summary

In this section, we utilized several new kinds of layers to implement a CNN, which performed better than the more simple model used in the last section. Hopefully the overall process of creating and training a model with prepared data is starting to become even more familiar.

3.6.1 Clear the Memory

Before moving on, please execute the following cell to clear up the GPU memory. This is required to move on to the next notebook.

```
In [24]: import IPython
app = IPython.Application.instance()
app.kernel.do_shutdown(True)
```

```
Out[24]: {'status': 'ok', 'restart': True}
```

3.6.2 Next

In the last several sections you have focused on the creation and training of models. In order to further improve performance, you will now turn your attention to *data augmentation*, a collection of techniques that will allow your models to train on more and better data than what you might have originally at your disposal.

